



PARAMETER ESTIMATION AND UNCERTAINTY QUANTIFICATION IN WATER RESOURCES MODELING

EDITED BY: Philippe Renard, Frederick Delay, Daniel M. Tartakovsky and
Velimir V. Vesselinov

PUBLISHED IN: *Frontiers in Environmental Science* and *Frontiers in Earth Science*



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-674-7

DOI 10.3389/978-2-88963-674-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

PARAMETER ESTIMATION AND UNCERTAINTY QUANTIFICATION IN WATER RESOURCES MODELING

Topic Editors:

Philippe Renard, Université de Neuchâtel, Switzerland

Frederick Delay, Université de Strasbourg, France

Daniel M. Tartakovsky, Stanford University, United States

Velimir V. Vesselinov, Los Alamos National Laboratory (DOE), United States

Numerical models of flow and transport processes are heavily employed in the fields of surface, soil, and groundwater hydrology. They are used to interpret field observations, analyze complex and coupled processes, or to support decision making related to large societal issues such as the water-energy nexus or sustainable water management and food production. Parameter estimation and uncertainty quantification are two key features of modern science-based predictions. When applied to water resources, these tasks must cope with many degrees of freedom and large datasets. Both are challenging and require novel theoretical and computational approaches to handle complex models with large number of unknown parameters.

Citation: Renard, P., Delay, F., Tartakovsky, D. M., Vesselinov, V. V., eds. (2020).

Parameter Estimation and Uncertainty Quantification in Water Resources Modeling.

Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-674-7

Table of Contents

- 04** *What We Talk About When We Talk About Uncertainty. Toward a Unified, Data-Driven Framework for Uncertainty Characterization in Hydrogeology*
Falk Heße, Alessandro Comunian and Sabine Attinger
- 24** *Groundwater Contaminant Transport: Prediction Under Uncertainty, With Application to the MADE Transport Experiment*
Aldo Fiori, Antonio Zarlenga, Alberto Bellin, Vladimir Cvetkovic and Gedeon Dagan
- 40** *Hydrogeophysical Parameter Estimation Using Iterative Ensemble Smoothing and Approximate Forward Solvers*
Corinna Köpke, Ahmed H. Elsheikh and James Irving
- 55** *Comparison Between Hydraulic Conductivity Anisotropy and Electrical Resistivity Anisotropy From Tomography Inverse Modeling*
Simon Gernez, Abderrezak Bouchedda, Erwan Gloaguen and Daniel Paradis
- 70** *Parallelized Adaptive Importance Sampling for Solving Inverse Problems*
Christoph Jäggli, Julien Straubhaar and Philippe Renard
- 85** *Efficient gHMC Reconstruction of Contaminant Release History*
David A. Barajas-Solano, Francis J. Alexander, Marian Anghel and Daniel M. Tartakovsky
- 98** *A Machine Learning Based Hybrid Multi-Fidelity Multi-Level Monte Carlo Method for Uncertainty Quantification*
Nagoor Kani Jabarullah Khan and Ahmed H. Elsheikh
- 113** *Using Meta-Models for Tsunami Hazard Analysis: An Example of Application for the French Atlantic Coast*
Vito Bacchi, Hervé Jomard, Oona Scotti, Ekaterina Antoshchenkova, Lise Bardet, Claire-Marie Duluc and Hélène Hebert
- 130** *Inversion Algorithm for Civil Flood Defense Optimization: Application to Two-Dimensional Numerical Model of the Garonne River in France*
Yann Richet and Vito Bacchi
- 146** *Combining Clustering Methods With MPS to Estimate Structural Uncertainty for Hydrological Models*
Troels Norvin Vilhelmsen, Esben Auken, Anders Vest Christiansen, Adrian Sanchez Barfod, Pernille Aabye Marker and Peter Bauer-Gottwein
- 161** *Heterogeneity and Prior Uncertainty Investigation Using a Joint Heat and Solute Tracer Experiment in Alluvial Sediments*
Richard Hoffmann, Alain Dassargues, Pascal Goderniaux and Thomas Hermans



What We Talk About When We Talk About Uncertainty. Toward a Unified, Data-Driven Framework for Uncertainty Characterization in Hydrogeology

Falk Heße^{1,2*}, Alessandro Comunian³ and Sabine Attinger^{1,2}

¹ Institut für Erd- und Umweltwissenschaften, Universität Potsdam, Potsdam, Germany, ² Department of Computational Hydrosystems, UFZ–Helmholtz Centre for Environmental Research, Leipzig, Germany, ³ Dipartimento di Scienze della Terra “A.Desio”, Università degli Studi di Milano, Milan, Italy

Keywords: Bayesianism, uncertainty analysis, hydrogeology, data science, opinion, prior derivation

A habit of basing convictions upon evidence and giving them only that degree of certainty which the evidence warrants, would, if it became general, cure most of the ills from which the world is suffering.
–Bertrand Russel

OPEN ACCESS

Edited by:

Frederick Delay,
Université de Strasbourg, France

Reviewed by:

Abderrahim Jardani,
Université de Rouen, France
Bart Rogiers,
Belgian Nuclear Research Centre,
Belgium

*Correspondence:

Falk Heße
falk.hesse@ufz.de

Specialty section:

This article was submitted to
Hydrosphere,
a section of the journal
Frontiers in Earth Science

Received: 30 November 2018

Accepted: 06 May 2019

Published: 18 June 2019

Citation:

Heße F, Comunian A and Attinger S
(2019) What We Talk About When We
Talk About Uncertainty. Toward a
Unified, Data-Driven Framework for
Uncertainty Characterization in
Hydrogeology. *Front. Earth Sci.* 7:118.
doi: 10.3389/feart.2019.00118

1. INTRODUCTION

Uncertainty is a central and unavoidable feature of decision problems that people face both in everyday life, as well as in virtually every field of science. Hydrogeology is no exception to that. In fact, the relevance of uncertainty to hydrogeology is particularly high, due to the high variability of many subsurface properties combined with a general scarcity of data. These factors have led to the development of stochastic hydrogeology where the subsurface properties are modeled as random variables (Gelhar, 1993; Rubin, 2003). Despite the wealth of research on stochastic modeling, a systematic investigation into the quantification of uncertainty and its impact on decision problems has remained limited. For example, Dagan (2002) noted that uncertainty “is a topic that has received little attention” and more recently, Kitanidis (2015) again pointed out that “it is somewhat surprising that this topic has not received more attention”. Instead, most discussions evolve around the specific topic of stochastic concepts, which is a closely related but ultimately an independent topic. For example, in a discussion forum, Zhang and Zhang (2004) solicited contributions discuss the perceived lack of applications that stochastic concepts have seen in hydrogeology. Of these contributions, only Ginn (2004) and Rubin (2004) discussed uncertainty briefly in the context of probability assessments. This lack may be explainable by the limited scope of that discussion forum, where only a fixed number of questions were to be answered by the contributors. However, more recently, Sanchez-Vila and Fernandez-Garcia (2016) organized a Special Issue that covered that same question but allowed for more space and personal involvement from the participants. While the topic of uncertainty was touched upon in all solicited contributions, none of them discussed its nature or aims. This neglect contrasts the closely related fields of hydrology (Mantovan and Todini, 2006; Vrugi et al., 2009; Clark et al., 2011) as well as geology (Wellmann and Regenauer-Lieb, 2012; Bond, 2015; de la Varga and Wellmann, 2016) which have and continue to have a broad and deep discussion about the nature, sources, aims and direction of uncertainty analysis. One of the main impediments for further progress in the field of uncertainty characterization, that has been identified, is the lack of a coherent terminology and framework (Montanari, 2007; Montanari et al., 2009).

A significant challenge for the establishment of a consistent framework of uncertainty analysis in hydrogeology is caused by the most prevalent workflow of inverse theory, or, inversion for short (Neuman, 2004; Tarantola, 2005; Carrera et al., 2005; Franssen Hendricks et al., 2009; Biegler et al., 2010; Menke, 2012; Linde et al., 2015; Zhdanov, 2015). This workflow, known as point estimation in inference, is based on a (regularized) data fitting or optimization approach which tries to identify a unique solution to the inverse problem. Typical examples in hydrogeology are goodness-of-fit criteria, i.e., the task of finding a single set of parameters such that the distance between predictions and observations is minimized (Indelman et al., 1996; Sánchez-Vila et al., 1999; Firmani et al., 2006; Schneider and Attinger, 2008; Riva et al., 2009; Coptý et al., 2011; Pechstein et al., 2016; Zech et al., 2016). Even when point-estimators are not applied during calibration and a probabilistic sampling of the parameter distribution is advised instead, often the analysis is performed again, by deriving single solutions through, e.g., averaging (Neuman, 2014). There seems to be a deep distrust of the generally non-unique nature of scientific inference. Strangely, no rationale for this distrust is ever asked for nor provided.

Since this workflow of inversion was not designed to account for uncertainties in the inference, it consequently exhibits a number of problems from that perspective. First, aiming for a single best estimate neglects all other parameter sets that are possible but may be less plausible. Neglecting possible states can be problematic for inference alone (Good, 2009a). For uncertainty analysis, point estimation is even more problematic since the use of a single best estimate implies absolute certainty. Second, these studies usually use the observed data only, without any reference to available background data and are therefore liable to the base rate fallacy (Barhillel, 1980; Kahneman et al., 1982).

Responses to this have been mixed and often inconsistent (Nearing et al., 2016). Arguably, the most obvious concern is centered around the so-called problem of equifinality (Beven, 2006), i.e., the observation that many, often diverging, parameter sets may provide nearly identical goodness-of-fit values. One response to this insight has been the development and application of the generalized likelihood uncertainty estimation (GLUE) framework (Beven and Binley, 1992). Although often described as an (informal) Bayesian method, it has been strongly criticized in the literature (Mantovan and Todini, 2006; Stedinger et al., 2008). Another seminal response has been the development of the differential evolution adaptive metropolis framework (DREAM, Vrugt et al., 2009; Laloy and Vrugt, 2012; Laloy et al., 2013). Unlike GLUE, DREAM is a fully Bayesian inference method and uncertainty estimator and has seen many applications in hydrology since its publication. Although DREAM has seen applications in hydrogeology too (Mariethoz et al., 2010; Hansen et al., 2012; Shi et al., 2014; Xu et al., 2017; Laloy et al., 2016; Hayek et al., 2018), no universally accepted Bayesian framework for uncertainty analysis currently exists. A rising number of Bayesian inversion methods for inference have been published over the years (see e.g., Cardiff and Kitanidis, 2009; Rubin et al., 2010; Shi et al., 2014; Elsheikh et al., 2014; Saley et al., 2016),

yet, the overall adoption of such methods has remained limited. This lackluster adoption rate can, e.g., be seen by the almost total absence of Bayesian methods in the aforementioned special issue on stochastic hydrogeology. Only Cirpka and Valocchi (2016) even care to mention this topic and only in the context of Bayesian model selection.

One problem that all the above-mentioned uncertainty frameworks share, is the lack of formalized prior derivation. Hydrogeology itself has only produced a small number of studies on this topic (Kitanidis, 2012; Li et al., 2017; Cucchi et al., 2019), which is sadly in line with the situation in many other fields. Since specifying the prior is the first step in any Bayesian analysis, high emphasis should be put on this step. However, coming up with universal and objective guidelines for every conceivable situation has proven to be elusive, so far (Earman, 1992; Scales and Tenorio, 2001). In fact, detractors of Bayesian inference are often criticizing the need to define a prior, claiming this step to be necessarily subjective and arbitrary (Kass and Wasserman, 1996; Ulrych et al., 2001; Kass, 2011). On the other hand, proponents of Bayesian inference cite the ability to include background knowledge in the form of prior probabilities as one of its major strengths (Jaynes, 1968).

To this date, the single most used inference framework in hydrogeology is arguably the Model-Independent Parameter Estimation and Uncertainty Analysis framework (PEST, Doherty, 2004). PEST itself constitutes a diverse set of optimization and estimation tools to calibrate a wide range of environmental models. The uncertainty framework allows users to estimate the predictive uncertainty of the model output using a somewhat inconsistent mixture of Bayesian and calibration techniques (Doherty, 2010). Such an eclectic approach to inference and uncertainty characterization may look intriguing. However, as Kitanidis (2015) pointed out, combining elements from two internally consistent systems is very risky.

In this manuscript, we will instead make the case for a single coherent data-driven framework for uncertainty characterization in hydrogeology. As we will argue below, this framework ought to be the Bayesian interpretation of probability, i.e., the system of probabilistic reasoning as developed by Ramsey (1931), Finetti (1975), Savage (1954), and others. When making this case, we will follow the reasoning of Pearl (1988b), who, while talking about Bayesian probability in the context of artificial intelligence, said the following:

Obviously, there are applications where strict adherence to the dictates of probability theory would be computationally infeasible, and their compromises will have to be made. Still, we find it more comfortable to compromise an ideal theory that is well-understood than to search for a surrogate theory, with only gut feeling for guidance. The merits of a theory-based approach are threefold:

1. The theory can be consulted to ensure that compromises are made only when necessary and that their damage is kept to a minimum.
2. When system performance does not match expectations, knowing which compromises were made helps identify the adjustments needed.

3. Compromised theories facilitate scientific communication; one need specifies only the compromise made, treating the rest of the theory as common knowledge.

To follow the advice of Pearl (1988b), we will start by describing such an ideal theory, outline the present challenges for its application and explain how to address them. In section 2, we will make the case for Bayesian probability by contrasting it to potential competitors. Next, in section 3, we will conceptualize the different forms of uncertainty using the Bayesian framework and identify the most relevant forms for hydrogeology. Finally in section 4, we will make a number of practical propositions by outlining what is currently missing, what compromises need to be made to make the Bayesian paradigm viable and what the relevant steps are that may help to reduce or even eliminate some of these compromises.

2. REASONING WITH UNCERTAINTY

A discussion about uncertainty should begin with a clear understanding of what is meant by this term. In a slightly ironic twist, the term uncertainty is far from being well-defined, both in everyday use as well as in the sciences. In fact, when scanning the hydrogeological literature, a wide range of, often conflicting, definitions, and conceptualizations are used (Hoffman and Hammonds, 1994; Hofer, 1996; Walker et al., 2003; Brown, 2004; Carrera et al., 2005; Refsgaard et al., 2007, 2012; Tartakovsky, 2013; Bond, 2015; Enemark et al., 2019). We will therefore start by providing an overview of different models that have been developed to conceptualize this term and facilitate both qualitative and quantitative reasoning. To avoid the often *ad-hoc* or gut driven nature of uncertainty analysis that is found in the literature, we will start by presenting frameworks that were developed in the field of epistemology, i.e., the field of philosophy concerned with the character of knowledge.

In general, uncertainty should be understood as a measure that describes the distance or gap between a current state of an agent and the one representing absolute certainty. The latter is formalized by the True and False statements found in classic logic, whereas the former extends this concept. Consequently, we will begin by describing models for such partial certainty as found in modern epistemic logic. As argued above, we will make the case for Bayesianism, i.e., the idea that certainty equals probability and vice versa. Only in the next step we will discuss a concept of uncertainty and make the case for the Kullback-Leibler divergence as the distance measure between current and absolute certainty.

2.1. Models for Reasoning With Certainty and the Case for Probabilities

In modern epistemology, certainty is defined over possible states of reality, collectively known as the set of *possible worlds* Ω (Halpern, 2003; Fagin et al., 2004). In Bayesianism, the equivalent term would be *possibility space* (Kruschke, 2010), whereas the equally labeled sample space Ω from probability theory may or may not have the same meaning depending on its interpretation (see below). This possibility space is now meant to contain,

next to the true state of affairs, all other possibilities that are compatible with a given set of constraints and data available to an epistemic agent. Such an epistemic agent may be a human, but with the advent of artificial intelligence, computational agents have increasingly become the focus of modern research (Russell and Norvig, 2009). The epistemic state of such an agent is then defined by a function that assigns weights to each possible world ω . This function is known as a certainty or credence distribution.

Despite its relatively short history, the field of modern epistemology has already developed a number of different measures to describe that certainty distribution. These measures differ both mathematically as well as conceptually. The latter difference can best be understood by viewing these different measures as extensions of classical logic. In logic, a simple True or False relationship exists between a statement and the reality that this statement is trying to describe. In these extensions, discussed below, this simple relationship becomes more flexible and allows for different kinds of degrees of certainty (Darwiche and Pearl, 1997).

The most common approach is to conceptualize such a gradual degree of certainty as stemming from uncertain or incomplete knowledge. This means that, similar to classical logic, a given statement is objectively either true or false in reality. However, due to the agent's limited knowledge, she cannot fully determine its veracity and has to assign a limited certainty to it. This is known as a *degree of belief* and it is described by a number between 0 and 1, corresponding to False and True in classic logic. It can be shown that such degrees of belief follow the rules of probability theory, i.e., the certainty of an agent is to be described by a probability measure (Savage, 1954; Lindley, 1987). To illustrate this model; consider a statement about the mean conductivity of a given sample being above a given threshold, say, 10^{-3} m/s. The certainty of this statement could then be ascertained if we would have access to a well-calibrated histogram of conductivity values of samples from the aquifer the sample was taken from.

An extension of this model is derived by the additional inclusion of uncertainty stemming from ignorance. Using the now (in)famous epistemology of the former US Secretary of Defense Donald Rumsfeld (2002), probability describes the known unknowns whereas ignorance is about the unknown unknowns. To illustrate this concept, let us consider a revised version of the above problem: Suppose that our knowledge base is now less certain such that the sample is only with 90% probability from the aforementioned aquifer, but with 10% probability of some unknown provenience. This leaves a 10% certainty gap in our reasoning system. The seminal work of Dempster (1968) demonstrated a way of handling this gap, whereas the later extensions of Shafer (1976) made this calculus into a full reasoning and inference system. The resulting reasoning framework is the Dempster-Shafer theory, evidence theory or theory of belief functions. Like in the example above, Dempster-Shafer theory is often praised for being able to combine evidence from different sources with different kinds of uncertainty attached to it. Its applicability has, however, been limited due to a number of criticisms (Pearl, 1988b,a, 1990). Note that other fields use similar concepts, with sometimes very different notations.

In economy, for instance, uncertainties from lack of knowledge are called risk, whereas uncertainties from ignorance are called Knightian uncertainties or simply uncertainties (Knight, 1921). On the other hand, political science and decision theory often describe uncertainties as a lack of knowledge and ignorance, shallow, and deep uncertainty, respectively (Walker et al., 2013).

An alternative approach is to model uncertainty stemming from uncertain truth. This means that the veracity of a given statement may never be fully determined, even in cases of complete knowledge. Using the above example, this could be the case if the statement is altered such that conductivity of the sample is said to be large. Even if all relevant data on this sample are gathered, e.g., some laboratory testing may determine the conductivity being 10^{-3} m/s, no definitive certainty can be determined. A statement that the conductivity of the sample is large given that the conductivity is 10^{-3} m/s may be considered as sort-of-true. The point is that our limited degree of certainty is not caused by limited knowledge but by some vagueness or fuzziness in the statement itself. Measures that are able to describe such situations are confusingly called possibility measures, although no particular connection to above possibility space exists, and the related mathematical framework is, more appropriately, called fuzzy logic (Zadeh, 1978; Dubois and Prade, 1988). Reasoning systems that employ fuzzy logic have seen wide-ranging applications in engineering, modern logic, artificial intelligence systems etc., which testifies to their versatility and usefulness.

Despite the existence of these comparably sophisticated systems for reasoning under uncertainty, the simplest approach of probabilistic reasoning has seen a strong resurgence in the last decades with the introduction of Bayesian networks (Pearl, 1988b; Neapolitan, 1990). These networks are probabilistic graphical models that represent the, typically causal, relationship between different physical processes.

Using the above statements, we would assert that probabilistic reasoning is the most-suited framework for reasoning under uncertainty in hydrogeology. Probability represents a simple, yet very flexible tool that is able to capture most of the problems encountered in this field. Fuzzy logic, although often employed in engineering, does not offer much additional benefit since most evidence being used is numerical in nature and therefore has virtually no fuzziness associated with it. Contrary to that, Dempster-Shafer theory does offer relevant benefits as an uncertainty framework, which do however, need to be considered in context. The first benefit is simply due to the fact that many situations of subsurface analysis do include an element of deep uncertainty, in particular the topic of structural uncertainty. To address that, we will deal with this problem exhaustively below and describe a way how to turn this deep into shallow uncertainty. In this way we are making the topic of structural uncertainty fully amendable to probabilistic analysis. Second, as pointed out by Rubin et al. (2018), the Dempster-Shafer theory can be helpful in accounting for unknown unknowns that result from the interaction of hydrogeological problems with societal developments in general (Walker et al., 2013; Maier et al., 2016). This second benefit is very important but does not necessarily conflict with our notion of probability as a default

system for uncertain reasoning. Since Dempster-Shafer theory is a full generalization of probability theory, it is easy to embed a fully probabilistic analysis within a larger framework. In addition to these benefits, the Dempster-Shafer theory has a number of drawbacks that make reasoning with it counter intuitive and hamper its applicability for real-world problems. In particular, it is not possible to employ it on top of existing techniques compared to probability theory or even fuzzy logic and its applicability to decision theory remains controversial. Although the number of applications in earth sciences is rising, it is still a niche theory with only few practical applications (Malpica et al., 2007). In summary, we propose to use probabilistic reasoning as the main tool of uncertainty analysis, while being aware of its limitations, and being prepared to account for other types of uncertainty by embedding probability measures within a larger analysis possibly using the Dempster-Shafer theory.

2.2. On the Interpretation of Probability

Owing to the seminal work of Kolmogorov (1933), modern probability theory is fully grounded in set theory and as such, it is as well founded and defined as any other field of mathematics (Kallenberg, 2002). Yet, unlike many other mathematical disciplines, there is no clear consensus about where to locate probability in real-world situations.

Roughly speaking, two different interpretations of probability can be distinguished; physical as well as epistemic probability (Figure 1). The first interpretation regards probability as an actual property of physical systems comparable to, e.g., mass, energy and momentum (Figure 1, upper right corner). This is best captured in its most widely applied form; frequentism, where the probability of an event is equated with the relative frequency of this event in an often-repeated random experiment (Neyman and Pearson, 1928, 1933). This definition has garnered wide support in the sciences, due to its clear and lucid formulation (von Mises, 1982). On the other side, the epistemic interpretation regards probability as an intrinsic property of epistemic agents (Figure 1, upper left corner). This means that, unlike mass, energy or momentum; probability is not a property of a physical system but of the epistemic state of an agent that is trying to reason about said system (Finetti, 1975; Savage, 1954; Jeffrey, 1992).

2.2.1. The Case for Bayesianism

The discussion about the best or most appropriate interpretation for any given situation is still ongoing and we do not want to uncritically favor any side. For our topic, however, the epistemic interpretation, called Bayesian interpretation, of probability seems to be the only appropriate one. The main rationales for its use shall be discussed in the following.

First, the epistemic interpretation is simply the more comprehensive interpretation of the two. In fact, it is a full generalization of the physical interpretation, since it is able to cover all cases described by the latter and then some. This is not the case for the frequentist interpretation since its application is constrained to cases where physical frequencies are available. This inclusiveness of Bayesianism is often obfuscated by calling it the subjective interpretation of probability, as

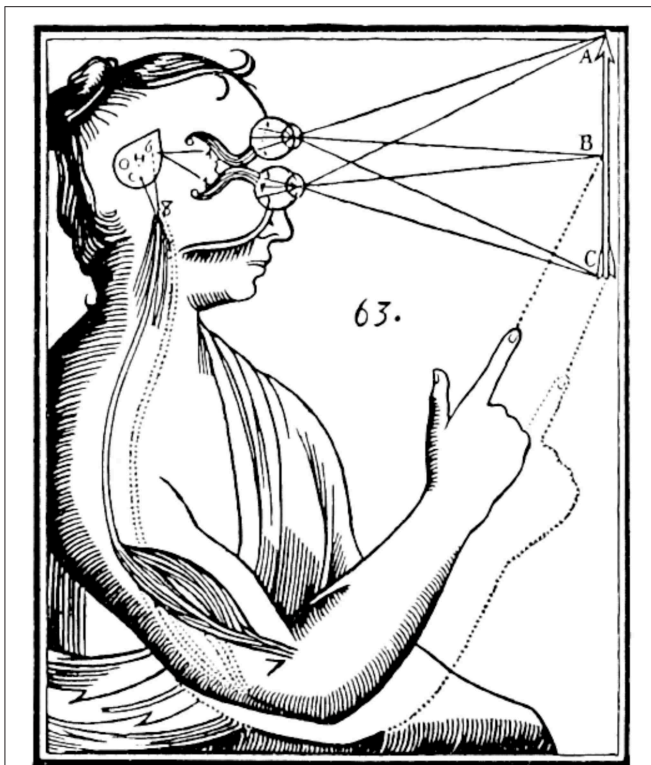


FIGURE 1 | Illustration from Descartes (1662) visualizing the difference between physical objects (upper right corner) and their epistemic representations (upper left corner).

opposed to the objective interpretation of frequentism. But such a characterization is misguided since epistemic probabilities can be subjective, objective and everything in between (Berger, 2006; Williamson, 2010).

Second, the concept of a long-running sequence of random experiments is ill defined in the context of hydrogeology. This is not to deny the relationship that relative frequencies of, say, conductivity values from other sites have on the characterization of a given site. As described above, most Bayesians would agree that such frequencies should always be used when available (Rubin, 1984). The difference to frequentism is not the importance of observed frequencies but the role they play in defining probability. Within the context of hydrogeological site characterization, we would argue that equating those frequencies with probability, and the use of frequentist methods therefore as well, is rather contrived (Renard, 2007).

Third, uncertainty is a property of knowledge and therefore fundamentally epistemic. This fact is often obfuscated by the practice of separating uncertainty into so called aleatoric and epistemic uncertainty, somewhat mimicking the above distinction between physical and epistemic probability (Hoffman and Hammonds, 1994; Helton and Burmaster, 1996; O'Hagan et al., 2006; Gong et al., 2013). Generally speaking, epistemic uncertainty is said to be reducible by collecting more data whereas aleatoric uncertainty is caused by intrinsic randomness

which cannot be further reduced by data. Typically, the latter is illustrated by referring to activities like throwing a die or tossing a coin. The problem with such examples is that both these activities are demonstrably deterministic without any intrinsic randomness (Diaconis et al., 2007). It may be argued that the physical world itself exhibits pure randomness on the quantum level of reality. This notion, associated with the Copenhagen interpretation of quantum mechanics was dominant for the better part of the 20th century but has become marginalized more recently in favor of the Everett (Deutsch, 1999; Sebens and Carroll, 2016) and Bayesian (Schack et al., 2001; Fuchs and Schack, 2013) interpretations. Whatever the case may be, within the context of the macroscopic physical laws relevant to hydrogeophysics, these debates are completely immaterial. On the macroscopic level, the laws of classical physics exhibit no randomness, which can therefore not suddenly manifest in situations that are fully determined by these laws.

The last major reason, for why frequentism does not provide an adequate framework for uncertainty analysis concerns the nature of frequentist inference itself. Following Royall (1997), any statistical inference, as well as any other form of evidential assessment, can be thought of as addressing a series of three questions; (i) What does the evidence say?; (ii) What should I believe?; and (iii) What should I do? Uncertainty itself concerns the second question, i.e., the question of belief but frequentist inference actually does not address this question at all (see **Table 1**). To explain why, let us start at question (i); the question of evidence. Evidence is central to the field of inference but surprisingly difficult to pin down (Feldman and Conee, 1985; Achinstein, 2003; Dougherty, 2011). In general, the evidence of some observations are those aspects of it which justify or lend credence to a hypothesis under question. The most important theoretical advance for the quantification of this notion came in the form of the Likelihood Principle (LP, Barnard et al., 1962), stating that all the evidence of the data is contained in their likelihood. Although some criticism exists, the LP is broadly accepted in the field of epistemology, due to being derived from extremely simple axioms (Barnett, 1999; Good, 2009b; Bandyopadhyay and Forster, 2011; Grossman, 2011). This first step alone, therefore, puts some pressure on frequentist hypothesis testing since it does not meet this criterion. In contrast to that, frequentist estimation, like calibration, parameterization, regression etc.; does not necessarily conflict with this principle. Here, the problem comes in the second step, i.e., using the evidence to justify belief. Frequentism does not deal with belief but uses the evidence, or some proxy thereof, to jump directly to decisions. As outlined above, decisions are made by deriving point estimates, for instance by applying a significance criterion to a p -value or an optimality criterion to some estimation procedure (**Table 1**). Some optimality criteria can be derived on evidential grounds, like the Maximum Likelihood (ML) estimator. While ML estimators are widely used, many other estimators exist, which typically divert from ML by trying to reproduce only certain features of the data or contain some application-specific reasoning (Krause et al., 2005; Pushpalatha et al., 2012; Bennett et al., 2013; Moriasi et al., 2015). This work flow, which forms the blueprint of most inference techniques

TABLE 1 | Comparison of the different concepts used in frequentism and Bayesianism.

		Given the data, what is the evidence?	Given the evidence, what is the belief?	Given the belief, what is the decision?
Frequentism	Hypothesis testing	Sampling distribution		Significance criteria
	Estimation	Likelihood		Optimality criteria
Bayesianism		Likelihood	Bayes' theorem	Utility function

in hydrogeology, is in contrast to the principles of Bayesian inference. First, Bayesian inference meets the LP by using only the likelihood to assess the evidential support of the data. As mentioned above, adherence to the LP is not unique to Bayesianism, but a necessary prerequisite. The most important difference comes in the next step, when Bayesian inference uses Bayes' theorem to compute the belief that follows from the evidence. This step is simply a conclusion from the axioms of probability theory and a number of more recent studies have demonstrated how updating through Bayes' theorem does indeed maximize the epistemic accuracy of an agent (Greaves and Wallace, 2006; Leitgeb and Pettigrew, 2010; Easwaran, 2013). However, the application of Bayes' theorem requires the derivation of the prior probability, which is regularly criticized. We will deal with this question in more detail below and continue for now with the above schematic. Having determined the belief given the evidence concludes the inferential part of the statistical analysis. This, however, leaves open the last step of the analysis, which is to make an informed decision. In above terms, this can mean to decide which parameter θ to use or which hypothesis to accept. In Bayesianism, decision making is done by maximizing the expected utility. Like the two other steps, the decision making through maximizing the expected utility is derived from simple axioms making it the most well-subscribed paradigm in decision making. What is important for us is that this last step is independent of the other two and the specification of the utility function is therefore left to the decision maker. This clear separation of inference and decision makes Bayesianism so relevant to uncertainty analysis.

Combined, reasons like this have led to the strong rise Bayesian methods have seen in many fields like physics (von Toussaint, 2011), biology (Huelsenbeck et al., 2001), environmental science (Clark, 2005), clinical research (Berry, 2006), genetics (Beaumont and Rannala, 2004), psychology (Wagenmakers, 2007), cognitive science (Clark, 2015), and many more. In addition, these reasons have made Bayesianism the leading paradigm in the field of philosophy of science (Howson and Urbach, 2005; Bandyopadhyay and Forster, 2011; Easwaran, 2011a).

2.3. Bayesian Uncertainty Analysis

Having established Bayesianism as the most appropriate framework for uncertainty analysis, we will quickly restate the basic properties of Bayesian inference and prediction. In addition to that, we will demonstrate how the Bayesian framework provides an axiomatically based definition of uncertainty and therefore allows a quantitative assessment of uncertainty reduction as provided by the inference.

2.3.1. Inference

Inference is defined as the process of characterizing a probability function using data. In Bayesian inference, this probability is defined over the possibility space Ω (Kruschke, 2010). As the name implies, Ω is supposed to contain all states that are possible for a given situation, i.e., states whose probability cannot be set *a priori* to zero. To avoid the curse of dimensionality, this space is typically approximated by a parsimonious parametric model, which is fully determined by the specification of its parameters θ . The initial certainty for these parameters θ , is the aforementioned prior $p(\theta)$. The likelihood of each parameter set θ is determined by the data generating process, which in hydrogeology is usually defined through partial differential equations. Combining prior and likelihood, Bayes' theorem can now be used to determine the probability conditioned on the data z called the posterior

$$p(\theta|z) = \frac{p(z|\theta)}{p(z)}p(\theta). \quad (1)$$

The only missing element in Equation (1) is the probability of the data $p(z)$ often called the marginal likelihood or the evidence. However, for most scenarios, this probability is only a normalization constant. It can therefore be omitted and Equation (1) can be computed by normalizing $p(z|\theta)p(\theta)$. This latter form is often used in the literature, since modern sampling methods are versions of the Markov-Chain-Monte-Carlo method, which guarantees this normalization by design.

Looking at this workflow, it becomes clear how Bayesian inference is the general probabilistic framework for the inverse problem with the likelihood being the Bayesian representation of the relevant forward problem.

2.3.2. Prediction

Using the conditioned, i.e., posterior, probability $p(\theta|z)$, the predictive probability for new unobserved data, i.e., predictions, z^* is given by

$$p(z^*|z) = \int_{\Theta} p(z^*|\theta)p(\theta|z)d\theta. \quad (2)$$

Bayesian prediction of new data z^* given the old data z is therefore achieved by marginalizing the posterior probability $p(\theta|z)$ times the predictive probability $p(z^*|\theta)$ of the model as defined by θ .

This influence of the parameterized space of possible worlds Θ on both Bayesian inference and prediction also provides a formal description of two closely related epistemological problems in science; namely the Theory-ladenness of Science and the Duhem-Quine Hypothesis. The former roughly states that scientific

inference is always affected by, usually implicitly held, beliefs of the investigator and is strongly associated with the works of Kuhn (1962) and Feyerabend (1975). The latter is more general and states that scientific inference is always under-determined by our observations and additional assumptions are needed to make sensible conclusions. Next to Duhem (1906) and Quine (1951), important contributions to the development of this notion came, e.g., from Van Fraassen (1980), Laudan (1990), and Stanford (2001). Note that the Bayesian framework does not solve this problem but does make the impact of Θ on the findings transparent. This means that all theoretical presuppositions are contained in the definition of Θ and only influence inference and prediction by virtue of its choice and through the aforementioned equations.

2.3.3. Uncertainty

In addition to properly describing the change of beliefs due to new evidence, Bayes' theorem provides a mathematically rigorous way to characterize the uncertainty represented in a probability distribution as well as the uncertainty reduction achieved during the inference. This is possible due to the intricate relationship between the concept of information in information theory and the way probabilities are updated in Bayesian inference (Ebrahimi et al., 2010). According to Shannon (1948), the information of a particular value of θ_i is given by $I(\theta_i) = -\log_2(p(\theta_i))$, with \log_2 being the logarithm of base 2. Due to this choice, information is usually measured in bits, with other bases simply leading to other units. To characterize the information content of a probability function of a discrete variable, Shannon (1948) introduced the expected value of information

$$H = \int p(\theta)I(\theta)d\theta = - \int p(\theta) \log_2(p(\theta))d\theta. \quad (3)$$

As recounted by Tribus and McIrvine (1971), Shannon initially called this quantity uncertainty but was unfortunately convinced by John von Neuman to use the term entropy instead. In addition, the Shannon entropy only describes the uncertainty with respect to a state of complete ignorance which is implicitly defined in his first and second axioms. Due to this limitation, the Shannon entropy is not suited for Bayesian inference, where updating from arbitrary priors to posteriors is possible. To that end, Equation (3) needs to be amended such that the uncertainty with respect to other degrees of certainty can be described. This extended concept is known as the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) but should be more appropriately called relative entropy/uncertainty or information

$$D_{KL} = \int p(\theta) \log_2 \left(\frac{p(\theta)}{q(\theta)} \right) d\theta. \quad (4)$$

Due to its positive sign, the KL divergence is the *negative relative entropy* of p with respect to q . It therefore belongs to the class of entropy measures meaning that uncertainty is the entropy of a probability function. Such measures have several advantages compared to, say, variance-based measures like Sobol indices that are often used in the literature (Ebrahimi et al., 2010). Compared to the Shannon entropy, this quantity is more fundamental in

several ways. First, D_{KL} can be easily extended to continuous as well as multidimensional variables. Second, D_{KL} describes the uncertainty as expressed in one distribution with respect to another and therefore reduces to the Shannon entropy in the marginal case of a flat q . Finally, D_{KL} connects the concept of information with the updating from the prior to the posterior in Bayesian inference. If p and q are identified with the posterior and prior distribution, respectively, then D_{KL} is a measure for the information and therefore the uncertainty reduction achieved during the inference (Hou, 2005; Tang et al., 2016).

2.4. Challenges of Bayesianism

Although Bayesianism has become such a popular position in the sciences, it has also received its fair share of criticism (Gelman, 2008; Easwaran, 2011b). These criticisms often include rather formal issues like the problem of logical omniscience and the problem of old evidence (Garber, 1983). Others, however, are more relevant to its applicability and should therefore be discussed in the following. Looking at Equation (1), we see that Bayesian inference consists of determining three expression only. Next to the likelihood, which is also often used outside of Bayesian inference, two expressions are peculiar to it and therefore need to be looked at in detail.

First, let us look at the marginal likelihood in Equation (1), which poses the biggest computational problem. Since a direct computation of this expression generally involves multi-dimensional integrals, Bayesian inference was, for a long time, confined to a comparably small number of simple cases. Nowadays, Markov-Chain-Monte-Carlo (MCMC) methods are used (Gelfand and Smith, 1990; Tierney, 1994; Chib and Greenberg, 1995), which circumvent the computation of these often-intractable integrals by sampling directly from the non-normalized posterior. Implementations of MCMC samplers exist as either standalone versions (Lunn et al., 2009; Gelman et al., 2015; Depaoli et al., 2016) or they are implemented for popular languages like R (Martin et al., 2011; Lindgren and Rue, 2015; Denwood, 2016) and Python (Patil et al., 2010; Foreman-Mackey et al., 2013).

Second, let us turn to the prior in Equation (1), which is both a theoretical and computational problem. The theoretical problem, known as the problem of the priors (Osherson et al., 1993), is caused by the limited number of restrictions that Bayesianism puts on a reasoning system to be rational. Taken to the extreme, an agent would be free to believe anything as long as there is no contradiction to the axioms of probability (Romeijn, 2017). In reality, however, common sense dictates that a sound opinion is constrained by additional sources of evidence. Responses from Bayesians to this obvious clash have been mixed, with roughly two extreme camps existing; subjective Bayesianism and objective Bayesianism. Subjective Bayesians stick to the purely theoretical principles and try to mitigate the conflict through expert elicitation, i.e., the application of formal rules to elicit expert opinions on a given topic and turn them into prior distributions (O'Hagan et al., 2006; Albert et al., 2012). On the other end, we find objective Bayesians who claim that additional conditions on rational beliefs, in particular prior beliefs, are necessary. These additional conditions revolve around the principle of equivocation, which means that, if no

evidence favors one possibility, a rational agent should initially equivocate between all possibilities (Williamson, 2010). A more sophisticated version of this idea is the Maximum Entropy (ME) method (Jaynes, 1957a,b), where non-flat priors are possible if some additional, often physically based, constraints or symmetry arguments make some possibilities less credible from the start. In practice, most statisticians, scientist and engineers strive for objectivity but use primarily tried-and-tested approaches, which strike a balance between objectivity and applicability. While the ME method has found some application in hydrogeology (Woodbury and Urych, 1993), most studies rely on using flat or extremely wide distributions over some parameter range derived from the literature (Woodbury and Urych, 1996, 2000; Marchant and Lark, 2007; Diggle and Ribeiro, 2007; Murakami et al., 2010; Laloy et al., 2013; Shi et al., 2014; Geiges et al., 2015; Mara et al., 2017; Hayek et al., 2018). Unfortunately, this procedure is not as harmless as often perceived, since neither are ranges particularly objective nor are flat priors necessarily uninformative (a property which is often used as a proxy for objectivity) (Gelman et al., 2017). In general, current use of objective and uninformative priors in hydrogeology is seriously lacking compared to the standards established in statistics. To give an example of the latter, let us consider so called reference priors. These priors are based on the idea to systematically minimize the impact on the inference. If properly done, the results would then be dominated by the data alone (Bernardo, 1997). Another example is the use of the data themselves to determine the best prior. This method is called empirical Bayes (Carlin and Louis, 2000; Malinverno and Briggs, 2004), since it uses only the empirically available data for the inference. These methods have been very successful in statistics due to their ease of application and apparent objectivity (Valakas and Modis, 2016) but have been heavily criticized on theoretical grounds. The main critique is that by using the data for both the likelihood and the prior, empirical priors—and to a lesser extend reference priors as well—use the data twice. This is a clear violation of the LP, which is generally seen as a necessary element of Bayesianism (Lindley, 1987; Good, 2009b).

A final point concerns Bayesian decision theory. Including this topic under challenges may seem strange given that the Bayesian framework enjoys the best integration with decision theory of any inferential framework. Beginning with the seminal works of von Neumann and Morgenstern (1944) and Savage (1954), Bayesianism has become the de-facto standard in modern decision theory (Berger, 1985; Jeffrey, 1992; Bernardo and Smith, 2000; Robert, 2001; Koehler and Harvey, 2004; Baron, 2004; Parmigiani and Inoue, 2009; Gilboa, 2009). It should consequently be one of its biggest assets. Alas, that is not what we see. Instead, very little effort has been devoted to connecting Bayesian inference with any of the established models from decision theory. To substantiate this pessimism, we simply refer here to the review of Tartakovsky (2013), who gives an excellent overview of this topic yet still fails to find more than a handful of studies, which apply Bayesian decision theory to hydrogeology. As explained above, this lack is of little consequence for the specific topic of uncertainty analysis. It is, however, clearly a challenge for Bayesian inference in general. Looking at **Table 1**,

current practice means to only implement the first and second step out of all three. While not the focus of this manuscript, we opine that a full appreciation of the Bayesian framework will only become a reality once all its elements are common knowledge and regularly applied.

3. WHAT KIND OF UNCERTAINTIES ARE WE TALKING ABOUT

To organize the different forms and sources of uncertainties, a wide range of often conflicting notations is used in the literature (Hoffman and Hammonds, 1994; Hofer, 1996; Walker et al., 2003; Brown, 2004; Carrera et al., 2005; Refsgaard et al., 2007; Kwakkel et al., 2010; Biegler et al., 2010; Refsgaard et al., 2012; Guillaume et al., 2012; Tartakovsky, 2013; Caers et al., 2014; Bond, 2015; Enemark et al., 2019). In the following, we are going to describe and contextualize these notations using the formalism established above.

3.1. Understanding Uncertainties Using the Framework of Bayesianism

A common approach is to separate uncertainties into epistemic and aleatoric uncertainties (Kiureghian and Ditlevsen, 2009; Beven and Young, 2013; Bond, 2015). As already explained above, all uncertainties in Bayesianism are necessarily epistemic and aleatoric uncertainties simply do not exist in this framework. In our opinion, this distinction can make sense for practical applications because, due to often highly non-linear processes, reality always has a tipping point—sometimes sudden, sometimes more gradual—beyond which the additional collection of data becomes too costly to be reasonably entertained. Identifying such tipping points is important for any engineering task in order to estimate sensible directions for the additional data gathering. This notion is often implicitly confirmed by authors who otherwise seem to argue for the physical presence of aleatoric uncertainty within macroscopic phenomena. Fox and Ülkümen (2011) for example admit as much when saying "Aleatory uncertainty is attributed to outcomes that for practical purposes cannot be predicted and are therefore treated as stochastic". Another situation, where the use of aleatoric uncertainty seems justified, is in the presence of so-called statistical uncertainty (Beven and Young, 2013). This means that the statistical variation of a given population, say, the conductivity values of an aquifer, puts an "inherent" limit on how much the uncertainty can be reduced. From a Bayesian perspective, such reasoning is simply false. Kiureghian and Ditlevsen (2009), for instance, articulate this problem when stating that "The distinction between aleatory and epistemic uncertainties is determined by our modeling choices." To illustrate this point and get at the root of this prevailing misunderstanding, let us look at the already used example of conductivity values of an aquifer. Using, e.g., the whole aquifer as the population, there is indeed an intrinsic limit of how much the uncertainty can be reduced through sheer data collection. It would, therefore, seem that this uncertainty fits the above definition of being aleatoric. However, as Kiureghian and Ditlevsen (2009) pointed out, there is no metaphysical reason

to model the whole aquifer as a single statistical population. If enough data are collected, the aquifer can easily be split up into, say, its hydrofacies, each of which would now have a much-reduced statistical uncertainty. This process of fine graining our statistical model, depending on the amount of data, can be repeated ad infinitum, which shows that no statistical variation is ever intrinsic to reality but only determined by our model. The last class of examples, which are often used to demonstrate aleatoric uncertainty are actually cases of deep or Knightian uncertainty (Fox and Ülkümen, 2011; Beven and Young, 2013). Deep uncertainty is certainly an important and even dominant form of uncertainty in everyday situations. However, as explained above, it should be modeled by the Dempster-Shafer framework and treating it within a probabilistic context is necessarily error prone. In summary, aleatoric uncertainty, as used in the literature, is an inconsistent mixture of several distinct concepts, with varying levels of usefulness.

The second differentiation is to separate uncertainties into inferential and predictive uncertainties. Within Bayesianism, these uncertainties are simply defined via Equations (1) and (2), respectively. This means that the inferential uncertainty is the D_{KL} of the posterior vs. the prior distribution, whereas the predictive uncertainty is given as the D_{KL} of the predictive distribution with the data vs. without them.

Another important differentiation is to separate uncertainties into input and parametric uncertainties (Refsgaard et al., 2012; Tartakovsky, 2013). Within the context of Bayesianism, input uncertainty is simply the uncertainty that is passed down from receiving nodes in a Bayesian network. This means, the input uncertainty of a given node is the combined uncertainty of its parent nodes. On the other hand, parametric uncertainty is uncertainty in the parameters θ of the used parametric model and is therefore identical to the inferential uncertainty given by Equation (1).

The last concept to be discussed is the topic of structural and conceptual uncertainty. Both these terms are used in sometimes overlapping and sometimes conflicting ways (Refsgaard et al., 2012; Tartakovsky, 2013; Enemark et al., 2019). In general, it is not even clear whether these two terms do differ in meaningful ways. From a Bayesian perspective, both refer to the necessary approximation of the possibility space Ω by a lower-dimensional parametric subspace Θ . Since describing Ω , e.g., the conductivity field of a real-world aquifer, in full detail is both impossible, due to the scarcity of data, as well as numerically intractable, such approximations will always be necessary. From our perspective, it can be beneficial to use two different terms in order to distinguish between the uncertainty expressed between the different parametric models, called structural models in the following, and the uncertainty being expressed within any such given structural model. In the following, we will focus on the latter and use the term structural uncertainty to describe this category.

Having used the Bayesian framework to put the different forms of uncertainty into a proper context, we will finish this section by ranking these different forms according to their relevance for hydrogeological modeling. Since such a ranking is strongly dependent on the context, we will shortly

discuss each form individually. First, parametric uncertainty is doubtless one of the dominant forms regardless of the situation. Bayesian inference is perfectly suited to handle it, provided that proper priors are provided. Structural uncertainty is also very important, due to the often-pronounced spatial pattern of many aquifers. Handling this form of uncertainty as well as structural modeling in general is comparably underdeveloped. Instead, structural uncertainty is often recast as a form of process uncertainty, in particular in the case of transport processes (Neuman and Tartakovsky, 2009). While such alternative process models may be important for pure modeling purposes, we are very skeptical about their use in uncertainty analysis. In addition to this, the parameters of such models are often pure convenience parameters and they are difficult or impossible to condition on point measurements. As mentioned above, conceptual uncertainty is not going to be the focus of this paper. Instead, we are going to describe in the following the range of structural models, which are used in hydrogeology and try to identify the most relevant paradigms. Finally, input uncertainty is technically not part of hydrogeology since the uncertainty is passed down from the parent nodes of the Bayesian network. This is quite apparent in case of groundwater recharge, where the input uncertainty is the accumulated uncertainty of the meteorological, land surface, and soil compartments of a complex hydrological model.

3.2. On the Role of the Structural Model

At this point, we have identified structural and parametric uncertainty as the most relevant categories of hydrogeological uncertainty. While in theory both are of similar importance, the way to handle them in practice is very different. Parametric uncertainty can be reduced by collecting more data and is primarily a question of data acquisition. On the other hand, structural uncertainty is connected to the model for the subsurface heterogeneity itself. It is therefore much harder to quantify, which has wide ranging ramifications for uncertainty analysis. To investigate this problem in more detail, we will first present and discuss the most common paradigms for generating subsurface structures.

The most famous of these paradigms is the Gaussian process (GP) model, also known as Multivariate Gaussian or a Gaussian random field (Rasmussen and Williams, 2006). In its basic form, a GP is a very parsimonious model and can therefore be applied in situations where only few data are available, as often the case in hydrogeology. In addition, a GP is hierarchical by nature and therefore scales well with the amount of data available, i.e., the dimensionality of Θ can be arbitrarily matched with the amount of data available for the inference (Gelfand and Schliep, 2016). However, employing a GP as the structural model for a conductivity field makes a number of strong assumptions about the properties and characteristics of the conductivity field, some of which have been strongly criticized (Gómez-Hernández and Wen, 1998; Zinn and Harvey, 2003; de Marsily et al., 2005; Linde et al., 2015). The most important of these criticisms concerns the inability of GPs to reproduce long-ranging high-conductivity structures, which are reported to exist in many real-world aquifers (Abelin et al., 1991; Zheng and Gorelick, 2003;

Kerrou et al., 2008). As a result, using a GP as the structural model for an aquifer will lead to a failure to (i) detect the presence of such features as well as (ii) to predict certain behaviors of, say, break-through curves (Heße et al., 2015; Savoy et al., 2017).

To improve on some of the limitations of GPs, truncated pluri-Gaussian models have been developed (Le Loc'h and Galli, 1997; Galli et al., 1994; Emery, 2004; Armstrong et al., 2011). These methods can be seen as an extension of the Gaussian paradigm by embedding Gaussian SRFs within a larger hierarchical framework. Here, hierarchy means that the Gaussian SRFs are used to create a larger spatial structure. This larger structure is typically representing distinct hydrogeological units, like hydrofacies or lithofacies. Despite some success in recent years (Emery, 2007; Mariethoz et al., 2009; Serrano et al., 2014), their overall geological realism has remained limited, and alternative paradigms have continued to attract considerable attention.

Surface-based modeling is one of these other paradigms (Caumon et al., 2009), often implemented in terms of implicit surfaces (Calcagno et al., 2008; Chilés et al., 2004). Although not particularly suited for modeling intricate structures and low scale heterogeneity, this paradigm provides a realistic representation of the main geological structures and allows to integrate a good range of information, including geological data (contact points, surface orientations, faults) but also information coming from geophysical surveys.

Geological structures can also be represented as geometrical objects using object-based models (Koltermann and Gorelick, 1996). Object-based methods are mainly based on geometrical considerations about the expected shapes.

Process-based methods on the other hand arguably provide the most realistic representation of real world structures (Koltermann and Gorelick, 1996). Software implementations of this paradigm are available both as commercial and academic releases, and for unconditional simulations the computational costs are acceptable. Nevertheless, as is the case for object-based methods, process-based methods have difficulty in honoring all the observed conditioning data, which puts some limit on their applicability.

The last, modeling paradigms discussed here are the multiple-point statistic (MPS) based techniques (Guardiano and Srivastava, 1993; Strebel, 2002; Mariethoz and Caers, 2014). In contrast to geostatistical methods based on two-point statistics, these methods allow to reproduce more realistic structures, which better represent important features observed in real world aquifers like connectivity. Computational costs remain relatively high compared to other paradigms. Nevertheless, MPS simulation algorithms intrinsically honor all the observed conditioning data, and the flexibility of the technique allows for the incorporation of information coming from different sources, in a straightforward way.

Together, these paradigms form the basis of most of the models used for generating subsurface heterogeneity. In addition to the properties already mentioned, they also differ in how much they are amendable to a Bayesian framework (Table 2). The Gaussian paradigm is very well-suited, since most of its parameters are simple statistics of typical

hydrogeological variables (e.g., conductivity, porosity, storativity etc.). Consequently, there exists a direct way to derive these parameters from real-world measurements. The truncated pluri-Gaussian paradigms score lower in this regard since a crucial feature of this method is the derivation and application of the truncation rule. These rules usually ought to come from expert elicitation but little to no guidelines exist on how to formalize this process. The surface-based paradigm fares much better in this regard since the inference of subsurface structures is able to draw on observable features of the subsurface. In fact, the possibility to frame these paradigms within a Bayesian framework was already explored by Wellmann et al. (2018) and de la Varga et al. (2018). Next is the object-based paradigm, where quite often objects are drawn based on purely geometrical considerations. It is therefore not straightforward to use this paradigm in a Bayesian framework. Nevertheless, object-based methods can profit from the statistics about the morphological attributes of real-world geological objects (Gibling, 2006; Colombero et al., 2012). Process-based models, on the other hand, are built on a plausible physical model by mimicking the geological genesis of the subsurface. However, the parameters of these models are usually pure convenience parameters making the derivation of prior PDFs subjective. The last paradigm discussed above is MPS, which is not without problems from a Bayesian perspective. However, its overall aptitude is arguably higher than the two former paradigms. First of all, realizations of MPS models are easy to condition on point measurements, which is an important feature. Looking at the generating mechanism itself, we see that its parameters are simple convenient parameters that are not directly connected to physical principles. Despite this clear drawback, MPS realizations are based in training images, which means the method is based on observable features of the subsurface.

In conclusion, we would argue that two of the above presented paradigms stand out as the most viable candidates for Bayesian uncertainty analysis. First, the Gaussian paradigm, which scores high on almost every metric except geological realism (see Table 2). While this is only one point of many, it is arguably the most important one. At the same time, GPs score so high on the other metrics, in particular its wide use, that it should not be excluded. Furthermore, GPs are good candidates for subdomain models within a larger hierarchical modeling framework and can consequently form an important component of a larger and more realistic framework. The second relevant paradigm is MPS, which, in some sense, can be placed on the other end of the spectrum of the paradigms presented here. This is to say that models based on MPS have a high degree of geological realism but suffer from a lack of ready-made software tools, that they are comparably unknown to practitioners and that they are computationally demanding. However, both paradigms have the ability to incorporate a variety of data sources and the generation of heterogeneous structures is based on observable characteristics of the subsurface. In addition, if the MPS framework is used to generate only the categorical SRF of the different hydrogeological units, it can be seamlessly integrated with the Gaussian paradigm. With these two candidate paradigms in mind, we will continue our discussion on Bayesian uncertainty analysis.

TABLE 2 | Overview of the different paradigms for sub-surface structure generation applied in hydrogeology.

	Gaussian process	Truncated pluri-Gaussian	Surface-based models	Object-based models	Process-based models	Multi-point statistics
Software tools	Very high	Low	Low	Low	Low	Low
Realism	Very low	Low	Moderate	High	Very high	Very high
Comp. costs	Low	Moderate	Moderate	High	High	Very high
Math. theory	Very high	Moderate	Moderate	Very low	Very low	Very low
Versatile data	High	High	Very high	Very low	Low	High
Bayesian	High	Moderate	High	Low	Low	Moderate

4. TOWARD A DATA-DRIVEN UNCERTAINTY CHARACTERIZATION

As we have discussed earlier, Bayesianism is not free of practical and theoretical problems. Of these, the question of prior derivation was presented as the most pressing, with no universally accepted guideline for prior derivation existing in the literature. In this last part of the manuscript, we are going to lay out the current challenges and explain a possible solution by using data-driven priors.

4.1. Putting Prior Derivation on Solid Grounds

In this paper, we want to make the case for combining aspects of objective Bayesianism with frequentist reasoning (Rubin, 1984; Bayarri and Berger, 2004; Little, 2006). While we fully agree that objectivity is an important goal in any scientific or engineering enterprise, we do not think that this needs to be achieved by minimizing the impact of the prior as often argued. On the contrary! Priors can have both a strong impact on the analysis and being objective. This can be achieved if these priors are derived from well-defined empirical frequencies (see, e.g., Li et al., 2017 for a rare example for hydrogeology). Hydrogeological variables like conductivity and porosity are physical properties or can at least be soundly derived from them (Di Palma et al., 2017). As a result, their parametric and structural uncertainties can be calibrated against observed frequencies. This feature, that the parameters of stochastic hydrogeology relate to physical quantities, makes this mixture of Bayesian analysis and frequentist reasoning the natural choice for prior derivation. This notion has, e.g., been voiced by Gelman (2008) when saying that a hardcore Bayesian is someone “who would apply Bayesian methods to *all* problems” whereas a reasonable person “would apply Bayesian inference in situations where prior distributions have a physical basis or a plausible scientific model.”

Such basis should be a formalized knowledge base, i.e., a database $DB = (X_i, Z_i)_{i \leq n}$, containing all investigated and cataloged cases. In the field of hydrogeology, these cases should be identified with investigated and cataloged sites. Furthermore, n is the number of these sites, Z_i are the measurements of the target variable at each site i and $X_i = (X_i^1, \dots, X_i^m)$ is a vector containing the m cataloged characteristics/features of each site. In hydrogeology, such characteristics may include latitude, longitude, climatic properties, rock type, environment type, physiographic properties etc.

4.1.1. Prior Derivation Using Machine Learning

To derive prior distributions from an established database DB , a large variety of techniques can be used, including machine learning. Here, for brevity, we illustrate only a single example of these techniques, namely supervised feature learning (Figure 2). As described above, the used database DB contains, next to the measurements, the features associated with each site. Feature learning would allow to determine the most predictive features as well as the functional dependency for any give site. The result would be a function that maps the observable features of a site to the distribution of its conductivity values (or any other target variable). Common machine learning methods that can tackle such tasks include Random Forests (Breiman, 2001; Segal and Xiao, 2011), Gradient Boosting Trees (Elith et al., 2008), or Bayesian Additive Regression Trees (Chipman et al., 2010; Pratola et al., 2014; Kapelner and Bleich, 2016).

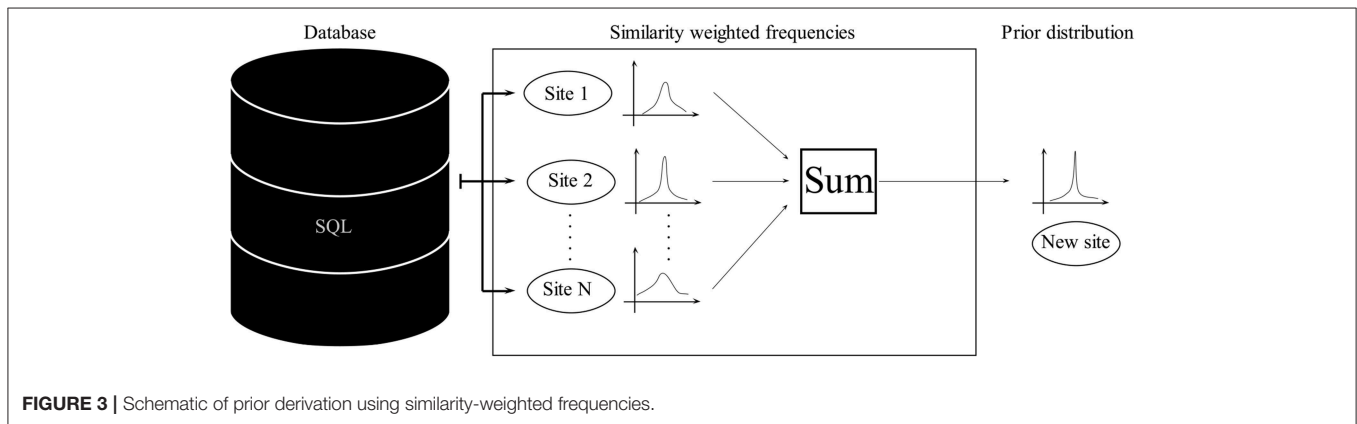
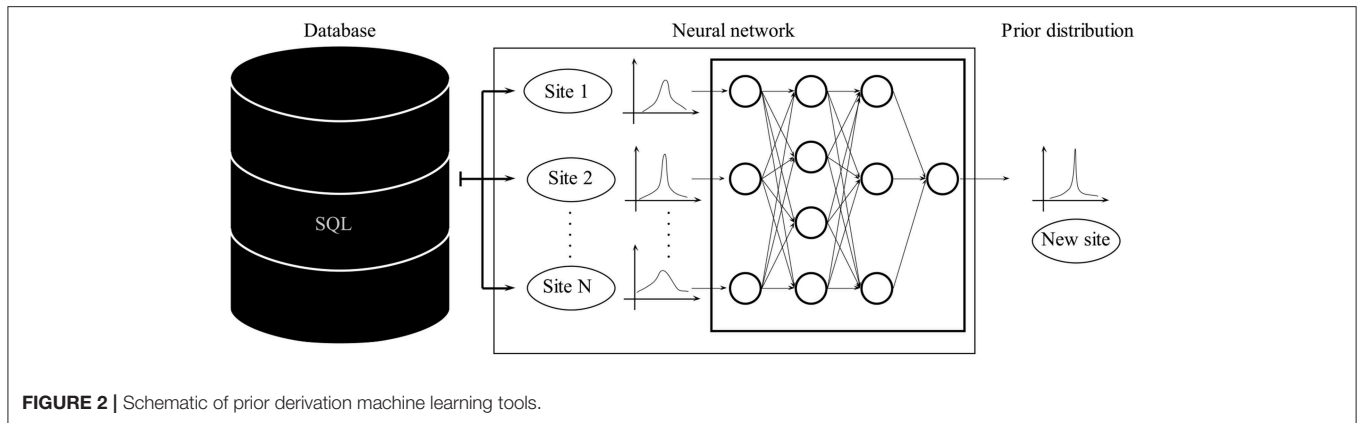
4.1.2. Prior Derivation Using Similarity-Weighted Frequencies

Machine learning tools like neural networks are well-established in data science. However, they can be extremely computationally expensive, need large amounts of data and the resulting models are notoriously hard to interpret.

Recent developments in the field of epistemology have provided a sound mathematical procedure to formalize the intuitive notion about how different cases can be made amenable to frequentist reasoning and therefore provide the basis for combining aspects of frequentism and Bayesianism (Billot et al., 2005; Gilboa et al., 2006, 2010). The basic idea behind it can be easily illustrated (Figure 3). Given a similarity function s , more on this later, the probability of the target variable at a new site $n + 1$ can be expressed as

$$p(Z_{n+1}^s) = \frac{\sum_{i \leq n} s(X_i, X_{n+1}) p(Z_i)}{\sum_{i \leq n} s(X_i, X_{n+1})}. \quad (5)$$

Since Equation (5) strongly depends on the specification of the similarity function $s(Z_i, Z_j)$, Gilboa et al. (2010) propose a simple *best fit* approach to find such a function, as the $s_{\text{opt}}(Z_i, Z_j)$ that best explains the given database. To handle the curse of dimensionality, a parametric model should be used. Following again Gilboa et al. (2010), we use an exponential model here for



demonstration

$$s(X_i, Y_j) = \exp \left(- \sqrt{\sum_{j \leq m} w_j (x^j - x'^j)^2} \right). \quad (6)$$

At this point, the task of finding the prior distribution for a target variable has first been transformed into finding s , and then—by virtue of a parametric model—transformed into finding the appropriate weights $\mathbf{w} = (w_1, \dots, w_m)$. Gilboa et al. (2010) propose a simple *best-fit* approach, i.e., finding \mathbf{w}_{opt} such that the sum of squared errors between all elements in the database is minimized. The principle behind this similarity function and the way it is used to estimate probability is identical to kernel density estimation but extended to the feature space of our database. The choice of the name similarity function, instead of kernel, was motivated by Gilboa et al. (2010) to emphasize the epistemic nature of the procedure.

4.1.3. Prior Derivation Using Bayesian Hierarchical Modeling

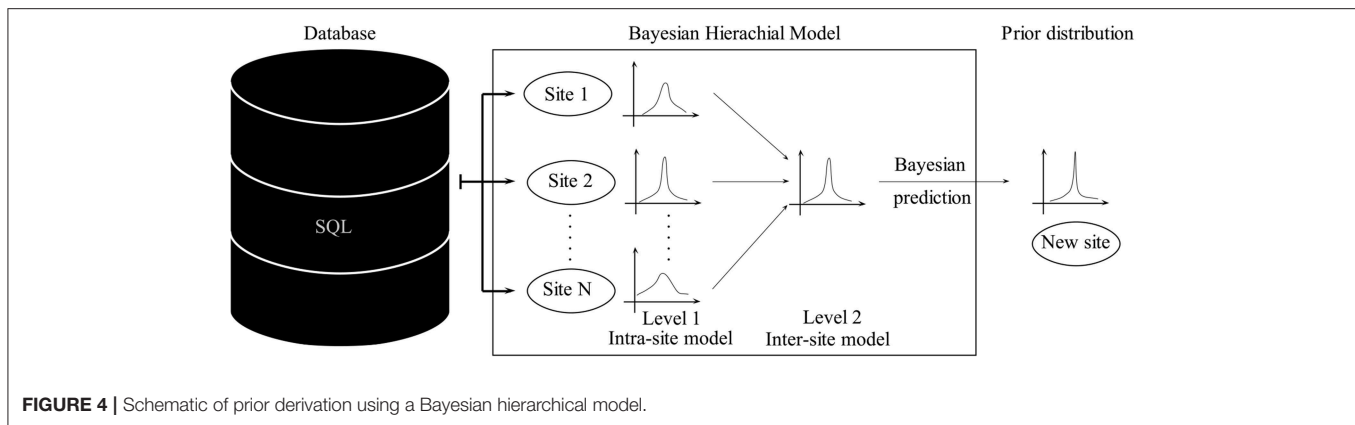
Although the above procedure is axiomatically elegant, it is not without flaws. One problem is that the used database might be too small to contain a relevant number of observations. Basing the prior on such a database would lead to attribute a zero probability to structures that have not yet been cataloged. Another problem is the incorporation of data on different scales like summary statistics of certain sites.

Such challenges can be handled by Bayesian hierarchical modeling, which provides a natural way to partially pool data from different sites by incorporating the structural dependencies between the data points. In hydrogeology, it is not immediately clear how data from different sites can be combined and jointly used for the inference. Hierarchical models facilitate a partial pooling by first assimilating data from each site independently and, in a next step, modeling each site as elements of a population of sites (**Figure 4**). Bayesian statistics is easy to adapt to such schemes by modeling the parameters of each site as being conditional on the statistics of the upper levels.

One drawback of Bayesian hierarchical modeling, compared to a similarity weighted approach, is that the prior distribution of a new, as of yet, unexplored site, is simply a random sample from all possible sites and therefore has a comparably large uncertainty. Narrowing down the uncertainties could be done by narrowing down the number of used sites by restricting the analysis to similar sites, only. Yet, this would necessitate the existence of a huge database, which we currently do not have (Cucchi et al., 2019).

4.2. Hydrogeological Data Science in the Context of Bayesian Uncertainty Analysis

At this point, it should become clear that, irrespective of the details of the aforementioned approaches, the performance of data-driven methods is not only determined by the sophistication



of the used method but is equally dependent on the amount and kind of data being used (Halevy et al., 2009). This means that a simple algorithm, having a lot of data, can easily outperform a highly complex one, with only a modest amount of data. Increasing the amount of available data is consequently of similar importance compared to the development of ever better algorithms.

In general, data generation in hydrogeology is quite costly compared to, say, hydrology, meteorology or land-surface modeling. To counter such challenges, much has been invested in the development and deployment of cost-effective methods for subsurface characterization. As a result, the total amount of data being collected every year is quite substantial. However, collecting these data and making them available to practitioners remains difficult. To demonstrate why this is a problem, we can use, e.g., the schematic proposed by Rogati (2017). Using the well-known hierarchy of human motivation (Maslow, 1943), Rogati (2017) promotes a hierarchy of needs in data science. According to this schematic, hydrogeology has focused most of its effort only on the first level of this hierarchy, where she puts data collection, data storage and data transformation. However, already at the second level, where she puts routines, protocols and infrastructure for moving and storing the data, hydrogeology is comparably underdeveloped. Given that the hierarchy proposed by Rogati (2017) has 6 levels in total, we can confidently state that hydrogeology has a long way to go before having a viable ecosystem for modern data-driven analysis.

This is not to say that there have not been efforts to provide standardized procedures for sharing and storing data in hydrogeology (Boisvert and Brodaric, 2011; Brodaric et al., 2018; Wojda et al., 2010). However, these data collection initiatives focus almost exclusively on *indirect* measurements like hydraulic heads or are restricted to some specific measurement sites. In contrast, the efforts to collect *direct* hydrogeological measurements have been so far been rather modest. For example, the largest open-access databases on the topic of hydraulic conductivity is the World-Wide Hydrogeological Parameter Database (WWHYPDA, Comunian and Renard, 2009), which contains a little bit over 20,000 measurements from approximately 50 different sites. This does only reflect a tiny fraction of the total amount of

data that has been collected on this topic and is not even close to what anyone would label as big data. This situation means that the field of hydrogeology is currently seriously under equipped for the deployment of any data-driven method in general and the derivation of data-driven priors in particular.

Focusing on the topic of data-driven priors, we can state that only a minimum number of tools currently exists. The WWHYPDA provides a modicum of data on hydraulic conductivity, which can be used to determine parametric uncertainty for a given site (Cucchi et al., 2019). However, the amount of data for other variables (e.g., porosity) is much lower and currently not sufficient for use in a Bayesian context. Concerning structural uncertainty, the situation is even less promising. Since the measurements cataloged in the WWHYPDA do currently not contain spatial coordinates, it is not possible to determine simple two-point statistics like a variogram or covariance function. This means that even for a simple paradigm like Gaussian SRFs, there are no databases from which prior distributions for, say, the parameters of a variogram function can be derived. It should come as no surprise, that the situation for more complex structural models is no better. In this last portion of the paper, we would therefore like to present a list of challenges as well as possible solutions for the field of Bayesian uncertainty analysis.

4.2.1. Parametric Uncertainty

Currently, the amount of data contained in WWHYPDA barely allows one to tackle the challenges of parametric uncertainty. The modest amount of data represented in WWHYPDA is probably caused by the lack of user contributions and/or policies that encourage the systematic and central publication of hydrogeological measurements. Probably, the best way to quickly populate databases like WWHYPDA with a sufficient amount of data would be to incorporate data spread in regional/national repositories, but also strongly interact with boards of the main scientific journals and set up a tighter interaction between authors and open data collection initiatives.

4.2.2. Structural Uncertainty (Gaussian Process Paradigm)

A Gaussian process is the simplest paradigm that can be used to describe structural uncertainty, with spatial structures defined by two-point statistics through variograms or covariance functions.

To the best of the authors' knowledge, there is no widely available open-source database, which provides practitioners with a catalog of either estimated variogram functions or measurements that allows one to estimate them.

The abilities of the WWHPDA could be expanded to that purpose with only a modest amount of effort. Compared to its current implementation, only the coordinates of the measurements need to be added as a feature.

4.2.3. Structural Uncertainty (Multiple-Point Statistics Paradigm)

In the text above, multiple-point statistics was identified as the overall most promising paradigm to realistically represent structural uncertainty. Some efforts have been made to share repositories of training images [see for example (2014), companion site of the book Mariethoz and Caers, 2014], in some cases focusing on some specific environments (Pushpalatha et al., 2008). In addition, other efforts were made to create database of analogs, with initiatives mainly sponsored by oil companies like the SafariDB (2019a), the Sedimentary Analogs Database and Research Consortium (2019b), the Fluvial Architecture Knowledge Transfer System (Colombera et al., 2012), CarbDB (Jung and Aigner, 2012), and WODAD (Kenter and Harris, 2006). In these cases, the access to the full functionalities of the database is very often restricted to the partner institution/companies. Therefore, when talking about open-access databases, the available resources are quite limited. In practice, a set of TIs that can even be remotely called representative of earth's subsurface structures simply does not exist, and the efforts made by consortium sponsored or private companies are often governed by very restrictive access policies. As a result, the task of building up an open-access knowledge base for Bayesian structural uncertainty analysis, or any other data-driven modeling efforts, has to start from scratch. Within the scope of this manuscript, we can neither detail the specific architecture of such a knowledge base nor explain the necessary steps to create one. We will, however, try to formulate a set of desiderata that such a data base should meet.

First, the data base should contain a representative sample of the structures encountered in the subsurface. Such a desideratum may sound obvious w.r.t. any statistical analysis. It does, however, need special attention due to being somewhat vague and elusive. For example, particularly in the case of three-dimensional case studies, TIs could come from high-resolution reconstructions of aquifer analogs (Bayer et al., 2011; Comunian et al., 2011; Bayer et al., 2015), but also be the result of a more or less complex simulation with object-based or process-based methods. Therefore TIs, in particular within a Bayesian context, can represent very different entities depending on how they were created.

Second, to derive prior distributions, the most predictive features of the cataloged sites must be reported as well. This desideratum is again somewhat weak, since it is not a priori clear what features of a site are most predictive of its subsurface structure.

In case a database meeting these desiderata becomes successful, the algorithm for prior derivation has to be adapted to the specifics of the MPS paradigm. As already discussed above, recasting this paradigm in a Bayesian framework is not straightforward, since the parameters for generating random realizations are derived *ad-hoc* and cannot, in general, be exchanged between different workflows. So, instead of deriving prior distributions over some parameters, the prior distribution should be defined over the TIs themselves. In this framework, one could provide a given prior distribution to each TI, for example based on the ranking procedure proposed by Pérez et al. (2014) computed using a portion of the available data. Then, uncertainty could be assessed by distributing the number of realizations for each TI proportionally to the prior computed in the previous step.

Probably, embedding or strongly connecting a database of TIs within a structure like WWHPDA, as argued by Comunian and Renard (2009), would improve its usefulness, because facies codes of categorical TIs could be directly linked to parameter distributions. Moreover, if included in the WWHPDA structure, TIs could be also organized in a more efficient and flexible way within the provided catalog of hydrogeological environments.

5. CONCLUSIONS

In this manuscript, we made the case for a unified data-driven framework for hydrogeological uncertainty analysis. Following Pearl (1988b), we attempted this by first identifying the most suitable theory for such a framework, motivating its use, explaining its properties, identifying the current challenges, showing what kind of approximations needed to be made to make the framework viable, and finally detailing a road map to fill the gaps which currently exist.

As the ideal theory for such a unified framework, we have argued for Bayesianism. This was done by contrasting the Bayesian framework with its most relevant competitors. In the realm of uncertain reasoning, these competitors include fuzzy as well as Dempster-Shafer reasoning, whereas in the realm of probabilistic reasoning, the main competitor is the frequentist framework. While all of these frameworks have their merits, the Bayesian framework is the only one combining a sound epistemology with wide-spread use and application. We then explained the main features, as well as challenges, of this framework and how it relates to the specifics of the field of hydrogeology.

With this foundation, we proceeded to contextualize the different forms of uncertainty, used in the literature. Of these forms, we identified structural and parametric uncertainty as the two most relevant. Parametric uncertainty is mainly a question of data collection, data preparation, and data dissemination. In

contrast, structural uncertainty is equally a conceptual challenge, i.e., the hydrogeological community is still lacking a paradigm for generating subsurface structures, which is widely accepted and used. To make this point, we reviewed the most common of these paradigms and discussed their most salient features. Special attention was put on their compatibility with a Bayesian framework. We tentatively identified two paradigms as the most relevant: Gaussian multivariate random fields as well as multiple-point statistics. While the Gaussian paradigm is mathematically elegant, it suffers from a lack of geological realism. On the other hand, the multiple-point statistics paradigm is very promising due to its conceptual clarity. To provide the data for the needed prior distribution, training images are the natural choice.

Finally, we make the point that the field of hydrogeology needs to increase its efforts in order to provide a large open-access databases with direct hydrogeological measurements. The examples given here are derived from specific applications, i.e., Bayesian uncertainty analysis, but the general point is independent of such concerns. As long as virtually all efforts are focused on the measurements themselves, we remain stuck on the first level of the hierarchy described by Rogati (2017). Without investing at least, a modicum of work into reaching the second level, all the advances of modern data science will remain restricted to extremely myopic and small data sets, at best,

and completely out of reach, at worst. From our perspective, the reasons for the inability of our community to “move up” remains somewhat of a mystery. After all, managing and maintaining a database is easy compared to the sophisticated network of measurement operations that are managed worldwide, and a minor redirection of resources from Level 1 to Level 2 of the schematic depicted by Rogati (2017) would make a big difference. In fact, we can hardly think of a more impactful way of how to spend one’s money when trying to improve hydrogeological data science. We therefore close this paper with a call for such an initiative.

AUTHOR CONTRIBUTIONS

As the main author, FH has worked on every section of the article. AC’s input was focused on the last portion of the manuscript dealing with the models for generating subsurface structures and the challenges of databases. SA provided overall guidance for the development of the main ideas presented here.

FUNDING

For the finalization of this manuscript, FH was financially supported by the Deutsche Forschungsgemeinschaft via grant HE 7028/2-1.

REFERENCES

- (2014). *Multiple-Point Geostatistics: Stochastic Modeling With Training Images*. Companion site of the book by G. Mariethoz and J. K. Caers section “TRAINING IMAGE LIBRARY.” Available online at: <http://www.trainingimages.org/training-images-library.html> (Online; accessed March 26, 2019).
- (2019a). *SafarIDB: The Geological Outcrop Database*. Available online at: <https://safaridb.com/> (Online; accessed March 26, 2019).
- (2019b). *The Sedimentary Analogs Database and Research Consortium*. Available online at: <http://sedimentaryanalogsdatabase.com/> (Online; accessed March 26, 2019).
- Abelin, H., Birgersson, L., Moreno, L., Widén, H., Ågren, T., and Neretnieks, I. (1991). A large-scale flow and tracer experiment in granite: 2. results and interpretation. *Water Resour. Res.* 27, 3119–3135. doi: 10.1029/91WR01404
- Achinstein, P. (2003). *The Book of Evidence*. Oxford Studies in the Philosophy of Science. Oxford University Press.
- Albert, I., Donnet, S., Guihenneuc-Jouyau, C., Low-Choy, S., Mengersen, K., and Rousseau, J. (2012). Combining expert opinions in prior elicitation. *Bayesian Anal.* 7, 503–531. doi: 10.1214/12-BA717
- Armstrong, M., Galli, A., Beucher, H., Loc’h, G., Renard, D., Doligez, B., et al. (2011). *Plurigaussian Simulations in Geosciences*. Berlin; Heidelberg: Springer.
- Bandyopadhyay, P. S., and Forster, M. R. (eds.) (2011). *Philosophy of Statistics*, Vol. 7 of *Handbook of the Philosophy of Science*. Amsterdam: Elsevier.
- Barhillel, M. (1980). The base-rate fallacy in probability judgements. *Acta Psychol.* 44, 211–233. doi: 10.1016/0001-6918(80)90046-3
- Barnard, G. A., Jenkins, G. M., and Winsten, C. B. (1962). Likelihood inference and time series. *J. R. Stat. Soc. Ser. A* 125, 321–372.
- Barnett, V. (1999). *Comparative Statistical Inference, 3rd Edn*. Chichester: Wiley.
- Baron, J. (2004). “Chapter: Normative Models of Judgment and Decision Making,” in *Blackwell Handbook of Judgment and Decision Making: Handbooks of Experimental Psychology*, eds D. J. Koehler and N. Harvey (Malden, MA: Blackwell Publishing Ltd), 19–37.
- Bayarri, M. J., and Berger, J. O. (2004). The interplay of bayesian and frequentist analysis. *Stat. Sci.* 19, 58–80. doi: 10.1214/08834230400000116
- Bayer, P., Comunian, A., Höyng, D., and Mariethoz, G. (2015). High resolution multi-facies realizations of sedimentary reservoir and aquifer analogs. *Sci. Data* 2:150033. doi: 10.1038/sdata.2015.33
- Bayer, P., Huggenberger, P., Renard, P., and Comunian, A. (2011). Three-dimensional high resolution fluvio-glacial aquifer analog: Part 1: Field study. *J. Hydrol.* 405, 1–9. doi: 10.1016/j.jhydrol.2011.03.038
- Beaumont, M., and Rannala, B. (2004). The Bayesian revolution in genetics. *Nat. Rev. Genet.* 5, 251–261. doi: 10.1038/nrg1318
- Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., et al. (2013). Characterising performance of environmental models. *Environ. Model. Softw.* 40, 1–20. doi: 10.1016/j.envsoft.2012.09.011
- Berger, J. (2006). The case for objective bayesian analysis. *Bayesian Anal.* 1, 385–402. doi: 10.1214/06-BA115
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. New York, NY; Berlin; Heidelberg; Tokyo: Springer.
- Bernardo, J., and Smith, A. (2000). *Bayesian Theory (Wiley Series in Probability and Statistics)*. Chichester: Wiley.
- Bernardo, J. M. (1997). Noninformative priors do not exist: a discussion. *J. Stat. Plan. Inference* 65, 159–189.
- Berry, D. (2006). Bayesian clinical trials. *Nat. Rev. Drug Discov.* 5, 27–36. doi: 10.1038/nrd1927
- Beven, K. (2006). A manifesto for the equifinality thesis. *J. Hydrol.* 320, 18–36. doi: 10.1016/j.jhydrol.2005.07.007
- Beven, K., and Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.* 6, 279–298. doi: 10.1002/hyp.3360060305
- Beven, K. and Young, P. (2013). A guide to good practice in modeling semantics for authors and referees. *Water Resour. Res.* 49, 5092–5098. doi: 10.1002/wrcr.20393
- Biegler, L., Biros, G., Ghattas, O., Heinkenschloss, M., Keyes, D., Mallick, B., et al. (eds.) (2010). *Large-Scale Inverse Problems and Quantification of Uncertainty*. Chichester: Wiley.

- Billot, A., Gilboa, I., Samet, D., and Schmeidler, D. (2005). Probabilities as similarity-weighted frequencies. *Econometrica* 73, 1125–1136. doi: 10.1111/j.1468-0262.2005.00611.x
- Boisvert, E., and Brodaric, B. (2011). Groundwater markup language (gwmml) – enabling groundwater data interoperability in spatial data infrastructures. *J. Hydroinformatics* 14:93. doi: 10.2166/hydro.2011.172
- Bond, C. E. (2015). Uncertainty in structural interpretation: lessons to be learnt. *J. Struct. Geol.* 74(Suppl. C), 185–200. doi: 10.1016/j.jsg.2015.03.003
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Brodaric, B., Boisvert, E., Chery, L., Dahlhaus, P., Grellet, S., Kmoch, A., et al. (2018). Enabling global exchange of groundwater data: groundwaternl2 (gwmml2). *Hydrogeol. J.* 26, 733–741. doi: 10.1007/s10040-018-1747-9
- Brown, J. D. (2004). Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief. *Trans. Inst. Br. Geogr.* 29, 367–381. doi: 10.1111/j.0020-2754.2004.00342.x
- Caers, J., Park, K., and Scheidt, C. (2014). *Modeling Uncertainty of Complex Earth Systems in Metric Space*, 1–25. Berlin; Heidelberg: Springer.
- Calcagno, P., Chilès, J., Courrioux, G., and Guillen, A. (2008). Geological modelling from field data and geological knowledge: part i. modelling method coupling 3d potential-field interpolation and geological rules. *Phys. Earth Planet. Inter.* 171, 147–157. doi: 10.1016/j.pepi.2008.06.013
- Cardiff, M., and Kitanidis, P. K. (2009). Bayesian inversion for facies detection: an extensible level set framework. *Water Resour. Res.* 45. doi: 10.1029/2008WR007675
- Carlin, B. P., and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis, 2nd Edn.* Number 69 in CRC Texts in Statistical Science. Boca Raton, FL; London; New York, NY; Washington, DC: Chapman & Hall.
- Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., and Slooten, L. J. (2005). Inverse problem in hydrogeology. *Hydrogeol. J.* 13, 206–222. doi: 10.1007/s10040-004-0404-7
- Caumon, G., Collon-Drouaillet, P., Le Carlier de Veslud, C., Viseur, S., and Sausse, J. (2009). Surface-based 3d modeling of geological structures. *Math. Geosci.* 41, 927–945. doi: 10.1007/s11004-009-9244-2
- Chib, S., and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *Am. Stat.* 49, 327–335. doi: 10.2307/2684568
- Chilès, J., Aug, C., Guillen, A., and Lees, T. (2004). “Modelling the geometry of geological units and its uncertainty in 3d from structural data: The potential-field method,” in *Orebody Modelling and Strategic Mine Planning*.
- Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). BART: bayesian additive regression trees. *Ann. Appl. Stat.* 4, 266–298. doi: 10.1214/09-AOAS285
- Cirpka, O. A., and Valocchi, A. J. (2016). Debates-stochastic subsurface hydrology from theory to practice: Does stochastic subsurface hydrology help solving practical problems of contaminant hydrogeology? *Water Resour. Res.* 52, 9218–9227. doi: 10.1002/2016WR019087
- Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.
- Clark, J. S. (2005). Why environmental scientists are becoming Bayesians. *Ecol. Lett.* 8, 2–14. doi: 10.1111/j.1461-0248.2004.00702.x
- Clark, M. P., Kavetski, D., and Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resour. Res.* 47. doi: 10.1029/2010WR009827
- Colombera, L., Felletti, F., Mountney, N. P., and McCaffrey, W. D. (2012). A database approach for constraining stochastic simulations of the sedimentary heterogeneity of fluvial reservoirs. *AAPG Bull.* 96, 2143–2166. doi: 10.1306/04211211179
- Comunian, A., and Renard, P. (2009). Introducing wwhypda: a world-wide collaborative hydrogeological parameters database. *Hydrogeol. J.* 17, 481–489. doi: 10.1007/s10040-008-0387-x
- Comunian, A., Renard, P., Straubhaar, J., and Bayer, P. (2011). Three-dimensional high resolution fluvio-glacial aquifer analog - Part 2: geostatistical modeling. *J. Hydrol.* 405, 10–23. doi: 10.1016/j.jhydrol.2011.03.037
- Copt, N. K., Trinchero, P., and Sanchez-Vila, X. (2011). Inferring spatial distribution of the radially integrated transmissivity from pumping tests in heterogeneous confined aquifers. *Water Resour. Res.* 47. doi: 10.1029/2010WR009877
- Cucchi, K., Heße, F., Kawa, N., Wang, C., and Rubin, Y. (2019). *Ex-situ* priors: a Bayesian hierarchical framework for defining informative prior distributions in hydrogeology. *Adv. Water Resour.* 126, 65–78. doi: 10.1016/j.advwatres.2019.02.003
- Dagan, G. (2002). An overview of stochastic modeling of groundwater flow and transport: From theory to applications. *Eos Trans. Am. Geophys. Union* 83, 621–625. doi: 10.1029/2002EO000421
- Darwiche, A., and Pearl, J. (1997). On the logic of iterated belief revision. *Artif. Intell.* 89, 1–29. doi: 10.1016/S0004-3702(96)00038-0
- de la Varga, M., Schaaf, A., and Wellmann, F. (2018). Gempy 1.0: open-source stochastic geological modeling and inversion. *Geosci. Model Dev. Discuss.* 2018, 1–50. doi: 10.5194/gmd-2018-61
- de la Varga, M., and Wellmann, J. F. (2016). Structural geologic modeling as an inference problem: a bayesian perspective. *Interpretation* 4, SM1–SM16. doi: 10.1190/INT-2015-0188.1
- de Marsily, G., Delay, F., Gonçalves, J., Renard, P., Teles, V., and Violette, S. (2005). Dealing with spatial heterogeneity. *Hydrogeol. J.* 13, 161–183. doi: 10.1007/s10040-004-0432-3
- Dempster, A. P. (1968). A Generalization of Bayesian Inference. *J. R. Stat. Soc. Ser. B Methodol.* 30, 205–247.
- Denwood, M. J. (2016). runjags: an r package providing interface utilities, model templates, parallel computing methods and additional distributions for mcmc models in jags. *J. Stat. Softw.* 71, 1–25. doi: 10.18637/jss.v071.i09
- Depaoli, S., Clifton, J. P., and Cobb, P. R. (2016). Just another gibbs sampler (JAGS): flexible software for MCMC implementation. *J. Educ. Behav. Stat.* 41, 628–649. doi: 10.3102/1076998616664876
- Descartes, R. (1662). *De Homine*. Leffen & Moyardus.
- Deutsch, D. (1999). Quantum theory of probability and decisions. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 455, 3129–3137. doi: 10.1098/rspa.1999.0443
- Di Palma, P. R., Guynnon, N., Heße, F., and Romano, E. (2017). Porous media flux sensitivity to pore-scale geostatistics: a bottom-up approach. *Adv. Water Resour.* 102(Suppl. C), 99–110. doi: 10.1016/j.advwatres.2017.02.002
- Diaconis, P., Holmes, S., and Montgomery, R. (2007). Dynamical bias in the coin toss. *SIAM Rev.* 49, 211–235. doi: 10.1137/S0036144504446436
- Diggle, P. J., and Ribeiro, P. J. (2007). *Model-based Geostatistics, 1st Edn.* New York, NY: Springer.
- Doherty, J. (2004). *PEST, Model Independent Parameter Estimation, User Manual, 5th Edn.* Oxford: Watermark Numerical Computing.
- Doherty, J. (2010). *Methodologies and Software for PEST-Based Model Predictive Uncertainty Analysis*. Watermark Numerical Computing.
- Dougherty, T. (ed.) (2011). *Evidentialism and Its Discontents*. Oxford: Oxford University Press.
- Dubois, D. and Prade, H. (1988). *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. New York, NY: Springer.
- Duhem, P. (1906). *La Théorie Physique: Son Objet et sa Structure*. Paris; Lyon: ENS Éditions.
- Earman, J. (1992). *Bayes or Bust*. Cambridge, MA: MIT Press.
- Easwaran, K. (2011a). Bayesianism i: introduction and arguments in favor. *Philos. Compass* 6, 312–320. doi: 10.1111/j.1747-9991.2011.00399.x
- Easwaran, K. (2011b). Bayesianism ii: applications and criticisms. *Philos. Compass* 6, 321–332. doi: 10.1111/j.1747-9991.2011.00398.x
- Easwaran, K. (2013). Expected accuracy supports conditionalization and conglomerability and reflection. *Philos. Sci.* 80, 119–142. doi: 10.1086/668879
- Ebrahimi, N., Soofi, E. S., and Soyer, R. (2010). Information measures in perspective. *Int. Stat. Rev.* 78, 383–412. doi: 10.1111/j.1751-5823.2010.00105.x
- Elith, J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. doi: 10.1111/j.1365-2656.2008.01390.x
- Elsheikh, A. H., Wheeler, M. F., and Hoteit, I. (2014). Hybrid nested sampling algorithm for bayesian model selection applied to inverse subsurface flow problems. *J. Comput. Phys.* 258, 319–337. doi: 10.1016/j.jcp.2013.10.001
- Emery, X. (2004). Properties and limitations of sequential indicator simulation. *Stochastic Environ. Res. Risk Assess.* 18, 414–424. doi: 10.1007/s00477-004-0213-5
- Emery, X. (2007). Simulation of geological domains using the plurigaussian model: new developments and computer programs. *Comput. Geosci.* 33, 1189–1201. doi: 10.1016/j.cageo.2007.01.006

- Enemark, T., Peeters, L. J., Mallants, D., and Batelaan, O. (2019). Hydrogeological conceptual model building and testing: a review. *J. Hydrol.* 569, 310–329. doi: 10.1016/j.jhydrol.2018.12.007
- Fagin, R., Halpern, J. Y., Moses, Y., and Vardi, M. (2004). *Reasoning About Knowledge*. Cambridge, MA: A Bradford Book.
- Feldman, R., and Conee, E. (1985). Evidentialism. *Philos. Stud.* 48, 15–34.
- Feyerabend, P. (1975). *Against Method: Outline of an Anarchist Theory of Knowledge*. London: New Left Books.
- Finetti, B. D. (1975). *Theory of Probability. A Critical Introductory Treatment*. Chichester: Wiley.
- Firmani, G., Fiori, A., and Bellin, A. (2006). Three-dimensional numerical analysis of steady state pumping tests in heterogeneous confined aquifers. *Water Resour. Res.* 42. doi: 10.1029/2005WR004382
- Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: The MCMC Hammer. *Publ. Astron. Soc. Pac.* 125, 306–312. doi: 10.1086/670067
- Fox, C. R., and Ülkümen, G. (2011). “Chapter Distinguishing Two Dimensions of Uncertainty,” in *Perspectives on Thinking, Judging, and Decision Making, 1st Edn*, eds W. Brun, G. Keren, G. Kirkeboen and H. Montgomery (Oslo: Universitetsforlaget), 21–35.
- Franssen Hendricks, H.-J., Alcolea, A., Riva, M., Bakr, M., van der Wiel, N., Stauffer, F., et al. (2009). A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments. *Adv. Water Resour.* 32, 851–872. doi: 10.1016/j.advwatres.2009.02.011
- Fuchs, C. A., and Schack, R. (2013). Quantum-bayesian coherence. *Rev. Mod. Phys.* 85:1693. doi: 10.1103/RevModPhys.85.1693
- Galli, A., Beucher, H., Le Loc’h, G., Doligez, B., and Group, H. (1994). “The pros and cons of the truncated gaussian method,” in *Geostatistical Simulations*, eds M. Armstrong and P. A. Dowd (Dordrecht: Springer), 217–233.
- Garber, D. (1983). “Old evidence and logical omniscience in bayesian confirmation theory,” in *Testing Scientific Theories*, Vol. 10 of *Minnesota Studies in the Philosophy of Science*, ed J. Earman (Minneapolis, MN: University of Minnesota), 99–131.
- Geiges, A., Rubin, Y., and Nowak, W. (2015). Interactive design of experiments: a priori global versus sequential optimization, revised under changing states of knowledge. *Water Resour. Res.* 51, 7915–7936. doi: 10.1002/2015WR017193
- Gelfand, A. E., and Schliep, E. M. (2016). Spatial statistics and Gaussian processes: a beautiful marriage. *Spatial Stat.* 18, 86–104. doi: 10.1016/j.spasta.2016.03.006
- Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85, 398–409.
- Gelhar, L. (1993). *Stochastic Subsurface Hydrology*. Englewood cliffs, NJ: Prentice-Hall.
- Gelman, A. (2008). Objections to Bayesian statistics. *Bayesian Anal.* 3, 445–449. doi: 10.1214/08-BA318
- Gelman, A., Lee, D., and Guo, J. (2015). Stan: a probabilistic programming language for Bayesian inference and optimization. *J. Educ. Behav. Stat.* 40, 530–543. doi: 10.3102/1076998615606113
- Gelman, A., Simpson, D., and Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy* 19:555. doi: 10.3390/e19100555
- Gibling, M. R. (2006). Width and thickness of fluvial channel bodies and valley fills in the geological record: A literature compilation and classification. *J. Sediment. Res.* 76, 731–770. doi: 10.2110/jsr.2006.060
- Gilboa, I. (2009). *Theory of Decision Under Uncertainty*. Cambridge: Cambridge University Press.
- Gilboa, I., Lieberman, O., and Schmeidler, D. (2006). Empirical similarity. *Rev. Econ. Stat.* 88, 433–444. doi: 10.1162/rest.88.3.433
- Gilboa, I., Lieberman, O., and Schmeidler, D. (2010). On the definition of objective probabilities by empirical similarity. *Synthese* 172, 79–95. doi: 10.1007/s11229-009-9473-4
- Ginn, T. R. (2004). On the application of stochastic approaches in hydrogeology. *Stochast. Environ. Res. Risk Assess.* 18, 282–284. doi: 10.1007/s00477-004-0199-z
- Gómez-Hernández, J. J., and Wen, X.-H. (1998). To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Adv. Water Resour.* 21, 47–61. doi: 10.1016/S0309-1708(96)00031-0
- Gong, W., Gupta, H. V., Yang, D., Sricharan, K., and Hero, A. O. (2013). Estimating epistemic and aleatory uncertainties during hydrologic modeling: an information theoretic approach. *Water Resour. Res.* 49, 2253–2273. doi: 10.1002/wrcr.20161
- Good, I. J. (2009a). “Chapter Rational decisions,” in *Good Thinking: The Foundations of Probability and Its Applications* (Dover Books on Mathematics), 3–15.
- Good, I. J. (2009b). *Good Thinking: The Foundations of Probability and Its Applications*. Minneapolis, MN: Dover Books on Mathematics.
- Greaves, H. and Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind* 115, 607–632. doi: 10.1093/mind/fzl607
- Grossman, J. (2011). “The likelihood principle,” in *Philosophy of Statistics*, Vol. 7 of *Handbook of the Philosophy of Science*, eds P. S. Bandyopadhyay and M. R. Forster (Amsterdam: North-Holland), 553–580.
- Guardiano, F. B., and Srivastava, R. M. (1993). “Multivariate geostatistics: Beyond bivariate moments,” in *Geostatistics: Troia ’92*, Vol. 1, ed A. Soares, A. (Dordrecht: Kluwer), 133–144.
- Guillaume, J. H. A., Qureshi, M. E., and Jakeman, A. J. (2012). A structured analysis of uncertainty surrounding modeled impacts of groundwater-extraction rules. *Hydrogeol. J.* 20, 915–932. doi: 10.1007/s10040-012-0864-0
- Halevy, A., Norvig, P., and Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intell. Syst.* 24, 8–12. doi: 10.1109/MIS.2009.36
- Halpern, J. Y. (2003). *Reasoning About Uncertainty*. Cambridge MA; London: MIT Press.
- Hansen, T. M., Cordua, K. S., and Mosegaard, K. (2012). Inverse problems with non-trivial priors: efficient solution through sequential gibbs sampling. *Comput. Geosci.* 16, 593–611. doi: 10.1007/s10596-011-9271-1
- Hayek, M., Younes, A., Zouali, J., Fajraoui, N., and Fahs, M. (2018). Analytical solution and Bayesian inference for interference pumping tests in fractal dual-porosity media. *Comput. Geosci.* 22, 413–421. doi: 10.1007/s10596-017-9701-9
- Helton, J. C., and Burmaster, D. E. (1996). Guest editorial: treatment of aleatory and epistemic uncertainty in performance assessments for complex systems. *Reliabil. Eng. Syst. Saf.* 54, 91–94. doi: 10.1016/S0951-8320(96)00066-X
- Heße, F., Savoy, H., Osorio-Murillo, C. A., Sege, J., Attinger, S., and Rubin, Y. (2015). Characterizing the impact of roughness and connectivity features of aquifer conductivity using bayesian inversion. *J. Hydrol.* 531(Pt 1), 73–87. doi: 10.1016/j.jhydrol.2015.09.067
- Hofer, E. (1996). When to separate uncertainties and when not to separate. *Reliabil. Eng. Syst. Saf.* 54, 113–118. doi: 10.1016/S0951-8320(96)00068-3
- Hoffman, F. O., and Hammonds, J. S. (1994). Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Anal.* 14, 707–712. doi: 10.1111/j.1539-6924.1994.tb00281.x
- Hou, T. Y. (2005). *Multiscale Computation of Fluid Flow in Heterogeneous Media*. Springer.
- Howson, C., and Urbach, P. (2005). *Scientific Reasoning: The Bayesian Approach, 3rd Edn*. Peru: Open Court.
- Huelsenbeck, J., Ronquist, F., Nielsen, R., and Bollback, J. (2001). Evolution - Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294, 2310–2314. doi: 10.1126/science.1065889
- Indelman, P., Fiori, A., and Dagan, G. (1996). Steady flow toward wells in heterogeneous formations: mean head and equivalent conductivity. *Water Resour. Res.* 32, 1975–1983. doi: 10.1029/96WR00990
- Jaynes, E. T. (1957a). Information theory and statistical mechanics. *Phys. Rev.* 106, 620–630. doi: 10.1103/PhysRev.106.620
- Jaynes, E. T. (1957b). Information theory and statistical mechanics. ii. *Phys. Rev.* 108, 171–190. doi: 10.1103/PhysRev.108.171
- Jaynes, E. T. (1968). Prior probabilities. *IEEE Trans. Syst. Sci. Cybern.* 4, 227–241. doi: 10.1109/TSSC.1968.300117
- Jeffrey, R. (1992). *Probability and the Art of Judgment*. Cambridge Studies in Probability, Induction and Decision Theory. Cambridge, MA; London: Cambridge University Press.
- Jung, A. and Aigner, T. (2012). Carbonate geobodies: hierarchical classification and database - a new workflow for 3d reservoir modelling. *J. Petrol. Geol.* 35, 49–65. doi: 10.1111/j.1747-5457.2012.00518.x
- Kahneman, D., Slovic, P., and Tversky, A., (eds.) (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, MA; London: Cambridge University Press.

- Kallenberg, O. (2002). *Foundations of Modern Probability*. New York, NY: Springer.
- Kapeller, A., and Bleich, J. (2016). bartMachine: machine learning with Bayesian additive regression trees. *J. Stat. Softw.* 70, 1–40. doi: 10.18637/jss.v070.i04
- Kass, R. E. (2011). Statistical inference: the big picture. *Stat. Sci.* 26, 1–9. doi: 10.1214/10-STS337
- Kass, R. E., and Wasserman, L. (1996). The selection of prior distributions by formal rules. *J. Am. Stat. Assoc.* 91, 1343–1370.
- Kenter, J. and Harris, P. (2006). “Web-based outcrop digital analog database (wodad): archiving carbonate platform margins,” in *AAPG International Conference* (Cambridge MA; London) 5–8.
- Kerrou, J., Renard, P., Franssen, H.-J. H., and Lunati, I. (2008). Issues in characterizing heterogeneity and connectivity in non-multigaussian media. *Adv. Water Resour.* 31, 147–159. doi: 10.1016/j.advwatres.2007.07.002
- Kitanidis, P. K. (2012). Generalized priors in bayesian inversion problems. *Adv. Water Resour.* 36, 3–10. doi: 10.1016/j.advwatres.2011.05.005
- Kitanidis, P. K. (2015). Persistent questions of heterogeneity, uncertainty, and scale in subsurface flow and transport. *Water Resour. Res.* 51, 5888–5904. doi: 10.1002/2015WR017639
- Kiureghian, A. D., and Ditlevsen, O. (2009). Aleatory or epistemic? does it matter? *Struct. Saf.* 31, 105–112. doi: 10.1016/j.strusafe.2008.06.020
- Knight, F. H. (1921). *Risk, Uncertainty, and Profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Co.
- Koehler, D. J. and Harvey, N. (eds.) (2004). *Blackwell Handbook of Judgment and Decision Making*. Malden, MA: Blackwell.
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Berlin; Heidelberg; New York, NY: Springer.
- Koltermann, C. E., and Gorelick, S. M. (1996). Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches. *Water Resour. Res.* 32, 2617–2658. doi: 10.1029/96WR00025
- Krause, P., Boyle, D. P., and Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* 5, 89–97. doi: 10.5194/adgeo-5-89-2005
- Kruschke, J. K. (2010). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Amsterdam: Academic Press.
- Kuhn, T. (1962). *The Structure of Scientific Revolutions, 1st Edn*. Chicago, IL: University of Chicago Press.
- Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* 22, 79–86.
- Kwakkel, J., Walker, W., and Marchau, V. (2010). Classifying and communicating uncertainties in model-based policy analysis. *Int. J. Technol. Policy Manage.* 10, 299–315. doi: 10.1504/IJTPM.2010.036918
- Laloy, E., Linde, N., Jacques, D., and Mariethoz, G. (2016). Merging parallel tempering with sequential geostatistical resampling for improved posterior exploration of high-dimensional subsurface categorical fields. *Adv. Water Resour.* 90, 57–69. doi: 10.1016/j.advwatres.2016.02.008
- Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., and Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage markov chain monte carlo simulation and polynomial chaos expansion. *Water Resour. Res.* 49, 2664–2682. doi: 10.1002/wrcr.20226
- Laloy, E., and Vrugt, J. A. (2012). High-dimensional posterior exploration of hydrologic models using multiple-try dream(zs) and high-performance computing. *Water Resour. Res.* 48. doi: 10.1029/2011WR010608
- Laudan, L. (1990). “Demystifying underdetermination,” in *Scientific Theories*, ed C. W. Savage (University of Minnesota Press), 267–97.
- Le Loc'h, G. and Galli, A. (1997). Truncated plurigaussian method: theoretical and practical points of view. *Geostat. Wollongong* 96, 211–222.
- Leitgeb, H., and Pettigrew, R. (2010). An objective justification of Bayesianism I: measuring inaccuracy. *Philos. Sci.* 77, 201–235. doi: 10.1086/651317
- Li, X., Li, Y., Chang, C.-F., Tan, B., Chen, Z., Sege, J., et al. (2017). Stochastic, goal-oriented rapid impact modeling of uncertainty and environmental impacts in poorly-sampled sites using *ex-situ* priors. *Adv. Water Resour.* 111, 174–191. doi: 10.1016/j.advwatres.2017.11.008
- Linde, N., Renard, P., Mukerji, T., and Caers, J. (2015). Geological realism in hydrogeological and geophysical inverse modeling: a review. *Adv. Water Resour.* 86(Pt A), 86–101. doi: 10.1016/j.advwatres.2015.09.019
- Lindgren, F., and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* 63, 1–25. doi: 10.18637/jss.v063.i19
- Lindley, D. V. (1987). The probability approach to the treatment of uncertainty in artificial intelligence and expert systems. *Stat. Sci.* 2, 17–24.
- Little, R. J. (2006). Calibrated Bayes: a Bayes/frequentist roadmap. *Am. Stat.* 60, 213–223. doi: 10.1198/000313006X117837
- Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The bugs project: evolution, critique and future directions. *Stat. Med.* 28, 3049–3067. doi: 10.1002/sim.3680
- Maier, H., Guillaume, J., van Delden, H., Riddell, G., Haasnoot, M., and Kwakkel, J. (2016). An uncertain future, deep uncertainty, scenarios, robustness and adaptation: how do they fit together? *Environ. Model. Softw.* 81, 154–164. doi: 10.1016/j.envsoft.2016.03.014
- Malinverno, A., and Briggs, V. (2004). Expanded uncertainty quantification in inverse problems: hierarchical Bayes and empirical Bayes. *Geophysics* 69, 1005–1016. doi: 10.1190/1.1778243
- Malpica, J. A., Alonso, M. C., and Sanz, M. A. (2007). Dempster-Shafer Theory in geographic information systems: a survey. *Expert Syst. Appl.* 32, 47–55. doi: 10.1016/j.eswa.2005.11.011
- Mantovan, P., and Todini, E. (2006). Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology. *J. Hydrol.* 330, 368–381. doi: 10.1016/j.jhydrol.2006.04.046
- Mara, T. A., Fajraoui, N., Guadagnini, A., and Younes, A. (2017). Dimensionality reduction for efficient Bayesian estimation of groundwater flow in strongly heterogeneous aquifers. *Stochast. Environ. Res. Risk Assess.* 31, 2313–2326. doi: 10.1007/s00477-016-1344-1
- Marchant, B. P., and Lark, R. M. (2007). The Matérn variogram model: implications for uncertainty propagation and sampling in geostatistical surveys. *Geoderma* 140, 337–345. doi: 10.1016/j.geoderma.2007.04.016
- Mariethoz, G., and Caers, J. (2014). *Multiple-point Geostatistics: Stochastic Modeling with Training Images*, 376. Chichester: Wiley.
- Mariethoz, G., Renard, P., and Caers, J. (2010). Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resour. Res.* 46. doi: 10.1029/2010WR009274
- Mariethoz, G., Renard, P., Cornaton, F., and Jaquet, O. (2009). Truncated plurigaussian simulations to characterize aquifer heterogeneity. *Groundwater* 47, 13–24. doi: 10.1111/j.1745-6584.2008.00489.x
- Martin, A. D., Quinn, K. M., and Park, J. H. (2011). MCMCpack: Markov Chain Monte Carlo in R. *J. Stat. Softw.* 42, 1–21. doi: 10.18637/jss.v042.i09
- Maslow, A. H. (1943). A theory of human motivation. *Psychol. Rev.* 50, 370–396.
- Menke, W. (2012). *Geophysical Data Analysis: Discrete Inverse Theory*. Amsterdam: Academic Press.
- Montanari, A. (2007). What do we mean by “uncertainty”? the need for a consistent wording about uncertainty assessment in hydrology. *Hydrol. Process.* 21, 841–845. doi: 10.1002/hyp.6623
- Montanari, A., Shoemaker, C. A., and van de Giesen, N. (2009). Introduction to special section on uncertainty assessment in surface and subsurface hydrology: an overview of issues and challenges. *Water Resour. Res.* 45. doi: 10.1029/2009WR008471.W00B00
- Moriasi, D. N., Gitau, M. W., Pai, N., and Daggupati, P. (2015). Hydrologic and water quality model: performance measures and evaluation criteria. *Trans. ASABE* 58, 1763–1785. doi: 10.13031/trans.58.10715
- Murakami, H., Chen, X., Hahn, M. S., Liu, Y., Rockhold, M. L., Vermeul, V. R., et al. (2010). Bayesian approach for three-dimensional aquifer characterization at the Hanford 300 Area. *Hydrol. Earth Syst. Sci.* 14, 1989–2001. doi: 10.5194/hess-14-1989-2010
- Neapolitan, R. E. (1990). *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. New York, NY: John Wiley & Sons, Inc.
- Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., and Weijs, S. V. (2016). A philosophical basis for hydrologic uncertainty. *Hydrol. Sci. J.* 61, 1666–1678. doi: 10.1080/02626667.2016.1183009
- Neuman, S. P. (2004). Stochastic groundwater models in practice. *Stochast. Environ. Res. Risk Assess.* 18, 268–270. doi: 10.1007/s00477-004-0192-6

- Neuman, S. P. (2014). The new potential for understanding groundwater contaminant transport. *Groundwater* 52, 653–U177. doi: 10.1111/gwat.12245
- Neuman, S. P., and Tartakovsky, D. M. (2009). Perspective on theories of non-fickian transport in heterogeneous media. *Adv. Water Resour.* 32, 670–680. doi: 10.1016/j.advwatres.2008.08.005
- Neyman, J., and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part i. *Biometrika* 20A, 175–240.
- Neyman, J., and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. Lond. Ser. A* 231, 289–337.
- O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., et al. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. Amsterdam: Wiley.
- Osherson, D., Shafir, E., and Smith, E. E. (1993). Ampliative inference: on choosing a probability distribution. *Cognition* 49, 189–210. doi: 10.1016/0010-0277(93)90004-F
- Parmigiani, G. and Inoue, L. (eds.) (2009). *Decision Theory: Principles and Approaches*. Probability & Mathematical Statistics. Chichester: Wiley.
- Patil, A., Huard, D., and Fonnesbeck, C. J. (2010). PyMC: Bayesian stochastic modelling in Python. *J. Stat. Softw.* 35, 1–81. doi: 10.18637/jss.v035.i04
- Pearl, J. (1988a). On probability intervals. *Int. J. Approx. Reason.* 2, 211–216.
- Pearl, J. (1988b). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann.
- Pearl, J. (1990). Reasoning with belief functions: an analysis of compatibility. *Int. J. Approx. Reason.* 4, 363–389.
- Pechstein, A., Attinger, S., Krieg, R., and Copty, N. K. (2016). Estimating transmissivity from single-well pumping tests in heterogeneous aquifers. *Water Resour. Res.* 52, 495–510. doi: 10.1002/2015WR017845
- Pérez, C., Mariethoz, G., and Ortiz, J. M. (2014). Verifying the high-order consistency of training images with data for multiple-point geostatistics. *Comput. Geosci.* 70, 190–205. doi: 10.1016/j.cageo.2014.06.001
- Pratola, M. T., Chipman, H. A., Gattiker, J. R., Higdon, D. M., McCulloch, R., and Rust, W. N. (2014). Parallel Bayesian Additive Regression Trees. *J. Comput. Graph. Stat.* 23, 830–852. doi: 10.1080/10618600.2013.841584
- Pushpalatha, R., Perrin, C., Moine, N. L., and Andréassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *J. Hydrol.* 420–421, 171–182. doi: 10.1016/j.jhydrol.2011.11.055
- Pyrce, M., Boisvert, J., and Deutsch, C. (2008). A library of training images for fluvial and deepwater reservoirs and associated code. *Comput. Geosci.* 34, 542–560. doi: 10.1016/j.cageo.2007.05.015
- Quine, W. V. O. (1951). Two dogmas of empiricism. *Philos. Rev.* 60, 20–43.
- Ramsey, F. P. (1931). "Chapter Truth and Probability," in *Foundations of Mathematics and other Logical Essays*, ed R. S. B. Braithwaite (New York, NY: Harcourt, Brace and Company), 156–198.
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Cambridge MA; London: MIT Press.
- Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højberg, A. L., and Trolborg, L. (2012). Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. *Adv. Water Resour.* 36, 36–50. doi: 10.1016/j.advwatres.2011.04.006
- Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., and Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process—a framework and guidance. *Environ. Model. Softw.* 22, 1543–1556. doi: 10.1016/j.envsoft.2007.02.004
- Renard, P. (2007). Stochastic hydrogeology: What professionals really need? *Ground Water* 45, 531–541. doi: 10.1111/j.1745-6584.2007.00340.x
- Riva, M., Guadagnini, A., Neuman, S. P., Janetti, E. B., and Malama, B. (2009). Inverse analysis of stochastic moment equations for transient flow in randomly heterogeneous media. *Adv. Water Resour.* 32, 1495–1507. doi: 10.1016/j.advwatres.2009.07.003
- Robert, C. P. (2001). *The Bayesian Choice*. New York, NY: Springer.
- Rogati, M. (2017). *The AI Hierarchy of Needs*.
- Romeijn, J.-W. (2017). "Philosophy of statistics," in *The Stanford Encyclopedia of Philosophy*, ed E. N. Zalta (Metaphysics Research Lab; Stanford University).
- Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm*. Monographs on Statistics & Applied Probability. Boca Raton, FL; London; New York, NY; Washington, DC: Chapman & Hall/CRC.
- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Stat.* 12, 1151–1172.
- Rubin, Y. (2003). *Applied Stochastic Hydrogeology*. New York, NY: Oxford University Press.
- Rubin, Y. (2004). Stochastic hydrogeology - challenges and misconceptions. *Stochast. Environ. Res. Risk Assess.* 18, 280–281. doi: 10.1007/s00477-004-0193-5
- Rubin, Y., Chang, C.-F., Chen, J., Cucchi, K., Harken, B., Heße, F., et al. (2018). Stochastic hydrogeology's biggest hurdles analyzed and its big blind spot. *Hydrol. Earth Syst. Sci. Discuss.* 2018, 1–36. doi: 10.5194/hess-2018-290
- Rubin, Y., Chen, X., Murakami, H., and Hahn, M. (2010). A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields. *Water Resour. Res.* 46. doi: 10.1029/2009WR008799
- Rumsfeld, D. H. (2002). *DoD News Briefing - Secretary Rumsfeld and Gen. Myers*. Online Transcript.
- Russell, S., and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach, 3rd Edn*. Englewood Cliffs, NJ: Pearson.
- Saley, A. D., Jardani, A., Ahmed, A. S., Raphael, A., and Dupont, J. (2016). Hamiltonian monte carlo algorithm for the characterization of hydraulic conductivity from the heat tracing data. *Adv. Water Resour.* 97, 120–129. doi: 10.1016/j.advwatres.2016.09.004
- Sanchez-Vila, X., and Fernandez-Garcia, D. (2016). Debates-Stochastic subsurface hydrology from theory to practice: why stochastic modeling has not yet permeated into practitioners? *Water Resour. Res.* 52, 9246–9258. doi: 10.1002/2016WR019302
- Sánchez-Vila, X., Meier, P. M., and Carrera, J. (1999). Pumping tests in heterogeneous aquifers: an analytical study of what can be obtained from their interpretation using Jacob's method. *Water Resour. Res.* 35, 943–952. doi: 10.1029/1999WR900007
- Savage, L. J. (1954). *The Foundations of Statistics*. New York, NY: Dover Publications.
- Savoy, H., Kalbacher, T., Dietrich, P., and Rubin, Y. (2017). Geological heterogeneity: goal-oriented simplification of structure and characterization needs. *Adv. Water Resour.* 109(Suppl. C), 1–13. doi: 10.1016/j.advwatres.2017.08.017
- Scales, J. A., and Tenorio, L. (2001). Prior information and uncertainty in inverse problems. *Geophysics* 66, 389–397. doi: 10.1190/1.1444930
- Schack, R., Brun, T. A., and Caves, C. M. (2001). Quantum Bayes rule. *Phys. Rev. A* 64. doi: 10.1103/PhysRevA.64.014305
- Schneider, C. L., and Attinger, S. (2008). Beyond Thiem: a new method for interpreting large scale pumping tests in heterogeneous aquifers. *Water Resour. Res.* 44. doi: 10.1029/2007WR005898
- Sebens, C. T., and Carroll, S. M. (2016). Self-locating uncertainty and the origin of probability in everettian quantum mechanics. *Br. J. Philos. Sci.* 67, 25–74. doi: 10.1093/bjps/axw004
- Segal, M. and Xiao, Y. (2011). Multivariate random forests. *Wiley Interdiscipl. Rev. Data Min. Knowl. Discov.* 1, 80–87. doi: 10.1002/widm.12
- Serrano, R. P., Guadagnini, L., Riva, M., Giudici, M., and Guadagnini, A. (2014). Impact of two geostatistical hydro-facies simulation strategies on head statistics under non-uniform groundwater flow. *J. Hydrol.* 508, 343–355.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton, NJ; London: Princeton University Press.
- Shannon, C. E. (1948). The mathematical theory of communication. *Bell Syst. Techn. J.* 27, 306–317.
- Shi, X., Ye, M., Curtis, G. P., Miller, G. L., Meyer, P. D., Kohler, M., et al. (2014). Assessment of parametric uncertainty for groundwater reactive transport modeling. *Water Resour. Res.* 50, 4416–4439. doi: 10.1002/2013WR013755
- Stanford, P. K. (2001). Refusing the devil's bargain: what kind of underdetermination should we take seriously? *Philos. Sci.* 68, S1–S12. doi: 10.1086/392893
- Stedinger, J. R., Vogel, R. M., Lee, S. U., and Batchelder, R. (2008). Appraisal of the generalized likelihood uncertainty estimation (glue) method. *Water Resour. Res.* 44. doi: 10.1029/2008WR006822
- Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol.* 34, 1–21. doi: 10.1023/A:1014009426274

- Tang, Y., Marshall, L., Sharma, A., and Smith, T. (2016). Tools for investigating the prior distribution in Bayesian hydrology. *J. Hydrol.* 538, 551–562. doi: 10.1016/j.jhydrol.2016.04.032
- Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: SIAM.
- Tartakovsky, D. M. (2013). Assessment and management of risk in subsurface hydrology: a review and perspective. *Adv. Water Resour.* 51, 247–260. doi: 10.1016/j.advwatres.2012.04.007
- Tierney, L. (1994). Markov-chains for exploring posterior distributions. *Ann. Stat.* 22, 1701–1728.
- Tribus, M., and McIrvine, E. C. (1971). Energy and information. *Sci. Am.* 224, 179–188.
- Ulrych, T. J., Sacchi, M. D., and Woodbury, A. (2001). A Bayes tour of inversion: a tutorial. *Geophysics* 66:55. doi: 10.1190/1.1444923
- Valakas, G., and Modis, K. (2016). Using informative priors in facies inversion: the case of c-isr method. *Adv. Water Resour.* 94, 23–30. doi: 10.1016/j.advwatres.2016.04.019
- Van Fraassen, B. (1980). *The Scientific Image*. Oxford: Oxford University Press.
- von Mises, R. (1982). *Probability, Statistics and Truth*. New York, NY: Dover.
- von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ; Oxford: Princeton University Press.
- von Toussaint, U. (2011). Bayesian inference in physics. *Rev. Mod. Phys.* 83, 943–999. doi: 10.1103/RevModPhys.83.943
- Vrugt, J., ter Braak, C., Gupta, H., and Robinson, B. (2009). Equifinality of formal (dream) and informal (glue) bayesian approaches in hydrologic modeling? *Stochast. Environ. Res. Risk Assess.* 23, 1011–1026. doi: 10.1007/s00477-008-0274-y
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychon. Bull. Rev.* 14, 779–804. doi: 10.3758/BF03194105
- Walker, W., Harremoës, P., Rotmans, J., van der Sluijs, J., van Asselt, M., Janssen, P., et al. (2003). Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integr. Assess.* 4, 5–17. doi: 10.1076/iaij.4.1.5.16466
- Walker, W. E., Lempert, R. J., and Kwakkel, J. H. (2013). *Deep Uncertainty*, 395–402. Boston, MA: Springer.
- Wellmann, J. F., de la Varga, M., Murdie, R. E., Gessner, K., and Jessell, M. (2018). Uncertainty estimation for a geological model of the sandstone greenstone belt, western australia - insights from integrated geological and geophysical inversion in a bayesian inference framework. *Geol. Soc. Lond. Spec. Publ.* 453, 41–56. doi: 10.1144/SP453.12
- Wellmann, J. F., and Regenauer-Lieb, K. (2012). Uncertainties have a meaning: information entropy as a quality measure for 3-d geological models. *Tectonophysics* 526–529(Suppl. C), 207–216. doi: 10.1016/j.tecto.2011.05.001
- Williamson, J. (2010). *In Defence of Objective Bayesianism*. Oxford: Oxford University Press.
- Wojda, P., Brouyère, S., Derouane, J., and Dassargues, A. (2010). Hydrocube: an entity-relationship hydrogeological data model. *Hydrogeol. J.* 18, 1953–1962. doi: 10.1007/s10040-010-0653-6
- Woodbury, A. D., and Ulrych, T. J. (1993). Minimum relative entropy: forward probabilistic modeling. *Water Resour. Res.* 29.
- Woodbury, A. D., and Ulrych, T. J. (1996). Minimum relative entropy inversion: theory and application to recovering the release history of a groundwater contaminant. *Water Resour. Res.* 32, 2671–2681.
- Woodbury, A. D., and Ulrych, T. J. (2000). A full-bayesian approach to the groundwater inverse problem for steady state flow. *Water Resour. Res.* 36, 2081–2093. doi: 10.1029/2000WR900086
- Xu, T., Valocchi, A. J., Ye, M., and Liang, F. (2017). Quantifying model structural error: efficient Bayesian calibration of a regional groundwater flow model using surrogates and a data-driven error model. *Water Resour. Res.* 53, 4084–4105. doi: 10.1002/2016WR019831
- Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.* 1, 3–28.
- Zech, A., Müller, S., Mai, J., Heße, F., and Attinger, S. (2016). Extending Theis' solution: using transient pumping tests to estimate parameters of aquifer heterogeneity. *Water Resour. Res.* 52, 6156–6170. doi: 10.1002/2015WR018509
- Zhang, Y.-K., and Zhang, D. (2004). Forum: the state of stochastic hydrology. *Stochast. Environ. Res. Risk Assess.* 18, 265. doi: 10.1007/s00477-004-0190-8
- Zhdanov, M. S. (2015). *Inverse Theory and Applications in Geophysics*. Amsterdam: Elsevier Science.
- Zheng, C. and Gorelick, S. M. (2003). Analysis of solute transport in flow fields influenced by preferential flowpaths at the decimeter scale. *Ground Water* 41, 142–155. doi: 10.1111/j.1745-6584.2003.tb02578.x
- Zinn, B., and Harvey, C. F. (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 Resour. Res.* 39. doi: 10.1029/2001WR001146

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Heße, Comunian and Attinger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Groundwater Contaminant Transport: Prediction Under Uncertainty, With Application to the MADE Transport Experiment

Aldo Fiori^{1*}, Antonio Zarlenga¹, Alberto Bellin², Vladimir Cvetkovic³ and Gedeon Dagan⁴

¹ Department of Engineering, Roma Tre University, Rome, Italy, ² Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy, ³ Department of Water Resources Engineering, Royal Institute of Technology, Stockholm, Sweden, ⁴ School of Mechanical Engineering, Tel Aviv University, Ramat Aviv, Israel

OPEN ACCESS

Edited by:

Daniel M. Tartakovsky,
Stanford University, United States

Reviewed by:

Gerardo Severino,
University of Naples Federico II, Italy
Felipe De Barros,
University of Southern California,
United States

*Correspondence:

Aldo Fiori
aldo@uniroma3.it

Specialty section:

This article was submitted to
Freshwater Science,
a section of the journal
Frontiers in Environmental Science

Received: 13 December 2018

Accepted: 21 May 2019

Published: 06 June 2019

Citation:

Fiori A, Zarlenga A, Bellin A,
Cvetkovic V and Dagan G (2019)
Groundwater Contaminant Transport:
Prediction Under Uncertainty, With
Application to the MADE Transport
Experiment. *Front. Environ. Sci.* 7:79.
doi: 10.3389/fenvs.2019.00079

Transport of solutes in porous media at the laboratory scale is governed by an Advection Dispersion Equation (ADE). The advection is by the fluid velocity U and dispersion by $D_{dL} = U\alpha_{dL}$, where the longitudinal dispersivity α_{dL} is of the order of the pore size. Numerous data revealed that the longitudinal spreading of plumes at field scale is characterized by macrodispersivity α_L , larger than α_{dL} by orders of magnitude. This effect is attributed to heterogeneity of aquifers manifesting in the spatial variability of the logconductivity Y . Modeling Y as a stationary random field and for mean uniform flow (natural gradient), α_L could be determined in an analytical form by a first order approximation in σ_Y^2 (variance of Y) of the flow and transport equations. Recently, models and numerical simulations for solving transport in highly heterogeneous aquifers ($\sigma_Y^2 > 1$), primarily in terms of the mass arrival (the breakthrough curve BTC), were advanced. In all cases ergodicity, which allows to exchange the unknown BTC with the ensemble mean, was assumed to prevail for large plumes, compared to the logconductivity integral scale. Besides, the various statistical parameters characterizing the logconductivity structure as well as the mean flow were assumed to be known deterministically. The present paper investigates the uncertainty of the non-ergodic BTC due to the finiteness of the plume size as well as due to the uncertainty of the various parameters on which the BTC depends. By the use of a simplified transport model we developed in the past (which led to accurate results for ergodic plumes), we were able to get simple results for the variance of the BTC. It depends in an analytical manner on the flow parameters as well as on the dimension of the initial plume relative to the integral scale of logconductivity covariance. The results were applied to the analysis of the uncertainty of the plume spatial distribution of the MADE transport experiment. This was achieved by using the latest, recent, analysis of the MADE aquifer conductivity data.

Keywords: solute transport, heterogeneous porous formations, breakthrough curve (BTC), uncertainty, MADE experiment, stochastic subsurface hydrology

1. INTRODUCTION

Aquifers pollution by various contaminants constitutes a major threat to fresh water resources all over the world. Unlike the accessible surface water bodies, groundwater pollution is detectable by wells which cover a limited zone and often respond after a large portion of the aquifer is already contaminated. Furthermore, the process is slow, occurring over periods of tens of years and cleaning by natural attenuation or remediation, whenever possible, is also slow. Under these circumstances, transport mathematical models, which may help analyzing field findings on one hand and predicting solute future spreading on the other, are of crucial importance.

There are various modes of quantification of transport. In this study we focus on characterization of solute plumes longitudinal spreading by the BTC (breakthrough curve) $M(x, t)$ at vertical control planes located at longitudinal distance $x = \text{const}$, normal to the mean flow direction. An alternative and related measure is the longitudinal mass distribution $m(x, t)$ as function of distance x , for a given time t . We limit the scope to inert solutes (tracers) and to constant mean head gradient J (natural gradient flow), a setup of interest for many applications and an essential first step toward analysis of more complex configurations. For the benefit of the reader not familiarized with the groundwater transport theory, we recapitulate in the following a few essential developments.

The traditional modeling of transport was based on column laboratory experiments (Bear, 1979) for which the macroscopic (at the pore, Darcy, scale) concentration $C(x, t)$ satisfies the advection dispersion equation (ADE)

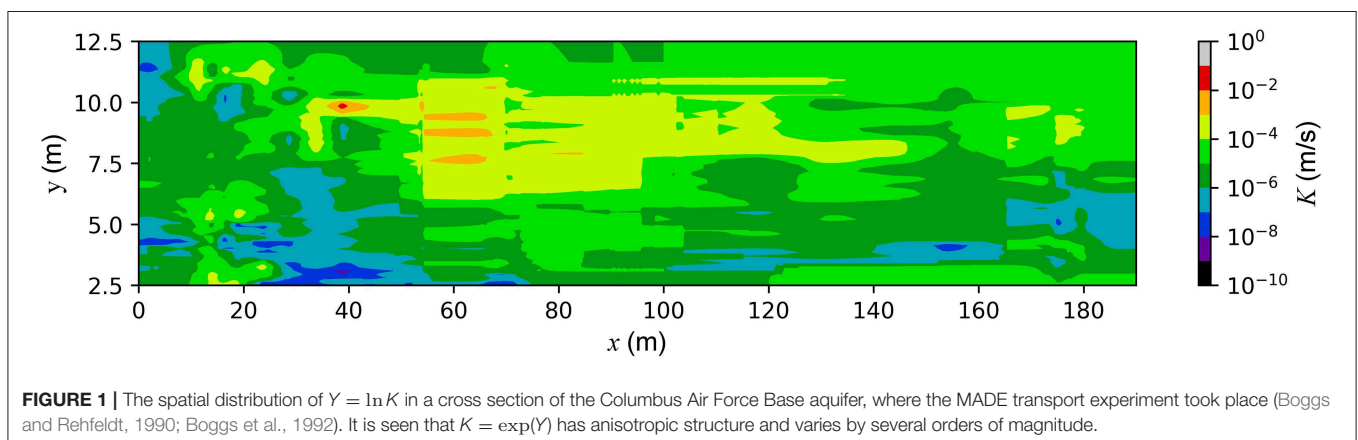
$$\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial x} = D_{dL} \frac{\partial^2 C}{\partial x^2} \quad (1)$$

where $U = q/n$ is the macroscopic flow velocity (q -specific discharge, n -effective porosity) at the Darcy's scale and D_{dL} is the longitudinal dispersion coefficient. For the large Peclet number $Pe_0 = Ud/D_0$ (d -pore scale, D_0 -molecular diffusion coefficient) encountered in applications it was found that $D_{dL} = \alpha_{dL} U$, where α_{dL} is the pore scale dispersivity, of order d (Bear, 1979).

Field findings (for a recent compendium see Zech et al., 2015) have revealed that the longitudinal α_L (derived for instance with the aid of the spatial moments of aquifer plumes) is larger than α_{dL} by orders of magnitude; α_L was coined as longitudinal macrodispersivity in the literature. The contrast has been attributed to the impact of aquifers heterogeneity, manifesting primarily in the spatial variation of the hydraulic conductivity $K(x)$. For illustration we present in **Figure 1** the spatial distribution of $Y = \ln K$ in a cross section of the Columbus Air Force Base aquifer, where the MACRO Dispersion Experiment (MADE) took place (Boggs and Rehfeldt, 1990; Boggs et al., 1992). It was obtained by interpolating among the measured values provided by Bohling et al. (2012) at a relatively dense set of points. A few significant features of the aquifer in **Figure 1** are worth mentioning: (i) K varies by orders of magnitude, (ii) the spatial distribution of K is seemingly erratic and difficult to be captured by smooth interpolators, (iii) the zones of different K are elongated in the horizontal direction, the aquifer being coined as anisotropic at the field scale (notice that for clarity of representation the scale of reduction is smaller in the vertical direction with respect to the horizontal one, such that these zones are more elongated than how they appear in the figure).

The above features have a dramatic impact upon the spreading of plumes of solutes. For illustration we represent in **Figure 2** the concentration spatial distribution obtained by a numerical solution of the advective transport equation (in 2D). The velocity field $\mathbf{V}(\mathbf{x})$ was derived by a numerical solution of the flow equation for a K field statistically similar to that of **Figure 1**, with logconductivity variance $\sigma_Y^2 = 6$, under conditions of constant mean head gradient J . The transport equation was solved by using a Smooth Particle Hydrodynamic (SPH) scheme, which is virtually free of numerical diffusion (Boso et al., 2013), for $Pe = I/\alpha_{dL} = 1000$ (I is the longitudinal integral scale of logconductivity).

A few qualitative features of distribution of C in **Figure 2** are: (i) for an initial rectangular pulse of constant $C = C_0$, the plume becomes highly fragmented with time and of progressive spreading, (ii) the plume splits, with quicker advancing “fingers” in zones of high K , and practically stagnant solute in regions of



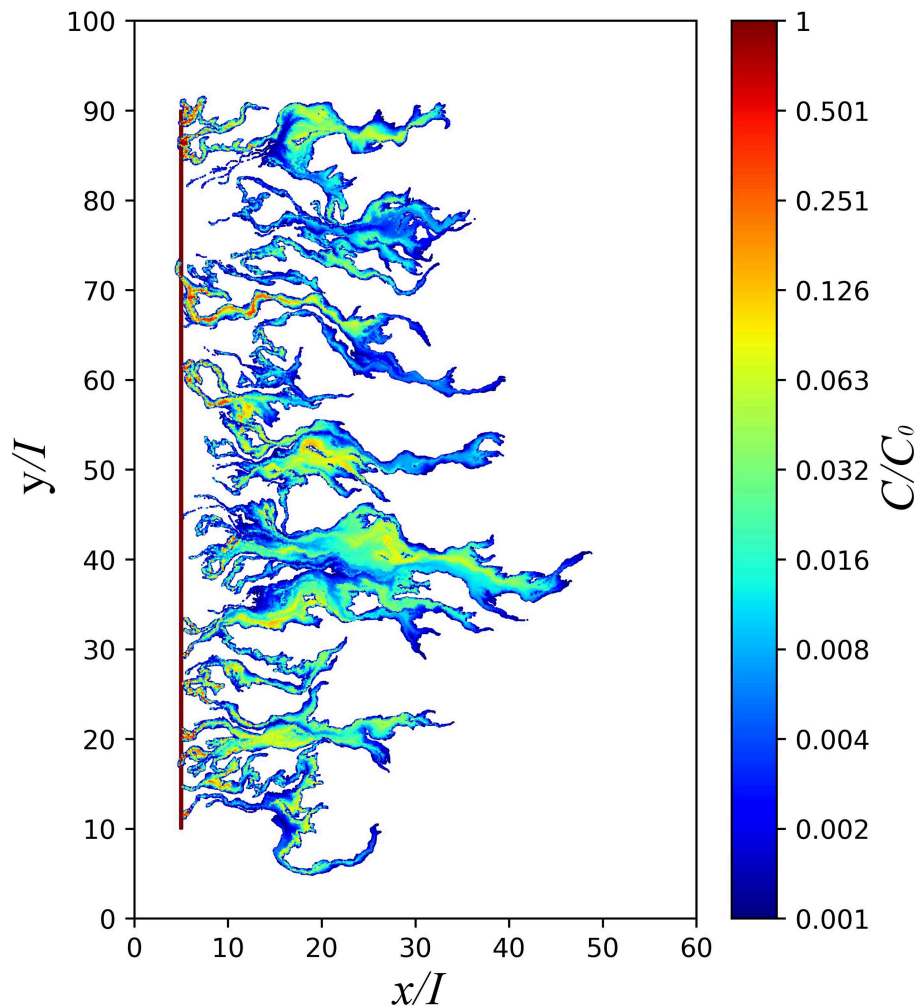


FIGURE 2 | Numerically simulated concentration field (in 2D) after an instantaneous injection within a thin (linear) source perpendicular to mean flow (indicated by a red vertical line at $x/I = 5$, with I the longitudinal integral scale of logconductivity), of dimensions $L_y = 80I$ and $L_x = 0.25I$ at time $tU/I = 10$; the other parameters are $\sigma_Y^2 = 6$ and $Pe = I/\alpha_{dL} = 1000$. The plume is seemingly erratic, with clear fingering patterns and stagnant regions.

low K , (iii) like the K field (see **Figure 1**) the plume is seemingly erratic and defies a representation by smooth functions but, at the same time, it makes the identification of point concentration at a given location an elusive goal, (iv) in contrast, global measures like the mass arrival at vertical planes over the entire domain (the BTC M) smooth out the variations and the extent of spreading can be quantified for instance by α_L , (v) it was found that the presence of the pore scale dispersion (primarily the vertical one) causes mixing which affects the local C , but has a minor effect on M (Fiori and Dagan, 2000) and (vi) space averages like M are the ones of interest in many applications, e.g., those in which the goal is to determine the mass of solute pumped by wells which intercept the plume (Fiori et al., 2016).

This state of affairs has motivated the emergence around 40 years ago (for a review see for instance the books by Dagan, 1989; Gelhar, 1993; Rubin, 2003) of a new discipline, namely stochastic subsurface hydrology. In its frame the hydraulic

logconductivity field $Y(\mathbf{x})$ is modeled as a stationary random space function whose univariate distribution and two point covariance are characterized by a few parameters: the geometric mean K_G , the variance σ_Y^2 and the horizontal I and the vertical $I_v < I$ integral scales, respectively. As a consequence, the steady Eulerian velocity $\mathbf{V}(\mathbf{x})$, is also a random space function, of constant mean $\mathbf{U}(U, 0, 0)$. The statistics of \mathbf{V} are obtained by solving the flow equations for conditions of constant mean head gradient $\mathbf{J}(J, 0, 0)$ and random $K(\mathbf{x})$. Similarly the concentration field $C(\mathbf{x}, t)$ is random and so are its global measures like the BTC $M(x, t)$. The latter is obtained by solving the transport equation with advection by the random \mathbf{V} and dispersion by the local pore scale dispersion tensor. Unfortunately, heterogeneity renders point concentration $C(\mathbf{x}, t)$ highly uncertain with a coefficient of variation that is controlled by pore-scale dispersion and reduces slowly with time (Fiori and Dagan, 2000). Uncertainty reduces considerably when global measures are used, such as $M(x, t)$,

or the sampling volume has dimensions comparable with the integral scales of Y , thereby making the ensemble mean $\langle C(\mathbf{x}, t) \rangle$ a reliable and robust measure, similarly to M (Bellin et al., 1994; Tonina and Bellin, 2008). If the point concentration is of interest, such as in risk assessment for example, uncertainty can be reduced by focusing on the probability that a given concentration threshold is exceeded, irrespective of the position where this occurs, rather than focusing on a given fixed location where the ensemble mean concentration provided a unreliable estimate (Bellin and Tonina, 2007). Another option is by conditioning on available data (e.g., Y , head, concentration etc), for which, however, several conditioning points are needed in order to significantly reduce uncertainty. In any case, as previously stated our study focuses on the BTC M which is relevant to many applications and is quite robust and less prone to uncertainty than point concentration.

A common adopted assumption is that transport is ergodic in the sense that M (or similar global attributes) in a realization can be exchanged with the ensemble mean $\langle M \rangle$. This is a basic tenet in many branches of physics and engineering, e.g., molecular diffusion driven by Brownian motion or effective properties of composite materials. It is justified by the large contrast between the microscopic length scales and the macroscopic ones of interest in applications. For groundwater transport ergodicity implies that the solute plume samples a sufficiently large aquifer volume compared to the integral scales so as to encounter zones of various K , representative of the entire population. Since typically I and I_v are of the order of meters and solute plumes of tens of meters, the contrast is not so large and ergodicity may be not obeyed and prediction of M by models is subjected to uncertainty. This feature differentiates transport by groundwater from the traditional pursuit of the effective properties solely, prevalent in the literature on heterogeneous media.

Similarly, it is common to assume in applications that the various parameters and variables like J , U , K_G , σ_Y^2 , I are known. In reality, in the field they are only estimated and subjected to uncertainty as well, impinging on the uncertainty of M .

The aim of the present study is to provide a discussion of uncertainty in modeling transport in three-dimensional heterogeneous aquifers, with application to the MADE transport experiment (Zheng et al., 2011) as a platform for discussion; we summarize what we have learned in the last two decades or so, in view of applications, with a particular focus on uncertainty due to plume sampling (i.e., non-ergodicity) and incomplete knowledge of parameters that are both important for MADE. A novel analytical formulation is also proposed for assessing uncertainty due to lack of ergodicity by the plume. It is emphasized that this is not a review paper and we build primarily on our developments in modeling flow and transport in three-dimensional heterogeneous formations.

The plan of the paper is as follows: section 2 provides an overview of concepts development and paper aims, recapitulating some of our recent developments in transport of ergodic plumes; section 3 addresses the modeling of uncertainty in the prediction of the BTC, the main topic of the paper; section 4 presents the application of the uncertainty analysis to the MADE-1

experiment, relying on the latest published data; finally, section 5 summarizes and concludes the study.

2. BACKGROUND AND MATHEMATICAL PRELIMINARIES

2.1. The K Structure

As already mentioned above, we limit the study to stationary random $Y(\mathbf{x})$. For sedimentary formations of concern here, the histogram of $Y = \ln K$ was found to fit a normal univariate distribution $f(Y)$, of mean $\langle Y \rangle = \ln K_G$ and variance σ_Y^2 (see for instance the analysis of MADE data by Fiori et al., 2015; Bohling et al., 2016) which we adopt here. At the lowest order, the spatial structure is captured by the two point covariance $C_Y(\mathbf{x}_1, \mathbf{x}_2) = \sigma_Y^2 \rho(\mathbf{r})$ where ρ is the autocorrelation and $\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2$ is the distance vector between the two points respect to which the covariance is computed. In turn, the assumed axisymmetric ρ is characterized by the finite horizontal I and vertical I_v integral scales, respectively. The Y field is often assumed to be multivariate normal (multi-Gaussian), and then the structure is completely characterized by $\langle Y \rangle$ and C_Y . This is a very convenient representation which is commonly adopted in numerical simulations or analytical approximations. For other types of structures higher multipoint correlations are required for a complete statistical characterization. Thus, in our past numerical simulations of 3D flow and transport (see for a summary our recent paper Jankovic et al., 2017) we have generated a few fields which share the same $f(Y)$ and I , I_v but differ at higher order as manifested for instance by the spatial connectivity of zones defined by classes of K . Thus, we considered besides the multivariate normal (multi-Gaussian) one, two Y fields devised by Zinn and Harvey (2003), obtained by transformations which led to more connected or disconnected zones of large conductivity. Besides, we investigated extensively flow and transport in a structure we coined as MIM (Multi-Indicator-Model): rectangular blocks which tessellate the space or spheroidal inclusions of dimensions $2I$ in the horizontal and $2I_v$ in the vertical directions, respectively, and are of independent $K = \exp(Y)$. Unlike the three previous ones, the connectivity of different classes of K values is the same.

It is worthwhile to note that even for thoroughly monitored aquifers like that of MADE, field data do not generally allow for determining statistical parameters beyond $f(Y)$ and ρ , i.e., K_G , σ_Y^2 , I , and I_v and even those are estimated within ranges of values.

2.2. Flow

We consider here steady flow governed by

$$\begin{aligned} \mathbf{q} &= -K \nabla h \text{ (Darcy's law),} \\ \nabla \cdot \mathbf{q} &= 0 \text{ (continuity)} \rightarrow \nabla \cdot (K \nabla h) = 0 \end{aligned} \quad (2)$$

where $h(\mathbf{x})$ is the head. With the assumed constant porosity n (which is much less variable than K), the velocity \mathbf{V} also satisfies $\nabla \cdot \mathbf{V} = 0$. The boundary condition of interest here is the one of constant head, such that the head gradient has components $\mathbf{J}(J, 0, 0)$. Consequently the mean velocity $\mathbf{U}(U, 0, 0)$ is also constant and its fluctuation $\mathbf{u}(\mathbf{x})$ is stationary. The relationship

$\mathbf{U} = K_{ef}\mathbf{J}/n$ defines the effective conductivity and its dependence on the structural parameters is the subject of a vast literature (see e.g., Renard and De Marsily, 1997).

Analytical approximate solutions of the statistics of the h and \mathbf{V} fields, as well as K_{ef} , were obtained in the literature (e.g., Dagan, 1989; Rubin, 2003) by a first order approximation in σ_Y^2 , presumably valid for weak heterogeneity. In contrast, based on numerical simulations, with values of the logconductivity variance up to $\sigma_Y^2 = 8$, we have recently presented (Zarlenga et al., 2018) the dependence of the K_{ef} components on σ_Y^2 for different $e = I_v/I$ and for the above different structures, as well as by the first order approximation. A striking result is that the first order approximation of the horizontal component K_{efh} is quite accurate for $\sigma_Y^2 < 2$, which reinforces what already observed for the velocity covariance function and the spatial plume moments in an early works by Bellin et al. (1992) and Salandin and Fiorotto (1998).

2.3. Transport (General)

We consider injection of a solute over an area A_0 in the plane at $x = 0$. A thin plume with total mass M_0 is modeled as a Dirac pulse of infinitesimal duration $\delta(t)$ at $x = 0$ (extension to arbitrary spatial or temporal initial distributed plumes is straightforward); following the definitions of Kreft and Zuber (1978) injection may be in resident or flux proportional mode. Thus, the initial mass density $m_0 = dM_0/d\mathbf{b}$ is constant and equal to M_0/A_0 for uniform initial resident concentration while it is given by the variable and random $m_0(\mathbf{b}) = [V_0(\mathbf{b})/\bar{V}_0](M_0/A_0)$ for the flux proportional one. Here $\mathbf{b}(b_y, b_z)$ is a coordinate in A_0 , $V_0(\mathbf{b}) = V_x(0, b_y, b_z)$ and $\bar{V}_0 = (1/A_0) \int_{A_0} V_0(\mathbf{b}) d\mathbf{b}$ is the mean velocity over A_0 , which for sufficiently large A_0 is equal to the mean velocity U .

We adopt the boundary condition of flux proportional initial condition which applies to many cases e.g., injection by wells as it was the case for field experiments including MADE. It simply states that solutes initially occupy preferential zones of high conductivity. We also adopt the detection in the flux proportional mode (Kreft and Zuber, 1978), i.e., the BTC defined by $M(x, t) = M_0 - \int_0^x \int \int nC(x, y, z, t) dx dy dz = \int_0^t dt \int \int nV_x(x, y, z) C(x, y, z, t) dy dz$, where integration in y, z is over the cross section of the domain. The issue related to injection and detection conditions and their impact on transport was extensively discussed in past work (see e.g., Kreft and Zuber, 1978; Dagan, 2017; Fiori et al., 2017). If it satisfies the ADE (1) with initial condition $M(0, t) = M_0 H(t)$ (where H is the Heaviside step function), the solution for the semi-bounded domain is given by the CDF of the Inverse Gaussian (IG) distribution (Kreft and Zuber, 1978):

$$\frac{M(x, t)}{M_0} = \frac{1}{2} \left\{ \operatorname{erfc} \left[\frac{x - Ut}{2(D_L t)^{1/2}} \right] + \exp \left(\frac{Ux}{D_L} \right) \operatorname{erfc} \left[\frac{x + Ut}{2(D_L t)^{1/2}} \right] \right\} \quad (3)$$

with D_L the longitudinal macrodispersion coefficient, whereas the relative mass flux is given by the Inverse Gaussian (hereinafter

IG) distribution

$$\frac{\mu(x, t)}{M_0} = \frac{1}{M_0} \frac{\partial M}{\partial t} = \frac{x}{2(\pi D_L t^3)^{1/2}} \exp \left[-\frac{(x - Ut)^2}{4D_L t} \right] \quad (4)$$

In the frame of random walk transport theory, the Inverse Gaussian distribution pertains to a first arrival process (see, e.g., Redner, 2001), i.e., the detection plane at x serves as an absorbing boundary. Note that IG is a special case of the more general Tempered One-Sided Stable distribution (TOSS) with exponent $1/2$ (Cvetkovic, 2011).

We proceed now with reviewing the results we obtained recently for ergodic transport in heterogeneous aquifers.

2.4. Transport (Summary of Results for Ergodic Plumes)

The starting point for our recent developments are the systematic accurate numerical simulations (see Jankovic et al., 2017, and reference therein) of flow and transport in 3D; they are recapitulated in the Appendix of Jankovic et al. (2017) and only briefly here. The K field was generated for a lognormal univariate distribution and two point covariances C_Y of integral scales I and I_v , with different values of the anisotropy coefficient $e = I_v/I$. The complete characterization of the structures was achieved by a variety of different models: multi-Gaussian, the previously mentioned connected and disconnected fields, spheroidal inclusions and rectangular blocks tessellating the space (MIM). The BTC was finally calculated by large-scale numerical simulations, for a variety of parameters (logconductivity variance, Peclet number, control plane distances etc.).

A striking result (Jankovic et al., 2017, Figure 3) was that for the various structures the bulk of M (say $M/M_0 < 0.95$) did not differ significantly among structures, proving indeed that M is a very robust measure. Furthermore, the simple model

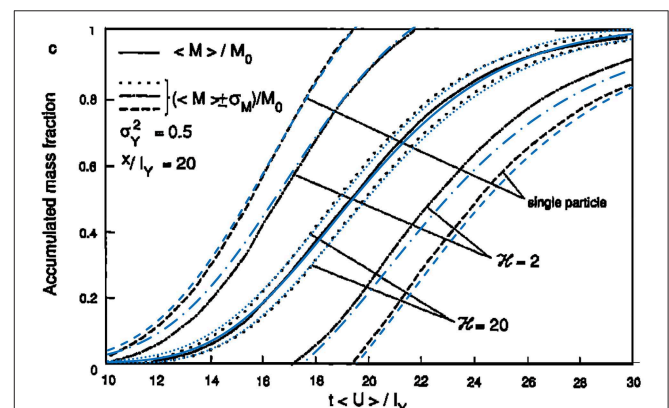


FIGURE 3 | Comparison of the proposed solution (11) with the results of Cvetkovic et al. (1992) (Figure 5C), for $\sigma_Y^2 = 0.5$, $x/l = 20$ and a few values of the size of the initial plume ($L_y/I = L_z/I = \gamma$). The figure displays the cumulative mass M and the bands $M \pm \sigma_M$ predicted by the present approach (blue lines) over the original Figure 5C of Cvetkovic et al. (1992).

(3), with the macrodispersivity given by the well-known first-order approximation $D_L = \alpha_L U$, $\alpha_L = \sigma_Y^2 I$, agreed well with the bulk of the BTC derived numerically (Fiori et al., 2017) while it underestimated the late arrival time of the tail of a few percents of M . However, the tail prediction is anyway quite imprecise. In particular, it was found that the IG model behaves similar to the MIMSCA (Multi Indicator Model Self Consistent Approximation) that we developed in the last 20 years (e.g., Dagan et al., 2003; Fiori et al., 2007), that is more accurate for the prediction of late mass arrival.

Thus, the IG model (3) is quite effective in capturing the behavior of the bulk of the BTC, for a wide range of flow and transport parameters; the model depends on a few parameters characterizing the permeability structure (σ_Y^2, I) and the flow (J, K_{eff}). It is emphasized that the first-order approximation was applied to deriving the longitudinal macrodispersivity α_L while $M(x, t)$ (3) itself depends non-linearly on σ_Y^2 and is different from the Gaussian distribution; the two coalesce for small σ_Y^2 or large tU/I .

Summarizing, the simple formula (3), with D_L given by the first-order approximation, is a very robust model that can be safely used in applications, e.g., as a screening tool for a preliminary assessment of the BTC. We note that a similar, simple approach, although with parameters based on numerical simulations, was proposed by Hansen et al. (2018).

We move now to the central topic of the present study, the uncertainty of prediction of the BTC.

3. UNCERTAINTY OF BTC PREDICTIVE MODELING

3.1. A Few Sources of Uncertainty

Transport predictions by the above modeling approach is prone to several sources of uncertainty, the major ones being:

- (i) Uncertainty in medium characterization, and its representation in general. Field data are generally scarce and even the identification of fundamental and basic quantities like K_G (geometric mean of K) and σ_Y^2 is difficult and error prone. This is even more true for higher order measures, like e.g., the covariance (or variogram) and I , based on the assumption of stationarity which is also hard to validate. As a matter of fact, multipoint correlations are needed for a full characterization of K , which is a prohibitive task for applications (Boso and Tartakovsky, 2016). Alternative models like facies identification by using the indicator variogram (Ritzi, 2000) requires identification of zones of connectivity of K classes, which again is feasible only for highly monitored aquifers. Fortunately, global measures like $\langle M \rangle$ are quite robust in 3D and adopting for instance the common multi-Gaussian structure model does not affect significantly the bulk of the BTC (Bianchi and Pedretti, 2017, 2018).
- (ii) Uncertainty due to the limited domain sampled by the plume. Thus, the initial plume may not be so large, as discussed in the previous section, and non-ergodicity issues may emerge (Kitanidis, 1988; Dagan, 1990; Andrićević and

Cvetković, 1998; Attinger et al., 1999). As a consequence, the quantities of interest, including M , are random, and uncertain due to non-ergodic behavior emerges. As it was found in the past and confirmed by the developments of the following section, uncertainty for small plumes can be quite large. Instead, the uncertainty related to the size of the sampling volume (Fiori et al., 2002; Bellin and Tonina, 2007; Severino et al., 2010) is not relevant for the transport scenario investigated here in which solute is detected at a large control plane at distance x from the source.

- (iii) Uncertainty in the modeling approach. This is a highly debated issue in the literature, and various models have been advanced during the years. However, our recent developments summarized in the previous section, based on accurate and systematic 3D numerical simulations which were not available in the literature for the high σ_Y^2 values, show that for global, upscaled, measures like $\langle M \rangle$, the ADE with upscaled first-order longitudinal macrodispersivity $\alpha_L = \sigma_Y^2 I$, solved under the conditions of flux proportional injection and detection, offers a simple and quite accurate, physically based solution.
- (iv) Parametric uncertainty. Besides the uncertainty of the adopted K model, the same is true for the other hydrological parameters controlling transport, like e.g., the mean velocity. Parametric uncertainty is expected to have an important impact on predictions because of the limitations of measurements and the significant impact of parameters, like U .
- (v) Additional uncertainty may stem from unsteadiness, spatial non-uniformity of the mean head gradient, partial knowledge of the plume initial condition etc. These sources of uncertainty are rather case specific and less prone to a general analysis (see e.g., Bellin et al., 1996; see e.g., Dagan et al., 1996).

According to the above discussion, the major sources of uncertainty for transport in mean uniform flow are likely (ii) and (iv), i.e., possible non-ergodicity of the plume and parametric uncertainty; this is particularly true for the MADE experiment, as shown in the sequel. Thus, in the following we shall focus on the uncertainty quantification originating from (ii) and (iv). Furthermore, we shall apply the uncertainty analysis to MADE, that is a well-known benchmark, very useful for a thorough discussion on uncertainty in applications.

3.2. Quantifying Uncertainty Due to Non-ergodic Effect

3.2.1. General

We follow here the theoretical framework and the notations of Cvetkovic et al. (1992) and Dagan et al. (1992). Focusing on the BTC M we may write for flux proportional injection over an area A_0

$$\frac{M(t; x)}{M_0} = \frac{1}{A_0} \int_{A_0} \frac{V_0(\mathbf{b})}{\bar{V}_0} H[t - \tau(x, \mathbf{b})] d\mathbf{b} \quad (5)$$

where we remind that V_0 is the local velocity at the location \mathbf{b} within A_0 and $\bar{V}_0 = (1/A_0) \int_{A_0} V_0(\mathbf{b}) d\mathbf{b} \approx U$. It is also reminded

that in (5) H is the Heaviside step function and $\tau(x, \mathbf{b})$ is the random travel time to the control plane at x of a fluid particle injected at $x = 0$ within A_0 . The travel time is related to the random velocity field by $d\tau/dx = (V_x[x, \eta(x), \zeta(x)])^{-1}$ where $y = \eta(x, \mathbf{b})$, $z = \zeta(x, \mathbf{b})$ are the equations of a streamline originating at \mathbf{b} within the plane A_0 at $x = 0$. In words, $(V_0(\mathbf{b})/\bar{V}_0) H[t - \tau(x, \mathbf{b})](d\mathbf{b}/A_0)$ in Equation (5) marks the contribution of the particle originating at \mathbf{b} to the mass that crossed the control section at x until the time $t = \tau$.

The expected mass $\langle M(t; x) \rangle$ is easily calculated from (5), yielding

$$\frac{\langle M(t; x) \rangle}{M_0} = \frac{1}{A_0} \left\langle \int_{A_0} \frac{V_0(\mathbf{b})}{\bar{V}_0} H[t - \tau(x, \mathbf{b})] d\mathbf{b} \right\rangle = G_1(t; x) \quad (6)$$

where $G_1 = \int_0^t f(\tau) d\tau$, is the cumulative travel time distribution, where τ is weighted by the injection velocity V_0 (Cvetkovic and Dagan, 1994). Here $f(\tau) = \int (V_0) f(\tau, V_0) d(V_0)$ is the marginal pdf of τ , with $f(\tau, V_0)$ being the joint pdf of τ and V_0 .

Formula (6) is the well-known result (Shapiro and Cvetkovic, 1988) that the mean mass arrival is the CDF of travel time. Furthermore, in view of the findings of section 2.3, it is seen that $G_1(\tau)$ is the CDF of the Inverse Gaussian distribution Equation (3) and the pdf $g_1 = dG_1/d\tau$ is the Inverse Gaussian (4).

After recapitulating these preparatory steps we move now, along Dagan et al. (1992), to the derivation of the variance of M

$$\frac{\sigma_M^2(t; x)}{M_0^2} = \left\langle \frac{M^2(t; x)}{M_0^2} \right\rangle - \frac{\langle M(t; x) \rangle^2}{M_0^2} \quad (7)$$

Considering the expression (5) for the mass M we may write

$$\left\langle \frac{M^2(t; x)}{M_0^2} \right\rangle = \frac{1}{A_0^2} \left\langle \int_{A_0} \int_{A_0} \frac{V_0(\mathbf{b}) V_0(\mathbf{b}')}{\bar{V}_0^2} H[t - \tau(x, \mathbf{b})] H[t - \tau(x, \mathbf{b}')] d\mathbf{b} d\mathbf{b}' \right\rangle \quad (8)$$

Thus, by taking advantage of the linearity of the ensemble mean operator and considering Equation (6) we arrive at

$$\frac{\sigma_M^2(t; x)}{M_0^2} = \frac{1}{A_0^2} \int_{A_0} \int_{A_0} G_2(t; x, \mathbf{b} - \mathbf{b}') d\mathbf{b} d\mathbf{b}' - G_1^2(t; x) \quad (9)$$

where G_2 is the bivariate travel time CDF, which is given by $G_2(t; x, \mathbf{b} - \mathbf{b}') = \int_0^t \int_0^t g_2(\tau, \tau'; x, \mathbf{b} - \mathbf{b}') d\tau d\tau'$, with $g_2(\tau, \tau'; x, \mathbf{b} - \mathbf{b}')$ being the marginal joint pdf of travel times τ, τ' of two particles injected at \mathbf{b} and \mathbf{b}' , respectively.

The general result (9) by Dagan et al. (1992) has served Cvetkovic et al. (1992) to effectively compute $\sigma_M^2(x, t)$ by adopting a few assumptions: the bivariate g_2 is lognormal, the travel time moments were derived from the velocity field by a first order approximation in σ_V^2 , A_0 is a square. Two quadratures, which were carried out numerically, were needed to complete the derivation.

The above approach can be generalized to compute the covariance of M (C_M) at two different times t_1, t_2 , leading to

$$\frac{C_M(t_1, t_2; x)}{M_0^2} = \frac{1}{A_0^2} \int_{A_0} \int_{A_0} G_2(t_1, t_2; x, \mathbf{b}' - \mathbf{b}'') d\mathbf{b}' d\mathbf{b}'' - G_1(t_1; x) G_1(t_2; x) \quad (10)$$

with $G_2(t_1, t_2; x, \mathbf{b} - \mathbf{b}') = \int_0^{t_1} \int_0^{t_2} g_2(\tau, \tau'; x, \mathbf{b} - \mathbf{b}') d\tau d\tau'$.

3.2.2. Simplified Derivation of $\sigma_M^2(x, t)$ and Comparison With Numerical Simulations

The derivation of the two particles covariance needed in (9) is complex and requires additional information, like e.g., the shape of the two particles covariance and its moments. Also, the calculation of σ_M^2 along (9) requires a few numerical quadratures, as done by Cvetkovic et al. (1992). We simplify the calculations by using the basic properties of the MIMSCA model which, as mentioned above, led to very good agreement with the numerical solution of $\langle M \rangle$.

Consider the covering of the input area A_0 by rectangles of sides $2I$ and $2I_v$ in the y and z directions, respectively. Following the MIMSCA model, the travel time of particles originating within an areal element in A_0 is the same whereas they are statistically independent for particles originating from different elements. This is the only property we use in the present derivation.

With the above assumption, the calculation of (9) can be considerably simplified. The detailed derivations are given in **Appendix A**, leading to the final result (A.8), that is reproduced here

$$\sigma_M^2 = M_0^2 \omega(\mathbf{L}) G_1(t; x) [1 - G_1(t; x)] \quad (11)$$

where the weight function ω is given by

$$\omega(\mathbf{L}) = \Omega\left(\frac{L_y}{I}\right) \Omega\left(\frac{L_z}{I_v}\right) \quad (12)$$

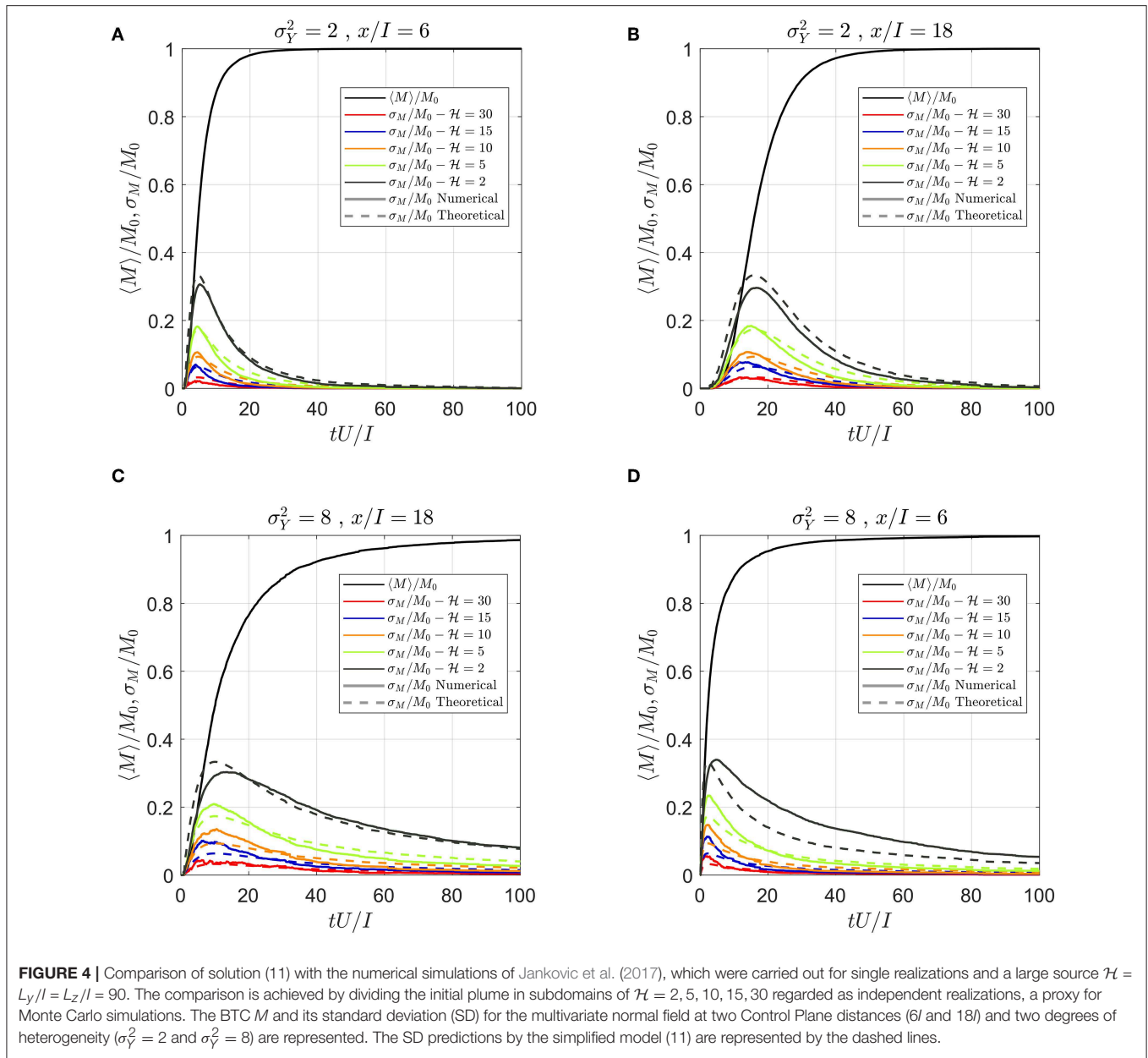
with L_y, L_z the sides of the rectangular injection area, i.e., $A_0 = L_y L_z$ and

$$\Omega(\ell) = \begin{cases} \left(1 - \frac{\ell}{6}\right) & \text{for } \ell < 2 \\ \frac{1}{\ell} \left(2 - \frac{4}{3\ell}\right) & \text{for } \ell > 2 \end{cases} \quad (13)$$

In particular $\omega \simeq 1$ for $A_0/(I I_v) \ll 1$ (maximal uncertainty for small source) and $\omega \simeq (I I_v/A_0)$ for $A_0/(I I_v) \gg 1$ (ergodic, practically deterministic).

Thus, σ_M^2 is given by an analytical expression supposed to apply to highly heterogeneous formations which separates the effect of spreading represented by the IG G_1 (3) with $D_L = \sigma_V^2 I U$, one hand, and the weight function ω accounting for the size of the injection area on the other hand.

As a first test of (11) we compare it in **Figure 3** with the results of Cvetkovic et al. (1992) described above ($\sigma_V^2 = 0.5$,



$x/I = 20$, $L_y/I = L_z/I = \mathcal{H}$) and it is seen that the agreement is very good in spite of the different methodologies. A more stringent test is carried out by comparison with the numerical simulations of Jankovic et al. (2017), which were carried out for single realizations for the large $\mathcal{H} = L_y/I = L_z/I = 90$. This was achieved by dividing A_0 in subdomains of $\mathcal{H} \simeq 2, 5, 10, 15, 30$ regarded as independent realizations, a proxy for Monte Carlo simulations. The results are displayed in **Figure 4**, where both the BTC and its standard deviation (SD) for the multivariate normal field at two Control Plane distances (6 I and 18 I) are represented. The SD predictions by the simplified model (11) are also displayed (dashed lines). It is seen that, in spite of the limited number of realizations for some cases (9 for $L_y = L_z = 30I$) the agreement is quite good, even for

the largest $\sigma_Y^2 = 8$. The behavior is very similar for other K structures examined (e.g., connected/disconnected and blocks, as described in section 2.4; not shown in the figure). Thus, the simple model proposed here can be an effective tool for the prediction of the BTC uncertainty due to the non-ergodic effect (i.e., finite size of the plume compared to the heterogeneity length scales).

It is worthwhile noting that for a small plume (ω close to unity) uncertainty affects its time of arrival at the control plane (de Barros et al., 2011; de Barros, 2018). In contrast, for a large plume the practically deterministic prediction reflects the spreading of the BTC. For intermediate cases the two effects are combined, and they are incorporated in the simple function ω (12).

3.3. Impact of Parametric Uncertainty

Even under the basic assumptions of stationarity, steadiness and mean flow uniformity the mean BTC (Equation 3 with $D_L = \sigma_Y^2 IU$) depends on a few parameters. Thus, $U = JK_{ef}/n$ depends, besides the mean gradient J , on K_G and σ_Y^2 since K_{ef}/K_G depends on σ_Y^2 as well as on the structure (Renard and De Marsily, 1997). Of course the knowledge of these parameters is not needed if K_{ef} is determined for instance directly by pumping tests and/or U by flowmeters. In all cases these parameters are affected by uncertainty due to measurement errors, insufficient data etc. The same is true for the parameters influencing the mean BTC namely σ_Y^2 and I besides U . Finally, the uncertainty quantified by σ_M^2 (11) depends on the additional parameter A_0 , reflecting the initial size of the plume.

The uncertainty of these parameters impacts that of M in addition to the non-ergodic effect. However, the magnitude depends on the availability of data and their precision, which is aquifer specific. In the following section we shall examine the impact of parametric uncertainty for the Columbus Air Force Base aquifer where the MADE experiment took place and for which a relatively large amount of data is available.

Nevertheless, we may make a few statements on the relative impact of various parameters based on the numerical simulations and theoretical developments. Thus, as mentioned above, Fiori et al. (2017) and Jankovic et al. (2017) have already found that $\langle M \rangle$ is quite insensitive to the structure (as characterized by connectivity), whereas U , and to a lesser extent heterogeneity σ_Y^2 , has a larger impact. Rather than a general discussion we defer the analysis to the MADE case in the following.

4. ANALYSIS OF THE MASS DISTRIBUTION AT THE MADE-1 EXPERIMENT

The first experiment conducted at the Columbus Air Force Base (MADE-1) represents the ideal platform for discussing the above issues regarding uncertainty of solute transport predictions in aquifers. In terms of the quantity and quality of investigations, regarding both aquifer characterization and plume monitoring, the MADE-1 experiment represents a benchmark for analyzing groundwater transport; it has motivated a large body of research work, from the testing of innovative measuring techniques to the development of novel theoretical frameworks. For such reasons, after more than 30 years, MADE is still providing insight and topics of discussion in the scientific community, as witnessed for instance by the recent 2015 AGU Chapman conference (Gómez-Hernández et al., 2016). Before discussing uncertainty, along the previous lines, we briefly recapitulate in the following the main features of the MADE-1 experiment, as well as the flow and transport parameters that shall be used in the present work, together with their uncertainty measures. More details can be found in the original papers (Boggs and Rehfeldt, 1990; Boggs et al., 1992) and the review (Zheng et al., 2011).

The experiment took place in a highly heterogeneous sedimentary aquifer at Columbus, Ohio (USA). A plume was injected in a relatively small area of the domain, and the plume movement, along the natural gradient, was

TABLE 1 | Mean and standard deviation (SD) for the MADE-1 parameters employed in the present work.

Symbol	Parameter	Mean	SD
σ_Y^2	Logconductivity variance [–] (*)	5.9	0.77
I	Horizontal integral scale [m] (*)	9.1	1.94
I_V	Vertical integral scale [m] (*)	1.8	0.33
K_G	Geometric mean of hydraulic conductivity [m/d] (*)	0.58	0.13
n	Porosity [–] (**)	0.31	0.08
J	Mean hydraulic gradient [–] (**)	0.0036	–
U	Mean velocity [m/d] (estimated)	0.026	0.009

Data are taken from Bohling et al. (2016) (*) and Boggs et al. (1992) (**); in the case of the data from Bohling et al. (2016), the SD was calculated from their 95 % confidence intervals after assuming a normal distribution for all parameters except K_G , for which lognormality was assumed. The SD for the mean velocity was estimated from Darcy formula $U = K_{ef}J/n$, assuming K_{ef}/K_G and J as deterministic.

continuously monitored for about 2 years by a dense network of multilevel samplers. The relevant transport quantity that was analyzed at the MADE-1 experiment is the longitudinal mass distribution $m(x; t)$, which was first analyzed and presented by Adams and Gelhar (1992). Six snapshots were analyzed, at $t = 49, 126, 202, 279, 370, 503$ days since injection. The mass distribution was derived by calculating the solute mass (after interpolation of concentration measurements) within a moving window of 10 m length, at spatial intervals of 5 m. The striking feature of m is its skewed shape, very much different from the presumed symmetrical Gaussian behavior, that was mainly caused by the highly heterogeneous velocity field induced by the complex aquifer system; such feature has motivated subsequently a flurry of theoretical developments to explain it. Despite this dense grid of samplers, mass recovery was incomplete, except for the snapshot at $t = 126$ d, and the mass recovery continuously decreased after it, down to 43% in the last snapshot at $t = 503$ d [the topic is discussed by Fiori (2014)].

In the following, the longitudinal mass distribution at MADE is modeled by the aid of the model (3). Following Adams and Gelhar (1992), the mass distribution is calculated within a moving window of $\Delta = 10$ m, i.e., $m(x; t) = (M(x + \Delta/2; t) - M(x - \Delta/2; t)) / \Delta$, with space intervals of 5 m ($x = 0, 5, 10, 15, \dots$). The parameters to be used in the model were inferred from different studies and are presented in **Table 1**; the standard deviation (SD) is also reproduced, when available. Of particular relevance is the analysis of Bohling et al. (2016) of the K values based on DPIL measurements, that superseded a previous analysis by same authors (Bohling et al., 2012). This study analyzed the conductivity field at an unprecedented detail and resolution, and constitute the best available conductivity analysis of MADE data so far.

The mean velocity is calculated by $U = K_G J / n$, with K_G, J, n, ϵ the geometric mean of K , the mean hydraulic gradient, the mean porosity and the effective conductivity ratio $\epsilon = K_{ef}/K_G$, respectively. Unfortunately, ϵ cannot be measured and it is variable, as a function of the particular conductivity structure at hand (the matter is discussed in Zarlenga et al., 2018). An

estimate of ϵ was provided by the formula (5) derived by Zarlenga et al. (2018) based on extensive 3D numerical simulations, obtaining $\epsilon = 3.93$; this results in the estimated $U = 0.026$ m/d. The SD of U is calculated by a perturbation approach over the parameters K_G, n , hence assuming J and ϵ as deterministic; as a consequence, the standard deviation of U appearing in **Table 1** is likely underestimated.

4.1. Prediction of Mass Distribution and Uncertainty Due to Non-ergodic Effects

Before embarking on the analysis of the spatial mass distribution (the snapshots), it is worthwhile to estimate the non-ergodic effect on the uncertainty of the BTC $M(x, t)$, although the latter was not determined experimentally. According to Equation (11) the maximal value of σ_M/M_0 is obtained by differentiating with respect to t and is reached for $G_1 = 1/2$. This leads to $(\sigma_M/M_0)_{\max} = \omega^{1/2}/2$. We show in the sequel that for MADE plume initial size the estimate is $\omega = 0.148$, i.e., $(\sigma_M/M_0)_{\max} = 0.19$. Thus, the width of the band $\langle M \rangle/M_0 \pm \sigma_M/M_0$ reaches its maximum at the time for which $\langle M \rangle/M_0 = 1/2$ and its size is ± 0.19 , diminishing to zero for $\langle M \rangle/M_0 \rightarrow 0, 1$. However, a direct comparison with the experimental results (the snapshots) needs reformulation in terms of spatial distribution.

Along the lines of section 4, the longitudinal mass distribution at MADE is given by

$$m(x; t) = \frac{M(x + \Delta/2; t) - M(x - \Delta/2; t)}{\Delta} \quad (14)$$

where m is the longitudinal mass distribution aggregated over the spatial interval $\Delta = 10$ m.

The expected value and variance of m for non-ergodic plumes can be derived with the same procedure of section 3.2.2, and detailed in **Appendix A**. The detailed derivations are given in **Appendix B**, and we reproduce here the final result

$$\begin{aligned} \frac{\langle m(x; t) \rangle}{M_0} &= \frac{G_1(x + \Delta/2; t) - G_1(x - \Delta/2; t)}{\Delta} \quad (15) \\ \sigma_m^2(x; t) &= \omega(L) \langle m(x; t) \rangle \left(\frac{M_0}{\Delta} - \langle m(x; t) \rangle \right) \end{aligned}$$

In (15), G_1 is given by (3), as previously explained, although any alternative model can be used. As discussed in Fiori et al. (2017), the time of the snapshots at the MADE experiment was very small in dimensionless terms, namely $tU/I = 0.15 - 1.0$, posing doubts regarding the use of a constant D_L in (3). Therefore, the pre-asymptotic D_L , as predicted by the first order approximation, was employed here; the issue is discussed in Fiori et al. (2017) and it is further elaborated in **Appendix C**, leading to the revised formula (C.3), for the travel time CDF G_1 that shall be used in the present analysis.

The formula (15) for σ_m^2 requires the vertical and transverse dimensions of the initial plume, L_y and L_z . The distances between the injection wells and the width of the screens in the wells are not reliable estimates of L_y and L_z as the plume underwent a significant expansion in all directions soon after the injection, as visible in the early snapshots. Figure 6a of Adams and Gelhar (1992) shows that after 9 days the vertical size of the plume was

around 8 m, much larger than the vertical size of the screen of the wells. Also, Figure 4a from the same paper suggests that, again after 9 days from the injection event, the size of the plume was already about 40 m wide. Thus, in the following we assume $L_y = 40$ m and $L_z = 8$ m as the initial sizes of the plume; such estimates are rather rough and uncertain, but there is no other way to accurately assess them. With those estimates of L_y, L_z , and those of I, I_v of (1), the variance reduction factor due to the finite size of the plume appearing in (15) is $\omega(L) = 0.148$ (Equations 12, 13).

Figure 5 displays the experimental longitudinal mass distribution at the MADE-1 experiment for the six snapshots presented by Adams and Gelhar (1992); black lines; the blue solid line depicts the theoretical $m(x; t)$ (Equation 15), while the dashed lines represent the bounds $m + \sigma_m$ (green) and $m - \sigma_m$ (orange). It is seen that the theoretical model captures quite well the experimental mass distribution at MADE; the result is not entirely new as a similar comparison was made in Fiori et al. (2017), although the updated estimates of Bohling et al. (2016) and the more accurate spatial aggregation over the Δ interval was made here for the first time. The direct comparison between experiment and theory is made difficult by the overestimation of mass in the first snapshot (around 200% of the injected mass was recovered) and the incomplete mass recovery for increasing time. Still, the model captures the peak and its timing quite accurately in most of the snapshots, all the approximations and uncertainties notwithstanding.

It is seen that the bands of uncertainty, described by the bounds $m \pm \sigma_m$ (dashed lines), are rather wide for all snapshots (note that m is subject to the constraint $\int_0^\infty m(t)dt = 1$). The bounds tend to increase with time; as a matter of fact the behavior is quite expected in view of the nature of the analytical solution (15): the broader is the distribution, the larger is the uncertainty. The wide bounds of uncertainty pose doubts regarding the applicability of analytical solutions, based on stochastic approaches, that implicitly assume ergodicity; in such cases, it is advisable to present results together with the bands of uncertainty, as done here. Although the bands can be rather wide, like the present MADE case, the representation of **Figure 5** in terms of prediction and bands of uncertainty may be of definite help in applications, for instance the case of risk assessment and plume management.

Uncertainty can in principle be constrained by some conditioning of the solution, e.g., based on available K data. Still, the impact of conditioning permeability at a point is generally limited to a domain of the order of the integral scales of Y (see e.g., Dagan, 1985), which has a minor effect on M for a large plume unless the grid of measurements is dense and covers the advancing plume. For small plumes conditioning may be more effective in reducing uncertainty if the measurements grid covers the trajectory. In any case, conditioning requires a theoretical model more complex than (3), posing additional computational burden that may not be otherwise required for simple preliminary (screening) analysis.

The relative good agreement of m (solid blue line) with experiments is quite surprising in view of the large uncertainty, as represented by the upper and lower limits (the dashed lines) of

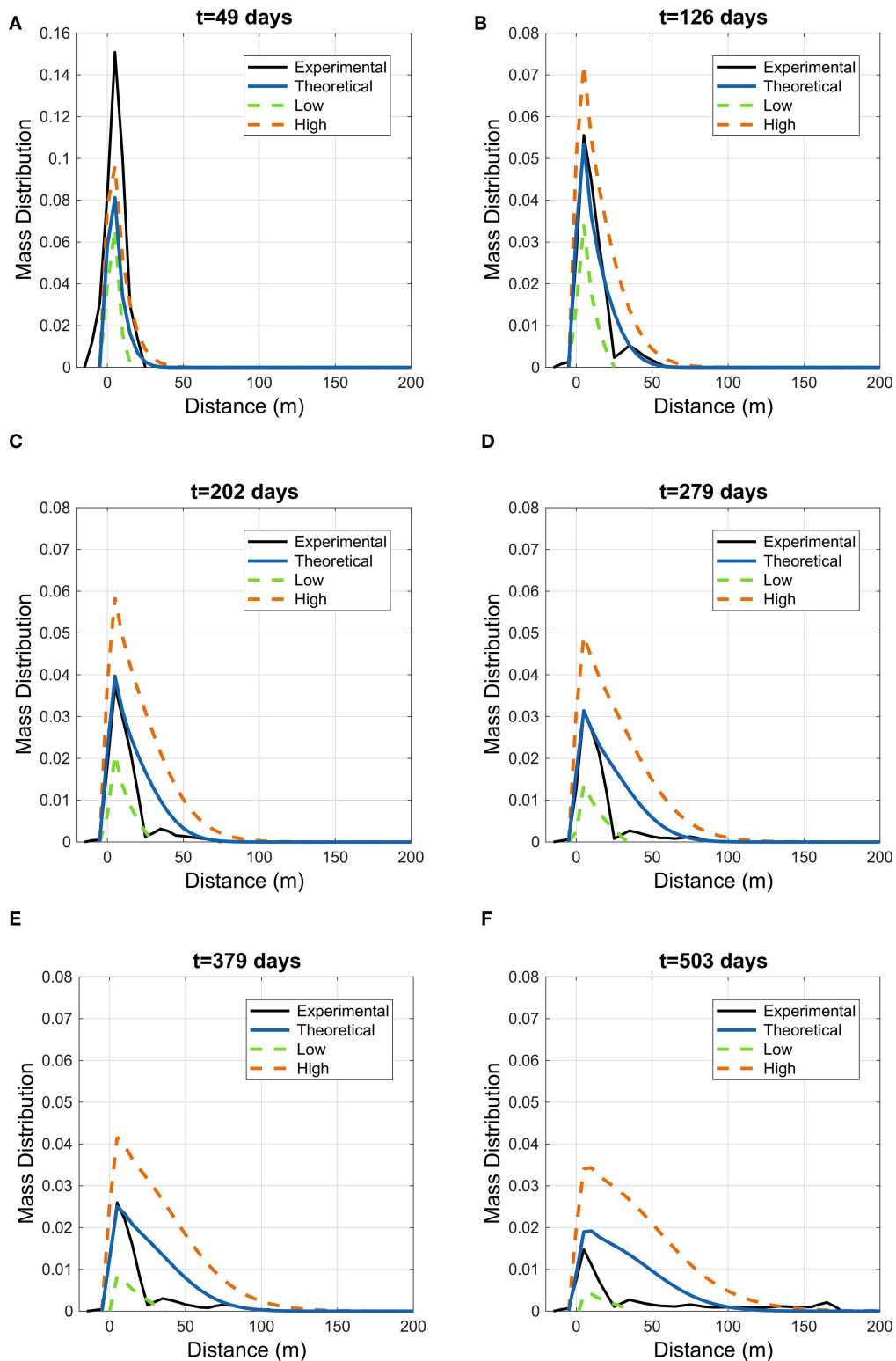


FIGURE 5 | Uncertainty due to non-ergodic effects: the experimental longitudinal mass distribution at the MADE-1 experiment for the six snapshots presented by Adams and Gelhar (1992); black lines); the blue solid line depicts the theoretical $m(x; t)$ (15), while the dashed lines represent the bounds for non-ergodicity uncertainty $m + \sigma_m$ (orange dashed line, "High") and $m - \sigma_m$ (green dashed line, "Low"). G_1 in solutions (15) is given by (C.3). Mass recovery in the experiments was 206, 99, 68, 62, 54, and 43% at snapshots $t = 49, 126, 202, 279, 370, 503$ days, respectively.

the figures; it may suggest that the initial size of the plume was indeed large enough to adequately sample the range of velocity variations in the aquifer, hence more in favor of ergodicity, and the estimates of L_y and L_z (that we recall were roughly estimated from the profiles of Adams and Gelhar, 1992) might perhaps be too conservative.

4.2. Parameter Uncertainty

Parameter uncertainty impact is assessed here by a simple first order analysis, that provides the variance of mass distribution m due to parametric uncertainty

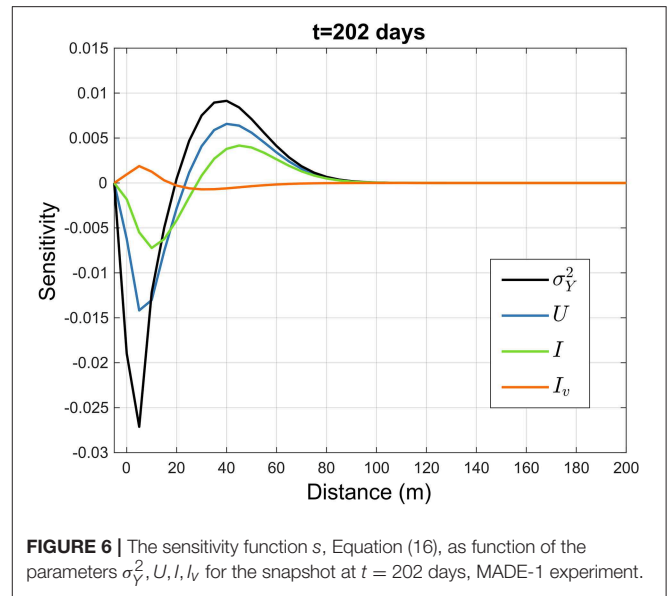
$$\sigma_m^2 = \sum_{i=1}^{N_p} \left(p_i \frac{\partial m}{\partial p_i} \right)^2 CV_{p_i}^2 = \sum_{i=1}^{N_p} s^2(p_i) CV_{p_i}^2 \quad (16)$$

where the function s is the sensitivity, and it represents the relative variation of the solution to changes in the generic parameter p_i ($i = 1, \dots, N_p$). The procedure is justified by the relatively small coefficient of variation of the parameters, that is below 0.35 for all parameters (see **Table 1**).

It is instructive to analyze first the sensitivity function $s(p_i)$, (Equation 16), as function of the generic parameter p_i . **Figure 6** illustrates the sensitivity pertaining to the relevant parameters σ_Y^2, U, I, I_v for the snapshot $t = 202$ d (the sensitivities for the other snapshots are similar). It is seen that the sensitivity displays an antisymmetric behavior, which is determined by the constraint that the area underneath the curve m is unitary. Hence, increasing a parameter has opposite effects in different segments of the mass distribution. The behavior is similar for all parameters except I_v , that contributes through the anisotropy ratio (see **Appendix C**) and hence has opposite effects with respect to I . The curves of **Figure 6** indicate that the most relevant parameter for uncertainty is the mean velocity U , followed by the logconductivity variance σ_Y^2 and the horizontal integral scale I ; the impact of the vertical scale I_v is rather small. This finding already suggests what are the parameters requiring a more careful and precise estimate in order to reduce uncertainty, with the mean velocity playing an important role; the issue was also mentioned in section 3.1.

The bands of parametric uncertainty $m \pm \sigma_m$, along the model (16), are represented in **Figure 7** for the six snapshots of the MADE-1 experiment. Comparison with **Figure 5** indicates that the parametric uncertainty effect is smaller than the one due to non-ergodic behavior (section 3.2). We remind, however, that the bands may be wider as the variance of the mean velocity U is expected to be larger than the one estimated here, and reproduced in **Table 1**; this issue was discussed in section 4.1. The behavior of the uncertainty bands observed in **Figure 7**, with a central area where the bands shrink, is easily explained by the antisymmetric shape of the sensitivity, as previously discussed. The width of the uncertainty bands increases with time, just like the case illustrated in **Figure 5**.

The analysis of parametric uncertainty is indeed a first and relative easy (if data are available) estimate of possible prediction errors. However, as shown here, it may be not the main source of uncertainty. It is worth noting that Cvetkovic et al. (2015) discussed the global sensitivity including mass transfer reactions.



5. SUMMARY AND CONCLUSIONS

Spreading of solute plumes in aquifers, as quantified for instance by the longitudinal macrodispersivity α_L , is much larger than the one observed in laboratory experiments (pore scale dispersion). This enhancement is caused by the spatial variability of the conductivity K , which in the context of stochastic subsurface hydrology, is modeled as a random space function. The paper considers flow which is uniform in the mean (natural gradient flow of velocity U) and inert solutes. Transport is quantified by the BTC $M(t, x)$ at control planes at x as well as the associated spatial longitudinal mass distribution $m(x, t)$. The logconductivity $Y = \ln K$ is modeled as stationary, of normal univariate pdf (parameters K_G and σ_Y^2) and axisymmetric covariance of horizontal I and vertical I_v integral scales. The latter are much larger than the pore scale, which explains the above findings. Flow and transport variables, solutions of the flow and transport equations, are consequently random as well.

Most of the transport models developed in the past, aiming at prediction of M , were underlain by the ergodic hypothesis, valid for plumes of large extent at the I, I_v scales. As a consequence, the one realization M is approximately equal to the ensemble mean $\langle M \rangle$.

The paper investigates the uncertainty of M (or m) in three dimensional formations, as quantified by the variance σ_M^2 . Among the various sources of uncertainty, we deal with two: primarily with the non-ergodic effect present for finite plumes, as encountered in many applications. Besides we consider the impact of uncertainty of parameters like U, K_G, σ_Y^2, I . Indeed, the latter are affected by measurement errors even in extensively monitored aquifers.

The non-ergodic effect on transport was investigated in the past by adopting the first-order approximation in σ_Y^2 in solving the flow and transport equations (weakly heterogeneous formations). One of our main aims here is to extend the analysis to highly heterogeneous aquifers for which $\sigma_Y^2 \leq 8$. The

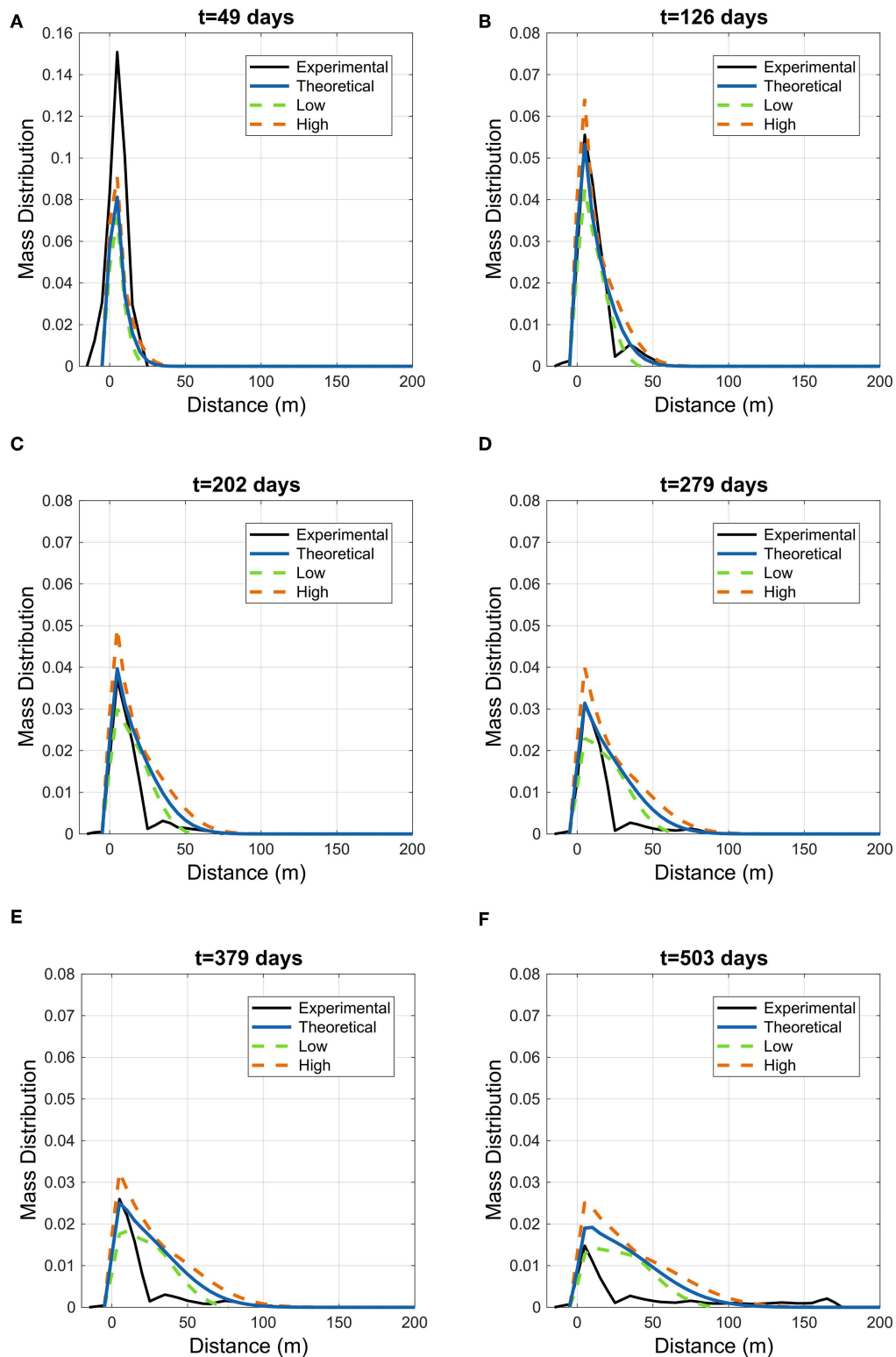


FIGURE 7 | Parametric uncertainty: the experimental longitudinal mass distribution at the MADE-1 experiment for the six snapshots presented by Adams and Gelhar (1992), black lines; the blue solid line depicts the theoretical $m(x; t)$ (15), while the dashed lines represent the bounds for parametric uncertainty $m + \sigma_m$ (orange dashed line, "High") and $m - \sigma_m$ (green dashed line, "Low"). G_1 in solutions (15) is given by (C.3).

investigation is based on our work in the last 15 years on ergodic transport, both by extensive and accurate numerical simulations not available in the past for three-dimensional configurations, as well as by simplified models. This was done for a few types of heterogeneous structures, differing in the connectivity of K classes. One of our main results was that $\langle M \rangle$ is a very robust predictor whose bulk can be modeled by the Inverse Gaussian CDF, with travel time variance given by the first order approximation.

The main novel theoretical contribution is the development of a simple analytical model to compute σ_M^2 . It combines the above Inverse Gaussian distribution with an analytical function of the size of the injected plume relative to the integral scale, covering the spectrum from small plumes to ergodic ones. The result is illustrated by depicting the bands of uncertainty delimited by $\langle M \rangle \pm \sigma_M$ (Figure 4) as dependent on σ_Y^2 . While the present contribution is limited to non-reactive transport, the methodology can be easily extended to reactive solutes, along the lines of Cvetkovic and Dagan (1994) and Fiori et al. (2002).

A major part of the paper is devoted to application of the concepts to the MADE aquifer ($\sigma_Y^2 \simeq 6$) transport experiment, which has become a platform for groundwater contaminant transport modeling in the last 30 years. We present the observed snapshots of $m(x, t)$, functions of x for a few values of t , as well as the bands of uncertainty related both to non-ergodic effects and uncertainty of parameters. The analysis relies on published recent analysis of field data, based on renewed characterization campaigns, and it represents a major overhaul of our previous analyses of the MADE-1 experiment. The results indicate that the most relevant parameter for uncertainty is the mean velocity U , followed by the logconductivity variance σ_Y^2 and the horizontal integral scale I . This finding suggests what are the parameters requiring a more careful and precise estimate in order to reduce uncertainty.

The main conclusion of the study is that, even for thoroughly characterized aquifers (like MADE) prediction of transport is affected by uncertainty; in particular, the major source of uncertainty for the MADE-1 experiment seems to be the non-ergodic behavior, i.e., the finite size of the plume with respect to the directional correlation scales of hydraulic conductivity. Uncertainty is prone to be even greater for the common, less detailed, sites data available in practice.

The above finding, and the general argumentation brought by the present work, enforces the conclusions from past work that estimating uncertainty of prediction should become an integral part of solving aquifer contamination problems, toward risk analysis. To that aim, characterization efforts should be directed toward reducing uncertainty of most influential parameters like the mean velocity U . Due to the prevailing scarcity of data in practice, it is advisable to use simple models, at least for screening scenarios.

The envisaged main future developments which may contribute to uncertainty and risk reduction are 2-fold. On one hand improvement of characterization technology may provide a detailed and large volume of data which may need analysis relying on Big Data treatment approach. On the other hand, numerical models of flow and transport in which the detailed aquifer architecture is based on conditioning on the large number of data should also be devised. At present, simple models like the ones presented here may serve for preliminary and screening analysis.

AUTHOR CONTRIBUTIONS

AF: derivation of the new analytical expression of variance reduction due to non-ergodic effects; application to the MADE transport experiment using updated field data; exploration of the impact of parameters uncertainty on that of the MADE experiment plume snapshots. AZ: calculation of the variance reduction and related figures; computation of the comparisons of the BTC variance; detailed computation of the updated solution of the mean mass spatial distribution at MADE Site and comparison with the experimental snapshots. AB: computation of the logconductivity profiles at the MADE (Figure 1); numerical simulation of a 2D flow and transport with MADE parameters values (Figure 2); discussion of Figures 1, 2 to serve as starting point for the new developments of the paper. VC: establishing the connection of the results of the present study with those derived previously in the literature; comparison of the new results with the previous ones in the literature; relating the methodology of the paper to random walk theory. GD: coordination of the preparation of the study; establishing and description of the connection with previous work of the authors; formulation and writing of the final text while blending the various contributions of the team members and formulating the conclusions.

ACKNOWLEDGMENTS

The authors are grateful to Francesca Boso for the numerical simulations of Figure 2. AB, AF, and AZ acknowledge funding from the Italian Ministry of Education, University and Research (MIUR) in the frame of the Departments of Excellence Initiative 2018–2022 granted to DICAM of the University of Trento (AB) and the Department of Engineering of Roma Tre University (AF and AZ).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fenvs.2019.00079/full#supplementary-material>

REFERENCES

- Adams, E. E., and Gelhar, L. W. (1992). Field study of dispersion in a heterogeneous aquifer: 2. Spatial moments analysis. *Water Resour. Res.* 28, 3293–3307.
- Andričević, R., and Cvetković, V. (1998). Relative dispersion for solute flux in aquifers. *J. Fluid Mech.* 361, 145–174.
- Attinger, S., Dentz, M., Kinzelbach, H., and Kinzelbach, W. (1999). Temporal behaviour of a solute cloud in a chemically heterogeneous porous medium. *J. Fluid Mech.* 386, 77–104.
- Bear, J. (1979). *Hydraulics of Groundwater*. New York, NY: McGraw-Hill, Inc.
- Bellin, A., Dagan, G., and Rubin, Y. (1996). The impact of head gradient transients on transport in heterogeneous formations: application to the borden site. *Water Resour. Res.* 32, 2705–2713.
- Bellin, A., Rubin, Y., and Rinaldo, A. (1994). Eulerian-lagrangian approach for modeling of flow and transport in heterogeneous geological formations. *Water Resour. Res.* 30, 2913–2924.
- Bellin, A., Salandini, P., and Rinaldo, A. (1992). Simulation of dispersion in heterogeneous porous formations: Statistics, first order theories, convergence of computations. *Water Resour. Res.* 28, 2211–2227.
- Bellin, A., and Tonina, D. (2007). Probability density function of non-reactive solute concentration in heterogeneous porous formations. *J. Contamin. Hydrol.* 94, 109–125. doi: 10.1016/j.jconhyd.2007.05.005
- Bianchi, M., and Pedretti, D. (2017). Geological entropy and solute transport in heterogeneous porous media. *Water Resour. Res.* 53, 4691–4708. doi: 10.1002/2016WR020195
- Bianchi, M., and Pedretti, D. (2018). An entrogram-based approach to describe spatial heterogeneity with applications to solute transport in porous media. *Water Resour. Res.* 54, 4432–4448. doi: 10.1029/2018WR022827
- Boggs, J., and Rehfeldt, K. R. (1990). *Field Study of Macrodispersion in a Heterogeneous Aquifer. 2. Observations of Hydraulic Conductivity Variability*. Technical Report 10308. Available online at: https://inis.iaea.org/collection/NCLCollectionStore/_Public/25/009/25009041.pdf?r=1&r=1
- Boggs, J. M., Young, S. C., Beard, L. M., Gelhar, L. W., Rehfeldt, K. R., and Adams, E. E. (1992). Field study of dispersion in a heterogeneous aquifer: 1. Overview and site description. *Water Resour. Res.* 28, 3281–3291.
- Bohling, G. C., Liu, G., Dietrich, P., and Butler, J. J. (2016). Reassessing the MADE direct-push hydraulic conductivity data using a revised calibration procedure. *Water Resour. Res.* 52, 8970–8985. doi: 10.1002/2016WR019008
- Bohling, G. C., Liu, G., Knobbe, S. J., Reboulet, E. C., Hyndman, D. W., Dietrich, P., et al. (2012). Geostatistical analysis of centimeter-scale hydraulic conductivity variations at the MADE site. *Water Resour. Res.* 48:2525. doi: 10.1029/2011WR010791
- Boso, F., Bellin, A., and Dumbser, M. (2013). Numerical simulations of solute transport in highly heterogeneous formations: a comparison of alternative numerical schemes. *Adv. Water Resour.* 52, 178–189. doi: 10.1016/j.advwatres.2012.08.006
- Boso, F., and Tartakovsky, D. M. (2016). The method of distributions for dispersive transport in porous media with uncertain hydraulic properties. *Water Resour. Res.* 52, 4700–4712. doi: 10.1002/2016WR018745
- Cvetkovic, V. (2011). The tempered one-sided stable density: a universal model for hydrological transport? *Environ. Res. Lett.* 6:034008. doi: 10.1088/1748-9326/6/3/034008
- Cvetkovic, V., and Dagan, G. (1994). Transport of kinetically sorbing solute by steady random velocity in heterogeneous porous formations. *J. Fluid Mech.* 265, 189–215.
- Cvetkovic, V., Shapiro, A. M., and Dagan, G. (1992). A solute flux approach to transport in heterogeneous formations: 2. Uncertainty analysis. *Water Resour. Res.* 28, 1377–1388.
- Cvetkovic, V., Soltani, S., and Vigouroux, G. (2015). Global sensitivity analysis of groundwater transport. *J. Hydrol.* 531, 142–148. doi: 10.1016/j.jhydrol.2015.07.035
- Dagan, G. (1985). Stochastic modeling of groundwater flow by unconditional and conditional probabilities: the inverse problem. *Water Resour. Res.* 21, 65–72. doi: 10.1029/WR021i001p00065
- Dagan, G. (1989). *Flow and Transport in Porous Formations*. Berlin: Springer-Verlag.
- Dagan, G. (1990). Transport in heterogeneous porous formations: spatial moments, ergodicity, and effective dispersion. *Water Resour. Res.* 26, 1281–1290.
- Dagan, G. (2017). Solute plumes mean velocity in aquifer transport: impact of injection and detection modes. *Adv. Water Resour.* 106, 6–10. doi: 10.1016/j.advwatres.2016.09.014
- Dagan, G., Bellin, A., and Rubin, Y. (1996). Lagrangian analysis of transport in heterogeneous formations under transient flow conditions. *Water Resour. Res.* 32, 891–899.
- Dagan, G., Cvetkovic, V., and Shapiro, A. (1992). A solute flux approach to transport in heterogeneous formations: 1. The general framework. *Water Resour. Res.* 28, 1369–1376.
- Dagan, G., Fiori, A., and Janković, I. (2003). Flow and transport in highly heterogeneous formations: 1. Conceptual framework and validity of first-order approximations. *Water Resour. Res.* 39. doi: 10.1029/2002WR001717
- de Barros, F. P. (2018). Evaluating the combined effects of source zone mass release rates and aquifer heterogeneity on solute discharge uncertainty. *Adv. Water Resour.* 117, 140–150. doi: 10.1016/j.advwatres.2018.05.010
- de Barros, F. P. J., Fiori, A., and Bellin, A. (2011). A simple closed-form solution for assessing concentration uncertainty. *Water Resour. Res.* 47:W12603. doi: 10.1029/2011WR011107
- Fiori, A. (2014). Channeling, channel density and mass recovery in aquifer transport, with application to the MADE experiment. *Water Resour. Res.* 50, 9148–9161. doi: 10.1002/2014WR015950
- Fiori, A., Berglund, S., Cvetkovic, V., and Dagan, G. (2002). A first-order analysis of solute flux statistics in aquifers: the combined effect of pore-scale dispersion, sampling, and linear sorption kinetics. *Water Resour. Res.* 38, 121–1215. doi: 10.1029/2001WR000678
- Fiori, A., Cvetkovic, V., Dagan, G., Attinger, S., Bellin, A., Dietrich, P., et al. (2016). Debates-stochastic subsurface hydrology from theory to practice: the relevance of stochastic subsurface hydrology to practical problems of contaminant transport and remediation. What is characterization and stochastic theory good for? *Water Resour. Res.* 52, 9228–9234. doi: 10.1002/2015WR017525
- Fiori, A., and Dagan, G. (2000). Concentration fluctuations in aquifer transport: a rigorous first-order solution and applications. *J. Contam. Hydrol.* 45, 139–163. doi: 10.1016/S0169-7722(00)00123-6
- Fiori, A., Janković, I., Dagan, G., and Cvetković, V. (2007). Ergodic transport through aquifers of non-gaussian log conductivity distribution and occurrence of anomalous behavior. *Water Resour. Res.* 43. doi: 10.1029/2007WR005976
- Fiori, A., Volpi, E., Zarlega, A., and Bohling, G. C. (2015). Gaussian or non-gaussian logconductivity distribution at the MADE site: is its impact on the breakthrough curve? *J. Contam. Hydrol.* 179, 25–34. doi: 10.1016/j.jconhyd.2015.05.004
- Fiori, A., Zarlega, A., Jankovic, I., and Dagan, G. (2017). Solute transport in aquifers: the comeback of the advection dispersion equation and the first order approximation. *Adv. Water Resour.* 110, 349–359. doi: 10.1016/j.advwatres.2017.10.025
- Gelhar, L. W. (1993). *Stochastic Subsurface Hydrology*. Englewood Cliffs, NJ: Prentice Hall.
- Gómez-Hernández, J. J., Butler, J. J., Fiori, A., Bolster, D., Cvetkovic, V., Dagan, G., et al. (2016). Introduction to special section on modeling highly heterogeneous aquifers: lessons learned in the last 30 years from the MADE experiments and others. *Water Resour. Res.* 53, 2581–2584. doi: 10.1002/2017WR020774
- Hansen, S. K., Haslauer, C. P., Cirpka, O. A., and Vesselinov, V. V. (2018). Direct breakthrough curve prediction from statistics of heterogeneous conductivity fields. *Water Resour. Res.* 54, 271–285. doi: 10.1002/2017WR020450
- Jankovic, I., Maghrebi, M., Fiori, A., and Dagan, G. (2017). When good statistical models of aquifer heterogeneity go right: the impact of aquifer permeability structures on 3d flow and transport. *Adv. Water Resour.* 100, 199–211. doi: 10.1016/j.advwatres.2016.10.024
- Kitanidis, P. K. (1988). Prediction by the method of moments of transport in a heterogeneous formation. *J. Hydrol.* 102, 453–473.
- Kreft, A., and Zuber, A. (1978). On the physical meaning of the dispersion equation and its solutions for different initial and boundary conditions. *Chem. Eng. Sci.* 33, 1471–1480.

- Redner, S. (2001). *A Guide to First-Passage Processes*. Cambridge: Cambridge University Press.
- Renard, P., and De Marsily, G. (1997). Calculating equivalent permeability: a review. *Adv. Water Resour.* 20, 253–278.
- Ritzi, R. W. (2000). Behavior of indicator variograms and transition probabilities in relation to the variance in lengths of hydrofacies. *Water Resour. Res.* 36, 3375–3381. doi: 10.1029/2000WR900139
- Rubin, Y. (2003). *Applied Stochastic Hydrogeology*. Oxford: Oxford University Press.
- Salandin, P., and Fiorotto, V. (1998). Solute transport in highly heterogeneous aquifers. *Water Resour. Res.* 34, 949–961.
- Severino, G., Comegna, A., Coppola, A., Sommella, A., and Santini, A. (2010). Stochastic analysis of a field-scale unsaturated transport experiment. *Adv. Water Resour.* 33, 1188–1198. doi: 10.1016/j.advwatres.2010.09.004
- Shapiro, A. M., and Cvetkovic, V. D. (1988). Stochastic analysis of solute arrival time in heterogeneous porous media. *Water Resour. Res.* 24, 1711–1718.
- Tonina, D., and Bellin, A. (2008). Effects of pore-scale dispersion, degree of heterogeneity, sampling size, and source volume on the concentration moments of conservative solutes in heterogeneous formations. *Adv. Water Resour.* 31, 339–354. doi: 10.1016/j.advwatres.2007.08.009
- Zarlenga, A., Janković, I., Fiori, A., and Dagan, G. (2018). Effective hydraulic conductivity of three-dimensional heterogeneous formations of lognormal permeability distribution: the impact of connectivity. *Water Resour. Res.* 54, 2480–2486. doi: 10.1002/2017WR022141
- Zech, A., Attinger, S., Cvetkovic, V., Dagan, G., Dietrich, P., Fiori, A., et al. (2015). Is unique scaling of aquifer macrodispersivity supported by field data? *Water Resour. Res.* 51, 7662–7679. doi: 10.1002/2015WR017220
- Zheng, C., Bianchi, M., and Gorelick, S. M. (2011). Lessons learned from 25 years of research at the MADE site. *Groundwater* 49, 649–662. doi: 10.1111/j.1745-6584.2010.00753.x
- Zinn, B., and Harvey, C. F. (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 Resour. Res.* 39. doi: 10.1029/2001WR001146

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Fiori, Zarlenga, Bellin, Cvetkovic and Dagan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Heterogeneity and Prior Uncertainty Investigation Using a Joint Heat and Solute Tracer Experiment in Alluvial Sediments

Richard Hoffmann^{1,2*}, Alain Dassargues¹, Pascal Goderniaux² and Thomas Hermans³

¹ Hydrogeology and Environmental Geology, GEO³, UEE, Liège University, Liege, Belgium, ² Geology and Applied Geology, Polytech Mons, University of Mons, Mons, Belgium, ³ Department of Geology, Ghent University, Ghent, Belgium

OPEN ACCESS

Edited by:

Frederick Delay,
Université de Strasbourg, France

Reviewed by:

Jean-Francois Girard,
UMR7516 Institut de Physique du
Globe de Strasbourg (IPGS), France
Bruno Cheviron,
National Research Institute of Science
and Technology for Environment and
Agriculture (IRSTEA), France

*Correspondence:

Richard Hoffmann
richard.hoffmann@uliege.be

Specialty section:

This article was submitted to
Hydrosphere,
a section of the journal
Frontiers in Earth Science

Received: 30 November 2018

Accepted: 26 April 2019

Published: 29 May 2019

Citation:

Hoffmann R, Dassargues A,
Goderniaux P and Hermans T (2019)
Heterogeneity and Prior Uncertainty
Investigation Using a Joint Heat and
Solute Tracer Experiment in Alluvial
Sediments. *Front. Earth Sci.* 7:108.
doi: 10.3389/feart.2019.00108

In heterogeneous aquifers, imaging preferential flow paths, and non-Gaussian effects is critical to reduce uncertainties in transport predictions. Common deterministic approaches relying on a single model for transport prediction show limitations in capturing these processes and tend to smooth parameter distributions. Monte-Carlo simulations give one possible way to explore the uncertainty range of parameter value distributions needed for realistic predictions. Joint heat and solute tracer tests provide an innovative option for transport characterization using complementary tracer behaviors. Heat tracing adds the effect of heat advection-conduction to solute advection-dispersion. In this contribution, a joint interpretation of heat and solute tracer data sets is proposed for the alluvial aquifer of the Meuse River at the Hermalle-sous-Argenteau test site (Belgium). First, a density-viscosity dependent flow-transport model is developed and induce, due to the water viscosity changes, up to 25 % change in simulated heat tracer peak times. Second, stochastic simulations with hydraulic conductivity (K) random fields are used for a global sensitivity analysis. The latter highlights the influence of spatial parameter uncertainty on the resulting breakthrough curves, stressing the need for a more realistic uncertainty quantification. This global sensitivity analysis in conjunction with principal component analysis assists to investigate the link between the prior distribution of parameters and the complexity of the measured data set. It allows to detect approximations done by using classical inversion approaches and the need to consider realistic K-distributions. Furthermore, heat tracer transport is shown as significantly less sensitive to porosity compared to solute transport. Most proposed models are, nevertheless, not able to simultaneously simulate the complementary heat-solute tracers. Therefore, constraining the model using different observed tracer behaviors necessarily comes with the requirement to use more-advanced parameterization and more realistic spatial distribution of hydrogeological parameters. The added value of data from both tracer signals is highlighted, and their complementary behavior in conjunction with advanced model/prediction approaches shows a strong uncertainty reduction potential.

Keywords: joint heat and solute tracer tests, density-viscosity dependent flow and transport, alluvial sediments, preferential flow paths, uncertainty investigation, distance-based global sensitivity analysis, principal component analysis

INTRODUCTION AND MOTIVATION

Heterogeneity in porous media, inducing preferential flow paths, and non-Gaussian effects, influences significantly subsurface transport (among others: Fuchs et al., 2009; Heeren et al., 2010). An improved imaging of these preferential pathways, in connection with reducing uncertainty in transport simulations and predictions, is crucial for answering future groundwater quality questions.

Innovative tracer test set-ups, along with relevant interpretations, are possible new ways for quantifying more realistically this heterogeneity and the associated uncertainty (Davis et al., 1980; Maliva, 2016). In this context, heat is considered as a complementary tracer, compared to conservative solute tracers (saline tracer or fluorescent dye). Heat is usually considered as a non-conservative tracer, allows information about advection-conduction processes to be obtained, and heat has a natural retardation and more diffusion linked to the heat capacity of the solid (Anderson, 2005). Using both, solute and heat, thus provides two tracer plumes that can be compared, allows more information about the solid matrix properties to be obtained, and enables the quantification of subsurface processes (immobile water, matrix contributions) with a better resolution (Anderson, 2005; Irvine et al., 2015). For example, Wildemeersch et al. (2014) combine heat and solute tracer experiments to assess the heterogeneity in an alluvial aquifer. Sarris et al. (2018) also give a recent application of jointly interpreting heat and solute tracer data. They show, in a deterministic way, how these innovative tracer tests can contribute to a high-resolution description of deposits, and a significant improvement of transport processes understanding. In the joint heat and solute tracer inversion by Sarris et al. (2018), heat and solute seem to be sensitive to hydraulic conductivity and porosity. In their case study, heat also shows a stronger sensitivity to vertical hydraulic conductivity, resulting in a more complex aquifer parametrization, and more realistic transport predictions.

Deterministic approaches are useful for process understanding. However, predictions based on an unique “best” model parametrization bear a lot of uncertainty and, thus, must be justified and used with care (Renard, 2007; Remonti and Mori, 2016). Deterministic approaches generally reduce heterogeneity, typically by replacing spatially distributed properties by averaged properties, leading to poorer predictions with underestimated uncertainty (Alcolea et al., 2006; Renard, 2007). Using additional information gained from joint heat and solute tracer tests adds more constraints to the inversion process. However, when the tracer behavior is getting more complex, deterministic approaches can quickly turn into ill-posed inverse problems (Zhou et al., 2014), reducing their predictive reliability. Classical deterministic inversion approaches could therefore be questioned.

To explain and adequately represent the observed variables, more advanced transport simulation and forecast approaches, such as stochastic methods (Ptak et al., 2004), are generally required. Monte-Carlo simulations may, for example, be used to explore uncertainty ranges and to consider heterogeneity (among others, Ptak et al., 2004; Renard, 2007; Ferré, 2017). However, the full stochastic inversion of hydrogeological data

when spatial uncertainty plays a key role remains difficult and time consuming, limiting the applicability of the methods (Renard, 2007). In this context, transdimensional inference (e.g., Sambridge et al., 2012) is a possible approach to combine the parsimony principle with stochastic inversion. In practice, transdimensional inference includes the number of parameters as an unknown, and therefore limits the complexity of the model to what is necessary to explain the data.

However, if field data is sparse and prior uncertainty is large, the transdimensional approach would also lead to an oversimplification of the model, which can be harmful for its predictive capability (referred to Hermans (2017), about the importance of realistic consideration of prior uncertainty). In contrast, a full stochastic approach allows realistically quantifying the uncertainty for transport predictions, instead of having a single deterministic inversion or multiple simulations with (partly) non-quantified approximations (among others, Caers, 2011; Hermans, 2017; Hermans et al., 2018; Scheidt et al., 2018). In combination with the use of complementary tracers such as heat, stochastic approaches and data analysis methods have a strong potential to learn more from collected data sets and falsify approximations done in conceptual models and prior estimations (Hermans et al., 2015a,b, 2016, 2018).

A further new potential for hydrogeological applications and transport predictions is Bayesian Evidential Learning (BEL) (Hermans, 2017; Hermans et al., 2018; Scheidt et al., 2018). BEL relies on a limited number of Monte-Carlo simulations sampling the prior distribution of model parameters, in order to analyze the global sensitivity of parameters (Park et al., 2016; Hermans et al., 2018) and falsify the prior distribution. In comparison to common single parameter sensitivity analysis, regionalized or global sensitivity analyses consider heterogeneous sources of model uncertainty (e.g., Park et al., 2016). The falsification step consists in obtaining consistency between the sampled simulation data (i.e., prior) and the reference data (Hermans et al., 2015a; Scheidt et al., 2018). BEL can also be used, if necessary, to identify a statistical relationship between historical and forecast variables (Hermans, 2017; Hermans et al., 2018; Scheidt et al., 2018). The common inversion step is thus replaced by finding the direct relationship between the prior (sampled simulation data) and the desired forecast, which only depends on the complexity of the subsurface (i.e., model) (Hermans, 2017). Using Monte-Carlo, samples from the prior distribution are generated and used to simultaneously simulate synthetic data and forecasts. Both outcomes are analyzed for detecting direct relationships. Following this innovative approach, cost-expensive inversion can be avoided by reformulating the prediction problem, and the likelihood directly in terms of the forecast (Hermans, 2017).

For addressing the uncertainty of transport predictions, for instance due to preferential flow paths, any forecasting approach should first consider realistic parameter uncertainty related to heterogeneity. In this paper and the context of BEL, “prior” is defined as the prior distribution of model parameters, according to the current knowledge of the field, and from which outcome samples are randomly drawn in Monte-Carlo methods (Rojas et al., 2009; Caers, 2011; Hermans, 2017).

In this paper, the prior uncertainty is investigated for a heterogeneous alluvial aquifer, using joint heat and solute tracer data, through a global sensitivity analysis. A performed tracer experiment (Wildemeersch et al., 2014; Hermans et al., 2015b) has been previously used by Klepikova et al. (2016), to calibrate a deterministic model through automatic inversion. The HydroGeoSphere code (HGS) was used allowing full 3D simulations (Therrien et al., 2010; Brunner and Simmons, 2012). HGS was used in conjunction with PEST as a parameter estimation tool for inversion (Doherty, 1994, 2003). Although those first analyses helped to understand groundwater flow and solute and heat transport in the aquifer, they also showed that the approximations impeded to explain simultaneously all observations. In Hermans et al. (2015a, 2018), additional stochastic approaches considering spatial uncertainty were successfully used in explaining parts of the experimental data, but they considered only limited data sets or parts of the whole aquifer system. In this paper, the current conceptual approximations (the prior) in the description of the alluvial deposits will be revisited with the goal to improve the heterogeneity characterization. Generating a prior, consistent with the observed heat and solute tracer test data, is a necessary step to being able to realistically predict transport in this complex geological setting.

Within this context, the objective of the paper is to take advantage of the complementary behavior of heat and solute tracers to better characterize the heterogeneity in the aquifer system. The aim is to formulate a more realistic prior and analyze its consistency, before moving toward more advanced prior analysis and direct predictions using the BEL framework. For this purpose, the variability of the tracer output signals will be analyzed (1) through a deterministic model, and (2) using Monte-Carlo simulations, followed by a global sensitivity analysis.

MATERIALS AND METHODS

Test Site: Hermalle-Sous-Argenteau

The test site of Hermalle-sous-Argenteau (HssA) in the north of Liege (Belgium) lies between the canal Albert and the Meuse River (**Figure 1A**) in an alluvial plain field with a groundwater natural gradient of around 0.06 %. Between the 20 m distant injection well (Pz09) and the pumping well (PP), there are three panels with 10 piezometers including 19 observation points (i.e., most piezometers are screened at two different levels, **Figure 1B**). The first panel is located at 3 m, the second at 8 m and the third at 15 m from the injection well. An evaluation of the borehole logs during drilling shows that the aquifer is mostly composed of sandy gravel. The sand matrix is finer in the top part and its proportion decreases in the bottom part (**Figure 1B**). In previous studies, the adopted conceptual model split the aquifer in two layers, an upper ($K = 2.38 \cdot 10^{-3} \text{ m s}^{-1}$) and a lower ($K = 4.67 \cdot 10^{-2} \text{ m s}^{-1}$) part (Klepikova et al., 2016; Hermans et al., 2018). The estimated bulk thermal

conductivity is in the range $\kappa_b = 1.37 \text{ W m}^{-1} \text{ K}^{-1}$ to $1.86 \text{ W m}^{-1} \text{ K}^{-1}$ (Klepikova et al., 2016).

The reference data set used in this study was described in Wildemeersch et al. (2014). A joint heat and solute tracer experiment was performed with a 24 h and 20 min continuous injection of heat ($\Delta T = 25.5 \text{ K}$) and naphthionate ($C = 5.48 \text{ mg L}^{-1}$) at the rate of $3 \text{ m}^3 \text{ h}^{-1}$, while $30 \text{ m}^3 \text{ h}^{-1}$ were extracted from the pumping well PP. Temperature distribution was the focus and therefore measured in all observation points, while the solute tracer was only measured in PP for validation purposes. Measurements in Pz13 and Pz17 are not used because these observation wells are uniformly screened all over the aquifer and do not allow separated measurements in both the upper and lower compartments (Klepikova et al., 2016; Hermans et al., 2018).

The observed heat tracer plume shows that the heat injected in the lower aquifer part tends to move upwards very quickly toward the first panel, then to be split and move downwards (Wildemeersch et al., 2014). The measured temperature in the second upper panel is significantly lower than in the first upper panel (for detailed measured reference data at all panels, we refer to the “**Supplementary Material Figures 1–4**”). This observed behavior is currently difficult to be simulated for all observation locations with just one deterministic model inverted with common methods, e.g., using pilot points (Klepikova et al., 2016).

Deterministic Porous Media Model

Based on the current conceptual two-layer aquifer model (each 3.5 m thick), Klepikova et al. (2016) developed a numerical model focusing on the heat transport simulation using HGS in finite difference mode (Therrien et al., 2010) and a pilot point approach (Doherty, 2003) to calibrate the model against heat data. This model is density dependent during the first 24 h only, as density effects were expected to be low afterwards (Ma and Zheng, 2009; Klepikova et al., 2016). This model describes a $40 \times 60 \times 7 \text{ m}$ volume of alluvial aquifer with a grid of 84,280 elements in total. No recharge was assumed for the duration of the experiment. Due to the high permeability of the gravel, the vertical leakage at the bottom of the model was considered as negligible compared to lateral input/output. The initial groundwater temperature was set to $T_{\text{ini}} = 13.48^\circ \text{C}$ according to the measured values before the experiment. The model was running under transient flow conditions due to the simulation of the tracer experiment. Peclet numbers of 300 for the upper part, and 14,000 for the lower part were computed, suggesting an expected advection-dominated transport (Klepikova et al., 2016).

Here, this model is extended with a simultaneous solute injection. In HGS, the injection is simulated selecting two nodes next to each other: at the first node, representing the screen location in the borehole, the solute tracer is injected and one node below the heat is injected, respecting the actual experimental set-up. Both injections are simulated using Neumann (2nd type) boundary conditions. For the solute injection, the prescribed mass injection rate is $4.3 \cdot 10^{-6} \text{ kg s}^{-1}$. For the heat injection, the prescribed injection rate is $8.3 \cdot 10^{-4} \text{ J K}^{-1} \text{ s}^{-1}$ (Klepikova et al., 2016). The grid is refined to 140,140 rectangular elements with

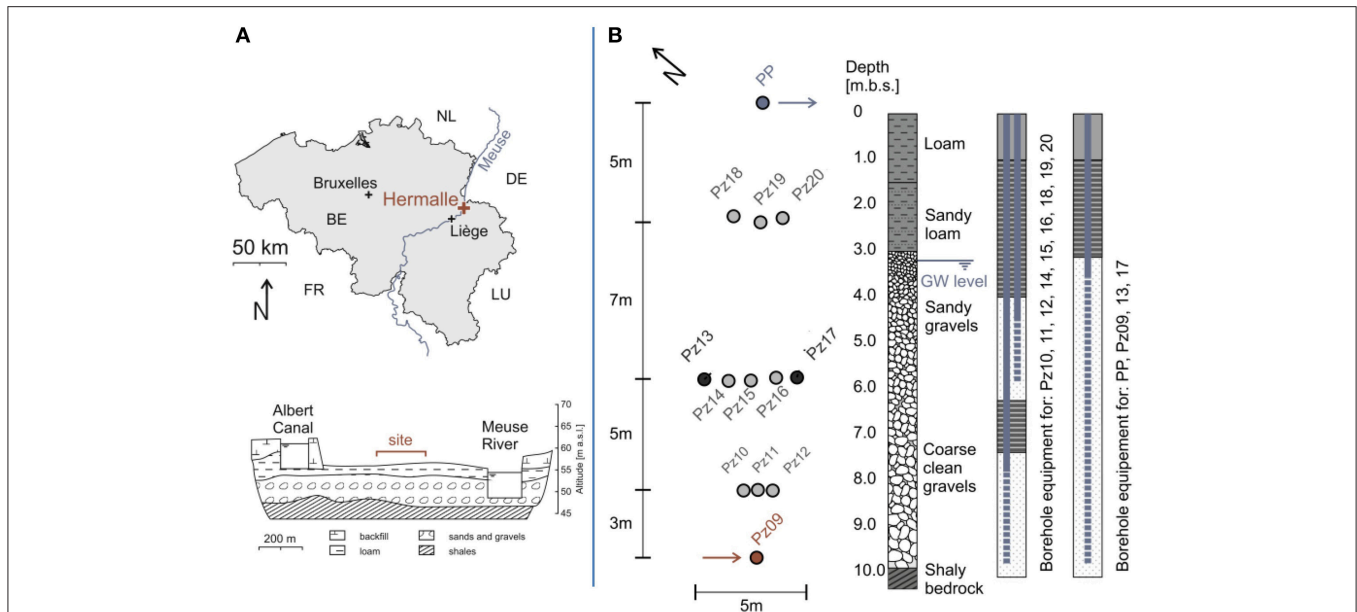


FIGURE 1 | (A) Overview of the field site location. **(B)** Borehole location showing the three panels between the injection well (Pz9) and the pumping well (PP), a typical log, and borehole equipment [modified figure according to Wildemeersch et al. (2014)].

14 numerical grid sublayers (each 0.5 m thick), allowing a better representation of the spatial heterogeneity and uncertainty. To account for the influence of immobile water on heat conduction, an apparent thermal conductivity is computed using:

$$\kappa_s' = \frac{\kappa_f (\theta - n_{\text{eff}}) + \kappa_s (1 - \theta)}{(1 - n_{\text{eff}})} \quad (1)$$

θ is the total porosity with 0.12, n_{eff} the effective porosity with 0.05, or otherwise called mobile water porosity, κ_f the fluid thermal conductivity with $0.59 \text{ W m}^{-1} \text{ K}^{-1}$ and κ_s the solid thermal conductivity estimated from previous works to around $1.43 \text{ W m}^{-1} \text{ K}^{-1}$. As temperature affects the groundwater density and dynamic viscosity, the injected heat influences both groundwater flow and transport simulations. In contrast to Klepikova et al. (2016), the new numerical implementation allows fully density-dependent, but also viscosity-dependent simulations.

Stochastic Prior Uncertainty Investigation

This study investigates the prior-uncertainty using Monte-Carlo simulations, followed by a global sensitivity analysis and prior falsification. The applied procedure in this study corresponds to the first, second and third steps of the BEL method as described by Hermans et al. (2018). In the Monte-Carlo simulations, part of the calibrated values from the deterministic model parameters are replaced by random values sampled from random uniform distributions (Table 1).

In each Monte-Carlo simulation step (prior sampling), a new HGS forward model is parameterized with randomly generated advection global parameters, like the $\log(K_{\text{mean}})$, the K variance, the porosity, the variogram ranges in X,Y,Z directions, the azimuth, and the gradient between the two main prescribed

head boundary conditions located upgradient from the injection well and downgradient from the pumping well. Fixed values [i.e., identical to those chosen by Klepikova et al. (2016)] are considered for dispersivity, thermal conductivity, specific heat capacity, specific storage, and bulk density (Table 1). To represent the K-distribution within each Monte-Carlo simulation more realistically, sequential Gaussian simulations are used following two scenarios.

Scenario A uses the prior distribution from Hermans et al. (2015a). This prior was not falsified by geophysical and hydrogeological data acquired during the experiment (Hermans et al., 2015a) in the middle panel. It was thus not tested against the whole available data set, as it is proposed here. In particular, it uses the same two-layer approximation and ignores any trend in the alluvial deposits grain size distribution. Random K_{mean} values between $10^{-3.5}$ and $10^{-2.5} \text{ m s}^{-1}$ in conjunction with a K-variance between 1 and 100 m s^{-1} are considered. Models are randomly generated without any additional constraint.

Scenario B considers the trend observed during drilling (Figure 1B) in the alluvial deposits, in which it is assumed that grain size distributions influenced the hydraulic conductivity values. Within 14 constrained sublayers, a vertical downwards increasing K-trend is considered in the geostatistical simulations. Within every Monte-Carlo step, for the 12 sublayers between the fixed top ($K_{\text{mean}} = 10^{-4} \text{ m s}^{-1}$) and bottom ($K_{\text{mean}} = 10^{-2} \text{ m s}^{-1}$) sublayer, new random generated mean values between $K_{\text{mean}} = 10^{-3.5}$ and $10^{-2.5} \text{ m s}^{-1}$, increasing downwards, are used. To account for field observations (see section Test site: Hermalle-sous-Argenteau, Wildemeersch et al., 2014; Hermans et al., 2015b) showing that the heat plume does not follow a straight path toward the pumping well, the possible occurrence

TABLE 1 | Parameter simulation ranges for the Monte-Carlo simulations (left column) and chosen fixed parameter values identical to those chosen by Klepikova et al. (2016) (right column).

Parameter	Simulation range U	Unit	Fixed parameter ⁺	Value	Unit
Log(K _{mean})	A: [−3.5 −2]	m s ^{−1}	Longitudinal dispersivity (upper / lower part)	1.5/3	m
	B: Trend [−4 −2] + [−3.5 −2.5]	m s ^{−1}	Transversal dispersivity (upper / lower part)	0.15/0.3	m
Variance K	[1 100]	m s ^{−1}	Solid thermal conductivity (apparent)	1.37*	W m ^{−1} K ^{−1}
Porosity	[0.05 0.12]	1	Fluid thermal conductivity	0.59	W m ^{−1} K ^{−1}
X	[1 to 8]	m	Solid specific heat capacity	1,000	J kg ^{−1} K ^{−1}
Y=Z	[0.1 0.5]	m	Water specific heat capacity	4,189	J kg ^{−1} K ^{−1}
Azimuth	[0 pi]	1	Specific Storage	10 ^{−4}	m ^{−1}
Gradient at prescribed head BC	[0.01 0.1]	%	Bulk density	1,950	kg m ³

⁺ The fixed values are taken from Klepikova et al. (2016).

*The solid thermal conductivity estimated at 1.43 W m^{−1} K^{−1} is replaced by an apparent value for simulation (see section: Deterministic porous media model).

[] Sampled from a random uniform distribution.

of local low hydraulic conductivity zones (flow barriers) in the aquifer is thus considered. The presence of loam lenses with low hydraulic conductivity are actually observed at some places in the Meuse river alluvial deposits and are here assumed as a possible origin/explanation for the observed behavior. Two flow barriers are placed in front of Pz14 and Pz17 in the upper part, and the third one in the lower aquifer part between the injection well and Pz11. The constrained hydraulic conductivity is four orders of magnitude lower. Within the sequential Gaussian simulation, the fixed constrained K-values are considered as hard data for the random simulations. The size of the potential flow barriers is thus dependent on variogram characteristics.

For each scenario, 250 simulations are generated through Monte-Carlo methods. This number was considered sufficient to model the variability in the prior data sets, while keeping the computational cost to a minimum, and to estimate the global sensitivity analysis (see below). The reason lies in the fact that the used approach analyzes the data response (temperature or solute curve) which is less complex than the model spatial heterogeneity, therefore requiring only a limited number of samples (Hermans et al., 2018).

A distance metric using the root of the square sums of the difference between each simulated $f(t_i)$ and observed $g(t_i)$ data over the same experimental time interval t_{Exp} (0–10 days), showing positive zero definition, symmetry, and triangle inequality, is used to compare the ability of different simulations to reproduce field data. It allows to identify the best realization at each observation point within the generated 250 realizations, separately for each scenario:

Best simulation at observation point means minimizing:

$$d_{ObsP} = \sqrt{\sum_{i=1}^{t_{Exp}} (f(t_i) - g(t_i))^2} \quad (2)$$

Best simulations are quantified calculating the root mean square error (RMSE) and the correlation coefficient (R^2).

Distance Based Sensitivity Analysis

A global sensitivity analysis reveals key information about model parameters most influencing the simulated data at observation points. With the output signals of the heat and solute realizations, the distance-based global sensitivity analysis (DGSA, Park et al., 2016) is applied, considering the global and spatial parameters of each simulation. DGSA can also identify conditional effects between pairs of parameters (Fenwick et al., 2014; Park et al., 2016). In DGSA, the sensitivity is defined by comparing the parameter cumulative distribution function (cdf) within k clusters to the original distribution. The number of clusters must be chosen so that there are enough simulations in each cluster while allowing sufficient discrimination between them (Hermans et al., 2018). The k clusters are computed using the k-medoid clustering technique applied on a multi-dimensional scaling map of the models. The latter is computed based on the metrics of equation (2). In DGSA, random parameters for each simulation are linked to the corresponding output signal produced by the forward model. We refer to Park et al. (2016) and Fenwick et al. (2014) for details.

Every sample of the prior distribution is parameterized using random generated global parameters, e.g., porosity, gradient, K_{mean} , and K-variance values and local parameter, i.e., the spatial random K-field, generated by geostatistical simulation. A local parameter is highly dimensional (number of elements) and therefore difficult to analyze using a sensitivity analysis. However, those local parameters can be reduced using Principal Component Analysis (PCA). PCA is one possibility to structure, simplify and visualize complex data sets by replacing multiple statistical variables with a limited, smaller, and approximated amount of linear combinations using the decomposition in eigenvectors (Krzanowski, 2000). With PCA, the first dimensions are explaining the average K distribution and, thus, larger scale heterogeneity, while higher dimensions will be characteristic of smaller-scale heterogeneity (e.g., Oware et al., 2013; Park and Caers, 2018). Therefore, the PCA's first dimensions represent the degree of heterogeneity in the aquifer. That is used to compute the score variables for each of the 250 simulations subsequently in the sensitivity analysis (Park and Caers, 2018).

Tracer Velocity Comparison

The modified deterministic model and the Monte-Carlo simulations are further used for a synthetic heat and solute tracer velocity comparison. The velocity comparison follows the approach of Irvine et al. (2015), but using here the peak times instead of the time of 50 % of tracer recovery in the breakthrough curve. The peak time is here preferred, due to the large uncertainty related to the missing solute tracer information between the injection and pumping wells. Irvine et al. (2015) equations are adapted by using, for the strong advective aquifer system of Hermalle-sous-Argenteau, as thermal retardation factor $R_{th} = 1$. An estimated thermal retardation factor based on a fixed specific heat capacity is in this study not sufficient to capture the difference between the two tracers as it cannot account for spatial heterogeneity. The calculations are:

$$v_{sol_{peak}} = \frac{x}{t_{sol_{peak}}} \quad (\text{modal velocity}) \quad (3)$$

$$v_{heat_{peak}} = v_{th_{peak}} \cdot R_{th} = \frac{x}{t_{heat_{peak}}} \cdot 1 \quad (4)$$

where $t_{sol_{peak}}$ [s] is the solute peak time of each prediction/simulation, x [m] is the shortest distance from the observation well to the injection point and $v_{sol_{peak}}$ [m s^{-1}] the corresponding modal velocity. In the heat case $t_{heat_{peak}}$ [s] is the peak time of each prediction/simulation, R_{th} the thermal retardation factor, $v_{th_{peak}}$ [m s^{-1}] the thermal front velocity using the peak time.

High K-zones (i.e., corresponding to preferential flow paths) resulting in a mismatch of solute and heat distributions, lead to different v_{heat} and v_{solute} values (i.e., diverging from a 1:1 line in a v_{heat} vs. v_{solute} diagram) inducing a decrease of the regression coefficient (Irvine et al., 2015).

RESULTS

Prior Uncertainty Investigation

The heat observations at panel 1, 2, 3, and the joint observed heat and solute information at the pumping well are investigated and used to attempt prior falsification. In a first step, the numerical model considers groundwater density and dynamic viscosity effects caused by the injected heat. In a second step, the multiple heat breakthrough curves and the solute one at the pumping well are simulated using the multiple realizations generated by the Monte-Carlo procedure in conjunction with both K-distribution scenarios. **Figure 2** shows the comparison of the deterministic solution of the density (basis model) and the density-viscosity dependent model with the reference data, the two simulation scenarios A and B and the individual best heat simulation for the upper screened part in Pz11-up, Pz15-up, and Pz19-up (observation points in the upper middle lane of Panel 1, 2, and 3).

The change of the dynamic viscosity, e.g., at the peak-time for Pz11-up about 25 % (upper screen), has a significant effect on the simulated temperature, while the effect of density is limited (0.02 %) (**Figures 2E,F, 3**). Accounting for this effect allows slight improvement to the fit with the observed heat breakthrough curve at Pz11-up and Pz15-up using a deterministic approach

(**Figures 2C–F**). The simulated peak, e.g., at Pz11-up is slightly improved, as well as the tailing, but the overall fit is still not satisfying for all three observation points.

Monte-Carlo realizations surround the real data set of the heat tracer at all three points (**Figure 2**). It highlights the influence of spatial parameter heterogeneity on the resulting breakthrough curves. However, the prior of scenario A seems to be falsified by the tailing part of the curve observed in Pz11-up and Pz15-up (**Figures 2C,E**). Clearly, scenario B considering the observed vertical downwards increasing K-trend, and describes the tailing part of the curve more realistically compared to the deterministic approach and scenario A (e.g., compare **Figures 2E,F**). For Pz15-up and Pz19-up the best simulation allows representing the reference data more accurately, compared to the deterministic solutions (**Figures 2A–D**). Again, scenario B gives more realistic solutions than ignoring any trend in the simulations.

Selecting the 10 best heat simulations from both prior scenarios for all observation points at panels 1 to 3, upper and lower screen, further confirms that considering a vertical downwards increasing K-trend in the simulations is a more realistic description of spatial heterogeneity (referring to “**Supplementary Material Figures 1–4**”). At panel 3 the observed data fluctuates around a temperature change of 0°C , without a significant peak. Thus, the solution simulation with $\Delta T = 0^\circ\text{C}$ is identified as the best one. The modeled $\Delta T = 0^\circ\text{C}$ is exactly zero for the simulation because it corresponds to a set of parameters where diffusion is larger than advection transport; the heat therefore does not reach panel 3. These results stress the need for a more realistic prior-uncertainty quantification and falsification of prior hypotheses. Here, a purely random K-field can be considerably improved by including sedimentological observations that the advocated procedure is capable of taking advantage of it.

Previous paragraphs do not integrate the joint heat and solute breakthrough curves at the pumping well (located 5 m downstream from panel 3). Here, the deterministic model solution calibrated on heat data only (Klepikova et al., 2016), including now both the density and viscosity changes, fails to predict the heat and solute tracers behaviors at the pumping well (**Figures 4A,B**). Note that the solution depending on density only is not realistic and is not shown).

Monte-Carlo simulations, with the random K-field without trend (scenario A), surround heat and solute observed data (**Figures 4A,B**), adding indications that spatial heterogeneity is necessary to generate realistic predictions. The best simulation for the observed heat signal using equation (2) ($R^2 = 0.96$, $\text{RMSE} = 0.01^\circ\text{C}$) is however different than the one for the solute signal ($R^2 = 0.99$, $\text{RMSE} = 1.2 \cdot 10^{-5} \text{ g L}^{-1}$). Thus, the best heat simulation poorly predicts the solute signal and vice versa (**Figures 4A,B**). For the random K-field with the vertical downwards increasing K-trend (scenario B), the heat breakthrough at the pumping well is not as well simulated as in the intermediate panels and the solute breakthrough concentrations are strongly underestimated, even though the time occurrence of the peak seems to be correctly predicted (**Figures 4C,D**). Near the pumping well, the tracer is intensively diluted due to the high pumping rate ($30 \text{ m}^3\text{h}^{-1}$), making

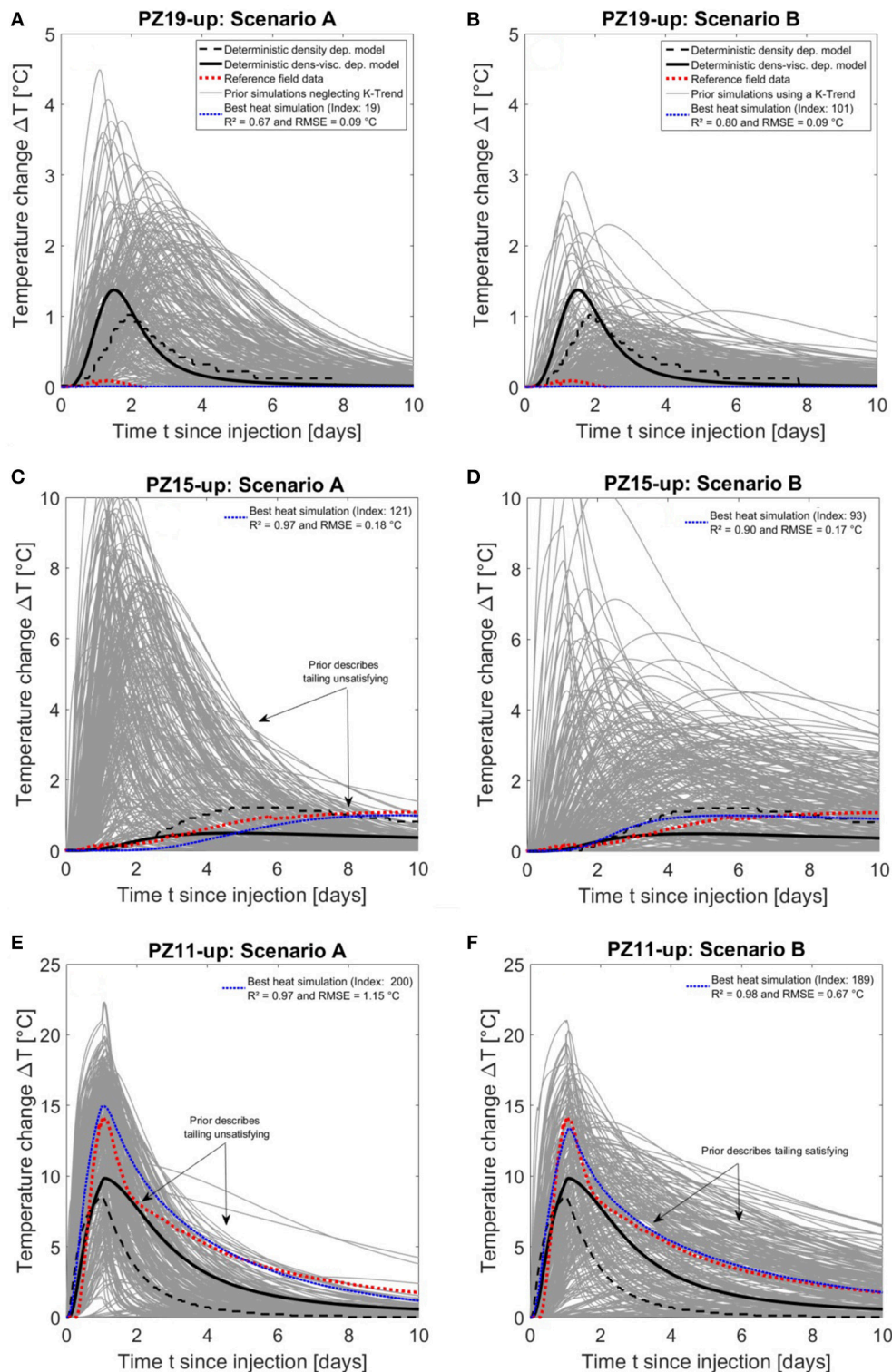


FIGURE 2 | Comparison between deterministic solution, real data and prior + best simulation for both random K-distribution scenarios for Pz19-up (Panel 3): **(A)** without K-trend (scenario A), **(B)**, with a downwards increasing K-trend (scenario B), for Pz15-up (Panel 2): **(C)** without K-trend, **(D)** with K-trend, and for Pz11-up (Panel 1): **(E)** without K-trend, **(F)** with K-trend (10 best heat simulations at each observation point are in **Supplementary File**). The index of the best solution refers its number within the 250 simulations.

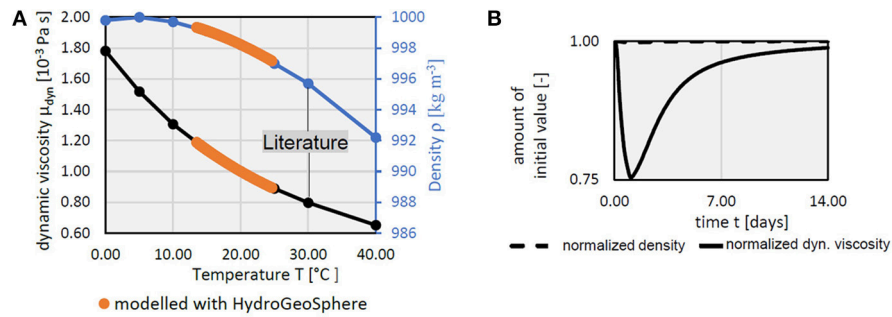


FIGURE 3 | Variation of density and water viscosity influenced by the injected heat observed at PZ11-up as absolute change (A) and relative change (B). (Hoffmann et al., 2018).

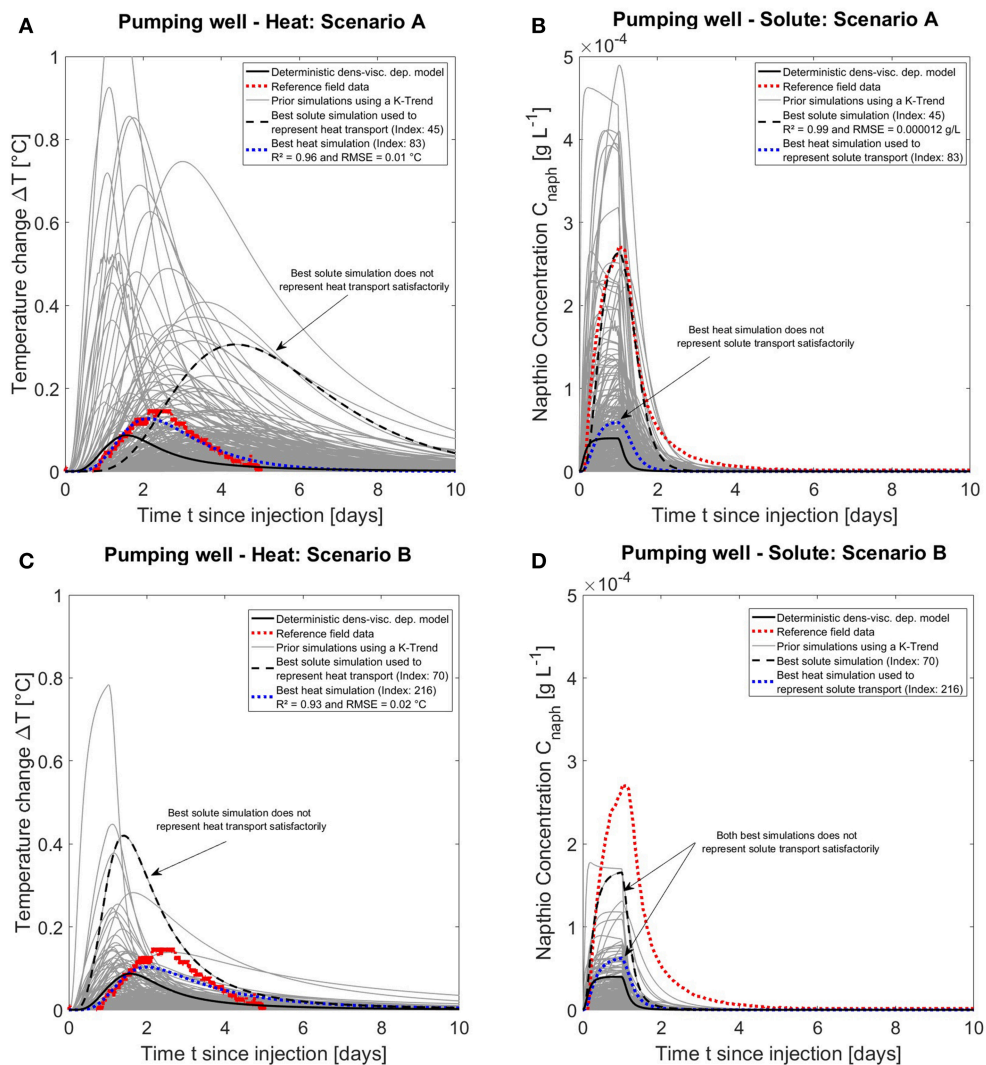


FIGURE 4 | Heat and solute simulation results at the pumping well ignoring any trend (A,B) and considering a downwards increasing K-trend (C,D) compared with the deterministic density-viscosity model. Best heat and solute simulation were used to visualize the complementary tracer (Note that the simulations which consider only the density dependence to temperature and not the viscosity dependence are not shown and rejected as they are not realistic enough).

the heterogeneity around this well crucial for explaining the breakthrough curve (see in Discussion section).

The K-fields corresponding to the best heat and solute simulations describing the tracer breakthrough curves at the pumping well using scenario A are largely heterogeneous (**Figures 5A,B**). However, for both tracers, K-distributions are slightly different, which explains in parallel to the less good modeling of the tailing (**Figure 2**), why scenario A is not suitable to adequately describe complementary tracer movement (**Figures 4, 5**). The injected heat forms a plume around the injection well enlarging with time by conduction and advection, while its temperature amplitude decreases (**Figure 5C**). Further, the solute prefers mainly the high hydraulic conductivity pathways, faster in the lower part than in the upper part (**Figure 5D**). Both tracers are best described by two different parameter distributions and only stochastic inversion could here result in simulations fitting both, while deterministic inversion finding one global parameterization, may tend to derive a smoothed parameter distribution poorly fitting the data. However, it should be stressed that this prior (scenario A) would not be able to reproduce the observations in intermediate panels. For the K-field and tracer distribution of scenario B (**Supplementary Material Figure 5**), the best solute tracer simulation displays a strongly heterogeneous model with preferential flow paths, while for heat, a more homogenous K-field respecting the observed trend in the borehole drillings sedimentology (**Figure 1B**) is found.

These single local best realizations, for each observation point, do not explain the full data set, further highlighting the role of local heterogeneity on the measured signals. This probably explains why the global fit of the deterministic solution is poor and highlights the need for more realistic priors instead of trying to find unique parameterization describing reality.

For the Hermalle-sous-Argenteau test site, the prior considering a vertical downwards increasing K-trend seems to better represent the overall hydraulic conditions (until panel 3) and constitutes a better prior assumption than neglecting any K-trend. However, it seems to be somewhat falsified between the last panel and the pumping well in terms of solute concentration amplitude. This suggests that the current parameterization still oversimplifies the heterogeneity of the deposits at a larger scale and cannot be used for inversion or prediction. New hypotheses should be formulated, or new data collected to identify the specific processes taking place between the third panel and the pumping well. Existing ERT transects (Hermans et al., 2015a,b) and newly acquired cross-hole GPR sections are promising tools to image heterogeneity patterns with a higher resolution and at a larger scale.

Distance Based Sensitivity Analysis

The 250 generated models from scenario B are further used for the distance-based global sensitivity analysis. The distance metrics (Equation 2) between pairs of Monte-Carlo simulations is calculated and used as starting point for DGSA. The sensitivity analysis investigates the global simulation parameter values within their given range (**Table 1**) and the spatial heterogeneity. To analyze spatial heterogeneity, the K-fields

from Monte-Carlo simulations are represented using orthogonal basis vectors computed through principal component analysis (PCA) with 250 observation rows and 140,140 corresponding cell K-values as columns (in total: 35,035,000 K values). Only the first 15 principal components are retained and used as an approximate measure of heterogeneity. While, those 15 first dimensions explain only 23 % of the total variance in K, the first three principal components together describe 9.6 % of the variance. The small amount of explained variance is related to the strongly varying K-values from one simulation to the other. The 15 corresponding PCA scores are further included in the sensitivity analysis to characterize the role of spatial heterogeneity.

The sensitivity analysis results for the heat signal at the panels 1, 2, and 3 in relation are presented in **Figures 6A,C,E**. **Figures 6B,D,F** show the corresponding classification of the 250 models in three clusters. Clusters are used to group the simulations according to their response (i.e., the first cluster contains simulations with high temperature far above the reference data, the second contains simulations with temperature below the reference data and the third group consists in simulations around the reference data). Using three clusters gives satisfactory results in this case.

The sensitivity of a parameter is computed based on the difference between its cumulative distribution function (cdf) in each of the cluster compared to the global cdf. Significant differences mean that the parameter is considered as sensitive. For all panels, the resampling quantile of the distance is $\alpha = 0.95$ (we refer to Park et al. (2016) for the detailed explanations of parameters used in DGSA).

The global “log(K_{var})” is the most sensitive parameter at Panel 1 and 2 (**Figures 6C,E**). Its sensitivity decreases within the third panel (**Figure 6A**) while “log(K_{mean})” is getting more sensitive (**Figures 6A,C,E**). The influence of the first component of spatial heterogeneity “PC1” is large at every distance from the injection well. The increasing sensitivity and influence of “PC1” with distance from the injection well is related to the hydraulic conductivity in the direction of the gradient, which indicates a strong link to preferential pathways (**Figures 6A,C,E**). For the first two panels, most PCA components are sensitive. This is in accordance with the previous results, showing that the introduced K-trend is crucial to explain the observed breakthrough curves (section Prior uncertainty investigation). The global parameter “gradient” shows a decreasing sensitivity with distance. The sensitivity of the gradient highlights the influence of uncertain boundary conditions on the simulations. The fluxes around the injection well are crucial to initiate the tracer transport (**Figures 6A,C,E**). The global “porosity” and the “azimuth” are, at all three panels, a much less sensitive parameter for the strong advective system at Hermalle-sous-Argenteau. The variance explained in the low dimensional space is relatively constant over all three panels, but the clusters are getting closer to each other (**Figures 6B,D,F**). The sensitivity analysis further indicates that the scale of heterogeneity playing a role on the tracer distribution at panel three is different. The vertical K-trend is not sufficient anymore to explain the observations. It underpins that more

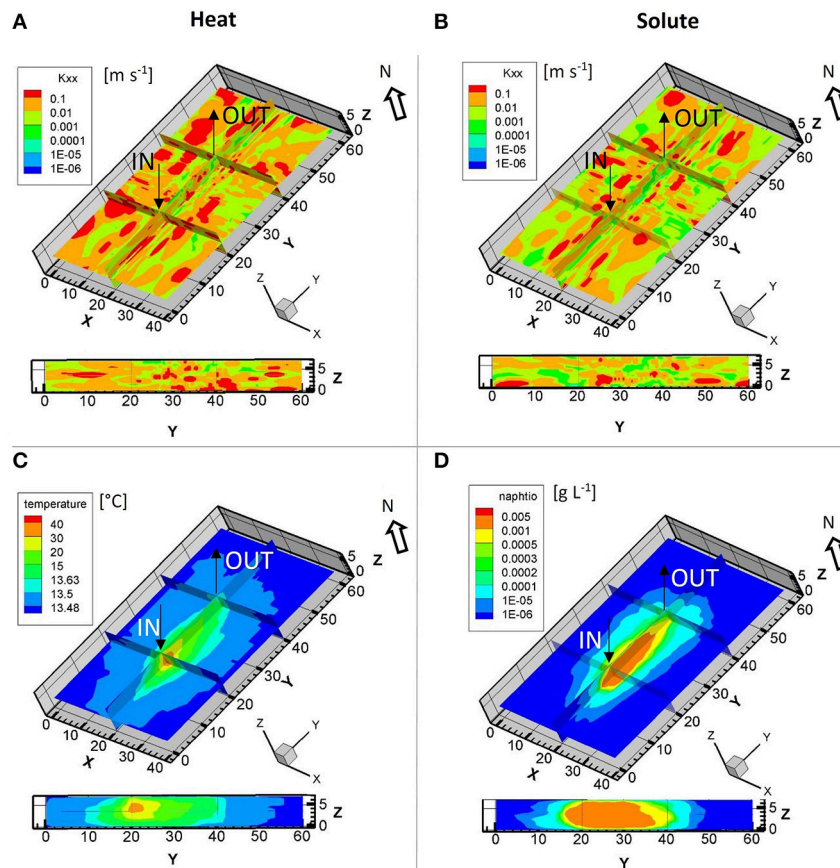


FIGURE 5 | Simulated results at the pumping well using a prior with no K-trend distribution (Scenario A): Spatial K-distribution as obtained for the picked up best simulation for **(A)** heat data and **(B)** for the best naphthionate simulation. Corresponding simulated tracer plumes for best simulation for **(C)** heat data and **(D)** for naphthionate data. Under each 3D visualization, a vertical 2D profile is shown along the gradient axis from the injection well through panels 1, 2, and 3, and down to the pumping well.

realistic imaged preferential pathways (e.g., using improved imaging methods like full-waveform GPR inversion) are probably necessary to understand the heterogeneity surrounding the pumping well.

The sensitivity analysis results at the pumping well using the heat and solute signals using the 250 simulation of scenario B are presented in **Figure 7**. For the heat and the solute signals, the “ $\log(K_{\text{mean}})$ ” is more sensitive than the “ $\log(K_{\text{var}})$ ” and the “PC1,” “PC2,” and “PC3.” The local heterogeneity is still sensitive and important to consider, but the results are less sensitive to small scale heterogeneity as mostly “PC1” to “PC4” are sensitive. For solute transport, “porosity” is also a sensitive parameter to be considered (**Figures 7A,B**), probably due to its direct effect on advection velocity. The analysis supports that simulating complementary tracer behavior requires a realistic description of heterogeneity.

The sensitivity analysis of the pumping well simulations using scenario B is extended by using alternatively the synthetic velocity ratio $v_{\text{heat}}/v_{\text{sol}}$ as prior response for the DGSA (**Figure 8**). Replacing the breakthrough curve as model response by the

velocity ratio $v_{\text{heat}}/v_{\text{sol}}$, averages the model response over the complete transport path.

As a reference, the field measured derived modal velocities at the peak time of heat (i.e., $R_{\text{th}} = 1 \geq \text{wavefront velocity}$) and solute at the pumping well are:

$$\frac{v_{\text{sol,peak}}}{v_{\text{th,peak}}} = \frac{2.05 \cdot 10^{-4} \text{ m s}^{-1}}{1.08 \cdot 10^{-4} \text{ m s}^{-1}} = 1.90 \quad (5)$$

Using the assumption of $R_{\text{th}} = 1$, for scenario B (**Figure 8A**), the obtained velocity ratios are less spread around the observed data. Many solute velocities forecasts have the same value, but with a different corresponding heat velocity (**Figure 8A**). This is an indication that the heat signal provides more information, while the variance of the solute velocity responses decreases.

Applying now the DGSA using this alternative prior response, the sensitivity of “porosity” is now less strong than in the solute case, but stronger than in the heat case (**Figures 7, 8B**). Then, similar to panel 1 to 3, the “ $\log(K_{\text{var}})$,” “gradient,” and “PC1” are the most sensitive parameter at the pumping well (**Figure 8B**). Interestingly, the velocity ratio seems to be not directly sensitive

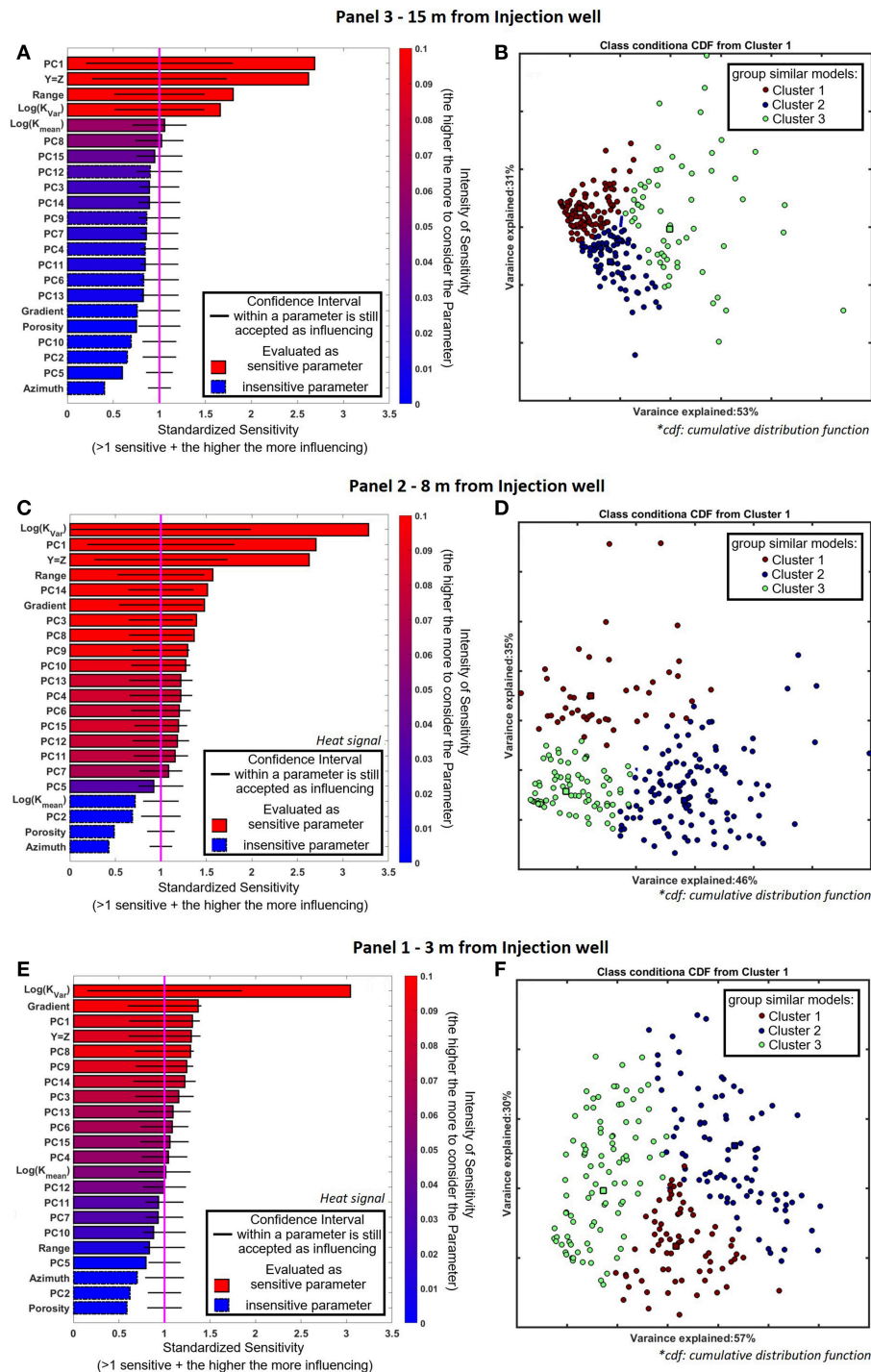
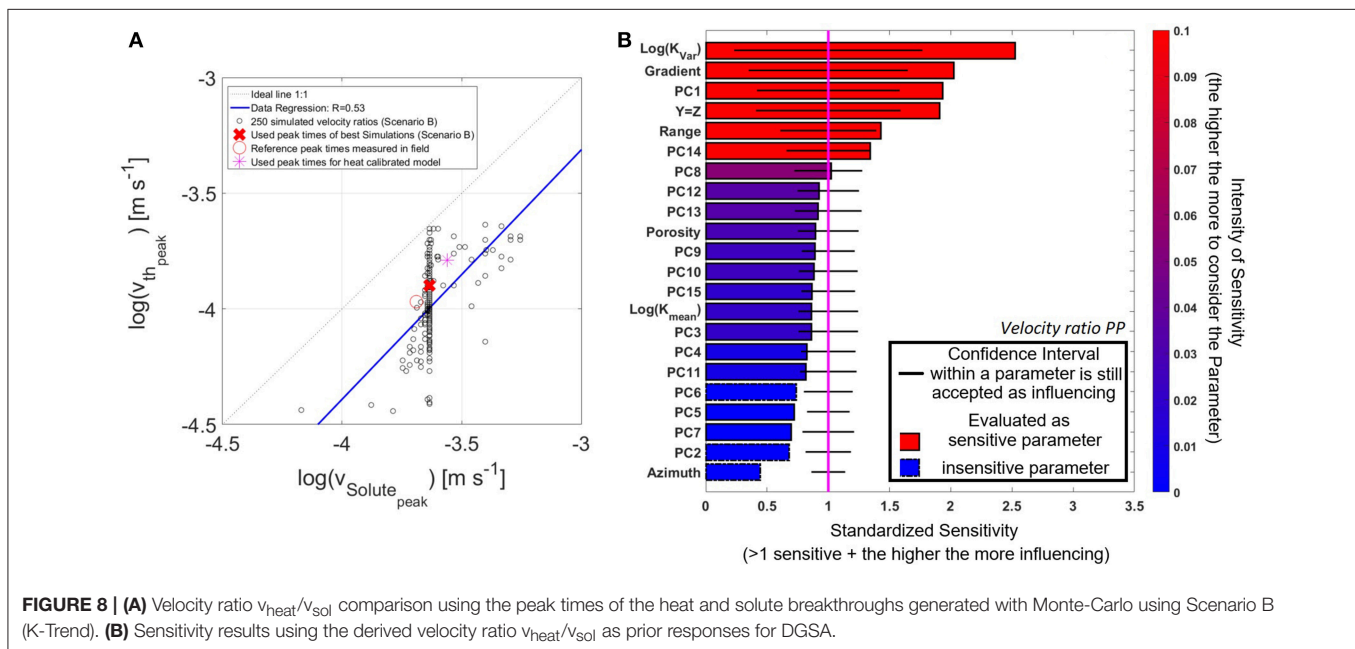
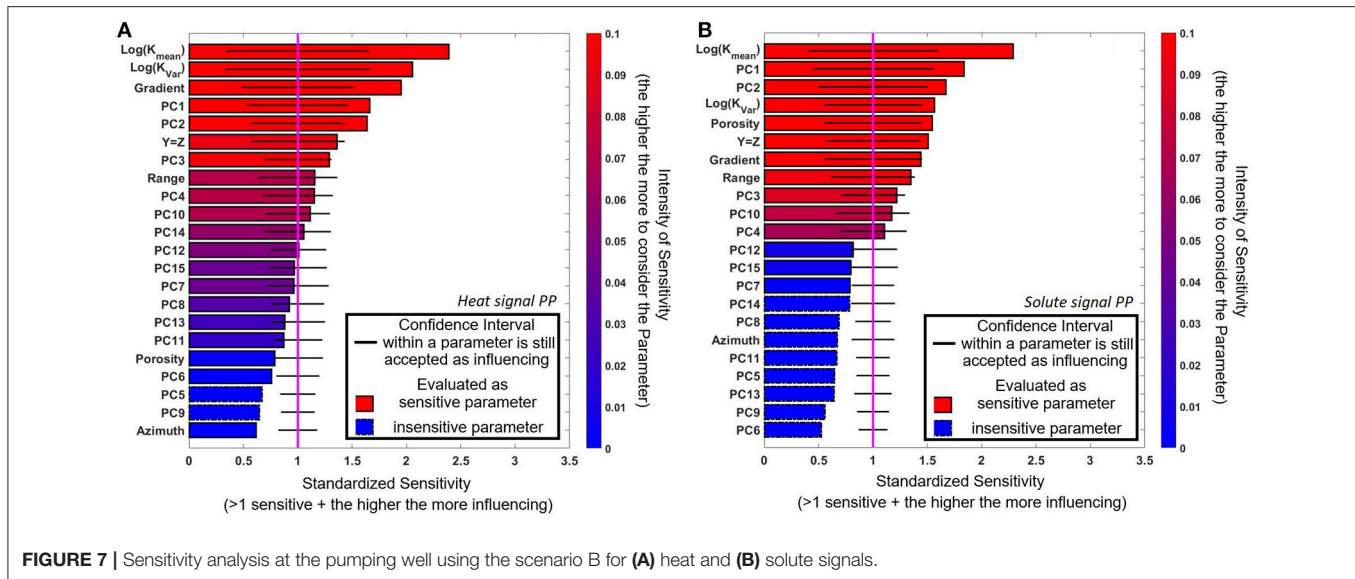


FIGURE 6 | Distance-based sensitivity analysis using the scenario B heat signals at (A) Panel 3, (C) Panel 2, (E) Panel 1, related to the corresponding cluster cumulative distribution function (CDF) for (B) Panel 3, (D) Panel 2, and (F) Panel 1. The square within each cluster in (B), (D), and (F) is the center of mass of each cluster.

to “log(K_{mean}),” but on the global heterogeneity (K_{var} and range), spatial heterogeneity (PC1) and fluxes (gradient). This is a clear indication that preferential flow paths, being the result of an interaction between K-heterogeneity and gradient, is the main reason for the variation in velocity ratio.

At this stage, it can be assessed that the joint heat-solute tracer experiment results are indeed better represented by a K-distribution with a vertical downwards increasing K-trend. The prior with the K-trend is the current best heterogeneity representation for Hermalle-sous-Argenteau test site between



the injection well and panel 3. However, this K-trend is not representative for the part of the simulated domain between the third panel and the pumping well as shown by the Monte-Carlo prior investigation and the DGSA results. This highlights how important a not-falsified prior is for robust decision making, and that every model containing approximations must be use with care for predictions. Furthermore, if the proposed prior seems valid at the local scale, it is not sufficient to explain all observations made at the site. The latter probably requires the inclusion of another level of heterogeneity, accounting for the change of behavior for the tracer.

DISCUSSION

In the field tracer experiment, the heat tracer arrives 1 day after the solute transport and the recovered energy at the pumping well is very low. This delay is a consequence of the different transport processes, mainly the retardation effect related to heat conduction in the solid phase and in the immobile water. For example, the heat tracer test provides useful information to better understand the matrix processes quantifying the immobile water part. This complementary behavior helps to better characterize the actual transport processes occurring in the aquifer, in particular the preferential flow paths.

The previous existing calibrated model was based on heat tracing data only. The new numerical model implemented in the framework of the presented study, showed that the dynamic viscosity has a strong impact on simulated temperature values even in a narrow temperature range. It clearly appeared that this model was failing to reproduce the observed solute concentrations at the pumping well. All attempts to find one single deterministic model fitting both tracer data failed, illustrating the difficulty to approximate solute advection and heat conduction/storage with one single (smoothed) spatial parameter distribution. Even if a global minimum was found, any prediction would remain based on a simplified model with limited prediction capabilities.

To overcome limitations of the deterministic approach, and to avoid full stochastic inversions, performing prior parameter uncertainty investigation using multiple Monte-Carlo realizations offers the possibility to generate more geologically realistic subsurface parameter distributions. Compared to transdimensional inference, which although stochastic in essence and would involve some degree of simplification or parsimony (Sambridge et al., 2012), keeping the full variability in the model is necessary to generate realistic predictions. Thus, in this study, the analysis of those simulations revealed that increasing the spatial heterogeneity of the alluvial deposits allows to better reproduce the observed breakthrough curves. The considered prior uncertainty generates a range of possible outcomes surrounding the observed data. The specific behavior of the breakthrough curves, such as the sharp decrease of temperature after the peaks, is much better reproduced. It is also clearly shown that approximations made in deterministic approaches (e.g., using smoothed K-distributions), strongly influence the results and contribute to higher uncertainty. Furthermore, it appeared that modeling the deposits with two separate layers did not allow the reproduction of the tailing part of the breakthrough curves, whereas a continuous distribution with a vertical downwards increasing trend was more able to model this behavior. However, it was also shown that the used vertical K-trend seems not to be appropriate, i.e., between the third panel and the pumping well. Investigating prior uncertainty here has greatly helped to update the previous conceptual ideas that were mostly based on simple investigations like borehole log description.

The proposed prior with the K-trend is consistent with all observation points (Panel 1 to 3) except the pumping well and Pz18, 19 in the lower aquifer part. Between panel 3 and the pumping well, there are likely preferential flow paths influencing the tracer behaviors, not properly described by the proposed prior. The latter was mainly built based on the high borehole density from intermediate panels. In the original log description of the pumping well (drilled in the 90's), there is no grain size trend described. One possibility is therefore that the strongly heterogeneous alluvial deposits cannot be described by a single simplified parameterization (here Gaussian simulations with a trend) but must include more heterogeneity at the larger scale (for example different vertical trend). It appears that lateral variations occur in the aquifer, stressing the need for a more global description of the heterogeneity, including larger scale sedimentological structures such as channels, and advanced integration of secondary data such as geophysical tomographies

(e.g., Hermans et al., 2015a). Indeed, geophysical data acquired on the site showed lateral variations in electrical resistivity related to gravel structures (Hermans and Irving, 2017). A trade-off between the acquisition of new data to refine understanding and cost-affordable field studies must be found. In this framework, the combination of hydrogeological testing (such as joint-heat tracer tests) with static and time-lapse geophysical data (such as GPR and ERT) at an early stage of site characterization is the key to acquire informative data sets at limited costs.

The prior uncertainty analysis also reveals that each specific temperature breakthrough observation is better reproduced by a different prior realization and, therefore, spatial parameter distribution. This clearly identifies spatial heterogeneity as having a major influence on the simulation results. The solute tracer breakthrough at the pumping well is better represented with models showing preferential flow paths, largely influencing advective-dispersive processes. In contrast, temperature observations in the intermediate panels and at the pumping well seem to be better represented with a slightly more homogeneous model, as conduction is indeed important. This might indicate that a significant part of the pore space is occupied by immobile water. This interpretation shows clearly that the previous conceptual model represented by Peclet numbers of 300 (in the upper layer) and 14,000 (in the lower layer) is not adequate. Furthermore, it shows that the use of a heat tracer alone is not necessarily a good choice to calibrate a model, especially if solute transport should be predicted. Trying to use one single deterministic model parametrization is limited here by two points: (1) complementary tracers cannot really predict each other with classic underlying simplifying assumptions and one parametrization and (2) heterogeneity patterns are complex, meaning that highly parameterized inversion might fail to converge toward a realistic solution.

Similarly, stochastic inversion or optimization techniques might be very complicated to tune to convergence in such a complex layout. Here starts the potential of advanced prediction approaches such as Bayesian Evidential Learning. An informative prior sampled by multiple realizations containing complementary tracer processes might be directly used for prediction if a statistical relationship can be found between data and prediction variables. This kind of approach is probably very promising for the future of hydrogeological modeling where the full, explicit inversion fails due to the lack of sufficient qualitative data to constrain the geometry of the deposits. Some uncertainty component might be irreducible and impossible to resolve through inversion methodologies. Approaches such as BEL, combined with an in-depth prior uncertainty analysis, can therefore be a good way to account for those in prediction uncertainty assessment in a computational efficient way.

The fact that a global sensitivity analysis shows different sensitivity patterns for heat and solute responses, here the porosity, is another indication of the complementarity between the tracers. If heterogeneity is more realistically represented by the K-Trend distribution (scenario B), heat seems, in comparison to solute, insensitive to porosity. Although, the Hermalle-sous-Argenteau site is characterized by a strongly advective system, the heat data set remains dominated by the effect of conduction. Heat is mainly stored around the injection

well and is only slightly and very slowly withdrawn from the reservoir at the pumping well. Some parts of the heat are transported fully through conduction (immobile water and solid matrix).

The presented study also shows that using the Euclidean distance for the distance-based global sensitivity analysis, might be of limited interest for data sets containing strong complementary behavior. It can result in different sensitivities between needed parameterization for the related output. An alternative proposition is to use the velocity ratio as a proxy for the model response, as it allowed to clearly identify preferential flow paths as the main explanation for the difference in tracer behaviors.

CONCLUSION

New innovative imaging methods, namely joint heat and solute tracer tests were combined with advanced field data analysis tools to better assess preferential pathways and associated uncertainty in complex alluvial deposits. This paper demonstrates the limitation of deterministic inversion approaches in capturing the complementary behavior of heat and solute tracers. To overcome those limitations, a prior-uncertainty investigation and a heat-solute velocity comparison are applied. Monte-Carlo simulations are used to investigate the range of simulated data and are complemented by a distance-based global sensitivity analysis. The main results are:

- 1) Heat injection with absolute measured temperature signals between 10 and around 40 °C, as observed for common heat tracer tests, requires considering dynamic viscosity effects in the simulations.
- 2) Although much effort has been done to calibrate a deterministic model on the complex heat tracer data, the underlying approximations always yield a too smooth K distribution, failing to predict the solute breakthrough curve.
- 3) In comparison to (over) simplified deterministic models, stochastic models allow for the relaxation of those approximations, and also consider K-fields in conjunction with heterogeneity. Thus, stochastic models using Monte-Carlo reproduce measurements significantly better. They can better reproduce specific behaviors of breakthrough curves (e.g., the tailing). A simple falsification procedure allows to easily reject prior hypotheses inconsistent with the data. It reveals that a single parameterization is not sufficient to fully describe the complex behavior observed at Hermalle-sous-Argenteau. Due to complex spatial heterogeneity and different behavior of the tracers, the simultaneous fitting of all observation points seems to be almost impossible using explicit inversion approaches. Approaches focusing on the prediction and avoiding model inversion (e.g., such as BEL) might be more successful.
- 4) The global sensitivity analysis reveals that heat seems to be less sensitive to advection parameters like porosity than solute, as was expected based on previous studies. For the strong advective system at Hermalle-sous-Argenteau, heat transport does not seem to be affected by porosity, as long as realistic heterogeneity is considered, using a vertical downwards increasing K-Trend distribution respecting the borehole sedimentology. Indicators linked to local spatial heterogeneity are sensitive parameters for both heat and solute transport, stressing the need to use an adequate prior description of the deposits, a prerequisite for any stochastic Bayesian inversion.
- 5) The tracer velocity comparison shows that the prior and the sampled Monte-Carlo simulations yield a better representation of the joint heat and solute behavior as observed on the field. This is a key point for further research steps in modeling and predicting the transport processes in this aquifer.

DATA AVAILABILITY

Note that the datasets analyzed for this study must be stored on the H+ Network (<http://hplus.ore.fr/en/enigma/data-hermalle>) as a mandatory part of the funding project (ENIGMA ITN). The files are ready and currently under upload. We guarantee, that they will be uploaded as it is a mandatory requirement. Codes for DGSA are freely available at <https://github.com/SCRFpublic>.

AUTHOR CONTRIBUTIONS

RH generated the results, did most of the writing and the layout of the contribution. TH was the main supervisor and motivator for the paper, contributed mostly to the redaction and writing part as well as assisted actively the modeling part. AD ensured the financial support of the joined tracer tests in Hermalle, did the supervision of the previous works (Wildemeersch et al., 2014; Klepikova et al., 2016), contributed in conceptual and methodological discussions, helped for the writing part and is the main promoter of the Ph.D. of RH. PG reviewed the paper from a global and external position and is the second promoter of the Ph.D. project of RH.

ACKNOWLEDGMENTS

This work is part of the Ph.D.-thesis of RH in the ENIGMA ITN framework. ENIGMA has received funding from European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement N°722028. We thank Jef Caers for our fruitful discussion about the results of the sensitivity analysis. Further, we are thankful for the review and the fruitful exchange with three reviewers and the associated Editor.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feart.2019.00108/full#supplementary-material>

REFERENCES

- Alcolea, A., Carrera, J., and Medina, A. (2006). Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Adv. Water Resources* 29, 1678–1689. doi: 10.1016/j.advwatres.2005.12.009
- Anderson, M. P. (2005). Heat as a ground water tracer. *Ground Water* 43, 951–968. doi: 10.1111/j.1745-6584.2005.00052.x
- Brunner, P., and Simmons, C. T. (2012). HydroGeoSphere: a fully integrated, physically based hydrological model. *Ground Water* 50, 170–176. doi: 10.1111/j.1745-6584.2011.00882.x
- Caers, J. (2011). *Modeling Uncertainty in Earth Sciences*. Wiley-Blackwell. doi: 10.1002/9781119995920
- Davis, S. N., Thompson, G. M., Bentley, H. W., and Stiles, G. (1980). Ground-water tracers—a short review. *Ground Water* 18, 14–23. doi: 10.1111/j.1745-6584.1980.tb03366.x
- Doherty, J. (1994). “PEST: a unique computer program for model-independent parameter,” in *Water Down Under*, Vol. 94. 551–554. Retrieved from <https://search.informit.com.au/documentSummary;dn=752715546665009;res=IELENG> (accessed May 10, 2019).
- Doherty, J. (2003). Ground water model calibration using pilot points and regularization. *Ground Water* 41, 170–177. doi: 10.1111/j.1745-6584.2003.tb02580.x
- Fenwick, D., Scheidt, C., and Caers, J. (2014). Quantifying asymmetric parameter interactions in sensitivity analysis: application to reservoir modeling. *Math Geosci.* 46, 493–511. doi: 10.1007/s11004-014-9530-5
- Ferré, T. P. A. (2017). Revisiting the relationship between data, models, and decision-making. *Groundwater* 55, 604–614. doi: 10.1111/gwat.12574
- Fuchs, J. W., Fox, G. A., Storm, D. E., Penn, C. J., and Brown, G. O. (2009). Subsurface transport of phosphorus in riparian floodplains: influence of preferential flow paths. *J. Environ. Q.* 38, 473–484. doi: 10.2134/jeq2008.0201
- Heeren, D. M., Miller, R. B., Fox, G. A., Storm, D. E., Halihan, T., and Penn, C. J. (2010). Preferential flow effects on subsurface contaminant transport in alluvial floodplains. *Trans. ASABE* 53, 127–136. doi: 10.13031/2013.29511
- Hermans, T. (2017). Prediction-focused approaches: an opportunity for hydrology. *Groundwater* 55, 683–687. doi: 10.1111/gwat.12548
- Hermans, T., and Irving, J. (2017). Facies discrimination with electrical resistivity tomography using a probabilistic methodology: effect of sensitivity and regularisation. *Near Surface Geophys.* 15, 13–25. doi: 10.3997/1873-0604.2016047
- Hermans, T., Nguyen, F., and Caers, J. (2015a). Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: workflow and case study. *Water Resources Res.* 51, 5332–5352. doi: 10.1002/2014WR016460
- Hermans, T., Nguyen, F., Klepikova, M., Dassargues, A., and Caers, J. (2018). Uncertainty quantification of medium-term heat storage from short-term geophysical experiments using bayesian evidential learning. *Water Resources Res.* 54, 2931–2948. doi: 10.1002/2017WR022135
- Hermans, T., Oware, E., and Caers, J. (2016). Direct prediction of spatially and temporally varying physical properties from time-lapse electrical resistance data: direct forecast from tl resistance data. *Water Resources Res.* 52, 7262–7283. doi: 10.1002/2016WR019126
- Hermans, T., Wildemeersch, S., Jamin, P., Orban, P., Brouyère, S., Dassargues, A., et al. (2015b). Quantitative temperature monitoring of a heat tracing experiment using cross-borehole ERT. *Geothermics* 53, 14–26. doi: 10.1016/j.geothermics.2014.03.013
- Hoffmann, R., Hermans, T., Goderniaux, P., and Dassargues, A. (2018). “Prior uncertainty investigation of density-viscosity dependent joint transport of heat and solute in alluvial sediments,” in *Computational Methods in Water Resources XXII - Bridging Gaps Between Data, Models, and Predictions* (Saint Malo). Retrieved from: https://www.irisa.fr/sage/jocelyne/CMWR2018/pdf/CMWR2018_paper_166.pdf (Abstract No. 166).
- Irvine, D. J., Simmons, C. T., Werner, A. D., and Graf, T. (2015). Heat and solute tracers: how do they compare in heterogeneous aquifers? *Groundwater* 53, 10–20. doi: 10.1111/gwat.12146
- Klepikova, M., Wildemeersch, S., Hermans, T., Jamin, P., Orban, P., Nguyen, F., et al. (2016). Heat tracer test in an alluvial aquifer: field experiment and inverse modelling. *J. Hydrol.* 540, 812–823. doi: 10.1016/j.jhydrol.2016.06.066
- Krzanowski, W. J. (2000). *Principles of Multivariate Analysis: A User's Perspective*, Revised Edn. New York, NY: Oxford University Press.
- Ma, R., and Zheng, C. (2009). Effects of density and viscosity in modeling heat as a groundwater tracer. *Ground Water* 48, 380–389. doi: 10.1111/j.1745-6584.2009.00660.x
- Maliva, R. G. (2016). *Aquifer Characterization Techniques*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-32137-0
- Oware, E. K., Moysey, S. M. J., and Khan, T. (2013). Physically based regularization of hydrogeophysical inverse problems for improved imaging of process-driven systems. *Water Resources Res.* 49, 6238–6247. doi: 10.1002/wrcr.20462
- Park, J., and Caers, J. (2018). “Sensitivity analysis of reservoir forecasts with both global and local model variables,” in *SCERF Annual Meeting* (Stanford, CA).
- Park, J., Yang, G., Satija, A., Scheidt, C., and Caers, J. (2016). DGSA: a matlab toolbox for distance-based generalized sensitivity analysis of geoscientific computer experiments. *Comput. Geosci.* 97 (Suppl. C), 15–29. doi: 10.1016/j.cageo.2016.08.021
- Ptak, T., Piepenbrink, M., and Martac, E. (2004). Tracer tests for the investigation of heterogeneous porous media and stochastic modelling of flow and transport—a review of some recent developments. *J. Hydrol.* 294, 122–163. doi: 10.1016/j.jhydrol.2004.01.020
- Remonti, M., and Mori, P. (2016). The stochastic approach in groundwater modeling for the optimization of hydraulic barriers: the stochastic approach in groundwater modeling for the optimization of hydraulic barriers. *Remediat. J.* 26, 109–120. doi: 10.1002/rem.21462
- Renard, P. (2007). Stochastic hydrogeology: what professionals really need? *Ground Water* 45, 531–541. doi: 10.1111/j.1745-6584.2007.00340.x
- Rojas, R., Feyen, L., and Dassargues, A. (2009). Sensitivity analysis of prior model probabilities and the value of prior knowledge in the assessment of conceptual model uncertainty in groundwater modelling. *Hydrol. Process.* 23, 1131–1146. doi: 10.1002/hyp.7231
- Sambridge, M., Bodin, T., Gallagher, K., and Tkalcic, H. (2012). Transdimensional inference in the geosciences. *Philos. Trans. R. Soc. A* 371:20110547. doi: 10.1098/rsta.2011.0547
- Sarris, T. S., Close, M., and Abraham, P. (2018). Using solute and heat tracers for aquifer characterization in a strongly heterogeneous alluvial aquifer. *J. Hydrol.* 558, 55–71. doi: 10.1016/j.jhydrol.2018.01.032
- Scheidt, C., Li, L., and Caers, J. (2018). *Quantifying Uncertainty in Subsurface Systems*. Hoboken, NJ; Washington DC: John Wiley and Sons & American Geophysical Union. doi: 10.1002/9781119325888
- Therrien, R., McLaren, R. G., Sudicky, E. A., and Panday, S. M. (2010). *HydroGeoSphere: a Three-Dimensional Numerical Model Describing Fully-Integrated Subsurface and Surface Flow and Solute Transport*. Waterloo, ON: Groundwater Simulations Group, University of Waterloo.
- Wildemeersch, S., Jamin, P., Orban, P., Hermans, T., Klepikova, M., Nguyen, F., et al. (2014). Coupling heat and chemical tracer experiments for estimating heat transfer parameters in shallow alluvial aquifers. *J. Contam. Hydrol.* 169, 90–99. doi: 10.1016/j.jconhyd.2014.08.001
- Zhou, H., Gómez-Hernández, J. J., and Li, L. (2014). Inverse methods in hydrogeology: evolution and recent trends. *Adv. Water Resources* 63, 22–37. doi: 10.1016/j.advwatres.2013.10.014

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hoffmann, Dassargues, Goderniaux and Hermans. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Comparison Between Hydraulic Conductivity Anisotropy and Electrical Resistivity Anisotropy From Tomography Inverse Modeling

Simon Gernez^{1*}, Abderrezak Bouchedda¹, Erwan Gloaguen¹ and Daniel Paradis²

¹ Institut National de la Recherche Scientifique, Eau-Terre-Environnement, Quebec City, QC, Canada, ² Natural Resources Canada, Geological Survey of Canada, Quebec City, QC, Canada

OPEN ACCESS

Edited by:

Philippe Renard,
Université de Neuchâtel, Switzerland

Reviewed by:

Renaud Toussaint,
Université de Strasbourg, France
Abderrahim Jardani,
Université de Rouen, France

*Correspondence:

Simon Gernez
simon.gernez@ete.inrs.ca

Specialty section:

This article was submitted to
Freshwater Science,
a section of the journal
Frontiers in Environmental Science

Received: 19 September 2018

Accepted: 06 May 2019

Published: 04 June 2019

Citation:

Gernez S, Bouchedda A, Gloaguen E and Paradis D (2019) Comparison Between Hydraulic Conductivity Anisotropy and Electrical Resistivity Anisotropy From Tomography Inverse Modeling. *Front. Environ. Sci.* 7:67. doi: 10.3389/fenvs.2019.00067

Hydrogeophysics is increasingly used to understand groundwater flow and contaminant transport, essential basis for groundwater resources forecast, management, and remediation. It has proven its ability to improve the characterization of the hydraulic conductivity (K) when used along with hydrogeological knowledge. Geophysical tools and methods provide high density information of the spatial distribution of physical properties in the ground at relatively low costs and in a non-destructive manner. Amongst them, the Electrical Resistivity Tomography (ERT) has been widely used for its high spatial coverage and for the strong theoretical links between electrical resistivity (ρ) and key hydrogeological parameters, such as K . Historically, ERT data processing was based on isotropic hypothesis. However, the unconsolidated aquifers in Canada reveal in most cases a strong anisotropic behavior for K both with *in situ* or laboratory measurements. Recently, electrical anisotropy has been considered model-wise, but it is seldom considered as an interpretation tool or in the characterization process of the anisotropy of K . In order to evaluate the potential of ERT to assess the anisotropy of electrical resistivity, we developed a forward and inverse modeling code. These codes have been validated and tested on a realistic synthetic case reproducing the behavior of a real aquifer extensively characterized, the site of Saint-Lambert-de-Lauzon in Quebec (Canada). On this site, innovative *in situ* hydraulic tomography has revealed a strong anisotropy, with up to three orders of magnitude between horizontal and vertical K components. In order to confirm the link between *in situ* K - and ρ -anisotropies, an ERT survey has been performed, using the same wells as for the hydraulic tomography. The inversion confirms a strong link between K - and ρ -anisotropies. It demonstrates the suitability of the anisotropic ERT approach coupled with well measurements to provide better estimates of K and its anisotropy at the scale of a site.

Keywords: hydrogeophysics, anisotropy, electrical resistivity tomography, hydraulic conductivity, modeling, groundwater

1. INTRODUCTION

Understanding groundwater flow and contaminant transport in the subsurface for water management and aquifer remediation generally requires a good knowledge of the spatial distribution of hydraulic properties within the aquifers. The hydraulic conductivity (K) is a key parameter to assess as it affects both the direction and velocity of flow and contaminant in aquifers. K can also vary over several orders of magnitude within a same geological unit, which highlights the importance of having accurate high-resolution and high-coverage estimates to reduce errors in groundwater flow and mass transport (de Marsily et al., 2005) and improve groundwater management. While several methods have shown their potential to estimate K at different scales (Butler, 2005), few have been focused on the characterization of its anisotropy that can greatly affect the outcomes of different hydrogeological *in situ* problems, such as groundwater recharge (e.g., Hart et al., 2006) well capture zone (e.g., Barry et al., 2009), and spreading of contaminant plumes (e.g., Falta et al., 2005).

Indeed, K -anisotropy can be obtained from laboratory permeameters on sediment or rock samples collected in the field (Wenzel and Fishel, 1942). However, the difficulties in the experimental procedures related to sample collection and manipulation may restrict reliable estimations of K -anisotropy for certain kinds of materials. Moreover, permeameter estimates may require an up-scaling to field conditions to be representative. In order to overcome this burdens, several authors have proposed different hydraulic tests in wells to estimate K -anisotropy, such as the dipole-flow test using in one well (Kabala, 1993; Zlotnik and Ledder, 1996; Xiang and Kabala, 1997; Zlotnik and Zurbuchen, 1998; Hvilshøj et al., 2000; Sutton et al., 2000; Zlotnik et al., 2001) or two wells (Goltz et al., 2008), the single-well vertical interference test (Burns Jr et al., 1969; Hirasaki et al., 1974; Onur et al., 2002; Sheng, 2009; Paradis and Lefebvre, 2013), and hydraulic tomography (Paradis et al., 2015a, 2016a,b).

While previous hydraulic tests were shown to provide invaluable estimates of K -anisotropy in real field conditions, these methods are time consuming to operate and can thus only provide very local information. In this study, we propose using geophysical data to complement hydraulic tests as geophysical methods can provide broad pictures of the subsurface in a considerably shorter amount of time than hydraulic methods. Electrical methods, in particular direct current (DC) methods, are frequently used to infer porosity and K (Archie, 1942; Lesmes and Friedman, 2005). However, only a few studies have been done to study the anisotropy of the resistivity of unconsolidated sediments. Anisotropy of electrical conductivity (ρ) is a well-known phenomenon (Maillet, 1947) but its accurate *in situ* estimation has only been studied recently (Greenhalgh et al., 2010; Kenkel and Kemna, 2016; Gernez et al., 2018). Moreover, there is a theoretical equivalence between K -anisotropy and ρ -anisotropy in unconsolidated sediments were the electric current flows in the conductive saturated pores (Hubbard and Rubin, 2005). Recently, laboratory investigations have demonstrated strong similarities between ρ - and K -anisotropies on core samples (Adams et al., 2016). In addition, recent field works have shown that taking into account ρ -anisotropy in DC surveys leads

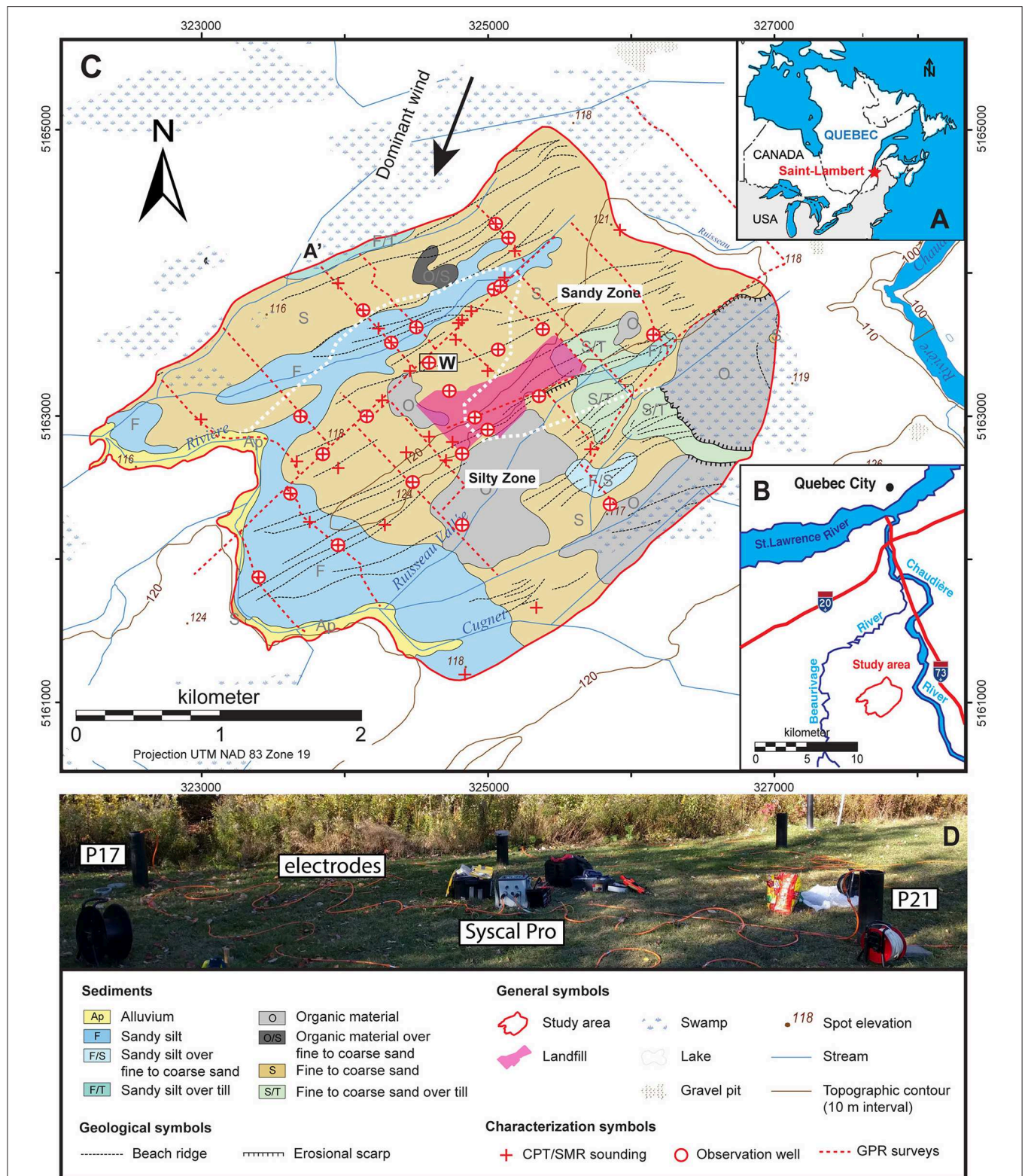
to more accurate estimations of both ρ values and structures (Pekşen and Yas, 2018), and have shown reasonable estimates of hydraulic anisotropy in slightly anisotropic aquifer systems (Yeboah-Forson and Whitman, 2014).

The objective of this paper is to demonstrate the ability of DC methods to quantify ρ -anisotropy and to illustrate how it compares with K -anisotropy in a real case study. After introducing the study area (section 2) and presenting theoretical considerations related to ρ -anisotropy (section 3), we provide methodological insights, through a synthetic case, related to DC data acquisition to ascertain the presence of ρ -anisotropy (section 4). Then, the methodology is applied for a real case study known to be highly heterogeneous, and ρ -anisotropy estimated through anisotropic inversion is compared to K -anisotropy obtained with hydraulic tests at the study site to strengthen the reliability of the proposed approach (section 5). This study exposes the capacity of DC surveys to improve hydrogeological characterization.

2. STUDY AREA AND EVIDENCES OF ANISOTROPIC CONDITIONS

The study area is located in Saint-Lambert-de-Lauzon (SLdL), 30 km south of Quebec City, Canada (Figure 1). The SLdL study area is a 12 km² sub-watershed surrounding a decommissioned sanitary landfill site in an unconfined granular aquifer. The surficial sediments composing the aquifer consists primarily of Late Quaternary sandy and silty sediments that were deposited in the receding Champlain Sea, which was an arm of the Atlantic Ocean that invaded the St-Lawrence Valley at the time of the last deglaciation (Bolduc, 2003). Deposition of the Saint-Lambert site was controlled mainly by longshore currents that redeposited in littoral and sublittoral settings that supplied the Chaudière River paleodelta. This geological depositional environment leads thus to sediment size ranging from fine sand to very fine silt with poor to very poor grain-size sorting. Furthermore, this environment shows minor proportions of clay (generally <20%). Clay in major proportions (>50%) is only present below the cross-section studied. The resulting superposition of finely layered sand and silt sediments create very heterogeneous distribution of sediments at centimetric to decametric scales along with more gradual lateral transitions in these littoral and sublittoral sediments as a result of changing energy levels along the Champlain Sea shorelines. The depth of the granular aquifer varies in depth between 0 and 22 m, the water table is generally within 2 m from the surface (Paradis et al., 2014; Tremblay et al., 2014).

This site has been extensively characterized by previous studies using different techniques, such as Ground Penetrating Radar (GPR) and resistivity surveys, Cone Penetrometer Test—Soil Moisture Resistivity soundings (CPT/SMR), hydraulic tests in wells and logging (Paradis et al., 2014; Tremblay et al., 2014). These data allowed to obtain valuable information on the structure of the aquifer system (aquifer and aquitard layers) including information on its heterogeneity. Particularly, several observations suggest that the heterogeneous nature of the sediments at a fine scale may induce anisotropy at larger scale posing challenges to the interpretation of flow and transport



processes in this environment. Anisotropy can be due to the microscopic scale organization of the minerals (micro or intrinsic anisotropy, e.g., crystals ordered structure or oblong grains) or, as is the case here, to the macroscopic structural elements of the ground (macro or extrinsic anisotropy, e.g., fractures or alternating heterogeneous beds). First, the comparison of 59 vertical hydraulic conductivity (K_V) estimates made on 15 cm undisturbed sediment samples with a laboratory permeameter to horizontal hydraulic conductivity (K_H) values obtained from high-resolution multi-level slug tests on similar intervals reveals a strong K -anisotropy even at this small scale. K -anisotropy (or the ratio of K_H on K_V , K_H/K_V) was indeed up to two orders of magnitude (Paradis and Lefebvre, 2013). Then, numerical inversion of vertical interference slug tests and hydraulic tomography experiments indicate that K -anisotropy should be considered to match hydraulic responses measured in wells (Paradis and Lefebvre, 2013; Paradis et al., 2016a). For a 60 cm vertical resolution of the numerical grid, K_H/K_V values ranged from near isotropy (1) to more than 100. Moreover, comparison of high-resolution cone-based ρ measurements (SMR probe) with collocated estimates of the ρ computed with surface-based surveys (ERT) revealed a bias between the two data sets (Ruggeri et al., 2014). For instance, the magnitudes of the SMR probe data are generally higher than those of the ERT surveys. This suggests that given the SMR probe sense essentially the horizontal component of ρ due to the configuration of its electrodes (spaced vertically by 9 cm), lower ρ values from ERT surveys are the results of the influence of the ρ -anisotropy induced by the heterogeneous nature of the sediments (section 3.2.2). Finally, those evidences motivate the need to develop a geophysical approach able to handle this anisotropy to provide insights about K -anisotropy in order to better characterize aquifer systems for groundwater flow and contaminant transport studies.

3. ELECTRICAL RESISTIVITY ANISOTROPY

3.1. Theoretical Considerations and Definitions

Electrical anisotropy refers to the directional dependence of electrical conductivity or resistivity which results in the directional dependence of the measured potential fields. This means that the current can preferentially flow in certain directions compared to others. Ohm's law establishes the relationship between an injected electric current in the ground and the induced potential field (Dey and Morrison, 1979). In order to take into account the 2D electrical anisotropy, the scalar conductivity σ in Ohm's law is replaced by the conductivity tensor $\vec{\sigma} = \begin{bmatrix} \sigma_H & 0 \\ 0 & \sigma_V \end{bmatrix}$ (or its inverse, the resistivity tensor $\vec{\rho} = \vec{\sigma}^{-1}$), with σ_H and σ_V being the conductivity values in the horizontal and vertical directions, respectively (Greenhalgh et al., 2009). Anisotropic Poisson's equation has the following expression in the 2.5D case, i.e., 2-D resistivity structure (plane invariance) and 3-D current flow (Zhou et al., 2009):

$$\nabla \cdot (\vec{\sigma} \nabla \tilde{\phi}) + k_y^2 \sigma_H \tilde{\phi} = -\frac{I}{2} \delta(\mathbf{r}(x, z) - \mathbf{r}_s(x_s, z_s)) \quad (1)$$

where $\tilde{\phi}$ is the potential in the frequential domain, k_y is the wavenumber, $\mathbf{r}(x, z)$ are the coordinates in the computational domain or on its boundaries, I is the current source intensity located at $\mathbf{r}_s(x_s, z_s)$ and δ is the Dirac function. The coefficient of anisotropy is defined as $\lambda = \sqrt{\sigma_H/\sigma_V} = \sqrt{\rho_V/\rho_H} \geq 1$: anisotropy increases as λ departs from the value of 1 ($\lambda = 1$ corresponds to isotropy). In this study, we will consider an H/V anisotropy. More complex geometries are handleable by the numerical modeling tool we developed to this end (AIM4RES), but they will not be investigated in this study.

3.2. Diagnosis of Electrical Anisotropy

The next sections aim at demonstrating the effects of electrical anisotropy on the interpretation of ERT data using isotropic ERT inversion. We take advantage of three particular effects to propose an electrical diagnosis to detect anisotropy on measured electric potentials. These effects are observable without the need for a complete characterization study, both in terms of field and numerical resources.

3.2.1. Importance of Data Acquisition Protocols

The measured electric potential field is linked to the amount of electric current passing through the different heterogeneous part of the ground. Hence, in the case of surface ERT measurements, a thin conductive anisotropic layer and a thicker less conductive isotropic layer can produce the same electric potential differences (equivalence principle, Maillet, 1947). In other words, it is impossible to distinguish between isotropic layer response from anisotropic layer response using surface ERT data. Consequently, isotropic ERT inversion (Loke, 2001; Bouchedda, 2010) of surface anisotropic data always converge to an equivalent resistivity model which is not representative of the true electrical state of the ground, leading to erroneous resistivity model of the earth. To overcome this problem, anisotropic ERT acquisition and inversion should be used. To address the data directionnality problem, we unavoidably need borehole electrodes along with surface electrodes. In that way, anisotropic ERT inversion requires an optimization of the acquisition protocol, in order to converge toward the true solution.

Nevertheless, in presence of anisotropy, isotropic inversion of ERT directional data leads to unrealistic solutions. It is explained by the fact that there is no physical isotropic solution fitting both surface and inhole data. To demonstrate this effect, two data experiments were simulated using only surface electrodes in the first one and both borehole and surface electrodes in the second one. The resistivity model consists of two horizontally anisotropic layers (Figure 2A). The first layer has a thickness $h = 4$ m with $\rho_H = 100 \Omega \cdot \text{m}$ and $\rho_V = 400 \Omega \cdot \text{m}$. The second layer is a semi-infinite space with $\rho_H = 10 \Omega \cdot \text{m}$ and $\rho_V = 40 \Omega \cdot \text{m}$. For the whole section, the anisotropy coefficient λ is 2.

The first experiment was performed using only surface Wenner array data. We assumed the convergence is reached as the RMSE values are very low (0.0026%), but the inverted model (Figure 2B.1) is not consistent with the true resistivity model (Figure 2A) neither in terms of amplitude of the resistivity nor in terms of geology. According to the theory, the resistivity of the upper layer appears to be $\sqrt{100 \cdot 400} = 200 \Omega \cdot \text{m}$ and the

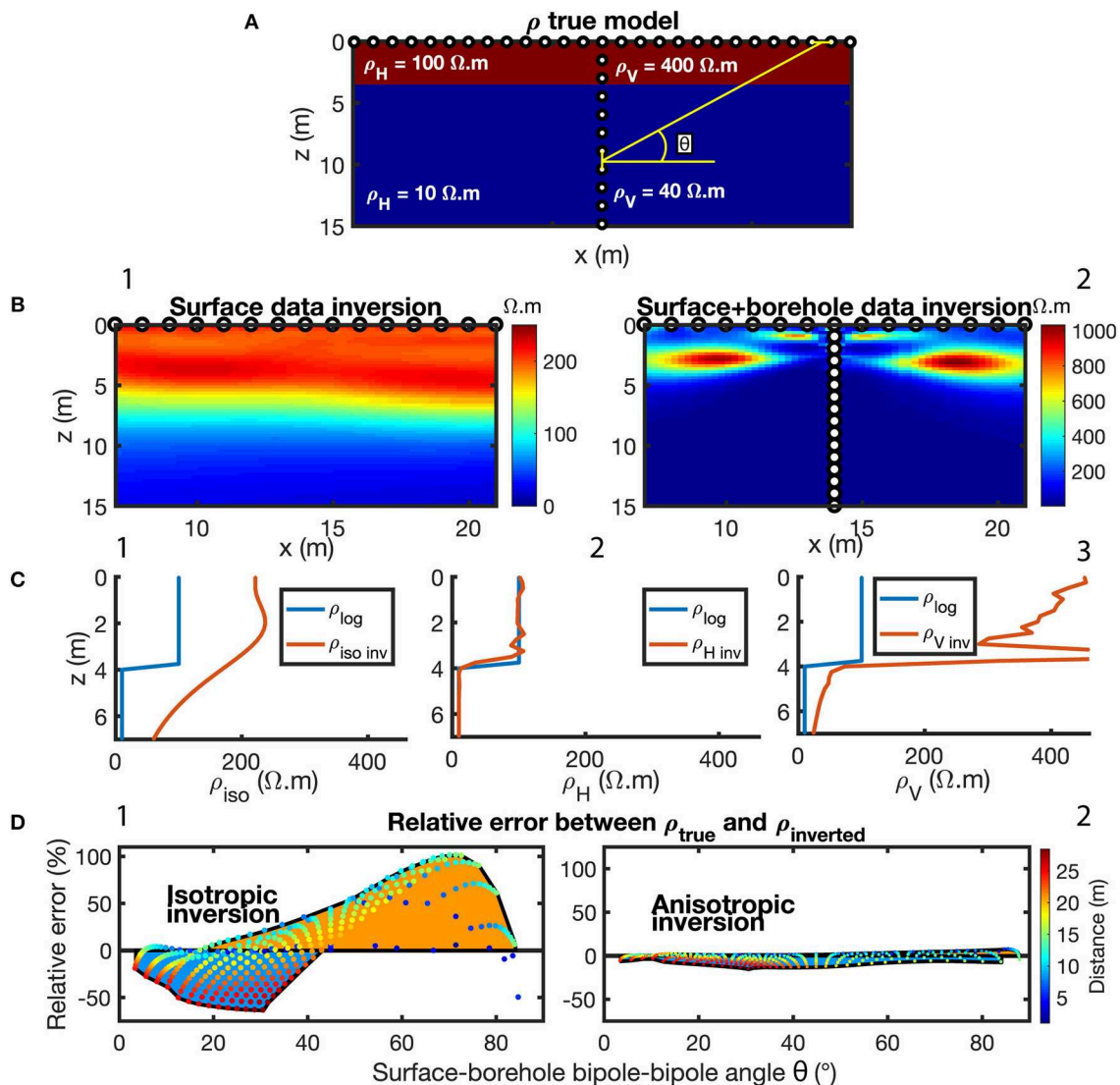


FIGURE 2 | (A) Synthetic model, constituted of an anisotropic layer over an anisotropic semi-infinite space. Upper layer: $\rho_H = 100 \Omega.m$ and $\rho_V = 400 \Omega.m$. Semi-infinite space: $\rho_H = 10 \Omega.m$ and $\rho_V = 40 \Omega.m$. $\lambda = 2$ for the whole model. Electrodes (white dots) are located at the surface and in-hole borehole at $x = 14$ m. In yellow is shown an example of surface-borehole measure angle θ . **(B)** (section 3.2.1) Isotropic inversions of potentials acquired using (B.1) only surface electrodes and (B.2) surface and borehole electrodes (borehole at $x = 14$ m). **(C)** (section 3.2.2) Comparison between logged (ρ_{log} , blue curve) and inverted (ρ_{inv} , red curves) resistivities: (C.1) isotropically inverted resistivity. (C.2) ρ_H from inversion vs. ρ_{log} . (C.3) ρ_V from inversion vs. ρ_{log} (borehole at $x = 20$ m). The logged resistivity is the direct resistivity measurement at $x = 20$ m, therefore the blue curve is the same for all (C.1–3) graphs. **(D)** (section 3.2.3) Relative error behavior as a function of the measure angle θ . The points are the error values, their color represents the associated dipole-dipole distance. Orange area represents a positive error, blue area represents a negative error. (D.1) Relative error from an isotropic inversion of data acquired on an anisotropic model, displaying a sigmoid shape. (D.2) Relative error from an anisotropic inversion of data acquired on an anisotropic model, relative error is close to zero and is not angle dependant (borehole at $x = 25$ m).

resistivity of the semi-infinite space appears to be $\sqrt{10 \cdot 40} = 20 \Omega.m$. In addition, the thickness of the upper layer appears to be $\lambda \cdot h = 8$ m.

In addition to the previous Wenner surface array, a dipole-dipole array in the borehole and a mixed surface-borehole array were added to the acquisition protocol in the second experiment. The isotropic ERT inversion result of these data sets are presented in the **Figure 2B.2**. It can be clearly seen that isotropic inversion of directional ERT data leads to unrealistic

solutions. Furthermore, the final misfit between measured data and predicted data is high (RMSE = 22.1%) for isotropic inversion in comparison to anisotropic inversion (0.002%), showing that directional data are unable to fit an isotropic solution. This can be used as an evidence of electrical anisotropy.

3.2.2. Effect of Anisotropy on ERT Measurements

In the case of horizontally anisotropic resistivity model, it has been pointed out by Maillet (1947) and

Keller and Frischknecht (1966) that the measurements made in the horizontal direction are equal to geometric mean of horizontal and vertical resistivity components, while the measurements made in the vertical direction are equal to the horizontal resistivity components. This is the paradox of electrical resistivity anisotropy: measurements along vertical profiles in the case of layered anisotropic model are sensitive to horizontal component as shown in our synthetic model example. For formal demonstration please see Lüling (2013). Indeed, electrical resistivity logs can be used as *in situ* measurements of the horizontal resistivity ρ_H that can be introduced as a constraint in the inversion system or employed in combination of surface ERT data to diagnose the anisotropy.

In our case, the electrical resistivity logging was measured using a CPT-SMR instrument which does not require a well installation, simplifying its implementation. The probe is 5 cm thick and 9 cm long. Note that the small probe diameter and the small electrodes separation make the hole effect negligible and the measured resistivity is only sensitive to ρ_H .

In order to demonstrate the effectiveness of resistivity well logs as anisotropy diagnosis tool, let us compare well logs resistivities and estimated resistivities from isotropic ERT inversion of surface data obtained on an anisotropic resistivity models. When isotropic ground is considered, both resistivities are expected to be similar. For horizontally anisotropic ground (as in SLdL), well log resistivities are equal to horizontal resistivity components whereas isotropic ERT inversion returns an equivalent resistivity model which combines both horizontal and vertical resistivity components (equivalence and paradox effects). In other words, the difference between the two resistivities can be very important depending on the value of anisotropy coefficient. Consequently, any difference between the two resistivities is an indication of the presence of anisotropy.

To illustrate the anisotropy diagnosis using synthetic model let consider the previous two layered anisotropic model (Figure 2A). Figure 2C.1 shows the comparison between the electrical resistivity logging (blue curve, e.g., obtained with a CPT-SMR logging at $x = 14$ m) and the corresponding resistivity (red curve) estimated using the isotropic ERT inversion of surface data. Both curves depart from each other, indicating the presence of anisotropy. Figures 2C.2, C.3 show the comparison between the same electrical resistivity logging and collocated horizontal and vertical resistivities obtained from anisotropic ERT inversion of surface and borehole data. As logged resistivity is carried along a vertical profile, it is only sensitive to the horizontal resistivity component of the ground and thus departs from the estimated vertical resistivity of anisotropic medium which confirms the validity of our methodological approach to quantify the anisotropy. Please see references for more details.

3.2.3. Relative Error vs. Array Angle

To assess the effect of anisotropy on data misfit error of isotropic ERT inversion, we consider the same two layered synthetic model as in sections 3.2.1 and 3.2.2. The data acquisition is simulated using only surface-borehole data. The current electrodes are

located in wells and the potential electrodes are located at the surface. Array configuration were made using 50 surface electrodes and 15 boreholes electrodes. Electrode spacing is 1 m. Centers of each bipole describe a skew line with the horizontal, forming an angle θ (Figure 2A). The simulated potential data (ϕ_{true}) are isotropically inverted. The data misfit relative error is computed as the normalized difference between the potential data calculated using the inverted model (ϕ_{calc}) and the simulated potential data (ϕ_{true}):

$$\frac{\phi_{calc} - \phi_{true}}{\phi_{true}} * 100 \quad (2)$$

Figure 2D displays scatter plots of data misfit relative error as function of array angle θ for isotropic and anisotropic ERT inversion. For angles between 0 and 45°, the relative errors are mostly negative, meaning that ϕ_{calc} are underestimated (blue area in Figure 2D.1). The errors become positive for angles between 45 and 90° (orange area in Figure 2D.1). This sigmoid error shape is expected when an isotropic inversion is used to invert ERT data of horizontally anisotropic media. The horizontal resistivity component ρ_H is lower than the vertical resistivity component ρ_V . At low acquisition angles, current flow is mainly driven by ρ_H , and isotropic inversion underestimate the apparent resistivity values. Conversely, current flow is mainly driven by ρ_V at high angles and isotropic inversion overestimate the apparent resistivity values. The underestimations (blue area at low angles) compensate overall the overestimations (orange area at high angles). The sigmoid shape arises for any borehole dipole depth. For a given angle value, the more space is integrated—i.e., the deeper the borehole dipole—, they higher the local relative error (as represented by the colored points in Figure 2D.1). The total mean error of the measures is −2%. The same relative error computation is made with anisotropically inverted data. It shows an error close to zero (−2.75%) and independent of array angles (Figure 2D.2). A sigmoid relative error shape between true data and calculated data resulting from isotropic ERT inversion is then a strong indication of anisotropy existence in the ground.

The points addressed by section 3.2 give various ways to detect electrical anisotropy by analyzing ERT data. It can be difficult to gather information from multiple sources on the field: lack of outcrops, incapacity of drilling numerous wells (e.g., as needed for hydraulic tomography), or even total absence of well. We propose this preliminary methodological qualitative study to ascertain the presence of electrical anisotropy, and then *a fortiori* to ascertain the presence of hydraulic anisotropy. A full quantitative anisotropic study, in terms of data acquisition and processing, represents more time, resources and costs than a common isotropic study. Nevertheless, processes comprehension and interpretations suffer greatly from the lack of trustful data, and anisotropy consideration might be unavoidable to produce better forecasts, reducing the uncertainties and then the risks on the investigation or engineering works.

The next sections methodologically demonstrate the ability of anisotropic ERT campaigns to quantify the electrical and hydraulic ground anisotropies.

4. ANISOTROPIC ELECTRICAL RESISTIVITY INVERSION FOR A SYNTHETIC CASE

Before starting the real case study, a synthetic electrical model is created on the basis of hydraulic tomography results. Forward modeling is performed on this model to generate synthetic electric potentials to simulate data acquisition in the field. After that, anisotropic ERT inversion is used to reconstruct ρ_H and ρ_V fields. The comparison between anisotropically inverted fields and the original synthetic model will allow to assess the robustness of the proposed approach to estimate ρ -anisotropy. The section describes the synthetic model (section 4.1), the optimal data acquisition protocol for anisotropic characterization of the subsurface (section 4.2), and the details of the forward and inverse modeling procedures (section 4.3) along with the performances of inversion (section 4.4).

4.1. Synthetic Model

The synthetic model used in this numerical experiment mimics the K fields model obtained from the hydraulic tomography experiment measured between wells P21 and P17 (see **Figure 1** for location) by Paradis et al. (2016a). The two wells are separated by 8 m and the aquifer thickness is 9 m, which corresponds to the approximate length of the wells. The K fields were directly transformed in σ values by increasing K by a factor 10^5 , which were inverted to obtain ρ values, to make it realistic of earth materials at the site (values between 10^2 and $10^5 \Omega.m$, **Figure 3**). This transformation leads to $\log(\rho_H) \in [1.30; 2.52] \log(\Omega.m)$ (**Figure 3A**), $\log(\rho_V) \in [2.44; 4.60] \log(\Omega.m)$ (**Figure 3B**), and $\log(\rho_V/\rho_H) \in [1; 3]$ (**Figures 3C,D**), which could be qualified as a moderate anisotropic field. For the synthetic simulation, 34 electrodes (black dots in **Figure 3**) were placed around the synthetic model: every 1 m inside the wells and every 0.5 m at the surface.

4.2. Optimal Data Acquisition Protocol

As in the isotropic case presented in section 3, it is crucial to adapt the data acquisition protocol given that electrode configurations are not necessarily sensitive to the same subsurface features. Different electrodes configurations do not have the same sensitivity to anisotropy (Wiese et al., 2009; Greenhalgh et al., 2010; Kenkel and Kemna, 2016). In particular, Bing and Greenhalgh (2000) have detailed the use of cross-hole ERT. Thus, configurations not sensitive to anisotropy should be avoided as using them will lead the inversion toward an isotropic (hence wrong) inverted solution. Using the synthetic model previously described (**Figure 3**), nine quadrupoles configurations were tested (**Figure 4**) to assess their ability to detect anisotropy (**Figure 5**). Those quadrupoles use different combinations of electrodes placed in wells and at the soil surface. The following arrays were tested:

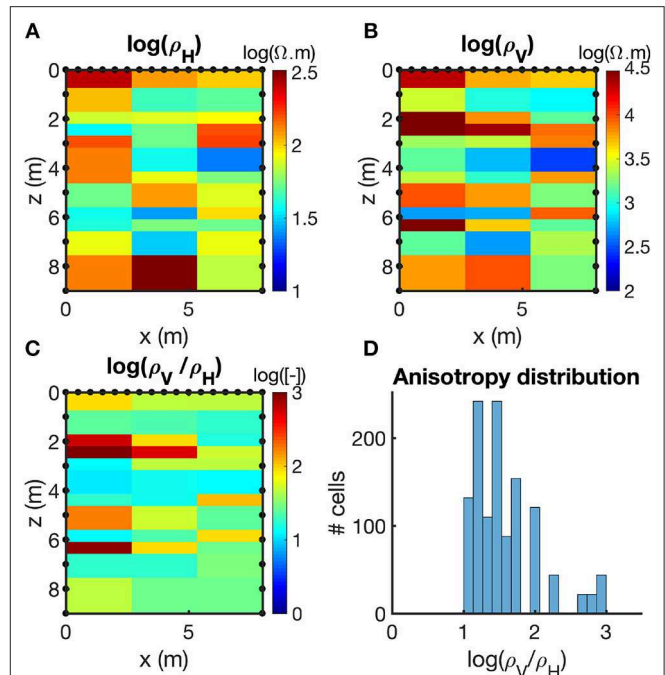


FIGURE 3 | Electrical resistivity synthetic model based on real case hydraulic study results from Paradis et al. (2016a). The horizontal (**A**) and vertical (**B**) resistivities ρ_H and ρ_V are the components of the anisotropic resistivity tensor $\bar{\rho}$. The anisotropy (**C**) shows locally up to three orders of magnitude. Black dots represent the electrodes locations. The distribution of anisotropy is represented on the histogram (**D**).

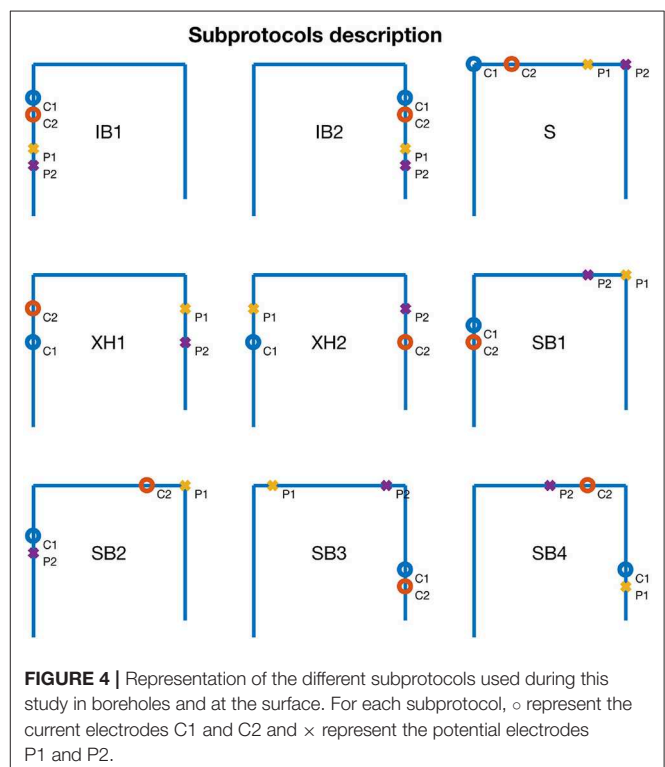


FIGURE 4 | Representation of the different subprotocols used during this study in boreholes and at the surface. For each subprotocol, \circ represent the current electrodes C1 and C2 and \times represent the potential electrodes P1 and P2.

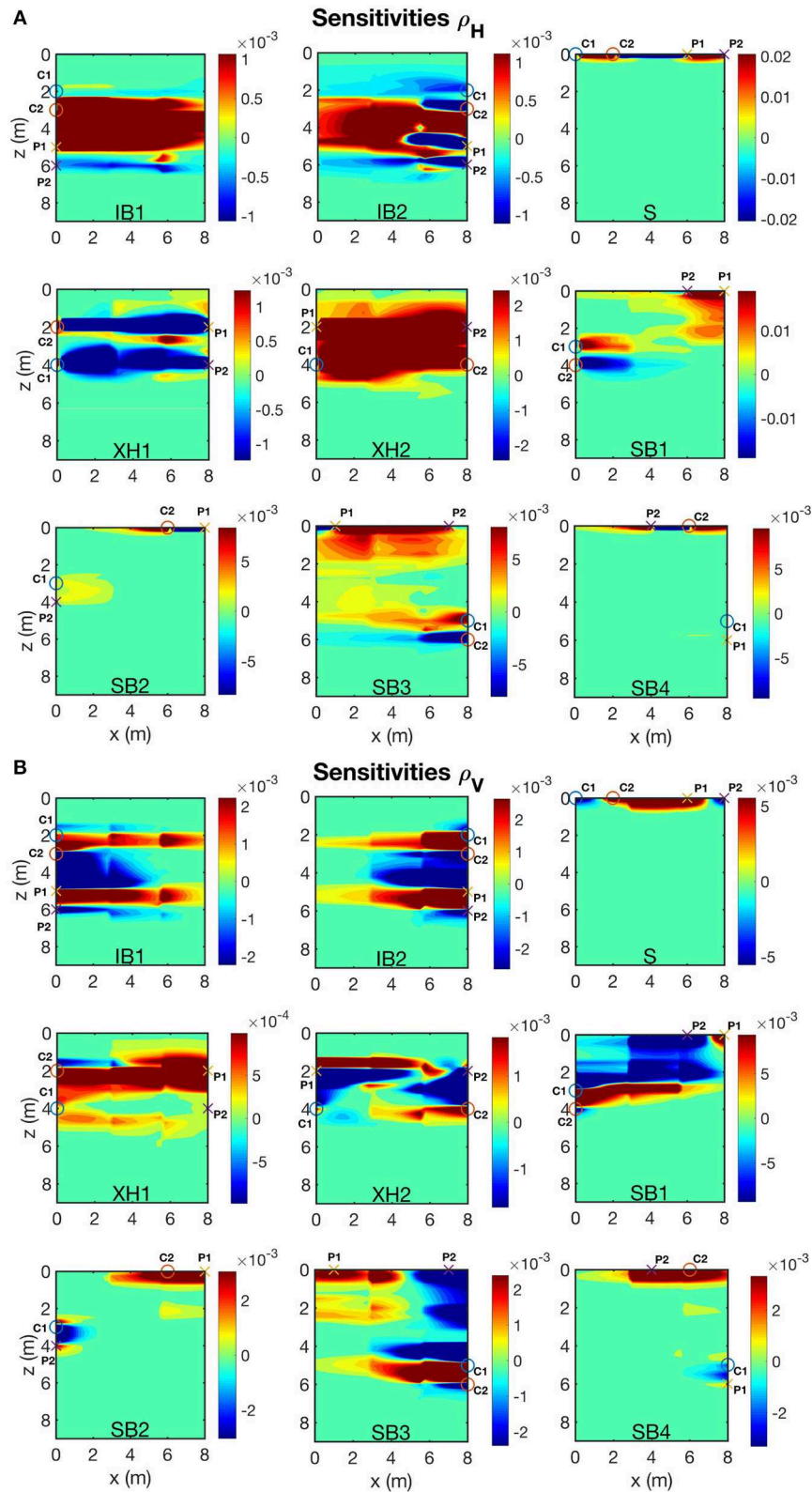


FIGURE 5 | The sensitivity pattern is obtained from the Jacobian matrix (matrix of the potential derivatives according to the resistivity values of the model) computation of the electrode quadrupoles on the synthetic model from **Figure 3**. **(A)** Sensitivity pattern relative to ρ_H and **(B)** Sensitivity pattern relative to ρ_V . Same subprotocols as **Figure 4**. \circ represent the current electrodes C1 and C2 and \times represent the potential electrodes P1 and P2.

- Two inline borehole protocols with the four electrodes in the left (IB1) or right (IB2) borehole,
- One surface protocol with the four electrodes at the surface (S),
- Two cross-hole protocols with the current electrodes in the same (XH1) or separated (XH2) boreholes,
- Two surface-borehole protocols using left (SB1, SB2) or right (SB3, SB4) borehole, with both current electrodes (or indifferently potential electrodes) in the borehole, the other two electrodes at the surface (SB1 or SB3) or with one current electrode and one potential electrode in the borehole, and one current electrode and one potential electrode at the surface (SB2 or SB4).

For each of these quadrupoles, the sensitivity of the electric potentials to ρ_H and ρ_V was analyzed using values of the Jacobian matrix, which is the matrix of the potential derivatives according to the resistivity values of the model (Greenhalgh et al., 2009).

As each quadrupole have a distinct sensitivity pattern, several observations can be made from **Figure 5**. First, inline borehole (IB1, IB2) and surface (S) configurations show larger sensitivities close to the location of the electrodes, which limit the area of investigation of the surveys performed with those quadrupoles. On the other hand, crosshole (XH1, XH2) and surface-borehole (SB1, SB2, SB3, SB4) quadrupoles are sensitive to much larger areas. However, the magnitude of the sensitivities is larger for S, SB1, and SB3 quadrupoles, which can better resolve ρ_H and ρ_V in the associated sensitive areas using those configurations. Then, the sensitivity patterns for symmetric configurations (IB1 and IB2, SB1 and SB3, SB2 and SB4) show similar behavior despite the electrodes being located in different materials due to the heterogeneous nature of the synthetic model. The contrast in ρ material seems thus to have less impact on sensitivities than the electrode configuration itself. Also, sensitivity patterns for ρ_H and ρ_V are different. This means that a quadrupole can be more influenced by one component of $\bar{\rho}$ than by the other. Quadrupoles configuration should be thus chosen accordingly to avoid bias in the measurements.

Given the previous observations, the S, IB1, IB2, XH1, XH2, SB1, and SB3 quadrupoles were then found the most informative and useful for an anisotropic inversion. IB1, IB2, XH1, XH2, SB1, and SB3 quadrupoles were chosen because they appear to be more sensitive to the central region of the investigated section, far from the surface and the wells. They provide information on the whole section. The S quadrupoles, even not significantly sensitive in depth, have been considered because they provide constraints for the model. This further constraints are particularly important since the surficial cells are not well-constrained by borehole electrode configurations and have an important effect on the inversion. Amongst the chosen subprotocols, electrode configurations might be sensitive to the same parts of the characterized section, incorporate redundancy. This redundancy is to be avoided in order to ease convergence of an inverted solution.

4.3. Forward and Inverse Modeling

In this section, the forward- and inverse-modeling (Dey and Morrison, 1979) adapted for anisotropic conditions (Gernez

et al., 2018) is used to compute both forward and inverse modeling of ERT data on a numerical grid made of 8,970 squared cells of 25×25 cm. The forward modeling on the synthetic model used a protocol made of 755 quadrupoles chosen to be sensitive to the anisotropy (IB1[73], IB2[63], S[39], XH2[220], SB1[118], SB3[242]). The synthetic potentials are transformed on equivalent apparent resistivities ρ_{app} and were then inverted to reconstruct ρ_H and ρ_V fields. For this reason, XH1 has not been considered since most of its apparent resistivities were negative. To reduce the risk of the model to converge toward a local minimum, homogeneous and anisotropic ρ values were also used to initialize the inverted model ($\ln \rho_H = 4.75 \ln \Omega.m$, $\ln \lambda = 6.2146$). A weak first-order Tikhonov constraint (α) on the vertical direction was used ($\alpha_V/\alpha_H = 0.5$) in order to promote horizontal structures (which is consistent with geological information and GPR data). This horizontal smoothing is used to favor a layered inverted model. Conversely, a ratio departing too much from 1 will show horizontal artifacts. By rule of thumb, we choose $0.1 < \alpha_V/\alpha_H < 1$. Refining is done by trial and error. A regularized iterative Gauss-Newton method was used to tackle the non-linear inverse problem.

4.4. Inversion Performances

The **Figure 6C** presents a histogram of the relative error between synthetic and inverted potential values after convergence of the model at the seventh iteration. With most of the relative error centered on zero and an overall low RMSE of 1.7%, the inversion is considered to fit almost perfectly the synthetic potentials.

Moreover, the **Figure 6A** (right) shows ρ fields resulting from the inversion. While the ρ fields are smoother than the synthetic model [**Figure 3**, **6A** (left)], the main features of the subsurface are reproduced, such as alternations of low and high ρ layers and overall range of ρ variations. Also, the analysis of the frequency distributions of synthetic and inverted ρ -anisotropy reveals similarities with quasi-normal distributions (**Figures 3D**, **6B**). Examination of rho profiles along the depth (**Figure 6D**) illustrates more specifically the good agreement between synthetic and inverted fast alternations between low and high values of rho-anisotropy: both the trends and magnitudes of the synthetic and inverted profiles are well-reproduced. Finally, inverted ρ_H matches well the logged rho (**Figure 6E**), in agreement with the paradox of electrical resistivity anisotropy detailed in section 3.2.2..

Given the above model performances, we have shown that anisotropic inverse modeling is able to reconstruct ρ fields, particularly ρ -anisotropy, even for a very challenging aquifer with moderate heterogeneity and anisotropy.

5. FIELD CASE STUDY: COMPARISON BETWEEN ELECTRICAL AND HYDRAULIC ANISOTROPIES

Through the synthetic study presented in section 4.1, we demonstrated the ability of our methodology to characterize an electrically anisotropic environment using an adapted ERT survey (acquisition and inversion), without further external

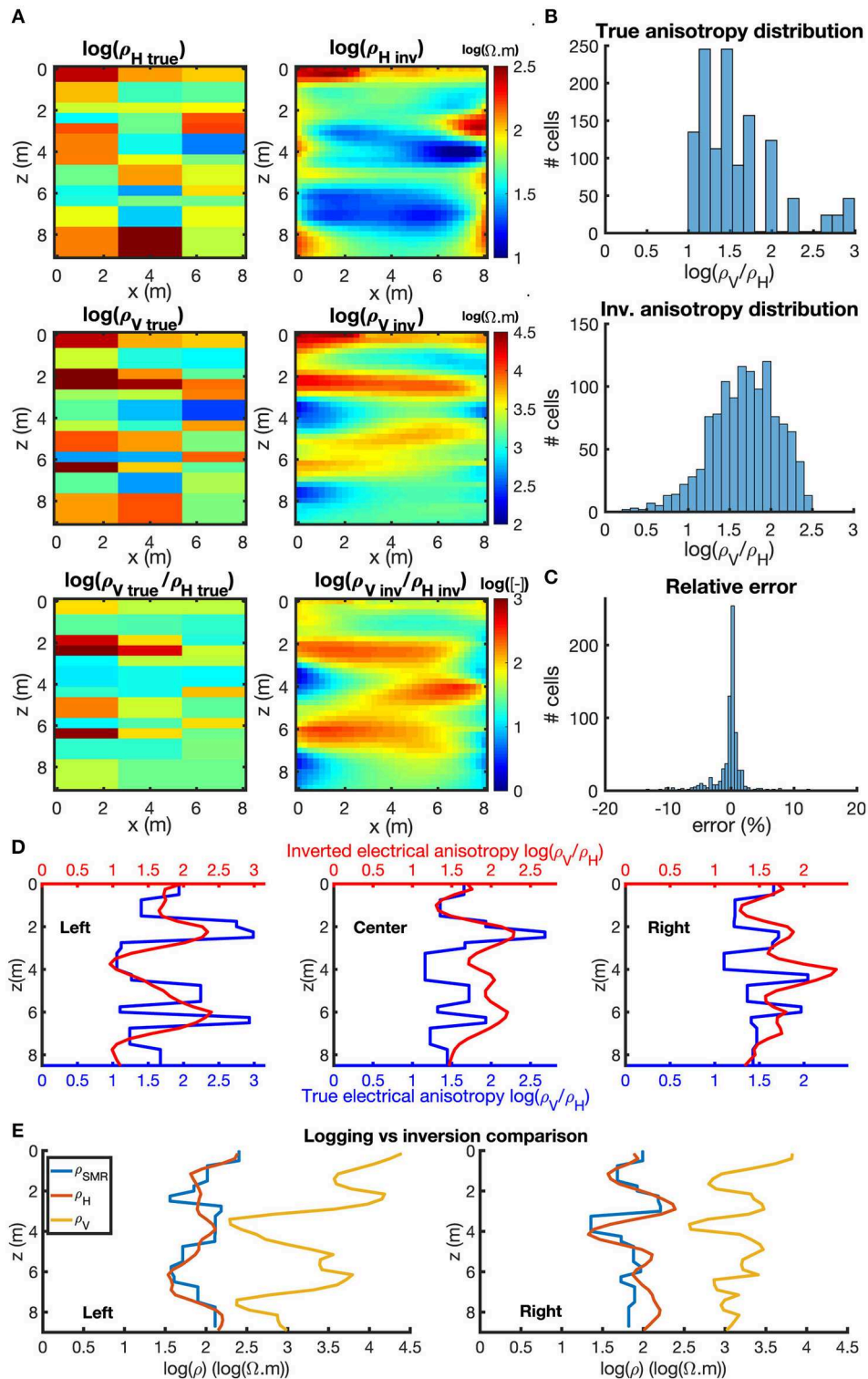


FIGURE 6 | Synthetic study anisotropic inversion results displaying logarithmic inverted resistivities. Unless otherwise stated, ρ_H and ρ_V represent the components of the inverted resistivity ρ_{Hinv} and ρ_{Vinv} . **(A)** True and inverted resistivity models (wells at $x = 0$ m and $x = 8$ m). **(B)** Histograms of the true and inverted anisotropy distributions. **(C)** Histogram of the relative error between measured and inverted ρ_{app} . **(D)** Comparison between true (blue curves) and inverted (red curves) anisotropies. To compensate the resolution difference, the inverted anisotropy is averaged on the 2.5 m around the wells and in the center of the modeled section. **(E)** Comparison between the logged (ρ_{SMR} , blue curves), and the inverted horizontal (ρ_H , red curves) and vertical (ρ_V , yellow curves) resistivities (corresponding to the left and right resistivities section). Notice the similarities between ρ_{SMR} and ρ_H .

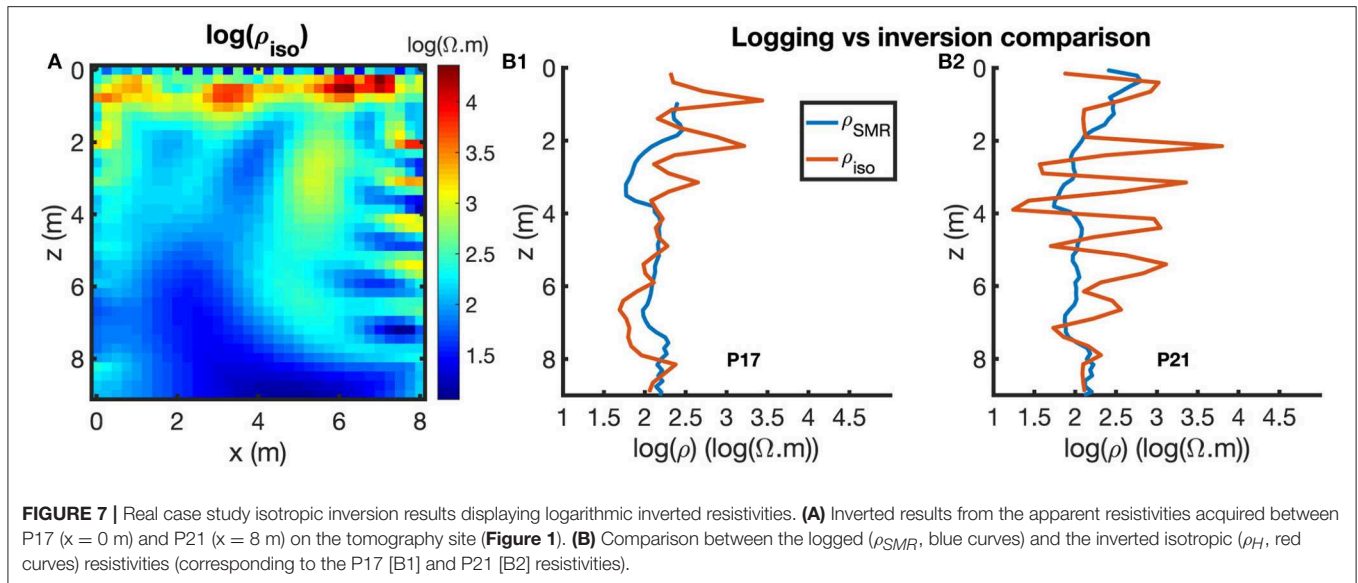


FIGURE 7 | Real case study isotropic inversion results displaying logarithmic inverted resistivities. **(A)** Inverted results from the apparent resistivities acquired between P17 ($x = 0$ m) and P21 ($x = 8$ m) on the tomography site (**Figure 1**). **(B)** Comparison between the logged (ρ_{SMR} , blue curves) and the inverted isotropic (ρ_{iso} , red curves) resistivities (corresponding to the P17 [B1] and P21 [B2] resistivities).

information. In this section, we want to verify with *in situ* measurements the possibility to characterize K -anisotropy from ERT anisotropic inversion.

The wells used for ERT were installed by direct-push technique in order to minimize skin effects around wells during testing (Paradis et al., 2011). Conventional well installation procedures indeed require the use of sand-pack to fulfill the space between the drilling hole and the screen, which may hinder the electrical response of the natural formation behind it. Direct-push well installation procedure allows the screen to be in direct contact with the aquifer with minimal disturbances to the surrounding sediments. The screen of the wells is open to the entire thickness of the aquifer allowing for multi-level hydrogeological and geophysical surveys. The screens ensure the free flow of water with slotted openings of 2.5 mm spaced vertically at every centimeter and covering over half of the circumference of the screens. The wells are also made of electrical insulator material (PVC) to ensure the integrity of electrical measurements.

The ERT setup is displayed in **Figure 1**. It consists of 17 inhole electrodes (9 in P17 and 8 in P21) and 17 surface electrodes located around the plane formed by P17 and P21 wells. P17 and P21 are separated by 8 m, electrodes separation is 1 m inside the wells and 0.5 m at the surface. Using this configuration, 18,936 electric potentials were measured with an IRIS Syscal Pro system. In addition, high resolution horizontal resistivity log data are available, acquired along P17 and P21 with the CPT-SMR probe. The SMR probe measures the resistivity using two ring electrodes 9 cm apart at a 1 kHz frequency to reduce polarization effects (Shinn et al., 1998; Paradis et al., 2015b). The log is 10.41 m deep at P17 and 9.96 m deep at P21, and ρ_H is measured over a 5 cm interval.

Before getting to the inversion, a quality control was done on the data using the reciprocal data. Interchanging the two electrodes inside a pair (current or potential) should only alter the

sign of the measured potential data. Alternatively, interchanging the two pairs (current with electrodes) should provide the same measured potential data by principle of reciprocity (Slater et al., 2000). During our survey, 5369 data has been acquired to that end. Amongst them, 89.6% of these data show a difference of $<15\%$, and that 84.5% show a difference of $<5\%$. These values show an overall good quality data set. From the whole data set, we can extract the data used for the inversion (section 5.2). Inverse modeling is computed on a numerical grid made of 8,970 squared cells of 25×25 cm, similarly to the synthetic case (section 4).

5.1. Anisotropy Diagnosis of Real Case Study ERT Data

In section 3.2 two different approaches were presented to assess the electrical anisotropy by analyzing ERT data. In the following, anisotropy diagnosis of real case study ERT data is studied by performing isotropic ERT inversion. After removing negative ρ_{app} data and poor quality data, 12,933 resistance data measurements were considered for the inversion. The isotropic ERT inversion converges after 10 iterations with an acceptable RMSE of 8.3%. Nevertheless, numerous erratic structures that do not correspond to the known geology of the site (section 2) appear on the isotropic resistivity image (ρ_{iso} , **Figure 7A**). More precisely, few small resistivity structures are close to the electrodes, where the inverted section is usually better resolved as shown in **Figure 2B.2**.

When ρ_{iso} is compared to the logged ρ in P17 and P21 wells (**Figure 7B**), ρ_{iso} shows very high frequency variations on both wells, and its values do not correspond to ρ values. As we have shown before (section 3 and **Figure 3C**), this is due to anisotropy. We then established the presence of anisotropy using a very fast approach in comparison to hydrogeological experiments. In fact, ERT data are carried out in less than a few hours, whereas several weeks are needed for anisotropic hydrogeological tomography data acquisition. Therefore, electrical anisotropy

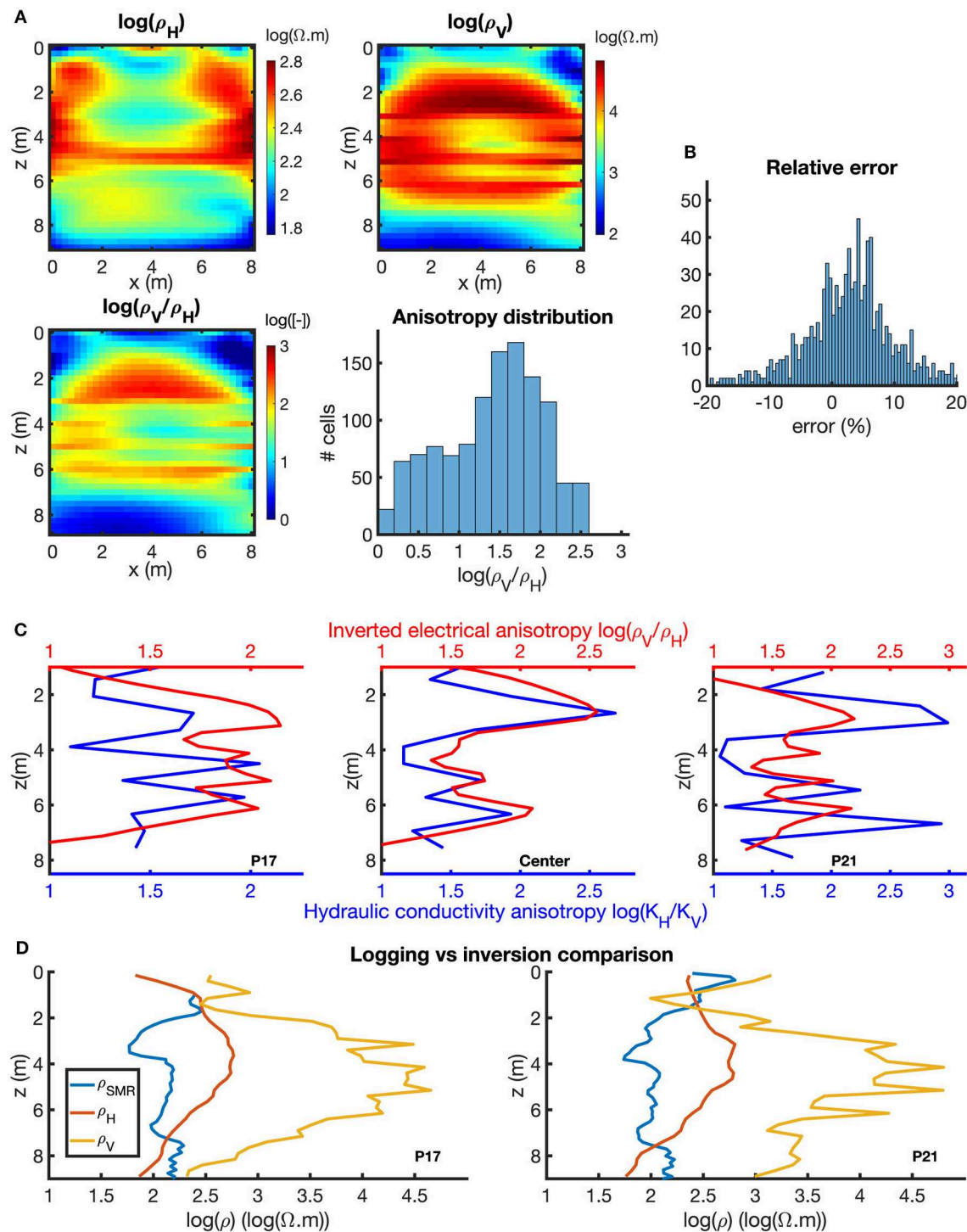


FIGURE 8 | Real case study inversion results displaying logarithmic inverted resistivities. **(A)** Inverted results from the apparent resistivities acquired between P17 ($x = 0$ m) and P21 ($x = 8$ m) on the tomography site (**Figure 1**). **(B)** Histogram of the relative error between measured and inverted ρ_{app} . **(C)** Comparison between inverted hydraulic (blue curves) and electrical (red curves) anisotropies. Hydraulic data starts at $z = 1$ m (saturated depth). To compensate the resolution difference, the inverted resistivities are averaged on the 2.5 m around the wells and in the center of the modeled section. **(D)** Comparison between the logged (ρ_{SMR} , blue curves), and the inverted horizontal (ρ_H , red curves) and vertical (ρ_V , yellow curves) resistivities (corresponding to the P17 and P21 resistivities).

diagnosis approaches can be used as an assisting tool to help taking hydraulic decisions.

5.2. Anisotropic Inversion of Anisotropic ERT Data

As for the synthetic case, XH1, SB2, and SB4 arrays are not considered. The used protocol is made of 975 measures (IB1[75], IB2[64], S[38], XH2[436], SB1[186], SB3[176]), based on the result obtained in the synthetic study. Homogeneous and anisotropic ρ values were used to initialize the inverted model with an initial ρ_H value corresponding to the median of the measured apparent resistivities and an initial anisotropy value of $\lambda = 10$. The convergence is reached after 6 iterations, after which misfit or RMSE slightly decrease, but do not improve significantly anymore. At this point, the inverted model starts integrating the noise held in the data, so further iterations are ignored. The relative error between the measured and the computed ρ_{app} from the inverted model is shown in **Figure 8B**. The histogram displays a slight bias of 4.3% in the relative error and a standard deviation of 10.83%. Unlike the synthetic case study, detailed information of the ground structure is unavailable, making hard the building of an optimized protocol. The chosen protocol was inspired from the synthetic case study since the latter was based on the hydraulic characterization of the ground. The residual error can be explained by the difficulty it met to reconcile its different sensitivities, whose preliminary examination is not achievable. Nevertheless, the final error (9.3%) combined with the relative error are considered low in a noisy real case study context.

The inverted sections are shown in **Figure 8A**. Both ρ_H and ρ_V sections show subhorizontal structures, as expected from our geological and hydrogeological knowledge of SLdL. The comparison between K (blue curve in **Figure 8C**) and ρ (red curve **Figure 8C**) anisotropies show strong similarities in their patterns and their amplitudes. It indicates the ability of anisotropic ERT inversion to characterize K -anisotropy. Finally, CPT-SMR resistivity is compared to anisotropic ERT inversion results (**Figure 8D**). Similarly to the synthetic case, the graphs display the collocated logged (ρ_{SMR} , blue curve), horizontal (ρ_H , red curve) and vertical (ρ_V , yellow curve) resistivities on wells P17 and P21. On the contrary to isotropic ERT inversion, horizontal resistivities at P17 and P21 are smooth which is consistent with CPT-SMR resistivities. However, there is a gap between the two curves. More precisely, horizontal resistivities are several times higher than CPT-SMR resistivities. This difference is due to the fully-screened wells effect on ERT data. CPT-SMR data measurement is carried out by direct-push before fully-screened well installation. Its coupling is very good, the electrodes being in direct contact with the undisturbed investigated underground. In the ERT case, acquisition is done using electrodes immersed into water in the screened well. Due to this aqueous environment, a part of the current is channelized along the well and affects the inverted model. The use of packers could prevent this channeling. Unfortunately, we were not able to implement this experiment during the campaign acquisition. Moreover, to our knowledge, the borehole effect

has been studied in the isotropic case when the electrodes are mounted on the electrically insulated borehole casing (Doetsch et al., 2010; Wagner et al., 2015; Lee et al., 2016). This effect is important only for large resistivity contrasts between the rock formation and borehole fluid and for large borehole diameters (Doetsch et al., 2010). Furthermore, sensitivity is very high close to the electrode. Impact of close objects or structures are important (Binley and Kemna, 2005). According to us, because the ERT method is very sensitive to the resistivity variations close to the electrodes, the screened borehole casing impacts the resistivity model. In our opinion, borehole effect in our case is substantially handled by anisotropic inversion because hydraulic tomography shows approximately the same anisotropy variations as electrical resistivity tomography. On the contrary, isotropic inversion shows a lot of artifacts around the boreholes.

If **Figures 8C,D** show the comparison between various linked parameters, a direct proportionality relation between ρ and K anisotropies is hard to achieve. The sensitivities of the geophysical and hydraulic methods are not the same. Moreover, both anisotropy sections are inverted sections, coming from two different inversions (in terms of grid size, regularization, etc.). To obtain a direct proportionality relation between ρ and K anisotropies, or even directly between ρ and K values, further investigations are needed to that goes beyond the framework of this work (use of packers to prevent current channeling, use 3-D finite elements code to more efficiently remove the well effects, etc.).

6. CONCLUSIONS

Hydraulic anisotropy has a major influence on the groundwater flow and mass transport. Its consideration is essential when it exists. Through this study, we pointed out: 1. the ability of isotropic ERT modeling to assess the presence of ρ -anisotropy, 2. the ability of anisotropic ERT modeling to quantify ρ -anisotropy and 3. the strong relationship existing between K - and ρ -anisotropies through an *in situ* survey. To achieve this work, we developed a new methodology based on an innovative anisotropic ERT modeling tool. To overcome the equivalence problem, electrodes were placed inside a fully screened borehole along with surface electrodes. Anisotropic ERT inversion is then carried out to estimate the ρ -anisotropic model. The latter suggest a strong link with the collocated K -anisotropic characterization: even though the setup used does not allow a direct proportionality relation, the proposed geophysical method is able to provide proxy of the *in-situ* hydraulic anisotropy.

In this study, we have shown that the anisotropic electrical resistivity surveys are helpful for anisotropic hydrogeologic parameters characterization, which paves the way for large scale hydrogeophysical characterization campaigns, even in challenging anisotropic environments. Integrated hydrogeophysical studies can therefore be powerful approaches in the understanding processes in order to produce more reliable forecasts.

AUTHOR CONTRIBUTIONS

EG funded the project. DP and EG conceived the idea of linking the *in situ* resistivity and hydraulic anisotropies. AB, DP, and EG supervised the project. AB and SG developed the physical and numerical theory, worked on data curation and analysis, and designed the computational modeling framework. AB encouraged further investigation on the data selected for the inversion, contributed to the interpretation of the results, and proposed further methodological experiments to improve data interpretation and final conclusion. SG conceived and carried out the synthetic and field experiments and wrote the manuscript

draft. SG, EG, AB, and DP provided review, editing, and final approving on the manuscript.

ACKNOWLEDGMENTS

The field data considered in this paper were acquired through the support of the Régie Intermunicipale de gestion des déchets des Chutes-de-la-Chaudière. The authors would like to acknowledge the participation of students for fieldwork. The authors would also like to thank the two reviewers that helped to improve the initial manuscript. This is LMS contribution number 20180391.

REFERENCES

- Adams, A. L., Nordquist, T. J., Germaine, J. T., and Flemings, P. B. (2016). Permeability anisotropy and resistivity anisotropy of mechanically compressed mudrocks. *Can. Geotech. J.* 53, 1474–1482. doi: 10.1139/cgj-2015-0596
- Archie, G. E. (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. *Petrol. Technol.* 146, 54–62.
- Barry, F., Ophori, D., Hoffman, J., and Canace, R. (2009). Groundwater flow and capture zone analysis of the Central Passaic River Basin, New Jersey. *Environ. Geol.* 56, 1593–1603. doi: 10.1007/s00254-008-1257-5
- Bing, Z., and Greenhalgh, S. (2000). Cross-hole resistivity tomography using different electrode configurations. *Geophys. Prospect.* 48, 887–912. doi: 10.1046/j.1365-2478.2000.00220.x
- Binley, A., and Kemna, A. (2005). “DC resistivity and induced polarization methods,” in *Hydrogeophysics*, eds Y. Rubin, and S. S. Hubbard (Dordrecht: Springer), 129–156.
- Bolduc, A.-M. (2003). *Géologie des Formations Superficielles*. Charny, QC: Commission géologique du Canada.
- Bouchedda, A. (2010). Inversion Conjointe des Données Électriques et de Radar en Forage. Ph.D. thesis, École Polytechnique de Montréal, Université de Montréal.
- Burns, W. A. Jr. (1969). New single-well test for determining vertical permeability. *J. Petrol. Technol.* 21, 743–752.
- Butler, J. J. (2005). “Hydrogeological methods for estimation of spatial variations in hydraulic conductivity,” in *Hydrogeophysics*, eds Y. Rubin, and S. S. Hubbard (Dordrecht: Springer), 23–58.
- de Marsily, G., Delay, F., Gonçalves, J., Renard, P., Teles, V., and Violette, S. (2005). Dealing with spatial heterogeneity. *Hydrogeol. J.* 13, 161–183. doi: 10.1007/s10040-004-0432-3
- Dey, A., and Morrison, H. (1979). Resistivity modelling for arbitrarily shaped two-dimensional structures. *Geophys. Prospect.* 27, 106–136.
- Doetsch, J. A., Coscia, I., Greenhalgh, S., Linde, N., Green, A., and Günther, T. (2010). The borehole-fluid effect in electrical resistivity imaging. *Geophysics* 75, F107–F114. doi: 10.1190/1.3467824
- Falta, R. W., Basu, N., and Rao, P. S. (2005). Assessing impacts of partial mass depletion in DNAPL source zones: II. Coupling source strength functions to plume evolution. *J. Contam. Hydrol.* 79, 45–66. doi: 10.1016/j.jconhyd.2005.05.012
- Gernez, S., Bouchedda, A., Gloaguen, E., and Paradis, D. (2018). “Comparison between hydraulic conductivity anisotropy and electrical resistivity anisotropy from tomography inverse modelling,” in *80th EAGE Conference and Exhibition 2018* (Copenhagen). doi: 10.3997/2214-4609.201800828
- Goltz, M. N., Huang, J., Close, M. E., Flintoft, M. J., and Pang, L. (2008). Use of tandem circulation wells to measure hydraulic conductivity without groundwater extraction. *J. Contam. Hydrol.* 100, 127–136. doi: 10.1016/j.jconhyd.2008.06.003
- Greenhalgh, S., Wiese, T., and Marescot, L. (2010). Comparison of dc sensitivity patterns for anisotropic and isotropic media. *J. Appl. Geophys.* 70, 103–112. doi: 10.1016/j.jappgeo.2009.10.003
- Greenhalgh, S. A., Zhou, B., Greenhalgh, M., Marescot, L., and Wiese, T. (2009). Explicit expressions for the fréchet derivatives in 3d anisotropic resistivity inversion. *Geophysics* 74, F31–F43. doi: 10.1190/1.3111114
- Hart, D. J., Bradbury, K. R., and Feinstein, D. T. (2006). The vertical hydraulic conductivity of an aquitard at two spatial scales. *Groundwater* 44, 201–211. doi: 10.1111/j.1745-6584.2005.00125.x
- Hirasaki, G. J. (1974). Pulse tests and other early transient pressure analyses for *in-situ* estimation of vertical permeability. *Soc. Petrol. Eng. J.* 14, 75–90.
- Hubbard, S. S., and Rubin, Y. (2005). “Introduction to hydrogeophysics,” in *Hydrogeophysics*, eds Y. Rubin, and S. S. Hubbard (Dordrecht: Springer), 3–21.
- Hvilshøj, S., Jensen, K., and Madsen, B. (2000). Single-well dipole flow tests: Parameter estimation and field testing. *Groundwater* 38, 53–62. doi: 10.1111/j.1745-6584.2000.tb00202.x
- Kabala, Z. J. (1993). The dipole flow test: a new single-borehole test for aquifer characterization. *Water Resour. Res.* 29, 99–107.
- Keller, G. V., and Frischknecht, F. C. (1966). *Electrical Methods in Geophysical Prospecting*. Pergamon Press.
- Kenkel, J., and Kemna, A. (2016). Sensitivity of 2-d complex resistivity measurements to subsurface anisotropy. *Geophys. J. Int.* 208, 1043–1057. doi: 10.1093/gji/ggw353
- Lee, K.-S., Cho, I.-K., and Kim, Y.-J. (2016). Borehole effect in 2.5D crosshole resistivity tomography. *J. Appl. Geophys.* 135, 212–222. doi: 10.1016/j.jappgeo.2016.10.015
- Lesmes, D. P., and Friedman, S. P. (2005). “Relationships between the electrical and hydrogeological properties of rocks and soils,” in *Hydrogeophysics*, eds Y. Rubin, and S. S. Hubbard (Dordrecht: Springer), 87–128.
- Loke, M. H. (2001). “Constrained time-lapse resistivity imaging inversion,” in *Symposium on the Application of Geophysics to Engineering and Environmental Problems 2001* (Denver, CO). doi: 10.4133/1.2922877
- Lüling, M. G. (2013). The paradox of anisotropy in electric logging: a simple proof and extensions to other physics domains. *Geophysics* 78, W1–W8. doi: 10.1190/geo2012-0123.1
- Maillet, R. (1947). The fundamental equations of electrical prospecting. *Geophysics* 12, 529–556.
- Onur, M., Hegeman, P. S., and Kuchuk, F. J. (2002). “Pressure-pressure convolution analysis of multiprobe and packer-probe wireline formation tester data,” in *SPE Annual Technical Conference and Exhibition* (San Antonio, TX: Society of Petroleum Engineers).
- Paradis, D., Gloaguen, E., Lefebvre, R., and Giroux, B. (2015a). Resolution analysis of tomographic slug test head data: two-dimensional radial case. *Water Resour. Res.* 51, 2356–2376. doi: 10.1002/2013WR014785
- Paradis, D., Gloaguen, E., Lefebvre, R., and Giroux, B. (2016a). A field proof-of-concept of tomographic slug tests in an anisotropic littoral aquifer. *J. Hydrol.* 536, 61–73. doi: 10.1016/j.jhydrol.2016.02.041
- Paradis, D., and Lefebvre, R. (2013). Single-well interference slug tests to assess the vertical hydraulic conductivity of unconsolidated aquifers. *J. Hydrol.* 478, 102–118. doi: 10.1016/j.jhydrol.2012.11.047
- Paradis, D., Lefebvre, R., Gloaguen, E., and Giroux, B. (2016b). Comparison of slug and pumping tests for hydraulic tomography experiments: a practical perspective. *Environ. Earth Sci.* 75:1159. doi: 10.1007/s12665-016-5935-4
- Paradis, D., Lefebvre, R., Gloaguen, E., and Rivera, A. (2015b). Predicting hydrofacies and hydraulic conductivity from direct-push data using a data-driven relevance vector machine approach: Motivations, algorithms, and application. *Water Resour. Res.* 51, 481–505. doi: 10.1002/2014WR015452

- Paradis, D., Lefebvre, R., Morin, R. H., and Gloaguen, E. (2011). Permeability profiles in granular aquifers using flowmeters in direct-push wells. *Groundwater* 49, 534–547. doi: 10.1111/j.1745-6584.2010.00761.x
- Paradis, D., Tremblay, L., Lefebvre, R., Gloaguen, E., Rivera, A., Parent, M., et al. (2014). Field characterization and data integration to define the hydraulic heterogeneity of a shallow granular aquifer at a sub-watershed scale. *Environ. Earth Sci.* 72, 1325–1348. doi: 10.1007/s12665-014-3318-2
- Pekşen, E., and Yas, T. (2018). Resistivity inversion of transversely isotropic media. *Turkish J. Earth Sci.* 27, 152–166. doi: 10.3906/yer-1702-6
- Ruggeri, P., Gloaguen, E., Lefebvre, R., Irving, J., and Holliger, K. (2014). Integration of hydrological and geophysical data beyond the local scale: application of bayesian sequential simulation to field data from the Saint-Lambert-de-Lauzon site, Québec, Canada. *J. Hydrol.* 514, 271–280. doi: 10.1016/j.jhydrol.2014.04.031
- Sheng, J. J. (2009). A new technique to determine horizontal and vertical permeabilities from the time-delayed response of a vertical interference test. *Transport Porous Media* 77, 507–527. doi: 10.1007/s11242-008-9274-0
- Shinn, J. D., Timian, D. A., Morey, R. M., Mitchell, G., Antle, C. L., and Hull, R. (1998). Development of a CPT deployed probe for *in situ* measurement of volumetric soil moisture content and electrical resistivity. *Field Anal. Chem. Technol.* 2, 103–109.
- Slater, L., Binley, A., Daily, W., and Johnson, R. (2000). Cross-hole electrical imaging of a controlled saline tracer injection. *J. Appl. Geophys.* 44, 85–102. doi: 10.1016/S0926-9851(00)00002-1
- Sutton, D., Kabala, Z., Schaad, D., and Ruud, N. (2000). The dipole-flow test with a tracer: a new single-borehole tracer test for aquifer characterization. *J. Contam. Hydrol.* 44, 71–101. doi: 10.1016/S0169-7722(00)00083-8
- Tremblay, L., Lefebvre, R., Paradis, D., and Gloaguen, E. (2014). Conceptual model of leachate migration in a granular aquifer derived from the integration of multi-source characterization data (St-Lambert, Canada). *Hydrogeol. J.* 22, 587–608. doi: 10.1007/s10040-013-1065-1
- Wagner, F. M., Bergmann, P., Rücker, C., Wiese, B., Labitzke, T., Schmidt-Hattenberger, C., et al. (2015). Impact and mitigation of borehole related effects in permanent crosshole resistivity imaging: an example from the ketzin CO₂ storage site. *J. Appl. Geophys.* 123, 102–111. doi: 10.1016/j.jappgeo.2015.10.005
- Wenzel, L. K., and Fishel, V. C. (1942). *Methods for Determining Permeability of Water-Bearing Materials, With Special Reference to Discharging-Well Methods, With a Section on Direct Laboratory Methods and Bibliography on Permeability and Laminar Flow*. US Geological Survey, United States Department of the Interior. doi: 10.3133/wsp887
- Wiese, T., Greenhalgh, S., and Marescot, L. (2009). DC resistivity sensitivity patterns for tilted transversely isotropic media. *Near Surf. Geophys.* 7, 125–139. doi: 10.3997/1873-0604.2009003
- Xiang, J., and Kabala, Z. (1997). Performance of the steady-state dipole flow test in layered aquifers. *Hydrol. Process.* 11, 1595–1605.
- Yeboah-Forsen, A., and Whitman, D. (2014). Electrical resistivity characterization of anisotropy in the biscayne aquifer. *Groundwater* 52, 728–736. doi: 10.1111/gwat.12107
- Zhou, B., Greenhalgh, M., and Greenhalgh, S. A. (2009). 2.5-D/3-D resistivity modelling in anisotropic media using Gaussian quadrature grids. *Geophys. J. Int.* 176, 63–80. doi: 10.1111/j.1365-246X.2008.03950.x
- Zlotnik, V., and Ledder, G. (1996). Theory of dipole flow in uniform anisotropic aquifers. *Water Resour. Res.* 32, 1119–1128.
- Zlotnik, V. A., and Zurbuchen, B. R. (1998). Dipole probe: design and field applications of a single-borehole device for measurements of vertical variations of hydraulic conductivity. *Groundwater* 36, 884–893.
- Zlotnik, V. A., Zurbuchen, B. R., and Ptak, T. (2001). The steady-state dipole-flow test for characterization of hydraulic conductivity statistics in a highly permeable aquifer: Horkheimer Insel site, Germany. *Groundwater* 39, 504–516. doi: 10.1111/j.1745-6584.2001.tb02339.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Gernez, Bouchedda, Gloaguen and Paradis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Parallelized Adaptive Importance Sampling for Solving Inverse Problems

Christoph Jäggli*, Julien Straubhaar and Philippe Renard

Stochastic Hydrogeology and Geostatistics Group, University of Neuchâtel, Neuchâtel, Switzerland

OPEN ACCESS

Edited by:

Yuri Fialko,
University of California, San Diego,
United States

Reviewed by:

Thomas Romary,
ParisTech École Nationale Supérieure
des Mines de Paris, Université de
Sciences Lettres de Paris, France
Shuai Zhang,
Lawrence Livermore National
Laboratory, United States Department
of Energy (DOE), United States
Sarah Minson,
Earthquake Science Center, United
States Geological Survey,
United States

*Correspondence:

Christoph Jäggli
christoph.jaeggli@gmail.com

Specialty section:

This article was submitted to
Solid Earth Geophysics,
a section of the journal
Frontiers in Earth Science

Received: 23 July 2018

Accepted: 26 October 2018

Published: 21 November 2018

Citation:

Jäggli C, Straubhaar J and Renard P
(2018) Parallelized Adaptive
Importance Sampling for Solving
Inverse Problems.
Front. Earth Sci. 6:203.
doi: 10.3389/feart.2018.00203

In the field of groundwater hydrology and more generally geophysics, solving inverse problems in a complex, geologically realistic, and discrete model space often requires the usage of Monte Carlo methods. In a previous paper we introduced PoPEX, a sampling strategy, able to handle such constraints efficiently. Unfortunately, the predictions suffered from a slight bias. In the present work, we propose a series of major modifications of PoPEX. The computational cost of the algorithm is reduced and the underlying uncertainty quantification is improved. Advanced machine learning techniques are combined with an adaptive importance sampling strategy to define a highly efficient and ergodic method that produces unbiased and rapidly convergent predictions. The proposed algorithm may be used for solving a broad range of inverse problems in many different fields. It only requires to obtain a forward problem solver, an inverse problem description and a conditional simulation tool that samples from the prior distribution. Furthermore, its parallel implementation scales perfectly. This means that the required computational time can be decreased almost arbitrarily, such that it is only limited by the available computing resources. The performance of the method is demonstrated using the inversion of a synthetic tracer test problem in an alluvial aquifer. The prior geological knowledge is modeled using multiple-point statistics. The problem consists of the identification of $2 \cdot 10^4$ parameters corresponding to 4 geological facies values. It is used to show empirically the convergence of the PoPEX method.

Keywords: adaptive importance sampling, machine learning, uncertainty quantification, bayesian inversion, monte carlo, multiple-point statistics, parallelization

1. INTRODUCTION

Inverse problems play a key role in almost all the geosciences. Indeed, this is often the only approach allowing to identify hidden structures of the interior of the earth and to estimate the physical properties of the buried rocks from indirect physical measurements at the surface or in a few boreholes. In groundwater hydrology, the aim is generally to infer the position of highly permeable or impermeable rocks and estimate their porosities and permeabilities from punctual measurements of state variables (e.g., hydraulic heads, tracer concentrations, water temperature, etc.). As for any geophysical problem, inverse methods are of utmost importance and a fundamental step in most quantitative hydrogeological studies (de Marsily et al., 2000; Carrera et al., 2005; Zhou et al., 2014) as well as many environmental modeling problems (Moles et al., 2003; Wainwright and Mulligan, 2005).

However, despite its huge significance and despite more than 50 years of research on this topic in geophysics and hydrology, current methods are still unable to solve certain types of problems efficiently. For instance, an open problem is to solve probabilistic inverse problems that involve discrete structures such as channels, lenses, karst conduits, or faults which cannot be represented by standard multi-Gaussian fields (Gómez-Hernández and Wen, 1998; Journel and Zhang, 2006). The identification and representation of such geological features is indispensable because it heavily controls fluid flow in the underground (Feyen and Caers, 2006). Using a wrong and smoothed representation of such discrete features is known to bias significantly the groundwater forecasts and corresponding uncertainty analysis (Gómez-Hernández and Wen, 1998; Kerrou et al., 2008).

To overcome this difficulty, different approaches have been developed and were recently reviewed by Linde et al. (2015). One general strategy is to construct first a probabilistic prior able to represent stochastic but geologically realistic structures and to embed it in the inverse method. Often, this geological prior can take only discrete values representing the rock types or some specific geological features.

Inverse methods relying heavily on continuity assumptions or simple statistical distributions (typically multi-Gaussian) are not capable to manage this type of problems. On the opposite, sampling algorithms can account for such complex setup (Oliver et al., 1997; Robert and Casella, 2004; Fu and Gómez-Hernández, 2008; Mariethoz et al., 2010a; Hansen et al., 2012; Laloy et al., 2016; Rubinstein and Kroese, 2016). These methods represent the solution of the inverse problem as a set of models (or samples) describing the posterior distribution. From this set of samples, one may approximate any quantity of interest such as mean values, maximum likelihood values, uncertainty bounds, or probabilities of characteristic events. Unfortunately, for most of these approaches, the computational effort is extremely demanding (Fu and Gómez-Hernández, 2008; Romary, 2010; Linde et al., 2015) and the challenge is to design an efficient sampling scheme able to deal with categorical information in the prior distribution.

In a previous paper (Jäggli et al., 2017), we proposed the Posterior Population Expansion (PoPEX) algorithm to expand iteratively an existing set of geological models. PoPEX was specifically designed for handling discrete parameter values, even if it can be applied to the continuous case as well. The discrete parameter fields can be generated with any geostatistical method.

In our previous paper and in this one, we use a multiple-point statistics technique for expressing the prior distribution because this allows the user of PoPEX to formulate its prior geological knowledge in the area where he is carrying out the inversion. This knowledge is expressed by providing a training image (TI). Multiple-point statistics (MPS) simulation techniques (Strebelle, 2002; Arpat and Caers, 2007; Honarkhah and Caers, 2010; Mariethoz et al., 2010b; Straubhaar et al., 2013) can learn the spatial patterns from the TI and can produce stochastic simulations that resemble the TI. The simulations can be conditioned by local values if they are known (hard data). The advantage of that approach is that it is flexible.

The same code can generate all kind of geological structures (channels, lobes, braided systems, fractures, etc.) and therefore it can be applied to a very wide range of inverse problems and applications.

Like most sampling techniques, PoPEX produces iteratively new parameter fields (the samples) using a geostatistical technique (see for example the book of Chilès and Delfiner, 2009), then runs the forward problem (in our case a groundwater flow and transport simulation, but it could be any forward operator), evaluates the misfit and likelihood for that solution, and accumulates novel knowledge. At each iteration, the geostatistical simulation algorithm is controlled by PoPEX: the general mechanism is to condition the simulation of the parameter fields with a set of punctual values (hard data) selected preferentially from previous models having a high likelihood.

This method proved to be very efficient on a synthetic example (Jäggli et al., 2017): a comparison with two existing Markov chain Monte Carlo (MCMC) methods showed that the method was able to considerably decrease the computational cost. But this study also allowed us to identify that the initial version of PoPEX produced slightly biased predictions.

In this paper, we revisit completely the core of the PoPEX algorithm. The overall goal is to improve the usability, accuracy and computational time. The most important contribution is to introduce a new strategy allowing to produce unbiased predictions. The bias happens because the generation of a new realization is influenced by all the previous models in the chain. This sampling strategy favors some realizations over others. When computing predictions, however, these correlations must be taken into account. In other words, we propose to consider the method as an *adaptive importance sampling* (AIS) (Naylor and Smith, 1988; Oh and Berger, 1992; Murphy, 2012) and suggest a simple technique to produce unbiased predictions. The additional computational cost is negligible and does not increase the overall running time. From this perspective, the method can be interpreted as an unsupervised machine learning scheme that aims to learn an optimal probability density which can be used in the AIS scheme. The class of inverse problems that can be addressed is very broad and goes beyond applications in the field of geostatistics. The only requirements are a forward problem solver, an inverse problem description (including the likelihood function), and a conditional simulation tool (e.g., any geostatistical method) that generates models according to the prior distribution.

On top of that, we show how the algorithm, together with all modifications, can be parallelized. We show that it scales perfectly in the considered example. Hence, the computational time is directly reduced by the number of parallel chains, without compromising the outcomes. This is a powerful result, because the main hindrance against the use of sampling strategies is the computational costs. With the proposed methodology, models can be produced in parallel. The only limitations concerns the number of available CPU's, or more precisely, the number of forward problem evaluations that can be run in parallel. Today, most research and engineering groups have access to high performance computer facilities, and therefore these requirements are not too restrictive.

The paper is organized as follows. Section 2 provides the required background related to the inverse problem and the general concepts of the method before explaining the details of the modified algorithm. A case study together with a convergence analysis is presented in section 3. Finally, in section 4, the advantages and limitations of the methodology are discussed and summarized.

2. METHODOLOGY

In this section, we first review the general definition of the inverse problem following the notations and approach from Tarantola (2005). Then we introduce the most important techniques constituting the base of PoPEX (Jäggli et al., 2017). As a consequence, the first part of this section mainly presents material that has been proposed and discussed elsewhere. It is toward the end of section 2.2 and in section 2.3 that we present the novel methods that constitute the core of this paper.

2.1. Inverse Problem

The general inverse theory presented by Mosegaard and Tarantola (2002) and Tarantola (2005) contains the commonly used Bayesian formulation as special case. Furthermore, it lives without the (problematic) notion of conditional probabilities (e.g., Borel's paradox) and alternatively uses the concept of *states of information*. In the following, we slightly enrich their explanations with a few comments specifically dedicated to the hydrogeological framework.

Solving an inverse problem is usually related to honoring a sparse set of observations $\mathbf{d}^{\text{obs}} = \{d_1^{\text{obs}}, \dots, d_m^{\text{obs}}\}$ called **data**. The nature of these observations can differ widely and may depend on the overall framework. When studying subsurface properties, they often represent measurements of state variables such as hydraulic heads, production data or contaminant concentration. Due to imperfect measuring devices, these quantities usually include uncertainties. It is common to use a finite set of parameters $\mathbf{m} = \{m_1, \dots, m_n\}$ to fully describe the physical system under study. Any possible collection of such values will henceforth be called a **model** or equivalently a **realization**. In this regard, a model can cover a vast number of physical and conceptual quantities, as, for instance, boundary conditions, hydraulic conductivity maps, or specific storage values. The collection of all possible models is called **model space** and is denoted by \mathcal{M} . In the hydrogeological framework, a common approach is to subdivide an aquifer into a finite number of volume elements (simulation grid) and characterize the hydraulic conductivity in each grid cell. In this case, the underlying model \mathbf{m} includes one parameter m_i per grid element, that defines the physical property in this small sub-domain. The choice of a set of representative dimensions is equivalent to the definition of a parametrization of \mathcal{M} . Note that for a given system, such a coordinate system is not unique. “Permeability,” for example, can be replaced by “resistivity,” “speed” with “slowness” or “frequency” with “period.”

In practice it is possible to observe parameters that can also be included in \mathbf{m} . Boreholes, for example, often provide cores, from which petrophysical values can be deduced with high precision.

If the model space is designed to describe the same quantities, we simply remove the corresponding degrees of freedom from any possible model \mathbf{m} , and reduce the number of dimensions in the model space \mathcal{M} .

In many fields, well-founded physical theories have been established in order to describe processes and interactions. They can be used to describe relations between the models and the observations. From a naïve point of view, it means that for a given model \mathbf{m} the error-free values of the corresponding data set \mathbf{d} can be predicted. This theoretical link between a model and the observable parameters is called the forward problem and described by $\mathbf{d} = \mathbf{g}(\mathbf{m})$. The function $\mathbf{g} = \{g_1, \dots, g_m\}$ denotes the **forward operator**. Tarantola (2005) formulated the probabilistic solution of an inverse problem as a non-negative measure function that combines two different states of information. Typically, these states of information are captured by the prior and the likelihood function. The prior distribution $\rho(\mathbf{m})$ describes any available information on the model parameters, that is independent of the data set. The likelihood function, $L(\mathbf{m})$, usually embeds the forward operator and is a probabilistic measure of how well a given model is able to explain the observations. The solution, called the **posterior distribution**, of an inverse problem is the conjunction of the prior and the likelihood operator such as

$$\sigma(\mathbf{m}) = c \rho(\mathbf{m})L(\mathbf{m}), \quad (1)$$

where c is a normalization constant. In the Bayesian framework, the posterior measure is considered to be the product of (conditional) probability distributions. The latter approach is contained in Equation (1) and applies under some regularity conditions. For this reasons, the formulation by Tarantola (2005) is more general.

2.2. Posterior Population Expansion (PoPEX)

It is worthwhile to recall several important concepts, that originally have been introduced by Jäggli et al. (2017). Afterwards, some small improvements will be suggested. These modifications just slightly influence the evolution of the sampling scheme, so that we decided not to rename the method and still call it Posterior Population Expansion (PoPEX). The general approach of the PoPEX algorithm is to generate a large number of models $\mathbf{m}_1, \dots, \mathbf{m}_N$ that represent the posterior probability density in Equation 1. From this approximation it is possible to compute posterior probabilities of events. The sampling procedure, however, requests to compute $\sigma(\mathbf{m}_k)$ for every $k = 1, \dots, N$, what can be highly intensive in terms of computational costs. For this reason, the main idea of the PoPEX method is to make the sampling as efficient as possible. Each generation of a new model \mathbf{m}_k is therefore guided by all the previous samples $\mathbf{m}_1, \dots, \mathbf{m}_{k-1}$. For doing so, information maps (denoted by P^k and $D(P^k||Q)$, see below) are computed iteratively and ensure that the sampling of \mathbf{m}_k is strongly guided by ‘good’ models with high posterior values. The transfer of information from $\mathbf{m}_1, \dots, \mathbf{m}_{k-1}$ to \mathbf{m}_k runs through a set of value restrictions imposed on the new model (denoted by HD^k , see below).

2.2.1. Set of Models \mathcal{M}^k

The underlying algorithm is able to examine many different types of uncertainties and parameter identification problems. It is possible, for example, to consider parameters concerning boundary and/or initial conditions, spatial heterogeneities, recharge time series, etc. The model set \mathbf{m} is then simply subdivided into different parts $\mathbf{m} = \{\mathbf{m}_1, \mathbf{m}_2, \dots\}$, where each $\mathbf{m}_i = \{m_{i1}, \dots, m_{ir}\}$ represents one specific parameter type. The only requirement is that samples representing that uncertainty can be generated from a *conditional simulation tool*.

In order to keep the following descriptions as simple as possible, we will only consider one type of model parameters and write $\mathbf{m} = \{\mathbf{m}_1\} = \{m_1, \dots, m_n\}$. This set will be used to describe spatial heterogeneities of hydraulic permeabilities and is generated by a pixel based MPS technique (Strebelle, 2002; Mariethoz et al., 2010b; Straubhaar et al., 2011). Such methods require a spatial subdivision of the computational domain into a finite number of $n \in \mathbb{N}$ elements (**pixels**). The union of all pixels is called the **simulation grid**. MPS generate realizations of a random variable by reproducing multiple-point statistics from a training image. Each realization can be associated to a model $\mathbf{m} = \{m_1, \dots, m_n\}$ by putting the MPS value from pixel j into the parameter m_j . In the example above, a variable m_j could then be linked to the constant permeability (or resistivity) in the j -th volume element of the computational domain. The term “linked” is used because it is not uncommon for the model parameters m_j to not contain permeability (or resistivity) values directly but only conceptual representatives of such. For the present work, it is assumed that the prior probability density ρ is precisely the distribution of the MPS random variable. Therefore, using the MPS machine to produce independent and unconditioned models is equivalent to drawing realizations from ρ . It is important to note that conditioning simulators work sequentially. This means that they start by randomly selecting a permutation ς over the set of indices $\{1, \dots, n\}$ that defines the order in which the components of a new model are treated. Whenever $m_{\varsigma(j)}$ is about to get informed, conditional simulation tools only consider previously simulated components and draw $m_{\varsigma(j)}$ according to the probability

$$\mathbb{P}(\cdot | m_{\varsigma(1)}, \dots, m_{\varsigma(j-1)}).$$

In other words, at this point of the simulation, $m_{\varsigma(j)}$ is considered to be independent of any uninformed component in \mathbf{m} .

Sampling a model space for solving an inverse problem, means to iteratively produce a finite number of N realizations

$$\mathbf{m}_1 \rightarrow \mathbf{m}_2 \rightarrow \dots \rightarrow \mathbf{m}_N,$$

that characterize (in some way) the posterior distribution. During this procedure, the likelihood function must be evaluated for every model in the chain. It is not uncommon that this computation is very demanding and represents the most important source of computational cost. After each iteration $k = 1, \dots, N$, the models can be assembled within the collection

$$\mathcal{M}^k = \{\mathbf{m}_1, \dots, \mathbf{m}_k\}, \quad (2)$$

while the normalized likelihood values

$$\tilde{L}(\mathbf{m}_j) = \frac{L(\mathbf{m}_j)}{\sum_{r=1}^k L(\mathbf{m}_r)}, \quad j = 1, \dots, k,$$

are joined in $\tilde{L}^k = \{\tilde{L}(\mathbf{m}_1), \dots, \tilde{L}(\mathbf{m}_k)\}$. The tilde notation indicates that a normalization has been applied, a convention that will be used throughout this paper. There are two different kinds of normalization that will be used. In the latter equation, the total weight was computed by summing all likelihood values from the previous iterations. This action must be renewed, whenever a new model \mathbf{m}_{k+1} is sampled. Secondly, we will define spatial maps. The normalization is then performed through all locational values, and the resulting map can be interpreted as a spatial probability density (c.f. Equation 5).

2.2.2. Probability Maps Q and P^k

The possible value range for each model parameter m_i depends on the TI. After defining a set of $s - 1$ threshold levels this range may be separated into s different categories, called **facies values** or simply **facies** and denoted by $\{f_1, \dots, f_s\}$. When working with discrete models, these categories usually define a one-to-one relation to the set of all possible values in the TI. From the facies values, it is possible to establish a collection of pixel-based indicator functions. If \mathbf{m} is a given model and each pixel $j \in \{1, \dots, n\}$ is represented by its center location \mathbf{x}_j , these functions are defined as

$$\mathbf{1}_{f_i}(\mathbf{m}; \mathbf{x}_j) = \begin{cases} 1 & \text{if } m_j \text{ belongs to category } f_i \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

Any linear combination of the quantities in Equation (3) can be interpreted as a map with constant value in each pixel. The concept of these indicator functions is very important throughout the present paper. If the precise pixel location \mathbf{x}_j is not relevant, we will henceforth omit its explicit notation. The indicator functions help to compute moments of the random vector that is associated to the MPS tool. Let q_i represent the pixel-wise probability of the model values to fall into category f_i . If $\mathbb{E}(\cdot)$ denotes the usual expectation operator, they read

$$q_i = \mathbb{E}(\mathbf{1}_{f_i}(\mathbf{m})), \quad i = 1, \dots, s.$$

The set $Q = \{q_1, \dots, q_s\}$ then collects all the *prior probability maps* for the facies categories. If the MPS machine is trained to produce stationary and unconditioned simulations, then the maps q_i are constant over the computational domain and equal the corresponding facies proportion in the training image. On the other hand, a set $\mathcal{M}^k = \{\mathbf{m}_1, \dots, \mathbf{m}_k\}$ can be used to define a second collection $P^k = \{p_1^k, \dots, p_s^k\}$ such that

$$p_i^k = \sum_{j=1}^k \mathbf{1}_{f_i}(\mathbf{m}_j) \tilde{L}(\mathbf{m}_j). \quad (4)$$

The superscript k in the notation p_i^k indicates the number of realizations that has been used in its computation. It is important to perceive the consequences of weighting the summands by the

normalized likelihood values $\tilde{L}(\mathbf{m}_j)$. If \mathbf{m}_{j_0} is a model with a large likelihood value (with respect to the other ones), this means that some facies patterns in \mathbf{m}_{j_0} may be very important. Therefore, the probability maps in Equation (4) are formed by weighting “good” facies patterns more heavily than “bad” ones. Consequently, these maps may be able to provide information that can be used to generate “good” models. But at this point it is unclear where this information can be found and how it could be used. The answer to this question lies in the relation between Q and P^k . The central idea of the PoPEX sampling is to consider and learn from all models $\mathbf{m}_1, \dots, \mathbf{m}_k$ before generating \mathbf{m}_{k+1} . This procedure can be split into two parts, that will be explained in the following.

2.2.3. Kullback-Leibler Divergence $D(P^k||Q)$

Kullback and Leibler (1951) introduced a measure called Kullback-Leibler divergence (KLD) to compare two probability distributions. It computes how a candidate probability diverges from an expected one. This is precisely what is needed to measure the information content of P^k with respect to Q . In other words, the Kullback-Leibler divergence can be used to identify pixel locations, where the facies probabilities in P^k are “extreme” with respect to Q . It is given by

$$D(P^k||Q) = \sum_{i=1}^s p_i^k \log \left(\frac{p_i^k}{q_i} \right). \quad (5)$$

Whenever $q_i > 0$ for all $i = 1, \dots, s$, this equation is well defined. But let's assume that there is $i \in \{1, \dots, s\}$ and a pixel \mathbf{x}_j with $q_i(\mathbf{x}_j) = 0$. This means that it is impossible for the MPS tool to produce a model \mathbf{m} where the value m_j falls into the i -th category. From Equation (4) it follows that $p_i^k(\mathbf{x}_j)$ must vanish as well. In short, $q_i(\mathbf{x}_j) = 0$ implies $p_i^k(\mathbf{x}_j) = 0$, and the corresponding terms in Equation (5) can be ignored. A brief comment on the prior maps q_i may help to enhance the meaning of Equation (5). If there is a large set of independent models $\{\mathbf{m}_1, \dots, \mathbf{m}_N\}$ that is distributed according to ρ , the law of large numbers (LLN) [c.f. Durrett (2010)] suggests to use approximations

$$q_i \approx \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{f_i}(\mathbf{m}_j). \quad (6)$$

From this perspective, the relation between p_i^k and q_i is easier to detect. Both definitions use the same indicator functions, but are weighted differently. $D(P^k||K)$ provides a pixel based information map, that indicates how surprising the facies patterns become, whenever they are weighted by the likelihood values. As mentioned earlier, it is possible to normalize the Kullback-Leibler divergence map spatially. The rescaled map is denoted by $\tilde{D}(P^k||Q)$ and can be interpreted as a discrete probability density defined over the pixel locations.

2.2.4. Hard Conditioning Data HD^k

We mentioned earlier, that each model must be generated by a “conditional simulation tool.” This means that it must be possible to condition (impose) some of the values in \mathbf{m} . Doing so allows fields that honor local data, commonly known as

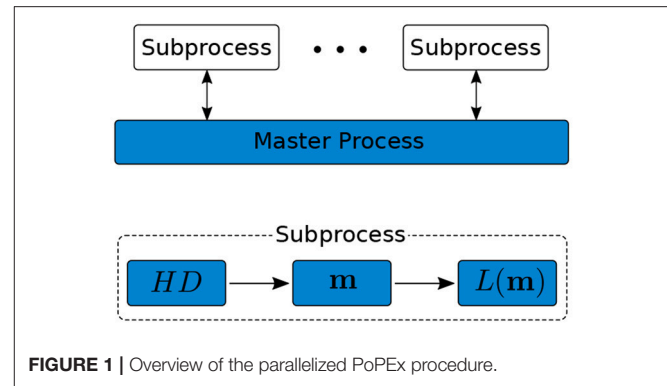


FIGURE 1 | Overview of the parallelized PoPEX procedure.

hard conditioning (HD) (Mariethoz and Caers, 2014), to be generated. The enforced value v together with the pixel location \mathbf{x} forms one conditioning object, denoted by (\mathbf{x}, v) . A reliable set of hard conditioning data may enhance the chance to generate a new model \mathbf{m}_{k+1} that provides a large likelihood value $L(\mathbf{m}_{k+1})$. Considering the previous explanations, it seems natural to sample a set $\{x_1, \dots, x_{n_k}\}$ of hard conditioning locations (where conditioning should apply) from the normalized Kullback-Leibler information $\tilde{D}(P^k||K)$. For every selected position \mathbf{x}_i , we can then sample a model index $j \in \{1, \dots, k\}$ according to \tilde{L}^k and extract the conditioning value (which value should be imposed) from $\mathbf{m}_j(\mathbf{x}_i)$. This produces a set of hard conditioning data $HD^k = \{(\mathbf{x}_1, v_1), \dots, (\mathbf{x}_{n_k}, v_{n_k})\}$.

So far, nothing original has been proposed. The modifications that we suggest now, concern the number of elements in HD^k . Jäggli et al. (2017) started with a set of unconditioned models, before fixing the number of conditioning points to a user defined parameter and leaving it unchanged. However, the statistical significance and robustness of the algorithm could certainly be increased by adding some “randomness” into this selection procedure. We suggest to change randomly the number of conditioning points in each iteration. For this, we suggest to fix an upper bound n_{\max} , and draw the number of conditioning points from a uniform distribution over the set $\{0, 1, \dots, n_{\max}\}$. The amount of hard conditioning data n_k thus may change in each iteration k . It is therefore possible to occasionally generate unconditioned realizations.

2.2.5. Parallelization of the Algorithm

Every loop of the PoPEX algorithm consists in four main steps: derive a set of hard conditioning points, generate a new model, compute its likelihood value and update the Kullback-Leibler divergence map. One strategy to parallelize this procedure is to encapsulate the first three steps in a subprocess separated from the last one. Then, a master process launches such subprocesses in parallel on other CPU's. Each subprocess is simply fed by the current available KLD map and performs the enclosed steps independently. After the result of a subprocess is communicated back to the master process, this latter updates the KLD map and launches another subprocess. A brief overview of this workflow is presented in Figure 1.

The pseudocodes of the parallelized PoPEx algorithm and the corresponding subprocesses are given in the algorithms 1 and 2, respectively. The variable “manager” appearing within the main algorithm is a FIFO (“first in first out”) queue of maximal length n_{par} that maintains the communication toward the subprocesses. FIFO stands for queues where new elements are appended at the tail (line 9) and removed from its head (line 11). In this regard, the lines 5–10 of algorithm 1 are designed to launch n_{par} parallel subprocesses (line 8) and retain corresponding handles (line 9). The lines 11–15 on the other hand, check the status of the first subprocess (line 12) and react accordingly. If it has terminated, their outputs are received (line 13) and the corresponding variables are updated (line 14). If it is still running however, the handle is sent to the back of the queue (line 17). The main motivation for appending the running subprocesses at the end is to rapidly detect and remove other jobs that have been completed. But as a consequence, reproducibility of the algorithm is not guaranteed. If reproducibility is crucial, we could simply change line 17 such that the processes are re-appended at the head of the queue and ensure that the first n_{par} workers are launched before lines 11–18 may apply.

Algorithm 1 PoPEx

```

1: Input:  $n_{\text{max}}, n_{\text{par}}, N$  and  $Q$ 
2:  $k \leftarrow 0$  and  $P^0 \leftarrow Q$ 
3: manager  $\leftarrow$  empty queue      # FIFO queue
4: while  $k < N$  do
5:    $n_m = \text{length}(\text{manager})$ 
6:   if  $n_m < n_{\text{par}}$  and  $k + n_m < N$  then
7:      $p \leftarrow$  new subprocess
8:      $p.\text{start}(\text{Subprocess}(\mathcal{M}^k, \tilde{L}^k, D(P^k||Q), n_{\text{max}}))$ 
9:     manager.append( $p$ )
10:  end if
11:   $p \leftarrow \text{manager.pop}()$ 
12:  if  $p.\text{ready}()$  then
13:     $(\mathbf{m}_{k+1}, L(\mathbf{m}_{k+1})) = p.\text{get}()$ 
14:    update  $\mathcal{M}^k, \tilde{L}^k$  and  $D(P^k||Q)$ 
15:     $k \leftarrow k + 1$ 
16:  else
17:    manager.append( $p$ )
18:  end if
19: end while

```

Algorithm 2 Subprocess

```

1: Input:  $\mathcal{M}^k, \tilde{L}^k, D(P^k||Q)$  and  $n_{\text{max}}$ 
2: Output:  $\mathbf{m}_{k+1}$  and  $L(\mathbf{m}_{k+1})$ 
3: sample  $n_k \sim U(0, n_{\text{max}})$ 
4:  $HD^k \leftarrow \text{hd}(n_k, \mathcal{M}^k, \tilde{L}^k, D(P^k||Q))$ 
5:  $\mathbf{m}_{k+1} \leftarrow \text{model}(HD^k)$ 
6:  $L(\mathbf{m}_{k+1}) \leftarrow \text{likelihood}(\mathbf{m}_{k+1})$ 

```

Calling “ $\text{hd}(n_k, \mathcal{M}^k, \tilde{L}^k, D(P^k||Q))$ ” within a subprocess (algorithm 2, line 4), uses the above strategy to compute a set of n_k hard conditioning couples. On the other hand, the

methods “ $\text{model}(HD^k)$ ” (line 5) and “ $\text{likelihood}(\mathbf{m})$ ” (line 6) are application dependent functions that generate a new model from a given set of conditioning data and compute the corresponding likelihood value.

In practice it might be unclear how to provide a suitable collection Q . Assuming that the involved modeling tool samples from the prior distribution, opens the door to approximate Q . Even before launching the PoPEx algorithm, we could produce a sufficiently large number of unconditioned models, and approximate Q by Equation (6). As the effort of generating a model is often negligible with respect to the computation of the likelihood value, the additional cost for approximating Q is unimportant. If there is a considerable effort required to generate a model, we could also consider to start with an initial guess of Q and iteratively improve it. However, changing Q along the sampling procedure may render the algorithm unstable.

2.3. Posterior Prediction of Events

Solving an inverse problem, not only serves to represent the posterior measure function, but also aims to compute the (posterior) probability of events $A \subset \mathcal{M}$. More generally, we would like to compute integrals with respect to σ , such as

$$\mu = \int_{\mathcal{M}} f(\mathbf{m}) d\sigma, \quad (7)$$

where $f(\cdot)$ is an operator that expresses some quantity of interest. Because the model space \mathcal{M} and the posterior measure function σ can be very complex, an analytical solution of these integrals is usually not available. The generic term importance sampling (IS) (Hesterberg, 2003; Robert and Casella, 2004; Liu, 2008; Rubinstein and Kroese, 2016) stands for a framework that provides approximations of such integrals by a weighted sum over a large number of realizations. Because it is often difficult or inefficient to directly sample from the distribution σ , importance sampling suggests instead, to draw realizations from a sampling distribution ϕ and weight the summands proportionally to the ratio $\sigma(\mathbf{m})/\phi(\mathbf{m})$. To find and use an appropriate sampling distribution ϕ however, can be challenging.

We propose to consider the PoPEx algorithm as a procedure, that iteratively learns and adapts the sampling distribution ϕ_k . During this procedure, all the previously generated realizations are combined and used to localize important regions in the model space. This is known as adaptive importance sampling (AIS) and has been introduced in a econometric framework (Naylor and Smith, 1988; Oh and Berger, 1992). The generation of a new model \mathbf{m}_{k+1} is understood as to randomly draw one sample according to ϕ_k . By construction, this distribution must include the random selection of HD^k as well as the conditional modeling tool. For each model in a chain of N realizations, we compute a weight ratio $w_k = \sigma(\mathbf{m}_k)/\phi_k(\mathbf{m}_k)$ and estimate the integral μ by

$$\hat{\mu} = \sum_{k=1}^N f(\mathbf{m}_k) \tilde{w}_k. \quad (8)$$

Again, the tilde notation was used to indicate normalized weights \tilde{w}_k such that

$$\tilde{w}_k = \frac{w_k}{\sum_j w_j}.$$

Several remarks are worth being considered. The computation of Equation (8) only uses normalized weights. Therefore, it is not required to know precisely the normalization constant of either σ or ϕ . Furthermore, the computation of the weights w_k can be simplified by using the factorization of σ (c.f. Equation 1). Each likelihood value $L(\mathbf{m}_k)$ is evaluated during the PoPEX procedure, so that for constructing the weights, it is sufficient to compute the ratio

$$\frac{\rho(\mathbf{m}_k)}{\phi_k(\mathbf{m}_k)}, \quad \text{for } k = 1, \dots, N. \quad (9)$$

Roughly speaking, this ratio compares the probability measure of generating a model \mathbf{m}_k with and without observed data.

Computation of the Sampling Weights

In every iteration of the PoPEX algorithm, a set of location-value pairs is derived and imposed as hard conditioning for the next model. We will show in this section, that when using a pixel-based MPS technique to generate the models, the sampling ratios in Equation (9) only depend on the sets HD^k . Let us consider $HD^k = \{(\mathbf{x}_1, v_1), \dots, (\mathbf{x}_{n_k}, v_{n_k})\}$ and distinguish two events that henceforth will be noted similarly:

1. The MPS scheme is assumed to follow the prior probability measure ρ . Recall that such a tool iteratively supplies pixels with simulation values from a training image. HD^k appearing within ρ will refer to the event where “the first n_k locations (met during the simulation) were $\{\mathbf{x}_1, \dots, \mathbf{x}_{n_k}\}$ and they obtained the values $\{v_1, \dots, v_{n_k}\}$.” Following this line, $\rho(\mathbf{m}|HD^k)$ expresses the conditional measure to draw \mathbf{m} , when the first n_k assignments imposed the values $\{v_1, \dots, v_{n_k}\}$ at the locations $\{\mathbf{x}_1, \dots, \mathbf{x}_{n_k}\}$.
2. The sampling distribution on the other hand takes the PoPEX iterations into account. HD^k appears within ϕ_k whenever we want to indicate that “during the k -th PoPEX iteration, HD^k has been produced and used as hard conditioning.” Accordingly, $\phi_k(\mathbf{m}|HD^k)$ measures the probability of sampling \mathbf{m} at iteration k , knowing that HD^k has been imposed.

Henceforth, we will only consider combinations \mathbf{m} and HD^k that produce strictly positive measure values (i.e., where the values on the n_k locations coincide). Furthermore, we will assume that all the conditioning binomials in HD^k are independent of each other. This assumption is reasonable if the conditioning locations are well separated. It is therefore necessary, that the number of conditioning points is adequate with respect to the simulation grid. The MPS processes involved behind $\rho(\mathbf{m}|HD^k)$ and $\phi_k(\mathbf{m}|HD^k)$ are the same. It follows that the two measure values must be equal, and thus

$$\frac{\rho(\mathbf{m})}{\phi_k(\mathbf{m})} = \frac{\rho(\mathbf{m})}{\rho(\mathbf{m}|HD^k)} \frac{\phi_k(\mathbf{m}|HD^k)}{\phi_k(\mathbf{m})}.$$

Using the definition of conditional probabilities, the ratio can be rearranged as

$$\frac{\rho(\mathbf{m})}{\phi_k(\mathbf{m})} = \frac{\rho(HD^k)}{\rho(HD^k|\mathbf{m})} \frac{\phi_k(HD^k|\mathbf{m})}{\phi_k(HD^k)}.$$

Standard techniques from the field of combinatorial probability allow to express all the above quantities. On the one hand, $\rho(HD^k|\mathbf{m})$ measures the probability of informing the first n_k pixels according to HD^k , when the sampled model is known. But knowing \mathbf{m} implies that the conditioning values in HD^k are given, so that we only need to compute the probability to meet the n_k conditioning locations (in any order) in the very beginning of the MPS simulation. If there are n pixels in the simulation grid, $\rho(HD^k|\mathbf{m})$ is given by

$$\rho(HD^k|\mathbf{m}) = \frac{n_k!(n - n_k)!}{n!}.$$

On the other hand, because the hard conditioning data is assumed to be independent, $\rho(HD^k)$ reads

$$\rho(HD^k) = \frac{n_k!(n - n_k)!}{n!} \prod_{j=1}^{n_k} \rho(v_j; \mathbf{x}_j),$$

where $\rho(v_j; \mathbf{x}_j)$ is the prior probability of meeting the value v_j at location \mathbf{x}_j . In section 2.2, the simulation values have been categorized into the set $\{f_1, \dots, f_s\}$. For a fixed HD^k , let us define an index-to-index map $r = r(j)$ such that $f_{r(j)}$ identifies the category of v_j . An approximation to $\rho(v_j; \mathbf{x}_j)$ can be obtained from Equation (6) by specifying $\rho(v_j; \mathbf{x}_j) \approx q_{r(j)}(\mathbf{x}_j)$. This simply suggests to find the map $q_{r(j)}$ that corresponds to the category of v_j and extract the probability value at \mathbf{x}_j .

Every iteration contains the following three steps. Select a number n_k , sample conditioning locations from $\tilde{D}(P^k|Q)$ and extract conditioning values by weighting the simulations according to the computed likelihood measures in \tilde{L}^k . They are performed independently such that the probability of selecting HD^k , knowing \mathbf{m} , is measured as

$$\phi_k(HD^k|\mathbf{m}) = \phi_k(n_k) \prod_{j=1}^{n_k} \tilde{D}(P^k|Q)(\mathbf{x}_j)$$

while similarly (with the hard conditioning data points being independent)

$$\phi_k(HD^k) = \phi_k(n_k) \prod_{j=1}^{n_k} \phi_k(v_j; \mathbf{x}_j) \tilde{D}(P^k|Q)(\mathbf{x}_j).$$

The value $\phi_k(n_k)$ is the probability of selecting n_k while the measure $\phi_k(v_j; \mathbf{x}_j)$ is the probability to draw a model (according to \tilde{L}^k) that presents the value v_j at location \mathbf{x}_j . This quantity can again be approximated by using the index-to-index relation $r(j)$ together with the categorical probabilities in P^k (c.f. Equation 4), such that $\phi_k(v_j; \mathbf{x}_j) \approx p_{r(j)}^k(\mathbf{x}_j)$. It is worthwhile to note that when

working with discrete models, where the categories $\{f_1, \dots, f_s\}$ have a one-to-one relation to the range of all simulation values in the training image, these approximations are exact. Finally, a computable ratio is provided by

$$\frac{\rho(\mathbf{m})}{\phi_k(\mathbf{m})} = \prod_{j=1}^{n_k} \frac{q_{r(j)}(\mathbf{x}_j)}{p_{r(j)}^k(\mathbf{x}_j)}. \quad (10)$$

This expression is very practical. All the quantities in Equation (10) are assembled during the PoPEx algorithm, so that the required effort for evaluating the ratio is negligible. Moreover, the expression is easily translated into log-probabilities, what can simplify the floating-point representation of the values. Although it only represents an approximation of the true ratio, often the assumptions are not too strongly violated, and the usage of the above equation is feasible. Finally, the weights w_k are computed by correcting the likelihood measure according to the hard conditioning data:

$$w_k = L(\mathbf{m}_k) \frac{\rho(\mathbf{m}_k)}{\phi_k(\mathbf{m}_k)} = L(\mathbf{m}_k) \prod_{j=1}^{n_k} \frac{q_{r(j)}(\mathbf{x}_j)}{p_{r(j)}^k(\mathbf{x}_j)}. \quad (11)$$

The ratios in the correction term compare the prior vs. the likelihood weighted probabilities of observing the selected values at the locations of the conditioning data. These quantities are directly available in the Q and P^k maps.

2.3.1. Degeneracy of the Sampling Weights

The estimator $\hat{\mu}$ in Equation (8) suffers from a degeneracy in the sense that the distribution of $W^N = \{w_1, \dots, w_N\}$ may become increasingly skewed when the dimension of \mathcal{M} grows large (Doucet et al., 2001; Robert and Casella, 2004; Liu, 2008). This means, that the weights may take small values with high probability, but occasionally become very large. Using such weights in Equation (8) would produce estimators that are dominated by very few samples. Several preventive techniques exist, and they often try to consider a reduced dimensionality in the computation of the weights (Doucet et al., 2001; Rubinstein and Kroese, 2016). The expression in Equation (10) uses a reduction technique by limiting the computation of the ratio to the hard conditioning data. But this expression only represents one part of the weights in Equation (11), so that the degeneracy problem still exists. A diagnostic that can be used to assess the skewness of the weights, is called the **effective sample size** (Owen, 2013) and defined as

$$n_e(W^N) = \frac{\left(\sum_{i=1}^N w_i\right)^2}{\sum_{i=1}^N w_i^2} = \frac{N\bar{w}^2}{\overline{w^2}}, \quad (12)$$

where $\bar{w} = (1/N) \sum_{i=1}^N w_i$ and $\overline{w^2} = (1/N) \sum_{i=1}^N w_i^2$. There is an obvious link between n_e and the variance of W^N . It suffices to notice, that an estimator of the variance is obtained by $\overline{w^2} - \bar{w}^2$. Strongly varying weights would give $\overline{w^2}/\bar{w}^2 \ll 1$ and therefore, $n_e \ll N$. In general, lowering the variance increases the effective number n_e . In practice, it is often hard to specify a bound under

which n_e is alarmingly small, because this strongly depends on the application.

We will now present a method that aims to soften the degeneracy by modifying the variance of the weights. The value of any positive weight $w_i > 0$ can be changed by exponentiation, $(w_i)^\alpha$, and we know that

$$\lim_{\alpha \rightarrow 0} (w_i)^\alpha = 1.$$

For a given $0 < \alpha < 1$, the variance can therefore be reduced by transforming the set W^N into

$$W_\alpha^N = \{(w_1)^\alpha, \dots, (w_N)^\alpha\}.$$

It is clear that in the limit $\alpha \rightarrow 0$, $n_e(W_\alpha^N)$ is equal to the total number of positive weights in W^N . Before computing an estimator $\hat{\mu}$ from W^N , we select an appropriate α , and use the weights in W_α^N instead. To make a good choice for α might depend on the application and can be challenging. We propose to define a lower bound l_0 and choose α such that

$$n_e(W_\alpha^N) = \max\{l_0, n_e(W^N)\}. \quad (13)$$

The idea of Equation (13) is to ensure that the computation of $\hat{\mu}$ is based on at least l_0 significant models. Furthermore, it assures that the growth rate of $n_e(W^N)$ and $n_e(W_\alpha^N)$ are equal for $n_e(W^N) > l_0$. This might be important for the asymptotic behavior of the method. Finally, we propose to use the pseudo code in the algorithm 3 to compute predictions. The computation

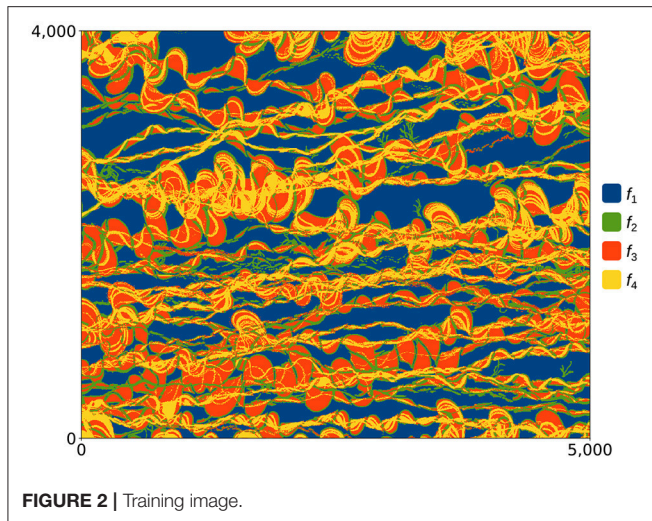
Algorithm 3 Prediction

- 1: **Input:** l_0 and $f(\cdot)$
 - 2: **Output:** $\hat{\mu}$
 - 3: compute α such that $n_e(W_\alpha^N) = \max\{l_0, n_e(W^N)\}$
 - 4: **for** $(w_i)^\alpha > 0$ **do**
 - 5: compute $f(\mathbf{m}_i)$
 - 6: **end for**
 - 7: compute $\hat{\mu}$
-

of α can be translated into a smooth, 1-dimensional optimization problem, and does not require a considerable effort. The most important effort usually goes into the evaluation of $f(\mathbf{m}_i)$. But all the weights are known in advance and therefore we can omit computations that are associated with zero weights. Furthermore, the iterations in the algorithm 3 are independent and can be performed simultaneously in parallel.

3. CASE STUDY AND RESULTS

In this section, we illustrate how PoPEx performs to solve an inverse problem with an example of a tracer test in a fluvial aquifer. We also consider the problem of quantifying the uncertainty related to the prediction of the capture zone of a pumping well in such geological environments.



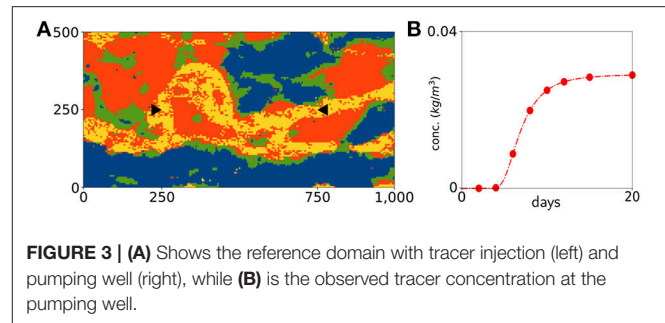
3.1. Problem Setup

For this example, the conceptual model for the geological heterogeneity is derived from a 3D simulation of the geological processes occurring in a fluvial plain using the FLUMY software (Lopez et al., 2009). This tool combines a process-based approach with a stochastic component. A meandering river crosses an alluvial valley with a given slope and causes erosion and deposition of sediments. Over time, the river migrates and alter the topography of the alluvial plain. This process generates complex geological patterns with a realistic and highly heterogeneous architecture. However, the thickness of the alluvial sediments is usually negligible with respect to the horizontal dimensions of the plain. It is therefore reasonable to reduce the complexity of the problem and neglect the vertical component of the flow. The parameter of interest is then the transmissivity of the aquifer.

Following this approach a 2-dimensional training image was generated. It is represented in **Figure 2**. The training image represents a domain of $5,000 \times 4,000$ m and is subdivided into $1,000 \times 800$ quadratic pixels. It was obtained by vertical integration of a 3D model generated with FLUMY. The resulting field was categorized into four facies types f_1, f_2, f_3 , and f_4 that represent transmissivity values of $10^{-5}, 10^{-3}, 10^{-2}$, and $10^{-1} (m^2/s)$, respectively. The drainage porosity and the specific storage were fixed uniformly to 0.2 and 10^{-6} .

We then consider a smaller area of size $1,000 \times 500$ m, discretized into 200×100 quadratic pixels. The area hosts a pumping well at the location (750,250) that extracts $15(l/s)$ of groundwater for a total duration of 20 days. The terrain is exposed to a natural slope of 4‰ in the x-direction, while the basin is closed at $y = 0$ (m) and $y = 500$ (m). Corresponding boundary conditions are: fixed head values of 4 (m) (left) and 0 (m) (right) together with no-flow on the upper and lower boundary. A constant tracer concentration of $1(kg/m^3)$ is enforced at (250,250) throughout the time period.

For any given model, the subsurface water flow together with the tracer expansion is computed by the GroundWater



simulation software (Arpat and Caers, 2007). At days 2, 4, 6, 8, 10, 12, 15, and 20 the solute concentration is recorded at the pumping well. This provides a set of 8 observations and represents all the data constraints used for conditioning the inverse problem.

For modeling the spatial structures, we used the training image shown in **Figure 2** and the DeeSse multiple-point statistics software (Straubhaar, 2011). An arbitrary seed was used to generate the reference domain in **Figure 3A**. The black triangles indicate tracer injection (left, pointing right) and pumping well location (right, pointing left), respectively. The tracer concentration at the pumping well resulting from this reference domain is shown in **Figure 3B**. The red dots indicate the extracted data that was used in the inverse procedure. This means that the entire reference domain in **Figure 3A** is unknown to the PoPEX algorithm. Its only task is to represent an unknown subsurface model and provide a sparse set of data points that can be used in the sampling procedure. For constructing the likelihood measure $L(\mathbf{m})$, we assume the observations to be independent and consider a multivariate normal distribution between the predictions $\mathbf{g}(\mathbf{m}) = \{g_1(\mathbf{m}), \dots, g_8(\mathbf{m})\}$ and observations $\mathbf{d}^{\text{obs}} = \{d_1^{\text{obs}}, \dots, d_8^{\text{obs}}\}$ with uniform standard deviation of $\sigma_L = 0.0015(kg/m^3)$. This represents 1.5‰ of the concentration at the injection point, and roughly 5% of the maximal concentration at the extraction location in the reference domain (c.f. **Figure 3B**). The subscript L distinguishes the standard deviation of the likelihood measure σ_L from the posterior density σ in Equation (1). Assuming an uniform and independent Gaussian behavior of $\mathbf{g}(\mathbf{m})$ around \mathbf{d}^{obs} , the density function of the likelihood measure is proportional to $\exp \left\{ -\frac{1}{2\sigma_L^2} \sum_i (g_i(\mathbf{m}) - d_i^{\text{obs}})^2 \right\}$.

3.2. Tracer Breakthrough Curve

The PoPEX method has been trained to run the above problem for a total of $N = 20,000$ models with $n_{\text{max}} = 25$. Three random realizations are shown in **Figure 4**.

The prior facies probabilities in Q were computed from 500 unconditioned MPS models. For each realization in the PoPEX chain, the algorithm computed the tracer concentration at the pumping well, extracted 8 data points and compared them to the reference data in **Figure 3B**. Together with the weights from Equation (11), the posterior distribution of the tracer breakthrough curve can be computed. **Figure 5** shows the 2.5 –

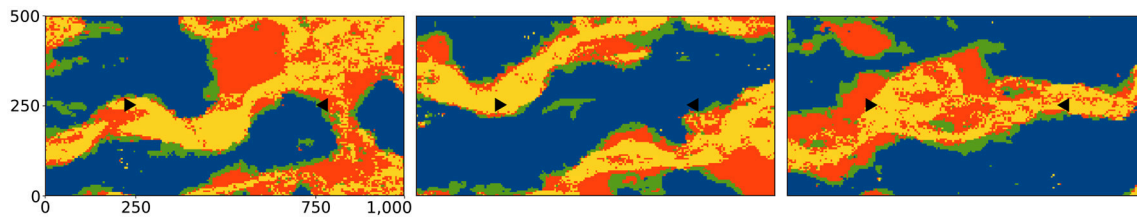


FIGURE 4 | Three random realizations from the PoPEX sampling. The above problem assigns a large likelihood value to the left, a very low likelihood value to the middle, and an average likelihood value to the right parameter map.

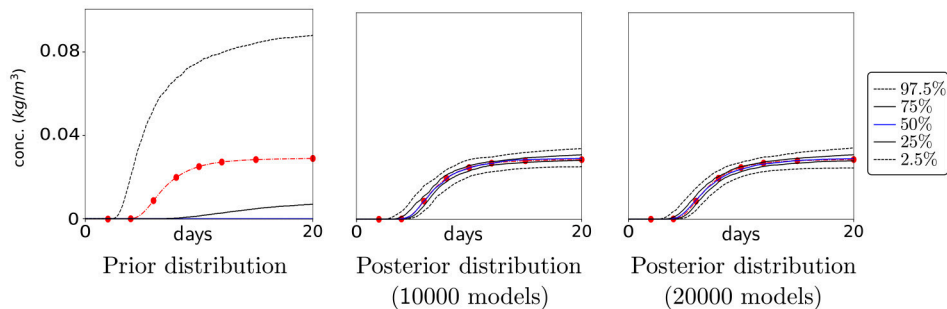


FIGURE 5 | Prior and posterior concentration probabilities at the pumping well. The curves indicate 2.5 – 97.5% region (dashed), 25 – 75% region (full) and average value (blue), while the red dots represent the reference concentration.

97.5% (dashed), 25 – 75% (full) and average (blue) curves of the prior and the posterior tracer concentration at the pumping well.

The red dots indicate the extracted reference data. It is clear that for any sampling strategy a critical measure is the required computational effort, which usually is proportional to the number of samples. For this reason, all results are shown for two different stages in the sampling procedure: after 10,000 and after 20,000 realizations. At a first glance, both estimations of the posterior probabilities are quite similar. This may be surprising when keeping in mind that the computational effort for the second estimation is twice as high. However, it can be seen that the probability lines are steadier and smoother in the last image. Both estimations of the 50% region (between the full lines) fully embeds the red reference data and follows the shape of the reference curve very precisely. The estimation of the posterior expectation (blue) almost matches the entire curve. The higher density of data points in the first 10 days, increases the relative importance of this period with respect to the second half. Thus, it is reasonable to allow less uncertainty in the beginning of the simulation. The more generous 95% regions (between the dashed curves) are still appropriate in reproducing the shape of the reference curve. This is even more significant when realizing that the prior distribution is far from being centered around the reference curve.

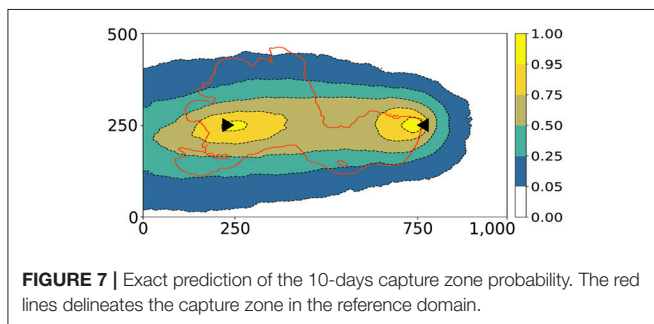
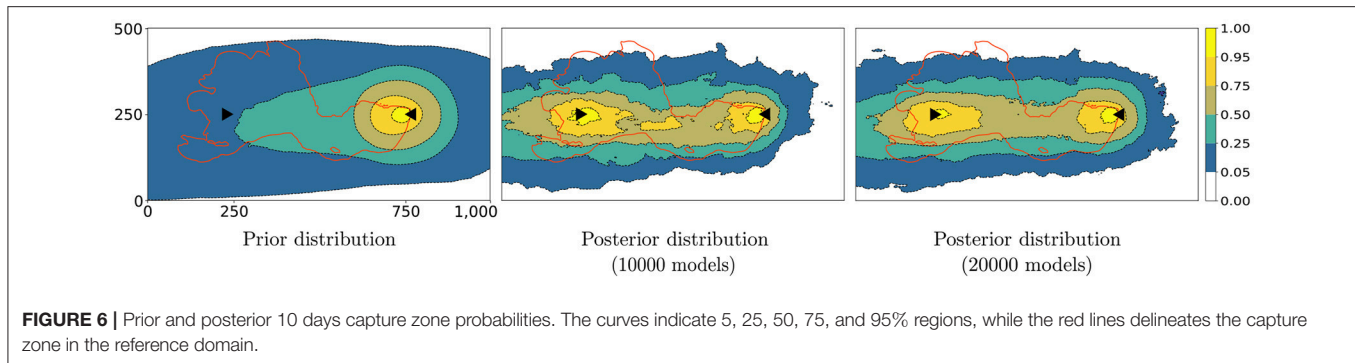
3.3. Predict 10-Days Capture Zone

In practice, when producing freshwater from an aquifer, it is often crucial to protect the resource and determine the capture

zone (Leeuwen et al., 1998). Here, we used the results of the PoPEX model chain for predicting the posterior probabilities of the 10-days capture zone. It means that for each location in the simulation grid, we computed a Bernoulli probability value for the water to be captured within 10 days. **Figure 6** shows the predicted probabilities for the prior distribution and for the posterior distribution after 10,000 and 20,000 iterations, respectively.

As expected, since the tracer is arriving in <10 days at the pumping well, the injection point is located within a region having a high probability to belong to the 10-days capture zone. This is already clearly visible in the map generated from 10,000 realizations. These results show the existence of a connected path of high transmissivity between the injection point and the pumping well. However, zones of lower probability are located in between these two points. This indicates that the position of the channel is not well identified from these tracer data alone. In the reference domain, shown in **Figure 3**, we can see that the yellow facies (with the largest transmissivity value) first shows a very tight upwards bend before heading almost directly toward the extraction well. The injected tracer will mostly follow the region with the largest transmissivity. Therefore, it will not take a direct path toward the well and its arrival time will be delayed. The only information that can be extracted from the observations is the delay. From the available data, it is therefore impossible to predict precisely water pathways that are far from the tracer injection and the algorithm is correctly informing us about that uncertainty.

It is interesting that the reference capture zone (red line) slightly passes outside the 95% region in the top section of the

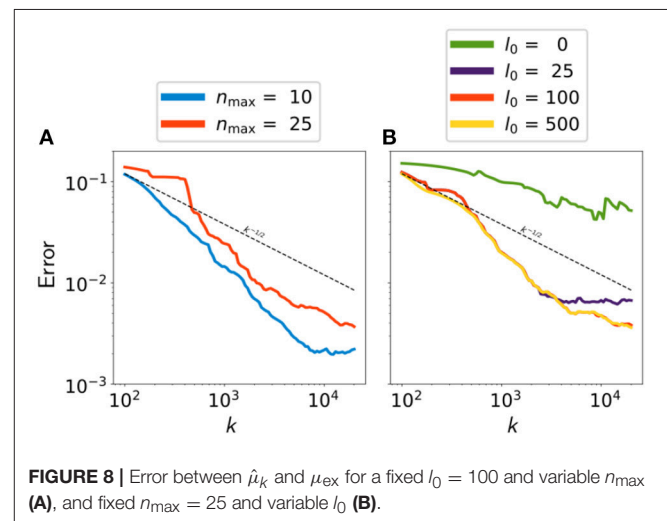


computational domain. This should not be interpreted as an inaccuracy of the PoPEX method, because it similarly appears in the approximation of the exact solution in **Figure 7**. However, it indicates that the training image (prior knowledge) together with the available observations (likelihood function) make the upwards extension of the reference zone very unlikely in terms of the posterior probability.

3.4. Convergence and Parallel Behavior

The synthetic inverse problem described above allows to compute exact predictions from a sufficiently large set of models. To do so, we put $n_{\max} = 0$ and generated an empirical reference set of 1,000,000 unconditioned realizations. From this large ensemble, any prediction can be computed accurately by using Equation (8) together with weights such that $w_k = L(\mathbf{m}_k)$ (c.f. Equation 11). As the reference set is sufficiently large, the degeneracy problem described in section 2.3 can be ignored. The resulting predictions are considered to be the exact solutions and are denoted by μ_{ex} . Although the number of realizations is very large, it is not unsoiled to call these solutions to be exact. Nevertheless, these are very accurate approximations of the true solution such that, in this work, we will call them “exact prediction” or “exact solution.” The corresponding prediction of the 10-days capture zone probability is shown in **Figure 7**.

Once an exact solution is available, we might be interested in the convergence speed of the PoPEX algorithm. Therefore, after each iteration $k = 1, \dots, N$, a prediction $\hat{\mu}_k$ is computed by using the algorithm 3 and compared to μ_{ex} . As mentioned earlier, these two maps define Bernoulli probability values for each point in the computational domain. It determines whether the groundwater



at the corresponding location belongs to the 10 days capture zone or not. A convenient distance between two Bernoulli probability maps $\hat{\mu}_k$ and μ_{ex} is the Jensen-Shannon divergence (JSD) [e.g., Lin (2006)] reading

$$J(\hat{\mu}_k || \mu_{\text{ex}}) = \frac{1}{2} \left(D(\hat{\mu}_k || m) + D(\mu_{\text{ex}} || m) \right),$$

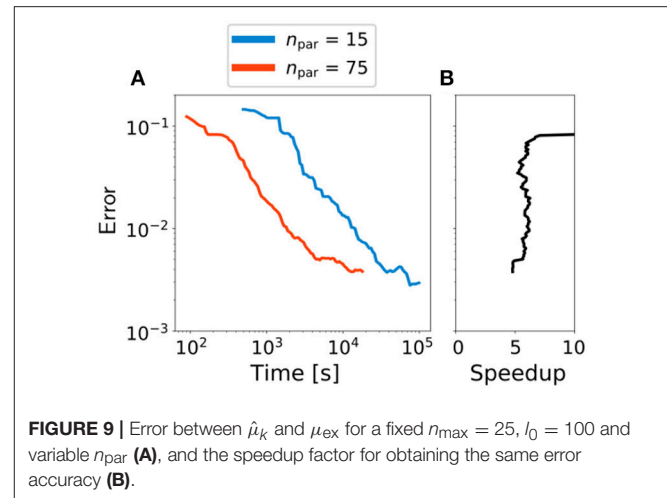
with $m = (\hat{\mu}_k + \mu_{\text{ex}})/2$ and D being the Kullback-Leibler divergence as in Equation (5). This distance measure is computed pointwise over the simulation grid and therefore defines one distance value per pixel. A (scalar) error value is then obtained by computing the spatial average of the Jensen-Shannon divergence map.

Figure 8 shows the evolution of error between $\hat{\mu}_k$ and μ_{ex} with respect to the iteration k . For increasing the statistical significance of the results, every curve represents the average performance of 5 similar runs with different initial seed. First, the minimum number of effective models l_0 has been fixed to 100 and we varied the maximum number of conditioning values $n_{\max} \in \{10, 25\}$. **Figure 8A** shows that the two convergence curves are quite similar. This is not surprising, because the PoPEX algorithm is designed to correct the influence of the hard conditioning by using Equation (10). It follows that for a reasonable hard

conditioning bound, the results are not highly sensible to the choice of n_{\max} . On the other hand, it can be seen from the blue curve that for $n_{\max} = 10$ and $k > 9,000$ the error reaches a “plateau.” This signifies that for a certain time, the PoPEX algorithm was not able to further improve the prediction or in other words, that the method could not find sufficiently important realizations. From such behavior it can be deduced that the learning effect must be reinforced by increasing n_{\max} . However, what is important is that the overall convergence rate of both curves well compares with the dashed line representing $k^{-1/2}$. This is significant because if we directly sample from the posterior probability distribution σ and k is the number of samples, the Central Limit Theorem (CLT) (Durrett, 2010) predicts a convergence rate of $k^{-1/2}$. Because the error curves represent the average performance of 5 PoPEX runs, it is not surprising that they slightly fluctuate and do not reproduce the theoretical rate of $k^{-1/2}$ precisely.

For the second experience, we fixed $n_{\max} = 25$ and varied $l_0 \in \{0, 25, 100, 500\}$. We recall that the choice of a large value for l_0 generally increases the effective number of weights but implies a risk to produce biased predictions. On the other hand, when the effective number of weights is too low, the predictions will be based on very few models and may be biased as well. It is therefore not surprising that for $l_0 = 0$ the approximation accuracy is very bad (green curve in **Figure 8B**). However, the remaining three convergence curves are highly similar for $k \leq 4,000$ where the magenta curve ($l_0 = 25$) reaches a “plateau” and has difficulties to further improve the approximations. As in the previous figure, the curves represent the average performance of 5 similar PoPEX runs with different initial seed. It follows that small fluctuations may arise and should not be overestimated. However, the stagnation of the curve with $l_0 = 25$ might be due out of a different reason. Whenever the parameter l_0 is small, the weights in W_{α}^k are more sensible to highly dominant values. This means that a model \mathbf{m}^{k_0} with very large weight w_{k_0} might dominate the prediction $\hat{\mu}^k$ for many iteration $k > k_0$ and therefore, the approximation error only slightly changes. So such a behavior indicates that l_0 should not be too small. We can again conclude by the fact for a reasonably large l_0 , the overall convergence rate compares very well with the theoretical rate of $k^{-1/2}$.

The last part of the results section is dedicated to a short analysis of the parallel scalability of PoPEX. We repeat the same exercise by first using $n_{\text{par}} = 15$ on a 64 CPU facility ($34.4(\text{Tflop/s})$), and then changing to $n_{\text{par}} = 75$ on 320 CPU's ($172(\text{Tflop/s})$). Therefore, between the first and the second procedure, the computational capacity has been increased by a factor of 5. The performances will be compared by measuring the total sampling time and by a convergence analysis similar to the one in **Figure 8**. We fixed $n_{\max} = 25$ and $l_0 = 100$ and ran PoPEX until 20,000 models have been sampled. All runs were performed 5 times with different initial seeds. The total runtime for the two setups was $27.51 \pm 1.521[h]$ and $5.00 \pm 0.397[h]$, respectively. This signifies an overall speedup factor of 5.5 ± 0.74 and therefore fully satisfies the expectations.



Considering the convergence analysis in **Figure 8** we are now interested in the speedup factor for obtaining the same approximation accuracy when predicting the 10-days capture zone probability. This means that in each iteration $k = 1, \dots, 20,000$, the approximation errors are again computed by a Jensen-Shannon divergence between the prediction and the exact solution. In **Figure 9A** however, we compare the approximation error vs. the elapsed time in (s).

It can be seen that the convergence rate of both curves are very similar and the obtained gain factor highly matches the increase of the computer resources. This becomes even more obvious in **Figure 9B**. It shows the observed speedup in time for obtaining the same approximation accuracy. This means that for any error value (y-axis) we computed the times (and the corresponding speedup factor) that were needed for reaching the considered approximation accuracy. From the relatively small statistical set of 5 chains per exercise, it is not surprising that there is a certain variability in the computations. However, it is evident that the curve significantly matches the predicted speedup factor of 5 and therefore underlines the exceptional scaling behavior of the PoPEX algorithm.

4. DISCUSSION

This paper presents a fast and efficient sampling method for solving inverse problems having a complex and discrete prior. The algorithm is parallelized and scales perfectly. This means that the number of samples computed in parallel is equal to the time reduction factor without compromising the quality of the results. Every sample involves two different main processes: generate a new model and compute the corresponding likelihood value. In this regard, the main concern for using the proposed method in practice is the number of such processes that can be run simultaneously. As there are many supercomputers publicly available however, handling a significant number of computations in parallel should not be a major issue.

Some important concepts of the above algorithm have originally been introduced by Jäggli et al. (2017) where the inverse method was named Posterior Population Expansion (PoPEX). In the present paper we suggest some minor changes concerning the sampling procedure and completely reconsider the method to compute predictions. Nevertheless, we decided to keep the name of the algorithm so that whenever the terminology *PoPEX* is used in the following, it refers to the algorithm as presented in this paper. PoPEX is capable to handle all the four different types of uncertainty distinguished by Sagar et al. (1975): spatial heterogeneities, initial conditions, boundary conditions and sources/sinks. The only requirement for the algorithm to be efficient, is that some uncertainties are modeled by conditional simulation tools.

As illustrated in the case study, a possibility is to use Multiple Point Statistics (MPS) to produce the conditional simulations of heterogeneity. But whenever MPS tools are used, a critical issue is to select an appropriate training image. In practice, it is therefore not uncommon to hesitate about this choice. With the above method, multiple training images can be included. This corresponds to a discrete choice that needs to be formulated in the inverse problem. PoPEX can iteratively learn which image is most appropriate and provide a posterior distribution of the training image selection issue.

PoPEX has been tested based on a two dimensional meandering channel aquifer of size $1,000 \times 500$ (m). A natural gradient of 4‰ and a groundwater extraction rate of 15(l/s) control the groundwater flow. Considering the high complexity of the categorical models, together with the small number of extracted data points, the method solved the inverse problem efficiently and produced accurate estimations of prediction uncertainty. After a very large computational effort, we were able to compute the exact solution and compare it with the predictions made by a PoPEX chain. It was shown empirically that the prediction converged to the exact solution very fast. The convergence speed was comparable with the theoretical rate of $k^{-1/2}$ predicted by the central limit theorem (where k is the number of samples). Furthermore, we demonstrated that the PoPEX results are not very sensitive to the choice of the two main input variables n_{\max} and l_0 . This is very convenient, because there is no uniform criterion for their optimal choice.

In section 2, we mentioned that PoPEX can be interpreted as an adaptive importance sampler (AIS). According to Oh and Berger (1992), the sampling distribution ϕ_k of an AIS technique should follow three properties:

- it should be easy to generate random samples from ϕ_k ;
- the tails of ϕ_k should not be sharper than the tails of $f * \sigma$;
- ϕ_k should mimic $f * \sigma$ well.

The first property depends on the conditional simulation tool entrained to generate new models and is usually satisfied. Regarding the third property, it can be shown that the sampling distribution that minimizes the variance of $\hat{\mu}_k$ in Equation (8) is proportional to $f * \sigma$. When working with prior distributions ρ that are fairly flat over the region where $f(\mathbf{m})L(\mathbf{m})$ is concentrated, taking a sampling distribution proportional to $f * L$

is nearly optimal (Oh and Berger, 1992). But as samples may be used to generate predictions for many different functions, PoPEX is trying to learn a sampling distribution according to the likelihood values $L(\mathbf{m})$ (c.f. Equation 4). However, the link between L and ϕ_k must not be too strong. Let's assume that for a sufficiently large k , the sampling distributions is approximately proportional to L^r for a given power $r > 1$. In this case we have

$$\frac{\sigma}{\phi_k} \propto \frac{\rho}{L^{r-1}}.$$

For a flat distribution ρ and an infinite model space \mathcal{M} this ratio might be unbounded so that the variance of Equation (8) is not finite.

The main limitation of the PoPEX method is that the likelihood values in Equation (8) must be evaluated and represented by a floating-point number. If the dimension of the data space is very large, it may happen that the numerical likelihood values are zero for most realizations. In this case, most of the indicator functions in Equation (4) are multiplied by zero and the learning process of the method is very slow. But if the number of observations is large, it is not uncommon that they are highly correlated. This means that it might be possible to trim the data set and project the observations onto a smaller data space. In other words, a possible strategy to overcome this issue would be to analyze the set of observations, extract a smaller amount of independent information and define an appropriate likelihood function. Alternatively, the likelihood function may be written as a *Gibbs field* (or *measure*) (Winkler, 2012), i.e.,

$$L(\mathbf{m}) = \frac{1}{C} \exp\{-H(\mathbf{m})\}.$$

Such a measure is induced by a normalization constant C and an energy function H . The latter is unique up to an additive constant and therefore, for finite model spaces as well as for Gaussian distributions we may assume that $H \geq 0$. Usually, for floating point operations it is easier to work with the energy $H(\mathbf{m})$ rather than with the unnormalized Gibbs measure $\exp\{-H(\mathbf{m})\}$ directly. During the evolution of the PoPEX algorithm, whenever P^k is computed from Equation (4), we could weight the indicator functions $\mathbf{1}_{f_i}(\mathbf{m}_j)$ proportional to

$$\frac{1}{1 + H(\mathbf{m}_j)}$$

rather than $\tilde{L}(\mathbf{m}_j)$. For the computation of ergodic predictions however, we would still need to compute the likelihood values. But considering the underlying floating point operations, it can be advantageous to learn from non-zero energy values $H(\mathbf{m})$ in order to obtain a sufficiently large number of non-zero likelihood values $L(\mathbf{m})$.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

All authors conceived of the presented idea and developed the theory. CJ designed and wrote the software in order to perform the computations. All authors designed the case study, discussed the results and contributed to the final manuscript.

REFERENCES

- Arpat, G. B., and Caers, J. (2007). Conditional simulation with patterns. *Math. Geol.* 39, 177–203. doi: 10.1007/s11004-006-9075-3
- Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., and Slooten, L. (2005). Inverse problem in hydrogeology. *Hydrogeol. J.* 13, 206–222. doi: 10.1007/s10040-004-0404-7
- Chilès, J.-P., and Delfiner, P. (2009). *Geostatistics: Modeling Spatial Uncertainty*, Vol. 497. Hoboken, NJ: John Wiley & Sons.
- de Marsily, G., Delhomme, J. P., Coudrain-Ribstein, A., and Lavenue, A. M. (2000). Four decades of inverse problems in hydrogeology. *Geol. Soc. Am. Spec. Pap.* 348, 1–17. doi: 10.1130/0-8137-2348-5.1
- Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. New York, NY: Springer.
- Durrett, R. (2010). *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. New York, NY: Cambridge University Press.
- Feyen, L., and Caers, J. (2006). Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations. *Adv. Water Resour.* 29, 912–929. doi: 10.1016/j.advwatres.2005.08.002
- Fu, J., and Gómez-Hernández, J. (2008). Preserving spatial structure for inverse stochastic simulation using blocking markov chain monte carlo method. *Inverse Probl. Sci. Eng.* 16, 865–884. doi: 10.1080/17415970802015781
- Gómez-Hernández, J. J., and Wen, X.-H. (1998). To be or not to be multi-Gaussian? a reflection on stochastic hydrogeology. *Adv. Water Resour.* 21, 47–61. doi: 10.1016/S0309-1708(96)00031-0
- Hansen, T. M., Cordua, K. S., and Mosegaard, K. (2012). Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling. *Comput. Geosci.* 16, 593–611. doi: 10.1007/s10596-011-9271-1
- Hesterberg, T. C. (2003). *Advances in Importance Sampling*. Ph.D. thesis, Stanford University.
- Honarkhah, M., and Caers, J. (2010). Stochastic simulation of patterns using distance-based pattern modeling. *Math. Geosci.* 42, 487–517. doi: 10.1007/s11004-010-9276-7
- Jäggli, C., Straubhaar, J., and Renard, P. (2017). Posterior population expansion for solving inverse problems. *Water Resour. Res.* 53, 2902–2916. doi: 10.1002/2016WR019550
- Journel, A., and Zhang, T. (2006). The necessity of a multiple-point prior model. *Math. Geol.* 38, 591–610. doi: 10.1007/s11004-006-9031-2
- Kerrou, J., Renard, P., Franssen, H.-J. H., and Lunati, I. (2008). Issues in characterizing heterogeneity and connectivity in non-multigaussian media. *Adv. Water Resour.* 31, 147–159. doi: 10.1016/j.advwatres.2007.07.002
- Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* 22, 79–86. doi: 10.1214/aoms/1177729694
- Laloy, E., Linde, N., Jacques, D., and Mariethoz, G. (2016). Merging parallel tempering with sequential geostatistical resampling for improved posterior exploration of high-dimensional subsurface categorical fields. *Adv. Water Resour.* 90, 57–69. doi: 10.1016/j.advwatres.2016.02.008
- Leeuwen, M., te Stroet, C. B. M., Butler, A. P., and Tompkins, J. A. (1998). Stochastic determination of well capture zones. *Water Resour. Res.* 34, 2215–2223. doi: 10.1029/98WR01552
- Lin, J. (2006). Divergence measures based on the shannon entropy. *IEEE Trans. Inf. Theor.* 37, 145–151. doi: 10.1109/18.61115
- Linde, N., Renard, P., Mukerji, T., and Caers, J. (2015). Geological realism in hydrogeological and geophysical inverse modeling: a review. *Adv. Water Resour.* 86, 86–101. doi: 10.1016/j.advwatres.2015.09.019
- Liu, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. New York, NY: Springer Publishing Company, Incorporated.
- Lopez, S., Cojan, I., Rivoirard, J., and Galli, A. (2009). *Process Based Stochastic Modelling: Meandering Channelized Reservoirs*. Hoboken, NJ: Wiley-Blackwell, 139–144.
- Mariethoz, G., and Caers, J. (2014). *Front Matter*. Hoboken, NJ: Wiley-Blackwell.
- Mariethoz, G., Renard, P., and Caers, J. (2010a). Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resour. Res.* 46:W11530. doi: 10.1029/2010WR009274
- Mariethoz, G., Renard, P., and Straubhaar, J. (2010b). The direct sampling method to perform multiple-point geostatistical simulations. *Water Resour. Res.* 46:W11536. doi: 10.1029/2008WR007621
- Moles, C. G., Mendes, P., and Banga, J. R. (2003). Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* 13, 2467–2474. doi: 10.1101/gr.1262503
- Mosegaard, K., and Tarantola, A. (2002). “16 - probabilistic approach to inverse problems,” in *International Handbook of Earthquake and Engineering Seismology, Vol. 81, Part A of International Geophysics*, eds P. C. J. William, H. K. Lee, H. Kanamori, and C. Kisslinger (Philadelphia, PA: Academic Press), 237–265.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.
- Naylor, J., and Smith, A. (1988). Econometric illustrations of novel numerical integration strategies for Bayesian inference. *J. Econom.* 38, 103–125. doi: 10.1016/0304-4076(88)90029-2
- Oh, M.-S., and Berger, J. O. (1992). Adaptive importance sampling in monte carlo integration. *J. Stat. Comput. Simul.* 41, 143–168. doi: 10.1080/00949659208810398
- Oliver, D. S., Cunha, L. B., and Reynolds, A. C. (1997). Markov chain monte carlo methods for conditioning a permeability field to pressure data. *Math. Geol.* 29, 61–91. doi: 10.1007/BF02769620
- Owen, A. B. (2013). *Monte Carlo Theory, Methods and Examples*. Available online at: <https://statweb.stanford.edu/~owen/mc/>
- Robert, C. P., and Casella, G. (2004). “Monte carlo statistical methods,” in *Springer Texts in Statistics* (New York, NY: Springer-Verlag).
- Romary, T. (2010). History matching of approximated lithofacies models under uncertainty. *Comput. Geosci.* 14, 343–355. doi: 10.1007/s10596-009-9166-6
- Rubinstein, R. Y., and Kroese, D. P. (2016). *Simulation and the Monte Carlo Method*, 3 Edn. Hoboken, NJ: Wiley.
- Sagar, B., Yakowitz, S., and Duckstein, L. (1975). A direct method for the identification of the parameters of dynamic nonhomogeneous aquifers. *Water Resour. Res.* 11, 563–570. doi: 10.1029/WR011i004p00563
- Straubhaar, J. (2011). *DeeSse Technical Reference Guide*. Technical Report, Centre d'hydrogéologie et géothermie, University of Neuchâtel, (Neuchâtel).
- Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O. (2011). An improved parallel multiple-point algorithm using a list approach. *Math. Geosci.* 43, 305–328. doi: 10.1007/s11004-011-9328-7
- Straubhaar, J., Walgenwitz, A., and Renard, P. (2013). Parallel multiple-point Statistics algorithm based on list and tree structures. *Math. Geosci.* 45, 131–147. doi: 10.1007/s11004-012-9437-y
- Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol.* 34, 1–21. doi: 10.1023/A:1014009426274

ACKNOWLEDGMENTS

We would like to thank Przemyslaw Juda for his contributions toward the generation of the conceptual training image. This work was founded by the Swiss National Science Foundation through the grant number 153637.

- Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation* (Philadelphia, PA: Society for Industrial and Applied Mathematics)
- Wainwright, J., and Mulligan, M. (2005). *Environmental Modelling: Finding Simplicity in Complexity*. West Sussex: John Wiley & Sons.
- Winkler, G. (2012). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Stochastic Modelling and Applied Probability. Berlin, Heidelberg: Springer.
- Zhou, H., Gómez-Hernández, J. J., and Li, L. (2014). Inverse methods in hydrogeology: evolution and recent trends. *Adv. Water Resour.* 63, 22–37. doi: 10.1016/j.advwatres.2013.10.014

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Jäggli, Straubhaar and Renard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Efficient gHMC Reconstruction of Contaminant Release History

David A. Barajas-Solano¹, Francis J. Alexander², Marian Anghel³ and Daniel M. Tartakovsky^{4*}

¹ Pacific Northwest National Laboratory, Computational Mathematics, Richland, WA, United States, ² Brookhaven National Laboratory, Computational Science Initiative, Upton, NY, United States, ³ Los Alamos National Laboratory, Computer, Computational, and Statistical Sciences Division, Los Alamos, NM, United States, ⁴ Department of Energy Resources Engineering, Stanford University, Stanford, CA, United States

We present a generalized hybrid Monte Carlo (gHMC) method for fast, statistically optimal reconstruction of release histories of reactive contaminants. The approach is applicable to large-scale, strongly nonlinear systems with parametric uncertainties and data corrupted by measurement errors. The use of discrete adjoint equations facilitates numerical implementation of gHMC without putting any restrictions on the degree of nonlinearity of advection-dispersion-reaction equations that are used to describe contaminant transport in the subsurface. To demonstrate the salient features of the proposed algorithm, we identify the spatial extent of a distributed source of contamination from concentration measurements of a reactive solute.

Keywords: source identification, contaminant transport, Markov Chain Monte Carlo, hybrid Monte Carlo, inverse problems, uncertainty quantification

OPEN ACCESS

Edited by:

Philippe Renard,
Université de Neuchâtel, Switzerland

Reviewed by:

Guillaume Pirot,
Universität Lausanne, Switzerland
Roseanna M. Neupauer,
University of Colorado System,
United States

*Correspondence:

Daniel M. Tartakovsky
tartakovsky@stanford.edu

Specialty section:

This article was submitted to
Freshwater Science,
a section of the journal
Frontiers in Environmental Science

Received: 12 November 2018

Accepted: 17 September 2019

Published: 11 October 2019

Citation:

Barajas-Solano DA, Alexander FJ,
Anghel M and Tartakovsky DM (2019)
Efficient gHMC Reconstruction of
Contaminant Release History.
Front. Environ. Sci. 7:149.
doi: 10.3389/fenvs.2019.00149

1. INTRODUCTION

An accurate reconstruction of the release history of contaminants in geophysical systems is essential to regulatory and remedial efforts. These efforts rely on measurements of pollutant concentration to identify the sources and/or release history of a pollutant. Unfortunately, available concentration data are typically sparse in both space and time and are corrupted by measurement errors. Source identification and reconstruction of release history are further complicated by both spatial heterogeneity of model parameters and their insufficient characterization, although we do not consider these effects in the present work.

Detailed reviews of the historic developments and state-of-the-art in the field of inverse modeling as related to contaminant source identification are presented in Atmadja and Bagtzoglou (2001b) and Hutchinson et al. (2017). The existing approaches can be subdivided into two broad classes: deterministic and probabilistic. Deterministic approaches include, but are not limited to, Tikhonov regularization of convolution integrals (Liu and Ball, 1999; Ito and Jin, 2015), least-square estimation from analytical approximations (Butcher and Gauthier, 1994), least-square solution of an optimal control problem (Gugat, 2012), the method of quasi-reversibility (Skaggs and Kabala, 1995; Bagtzoglou and Atmadja, 2003), and the backward beam equation method (Atmadja and Bagtzoglou, 2001a; Bagtzoglou and Atmadja, 2003). These approaches provide estimates of the release history from a source of known locations and are not designed for quantifying the uncertainty associated with these estimates. The robustness of these methods is highly sensitive to measurement errors, and their mathematical formulations are often fundamentally ill-posed.

While existing probabilistic approaches, such as random walk particle tracking for the backward transport equation (Bagtzoglou et al., 1992), minimum relative entropy (Woodbury and Ulrych, 1996), and adjoint methods (e.g., Neupauer and Wilson, 1999), alleviate some of these problems, others remain. For example, these and similar methods do not take advantage of the regularizing

nature of the measurement noise and, hence, are often ill-posed. Thus, the minimum relative entropy method treats concentration measurements as ensemble averages. Additionally, there are some outstanding issues with quantifying uncertainty (Neupauer et al., 2000) and the inability of many existing approaches to handle more than one observation point (Neupauer and Wilson, 2005).

Finally, most existing approaches to the reconstruction of release history are restricted to linear transport phenomena, that is, transport phenomena for which the transport equation is linear in the concentration, and thus are limited to migration of contaminants that are either conservative (all the references above) or exhibit first-order (linear) reaction rates (Neupauer and Wilson, 2003, 2004). This is because such approaches are based on either Green's functions (Skaggs and Kabala, 1994; Woodbury and Ulrych, 1996; Stanev et al., 2018) or analytically derived adjoint equations (Neupauer and Wilson, 1999, 2005). The use of Kalman filters for source identification (Herrera and Pinder, 2005) is formally limited to linear transport phenomena and Gaussian errors. While both limitations can be relaxed by employing various generalizations of the Kalman filter such as the extended and ensemble Kalman filter (e.g., Xu and Gómez-Hernández, 2016, 2018), these generalizations are known to fail if the nonlinearity is too strong. Bayesian optimization approaches (Piro et al., 2019), accelerated by the use of Gaussian process models as surrogates, provide a promising alternative to the Kalman filter since they impose no linearity requirements.

Purely statistical approaches to history reconstruction, such as the geostatistical inversion with Bayesian updating (Snodgrass and Kitanidis, 1997) and various machine learning techniques (Vesselinov et al., 2018, 2019), are applicable to nonlinear transport. Since this is achieved by ignoring governing equations, the reconstructed release histories could have non-physical characteristics, including negative concentrations. These problems have been alleviated by introducing additional constraints into an optimization functional and requiring the reconstructed field to be Gaussian (Michalak and Kitanidis, 2003, 2004a). Combining these geostatistical approaches with analytically derived adjoint equations (Michalak and Kitanidis, 2004b; Shlomi and Michalak, 2007) however brings back the linearity requirement.

We present an optimal reconstruction of contaminant release history that fully utilizes all available information and requires neither the linearity of governing transport equations nor the Gaussianity of the underlying fields. In section 2 we formulate the problem of reconstructing the contaminant release history from noisy observations. Section 3 introduces our general computational framework, which is further implemented in section 4 for various examples.

2. PROBLEM FORMULATION

2.1. Reconstruction of Contaminant Release History

We consider migration of a single chemically active contaminant in a porous medium $\Omega \subset \mathbb{R}^d$, $d \in [1, 3]$. We assume that reactive

transport is adequately described by the advection-dispersion-reaction equation with reaction term $R(c)$:

$$\frac{\partial c}{\partial t} = \nabla \cdot (\mathbf{D} \nabla c) - \nabla \cdot (\mathbf{u} c) - R(c) + r(\mathbf{x}, t), \quad \mathbf{x} \in \Omega, \quad t > 0, \quad (1)$$

together with corresponding boundary conditions. Here $c = c(\mathbf{x}, t)$ is the solute concentration at point \mathbf{x} and time t , \mathbf{u} is the average macroscopic pore velocity, \mathbf{D} is the dispersion coefficient tensor, and $r(\mathbf{x}, t)$ is the source function. Both the location and duration of the contaminant release, i.e., the source function $r(\mathbf{x}, t)$, can be unknown, but only the former source of uncertainty is treated in the computational examples of section 4, that is, we assume that $r(\mathbf{x}, t) = r(\mathbf{x})\delta(0)$.

Introducing the dimensionless quantities

$$c' = \frac{c}{c_0}, \quad x'_i = \frac{x_i}{x_0}, \quad i \in [1, d], \quad t' = \frac{t}{t_0}, \quad \mathbf{D}' = \mathbf{D} \frac{t_0}{x_0^2},$$

$$\mathbf{u}' = \mathbf{u} \frac{t_0}{x_0}, \quad r' = r \frac{t_0}{c_0},$$

and the normalized reaction term $R' = R t_0 / c_0$, we rewrite (1) in terms of dimensionless quantities,

$$\frac{\partial c'}{\partial t'} = \nabla' \cdot (\mathbf{D}' \nabla' c') - \nabla' \cdot (\mathbf{u}' c') - R' + r' \quad \mathbf{x}' \in \Omega', \quad t' > 0. \quad (2)$$

The non-dimensionalization of the reaction term is specific to its functional form. The non-dimensionalization of a particular case employed in this manuscript is discussed in section 4. For all the following discussions we will consider the dimensionless Equation (2) and we will drop the prime notation for denoting dimensionless quantities.

For the remainder of this work we assume that \mathbf{u} , \mathbf{D} , and the boundary conditions are known, while the source function $r(\mathbf{x}, t)$ is unknown. Our goal is to reconstruct the release history $r(\mathbf{x}, t)$ from concentration data $\tilde{c}_{mi} = \tilde{c}(\mathbf{x}_m, t_i)$ collected at points $\{\mathbf{x}_m\}_{m=1}^M$ at times $\{t_i\}_{i=1}^I$, and for known \mathbf{u} , \mathbf{D} , and boundary conditions.

Concentration measurements are corrupted by measurement errors. We assume that the measured concentrations \tilde{c}_{mi} differ from the true concentration by an additive measurement noise, so that

$$\tilde{c}_{mi} = c(\mathbf{x}_m, t_i) + \epsilon_{mi}, \quad (3)$$

where the errors ϵ_{mi} are zero-mean Gaussian random variables with covariance $\mathbb{E}[\epsilon_{mi}\epsilon_{nj}] = \delta_{ij}R_{mn}$, where $\mathbb{E}[\cdot]$ is the expectation operator, δ_{ij} denotes the Kronecker delta function, and the R_{mn} , $m, n \in [1, M]$, are components of the spatial covariance matrix $\mathbf{R} \in \mathbb{R}^{M \times M}$ of measurement errors. This treatment of measurement noise assumes that the measurements are well separated in time to neglect any temporal correlations, but the model can be easily extended to include temporal correlations. We use the additive error model (3) of Woodbury and Ulrych (1996) rather than the multiplicative error model of Skaggs and Kabala (1994) for the purpose of illustration only. Both models have similar effects on the accuracy of history

reconstruction (Neupauer et al., 2000) and can be handled by our approach.

In a typical situation, one has prior information (or a belief) about potential sources of contamination (a region Ω_c within the flow domain Ω) and a time period $[T_l, T_u]$ during which the release has occurred. Examples of Ω_c include spatially distributed zones of contamination (e.g., landfills) and a collection of point sources (e.g., localized/small industrial sites or storage facilities), some of which have contributed to contamination. The lower (T_l) and upper (T_u) bounds of the release interval might represent the time when a landfill became operational and the time when contamination has first been detected, respectively. In the absence of prior information about the release occurrence, one can assume a uniform random distribution of the release in $[T_l, T_u] \times \Omega$. We allow for an arbitrary number of measurement points and for either discrete or continuous-in-time measurements.

2.2. Likelihood Function

To simplify the exposition, we assume a spatially distributed chemical release at time $t = 0$ only, i.e., $r(\mathbf{x}, t) = c(\mathbf{x}, 0)\delta(t)$. Given the measurements $\bar{c}(\mathbf{x}_m, t_i)$ and the noise model (3), our goal is to determine the likelihood of a given release configuration $c(\mathbf{x}, 0)$. Unfortunately, the measurements, generally taken at later times, do not estimate directly the likelihood of a release configuration. Nevertheless, because the transport Equation (2) is deterministic, we can implicitly assess the likelihood of a given release configuration, $P[c(\mathbf{x}, 0)]$, from the probability (likelihood) of a given (computed) concentration history $c(\mathbf{x}, t)$. This likelihood can be expressed as (Alexander et al., 2005)

$$P[c(\mathbf{x}, 0)] \sim \exp\{-\tilde{H}_{\text{obs}}[c(\mathbf{x}, t)]\}, \quad (4)$$

where $\tilde{H}_{\text{obs}}[c(\mathbf{x}, t)]$ is the so-called “Hamiltonian” or log-likelihood function,

$$\tilde{H}_{\text{obs}}[c(\mathbf{x}, t)] = \frac{1}{2} \sum_{m,n=1}^{M,I} \Delta_{mi}(\mathbf{R}^{-1})_{mn} \Delta_{ni}, \quad (5)$$

where $\Delta_{mi} \equiv c(\mathbf{x}_m, t_i) - \bar{c}(\mathbf{x}_m, t_i)$, and \mathbf{R} is the covariance matrix of measurement errors, as defined in section 2.1. Since (2) uniquely determines the evolution of the solute concentration from its initial state $c(\mathbf{x}, 0)$, the Hamiltonian (5) is a nonlinear functional of the initial conditions $c(\mathbf{x}, 0)$, i.e., $\tilde{H}_{\text{obs}}[c(\mathbf{x}, t)] = H_{\text{obs}}[c(\mathbf{x}, 0)]$.

This formulation assumes that the measurement errors ϵ_{mi} are Gaussian and uncorrelated with the state of the system. Other distributions of the measurement noise and the stochasticity of governing equations can be handled as well (Alexander et al., 2005). The Hamiltonian for stochastic systems, which can represent, e.g., uncertain hydraulic conductivity and flow velocity that are treated as random fields, can be reformulated to explicitly include the transport equation (Alexander et al., 2005; Archambeau et al., 2007).

The contribution of highly fluctuating or unphysical initial conditions is reduced by adding a regularization term

$H_{\text{reg}}[c(\mathbf{x}, 0)]$ to the observation Hamiltonian (5) and replacing the likelihood function (4) with

$$P[c(\mathbf{x}, 0)] \sim \exp\{-H[c(\mathbf{x}, 0)]\}, \quad (6a)$$

where

$$H[c(\mathbf{x}, 0)] = H_{\text{obs}}[c(\mathbf{x}, 0)] + \gamma H_{\text{reg}}[c(\mathbf{x}, 0)], \quad (6b)$$

and the weight $\gamma > 0$ is a tuning hyperparameter. The regularization term H_{reg} is equivalent to a Bayesian log-prior distribution on the initial condition. The selection of an appropriate regularization Hamiltonian is particularly important for problems in which the observation Hamiltonian does not specify a proper probability distribution for $c(\mathbf{x}, 0)$ due to a lack of measurements. For a one-dimensional source profile, the squared gradient of the initial spatial profile can play the role of the regularization Hamiltonian. In higher dimensions, one can use a thin-plate penalty functional (Wahba, 1990).

A conceptual difference between our approach and maximum likelihood methods is worth emphasizing. Rather than sampling the Gibbs distribution $\exp\{-H[c(\mathbf{x}, 0)]\}$, as we do here, maximum likelihood methods minimize the Hamiltonian (6b) over $c(\mathbf{x}, 0)$. While standard maximum likelihood methods determine the mode and variance of the posterior distribution under a Gaussian approximation, the approach described below can be used to determine the mean and higher-order statistics, and it is valid even when the posterior distribution is multi-modal.

3. MONTE CARLO SAMPLING

In principle, one can sample the Gibbs distribution by using Markov-chain Monte Carlo (MCMC) (e.g., Michalak and Kitanidis, 2003). However, quite often, the disadvantage of local MCMC-based methods is their slow convergence. To improve convergence, we apply a Generalized Hybrid Monte Carlo (gHMC), which enables one to efficiently sample release configurations $c(\mathbf{x}, 0)$ with probability given by (6a).

3.1. Hybrid Monte Carlo (HMC)

Hybrid Monte Carlo (HMC) refers to a class of methods that combine Hamiltonian molecular dynamics with Metropolis-Hastings Monte Carlo simulations (see Neal, 1993 for an introductory survey). Specifically, a time-discretized integration of the molecular dynamics equations is used to propose a new configuration, which is then accepted or rejected by the standard Metropolis-Hastings Monte Carlo criteria. The change in total energy serves as the acceptance/rejection criteria.

In HMC one treats the log-likelihood function H in (6b) as the configurational Hamiltonian for a system of N “particles,” each of which has unit mass and generalized coordinates q_1, q_2, \dots, q_N . Each of these generalized positions corresponds to the solute concentration $c(\mathbf{x}, t)$ at a space-time point (\mathbf{x}, t) . In the following, the particle positions correspond to the initial concentration at time $t = 0$, e.g., $q_i = c(x_i, 0)$, $x_i = iL/(N-1)$, $i = 0, \dots, N-1$ for a contaminant release over the one-dimensional domain $[0, L]$.

At any given time, the state of the system is completely described by (\mathbf{q}, \mathbf{p}) , where $\mathbf{q} = \{q_i\}_{i=1}^N$ and $\mathbf{p} = \{p_i\}_{i=1}^N$. Here, the momentum of the i th particle, p_i , is $dq_i/d\tau = p_i$, where τ is the fictitious time of the molecular dynamics. The kinetic energy of the system of N particles is given by

$$H_K(\mathbf{p}) = \frac{1}{2} \sum_{i=1}^N p_i^2, \quad (7)$$

and the total Hamiltonian of the system is

$$\hat{H}(\mathbf{q}, \mathbf{p}) = H(\mathbf{q}) + H_K(\mathbf{p}). \quad (8)$$

It follows that the Hamiltonian dynamics are given by

$$\frac{dq_i}{d\tau} = p_i, \quad \frac{dp_i}{d\tau} = F_i, \quad F_i = -\frac{\partial H}{\partial q_i}, \quad (9)$$

where F_i is the force acting on the i th particle, that is to be computed from the governing transport equation. During the time interval $\Delta\tau$, the system evolves from its current state (\mathbf{q}, \mathbf{p}) to a new state $(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})$, which can be computed by discretizing the Hamiltonian dynamics (9). An example of such a discretization is the standard leapfrog method, which is written as

$$\tilde{q}_i = q_i + \Delta\tau p_i + \frac{\Delta\tau^2}{2} F(\mathbf{q}) \quad (10a)$$

$$\tilde{p}_i = p_i + \frac{\Delta\tau}{2} \{F(\mathbf{q}) + F(\tilde{\mathbf{q}})\}, \quad (10b)$$

for $i = 1, \dots, N$. Multiple leapfrog steps, i.e., multiple applications of Equation (10), can be performed. For the hybrid Monte Carlo method, the number of leapfrog steps S is larger than one. For $S = 1$ we obtain the Langevin Monte Carlo method (Neal, 1993). This completes the “proposal part” of HMC.

The remaining part of HMC consists of deciding whether to accept or reject the new state $(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})$. This is done by the Metropolis-Hastings sampling strategy, according to which the new state $(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})$ is accepted with probability

$$Q = \min \left\{ 1, \exp\{\hat{H}(\mathbf{q}, \mathbf{p}) - \hat{H}(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})\} \right\}. \quad (11)$$

The momenta variables $\tilde{\mathbf{p}}$ are resampled after each acceptance/rejection of the position variables according to a Gaussian distribution of independent variables $\sim \exp(-H_K)$. The time-marching and acceptance/rejection process represents one step in the Markov chain, and therefore one Monte Carlo sample. It is important to note that the update from (\mathbf{q}, \mathbf{p}) to $(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})$ does not conserve energy as a result of the time discretization. The extent to which energy is not conserved is controlled by the time step $\Delta\tau$. Detailed balance is achieved if the configuration obtained after evolving several steps is accepted with probability Q in (11). Thus, the Metropolis step corrects for time discretization errors.

As we have noted before, the method samples from the multivariate target distribution, $\sim \exp(-\hat{H})$, by computing a Markov chain. Sampling from this density allows us to estimate

the mean state (reconstructed initial configuration) and its variance. Markov chain sampling from the posterior distribution involves a *transient phase*, in which we start from some initial state and simulate the Markov chain for a period long enough to reach its stationary density, followed by a *sampling phase*, in which we assume that the Markov chain visits states from this stationary density. If the chain has converged and the sampling phase is long enough to cover the entire posterior distribution, accurate inferences about any quantity of interest are made by computing the sample mean, variance, and other desired statistics (Landau and Binder, 2009).

3.2. Generalized Hybrid Monte Carlo (gHMC)

In many cases, the generalized hybrid Monte Carlo (gHMC) of Toral and Ferreira (1994) can improve the efficiency of standard HMC by means of the nonlocal sampling strategy described in some detail below. For $\mathbf{q}, \mathbf{p} \in \mathbb{R}^N$, gHMC replaces the Hamiltonian dynamics in (9) with a more general formulation,

$$\frac{d\mathbf{q}}{d\tau} = \mathbf{A}\mathbf{p}, \quad \frac{d\mathbf{p}}{d\tau} = \mathbf{A}^\top \mathbf{F}, \quad (12)$$

where \mathbf{A} is a linear operator represented by a $\mathbb{R}^{N \times N}$ matrix. The corresponding leapfrog discretization is then given by

$$\mathbf{q}(\delta\tau) = \mathbf{q} + \delta\tau \mathbf{A}\mathbf{p} + \frac{\delta\tau^2}{2} \mathbf{A}\mathbf{A}^\top \mathbf{F}[\mathbf{q}], \quad (13a)$$

$$\mathbf{p}(\delta\tau) = \mathbf{p} + \mathbf{A}^\top \frac{\delta\tau}{2} \{\mathbf{F}[\mathbf{q}] + \mathbf{F}[\mathbf{q}(\delta\tau)]\}. \quad (13b)$$

The two formulations, (9) and (12), are identical if \mathbf{A} is the identity matrix. The goal is to find a matrix \mathbf{A} that leads to a significant reduction of the temporal correlations of the Markov chain without appreciably increasing the cost of the update due to matrix-vector multiplications.

In order to illustrate how the introduction of the matrix \mathbf{A} can help to reduce the correlations of the Markov chain, consider the problem with $\mathbf{q} \in \mathbb{R}^N$ and Hamiltonian

$$H(\mathbf{q}) = \frac{1}{2} (\mathbf{q} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{q} - \boldsymbol{\mu}),$$

so that the forcing is given by $-\boldsymbol{\Sigma}^{-1}(\mathbf{q} - \boldsymbol{\mu})$. For the case $\mathbf{A} = \mathbf{I}$, it can be seen from (13) that the different components q_i are updated at different rates, given by the covariance matrix $\boldsymbol{\Sigma}$. For a given $\delta\tau$, some components would be updated with long steps, while others would be updated with shorter steps.

The disadvantage of such a configuration is that too long of a step for a certain component might increase the total Hamiltonian enough to produce a rejection according to (11). If the rejection rate of the chain is too large, one would have to reduce $\delta\tau$, which affects all components. The issue of the rejection rate would be addressed, but then some components would be updated with very short steps, increasing their correlation. To solve this issue, one can remove the appearance of $\boldsymbol{\Sigma}$ altogether by choosing \mathbf{A} such that $\mathbf{A}\mathbf{A}^\top \boldsymbol{\Sigma} = \mathbf{I}$. If $\boldsymbol{\Sigma}$ is a valid covariance matrix, this is trivially accomplished by choosing

\mathbf{A} as the Cholesky factor of Σ . We therefore refer to \mathbf{A} as the “acceleration” matrix.

Unfortunately, in general, the Hamiltonian (6b) for our problem does not have a simple bilinear form for which an appropriate selection of acceleration matrix \mathbf{A} can be derived. Nevertheless, it stands to reason that one can build acceleration matrices for more complex systems to partially reduce the correlation of the Markov chain.

4. NUMERICAL EXPERIMENTS

In this section we illustrate how the framework outlined above can be applied to source identification problems. In the first case we study the implementation of HMC to contaminant transport problems with a nonlinear reaction term for different configurations of observations. In the second case we study a linear advection-dispersion problem and explore possible selections of the gHMC acceleration matrix.

4.1. Discrete in Space, Continuous in Time Measurements

We consider a one-dimensional transport with uniform velocity u and dispersion coefficient D . We employ as the regularization operator the ℓ_2 -norm of the gradient of the initial spatial distribution. Furthermore, we assume that the time of the contaminant release is precisely known and we impose no constraints on the total mass of the released contaminant. The measurements are taken continuously over the time interval $(0, T)$ at a subset J of discrete locations in the spatial domain. We assume that the measurement errors are uncorrelated in space and time and have the same variance σ_ϵ^2 at every point. This setup represents observations of contaminant breakthrough curves at a number of sampling locations.

The Hamiltonian corresponding to this setup is

$$H = \frac{1}{2\sigma_\epsilon^2} \int_0^T \sum_{j \in J} [c(\mathbf{x}_j, t) - \bar{c}(\mathbf{x}_j, t)]^2 dt + \gamma \int_\Omega |\nabla c(\mathbf{x}, 0)|^2 dx. \quad (14)$$

It is evaluated, together with its sensitivity with respect to the initial condition, by using a method-of-lines discretization of the concentration field $c(\mathbf{x}, t)$. Once the governing equation has been discretized into a system of ordinary differential equations (ODEs), one can compute the sensitivity $\nabla_{\mathbf{q}} H$ via the adjoint sensitivity method (Cao et al., 2003; Li and Petzold, 2004). The disadvantage of this approach is that it incurs two levels of numerical error: the integration error of the forward problem, which affects the initial condition of the adjoint problem, and the integration error of the backward problem. If these errors are significant, both the quality of the estimator and the rejection rate of the Markov chain can be compromised. Reducing the error requires one to decrease the time step used for integration in both directions, which would increase the computational cost per leapfrog step.

To partially alleviate this problem, we use a single-step ODE integration scheme for the forward problem and compute the sensitivity of H with respect to the initial condition via multiple

applications of the chain rule (Daescu et al., 2000). Let \mathbf{c}^i ($i = 0, \dots, I$) be a vector of discretized states evaluated at time $t = iT/I$ and $\bar{\mathbf{c}}^i$ be a vector of the measurements at time t_i in the elements corresponding to the J measurements' locations and zeros in the other elements. Let \mathbf{C} be a diagonal matrix with ones on the diagonal elements corresponding to the J subset of measurement locations and zeros in all other locations. We use this notation to rewrite the observation Hamiltonian and the sensitivity as

$$H_{\text{obs}}(\mathbf{c}, \mathbf{q}) = \frac{1}{2\sigma_\epsilon^2} \Delta t \sum_{i=1}^I (\mathbf{c}^i - \bar{\mathbf{c}}^i)^\top \mathbf{C} (\mathbf{c}^i - \bar{\mathbf{c}}^i),$$

and

$$\nabla_{\mathbf{q}} H_{\text{obs}} = \frac{1}{\sigma_\epsilon^2} \Delta t \sum_{i=1}^I \left(\frac{d\mathbf{c}^i}{d\mathbf{q}} \right)^\top \mathbf{C} (\mathbf{c}^i - \bar{\mathbf{c}}^i),$$

respectively, where $d\mathbf{c}^i/d\mathbf{q}$ denotes the Jacobian of \mathbf{c}^i with respect to \mathbf{q} . Using the chain rule, the sensitivity $\nabla_{\mathbf{q}} H_{\text{obs}}$ can be rewritten as

$$\nabla_{\mathbf{q}} H_{\text{obs}} = \frac{1}{\sigma_\epsilon^2} \Delta t \left(\frac{d\mathbf{c}^1}{d\mathbf{q}} \right)^\top \left[\mathbf{C} (\mathbf{c}^1 - \bar{\mathbf{c}}^1) + \left(\frac{d\mathbf{c}^2}{d\mathbf{c}^1} \right)^\top [\mathbf{C} (\mathbf{c}^2 - \bar{\mathbf{c}}^2) + \dots] \right].$$

This implies that the sensitivities can be evaluated by repeatedly computing products of the form $(d\mathbf{c}^{i+1}/d\mathbf{c}^i)^\top \mathbf{u}$. If these products can be computed exactly, then this approach provides the exact sensitivity of the (space-time discretized) system, which is useful for problems with costly forward and backward solutions. The disadvantage of this approach is that it is highly application-specific and restricts the selection of ODE solvers to a specific family. Details of the implementation of this discrete sensitivity analysis approach are presented in **Appendix**.

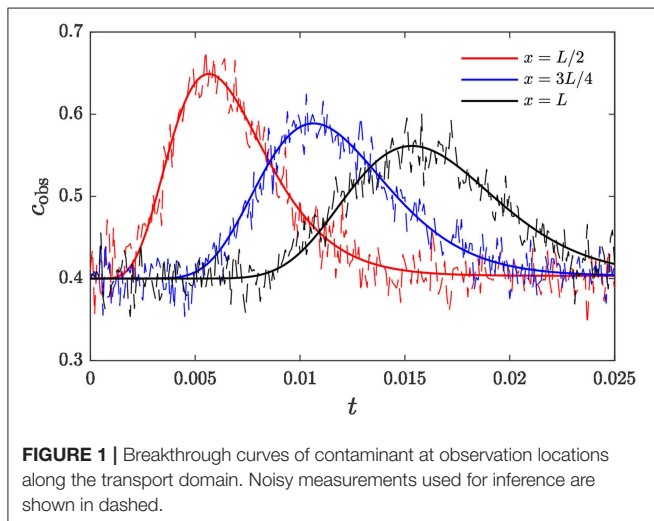
We test this formulation on a one-dimensional transport problem defined in the domain $[0, L]$, $L = 1$ with constant velocity u , dispersion coefficient D , and the reaction model (Lichtner and Tartakovsky, 2003)

$$R(c) = 2k(c^2 - c_{eq}^2), \quad (15)$$

corresponding to a nonlinear heterogeneous (precipitation/dissolution) reaction with equilibrium concentration c_{eq} . Here k denotes the dimensionless kinetic rate constant normalized by porosity, given by $k = \hat{k} t_0 c_0$, where \hat{k} is the kinetic rate constant normalized by porosity, with dimensions of inverse concentration times inverse time, and t_0 and c_0 are the time and concentration scales defined in section 2.1. The parameter values are set to $D = 1.0$, $u = 50.0$, $c_{eq} = 0.4$, and $k = 1.0$. Boundary conditions are $dc/dx = 0$ at $x = 0, L$. The transport equation is discretized employing a finite-volumes scheme consisting of $N = 128$ cells of uniform size Δx . Concentration measurements are taken at the $M = 3$ spatial locations $x = L/2$, $x = 3L/4$, and $x = L$ over the time period $(0, 2.5 \times 10^{-2})$ (**Figure 1**). The standard deviation of these measurements is set to $\sigma_\epsilon = 0.02$.

The release configuration is inferred using the HMC scheme, carried out with hybrid timestep $\Delta\tau = 0.17$, number of leapfrog steps $S = 5$, and regularizing parameter $\gamma = 0.05$. A total of 1×10^5 samples of the release profile were retained after the burn-in period, which are employed to compute the Monte Carlo estimates \bar{c}_0 and $\hat{\sigma}_{c_0}$ of the posterior mean and standard deviation, respectively. These estimates are shown in **Figure 2**. In particular, **Figure 2A** shows the posterior mean estimate \bar{c}_0 compared to the true release profile, c_0 . We also show the 95% confidence interval of the posterior mean estimate. It can be seen that the HMC scheme is able to infer the main features of the initial condition, namely the location of the release and the total mass of contaminant released. For comparison, we also compute the Bayesian *maximum a posteriori* (MAP) point estimate c_{MAP} of the release profile, also presented in **Figure 2A**. The MAP is given by

$$c_{\text{MAP}} = \arg \min_c H[c],$$



where $H[\cdot]$ is the Hamiltonian given in (6b). MAP estimation is similar to the method of Bayesian global optimization (BGO) (Pirrot et al., 2019) in that both aim to minimize the data misfit. BGO yields an estimate guaranteed to be the global minima over the search space, while MAP may converge toward local minima of the data misfit function. It can be seen that the MAP estimate and the posterior mean estimate are similar, although in general they need not coincide, as they correspond to different statistics. We note that, unlike MAP estimation and Bayesian global optimization, the HMC method is not limited to point estimates and can be used to quantify the uncertainty in the reconstruction. Nevertheless, MAP and Bayesian global optimization estimates are useful when quantifying uncertainty is not critical as their computational cost is smaller than that of HMC.

Figure 2B shows the posterior standard deviation. It can be seen that the posterior standard deviation is large, which is due to the dearth of data available and the ill-posedness of the inversion problem. We also note that the posterior standard deviation is the largest for $x = L$. Due to the strong advective velocity together with the outflow boundary condition, $c_0(L)$ is only informed by the observations at $x = L$ at early times.

4.2. Application of gHMC to Linear Transport

In order to study the construction of an acceleration matrix A appropriate for contaminant transport, we consider a 1-D advection-dispersion (no reaction) problem

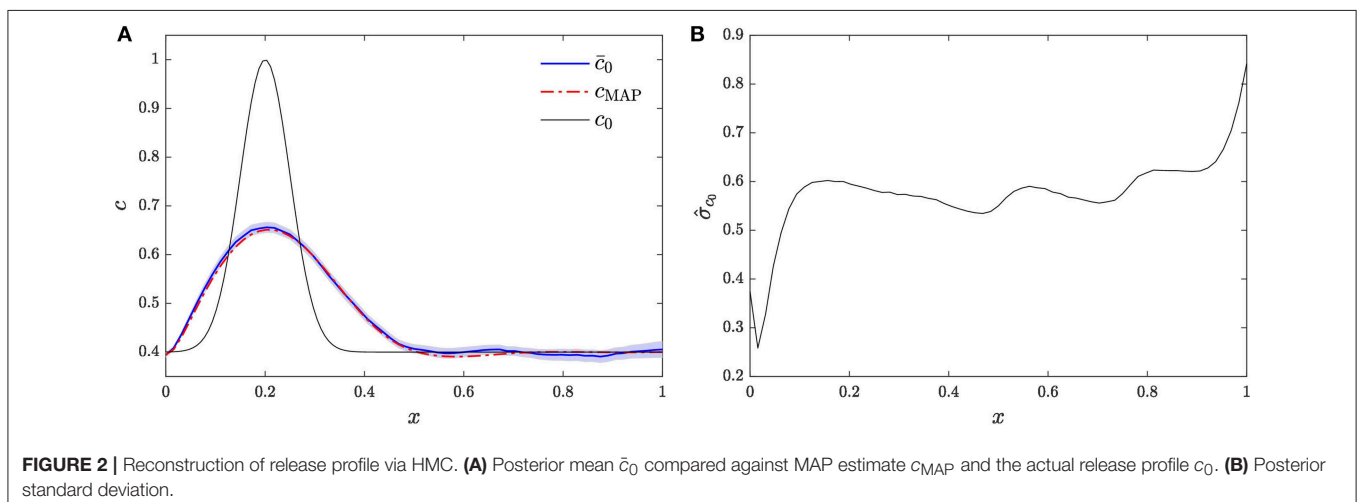
$$\frac{\partial c}{\partial t} + u \frac{\partial c}{\partial x} = D \frac{\partial^2 c}{\partial x^2}, \quad x \in [0, 2\pi], \quad t = (0, T], \quad (16)$$

with uniform coefficients u and D . This equation is subject to the periodic boundary condition

$$c(0, t) = c(2\pi, t), \quad (17)$$

and (unknown) initial condition

$$c(x, 0) = c_0(x). \quad (18)$$



Similar to the case studied in section 4.1, available concentration data consist of a set of measurements continuous in time on the interval $(0, T)$ collected at a subset J of the discrete locations x_j , $c_{\text{obs},j} = c_{\text{obs}}(x_j, t)$. The measurements are subject to space-time uncorrelated additive errors of equal variance σ_ϵ^2 .

4.2.1. Observation Hamiltonian

The state variable $c(x, t)$ is discretized into N functions $c_j(t) = c(x_j, t)$, where $x_j = 2\pi j/N$, $j = 0, \dots, N-1$ are N equidistant nodes along the domain $[0, 2\pi)$. We define the measurement Hamiltonian as

$$H_{\text{obs}} = \frac{1}{2\sigma_\epsilon^2} \sum_{j \in J} \int_0^T [c_j(t) - c_{\text{obs},j}(t)]^2 dt, \quad (19)$$

which defines the likelihood of the measurements given an initial release vector \mathbf{q} with components $q_j = c_{0,j}$. The solution to (16)–(18) can be represented in terms of its discrete Fourier transform (DFT)

$$\hat{c}_k = \frac{1}{N} \sum_{j=0}^{N-1} c_j e^{-ikx_j}, \quad k = -N/2, \dots, N/2 - 1, \quad (20)$$

which defines the N Fourier modes \hat{c}_k . The backward or inverse transform is given by

$$c_j = \sum_{k=-N/2}^{N/2-1} \hat{c}_k e^{ikx_j}, \quad j = 0, \dots, N-1. \quad (21)$$

Let \mathbf{c} denote a vector of discrete values c_j and $\hat{\mathbf{c}}$ denote its DFT. Then, (20) and (21) can be rewritten as

$$\hat{\mathbf{c}} = \frac{1}{N} \mathcal{F} \mathbf{c}, \quad \mathbf{c} = \mathcal{F}^* \hat{\mathbf{c}}, \quad (22)$$

where \mathcal{F} is the DFT matrix whose elements are

$$\mathcal{F}_{pq} = \omega^{(p-N/2)q}, \quad \omega = e^{-2\pi i/N} \quad (23)$$

and $(\cdot)^*$ denotes the Hermitian adjoint. By projection, (16) is discretized into the set of uncoupled ODEs for the Fourier modes

$$\frac{\partial \hat{c}_k(t)}{\partial t} = -(Dk^2 + iku)\hat{c}_k, \quad k = -N/2, \dots, N/2 - 1,$$

with initial conditions $\hat{c}_k(0) = \hat{q}_k$, where the \hat{q}_k , $k = -N/2, \dots, N/2 - 1$ are the components of $\hat{\mathbf{q}}$, the DFT of $\mathbf{q} \equiv \mathbf{c}(0)$. The solution to the ODEs is then given by

$$\hat{c}_k = \hat{q}_k \exp\{-(Dk^2 + iku)t\}. \quad (24)$$

Substituting (22) and (24) into (19) yields the following expression for the measurement Hamiltonian:

$$\begin{aligned} H_{\text{obs}} &= \frac{1}{2\sigma_\epsilon^2} (\hat{\mathbf{q}} - \hat{\mathbf{q}}_{\text{obs}})^* \left(\int_0^T \mathbf{B}^* \mathcal{F}_J \mathcal{F}_J^* \mathbf{B} dt \right) (\hat{\mathbf{q}} - \hat{\mathbf{q}}_{\text{obs}}), \\ &= \frac{1}{2\sigma_\epsilon^2} (\hat{\mathbf{q}} - \hat{\mathbf{q}}_{\text{obs}})^* \hat{\mathbf{G}}_{\text{obs}} (\hat{\mathbf{q}} - \hat{\mathbf{q}}_{\text{obs}}), \end{aligned} \quad (25)$$

where $\hat{\mathbf{q}}_{\text{obs}}$ is the DFT of $\mathbf{c}_{\text{obs}}(0)$, \mathcal{F}_J corresponds to the J columns of \mathcal{F} , $\mathbf{B}(t)$ is a diagonal matrix with elements $B_{kk} = \exp[-(Dk^2 + iku)t]$, and $\hat{\mathbf{G}}_{\text{obs}}$ is a Hermitian (semi)positive definite matrix. The measurement Hamiltonian specifies a multivariate normal distribution for $\hat{\mathbf{q}}$, and given that \mathbf{q} and $\hat{\mathbf{q}}$ are related via a linear transformation, it follows that the measurement Hamiltonian specifies a multivariate normal distribution for \mathbf{q} .

If the measurements are available at every node of the computational domain, i.e., if $\mathcal{F}_J = \mathcal{F}$, then $\mathcal{F}\mathcal{F}^* = N\mathbf{I}$ and (25) simplifies to

$$H_{\text{obs}} = \frac{N}{2\sigma_\epsilon^2} (\hat{\mathbf{q}} - \hat{\mathbf{q}}_{\text{obs}})^* \left(\int_0^T \mathbf{B}^* \mathbf{B} dt \right) (\hat{\mathbf{q}} - \hat{\mathbf{q}}_{\text{obs}}),$$

which is equivalent to

$$H_{\text{obs}} = \frac{N}{2\sigma_\epsilon^2} \sum_{k=-N/2}^{N/2-1} |\hat{q}_k - \hat{q}_{\text{obs},k}|^2 \hat{g}_k \quad (26)$$

where the coefficients \hat{g}_k are given by

$$\begin{aligned} \hat{g}_k &= \int_0^T |\exp\{-(Dk^2 + iku)t\}|^2 dt \\ &= \int_0^T e^{-2Dk^2 t} dt = \frac{1 - \exp(-2Dk^2 T)}{2Dk^2}. \end{aligned}$$

Note that all coefficients \hat{g}_k are real, symmetric ($\hat{g}_k = \hat{g}_{-k}$), and depend only on the dispersion coefficient D .

It follows from (25) that

$$\begin{aligned} \nabla_{\mathbf{q}} H_{\text{obs}} &= \frac{1}{\sigma_\epsilon^2} \mathcal{F}^* \hat{\mathbf{G}}_{\text{obs}} (\hat{\mathbf{q}} - \hat{\mathbf{q}}_{\text{obs}}) = \mathcal{F}^* \hat{\mathbf{G}}_{\text{obs}} \mathcal{F} (\mathbf{q} - \mathbf{q}_{\text{obs}}) \\ &= \mathbf{G}_{\text{obs}} (\mathbf{q} - \mathbf{q}_{\text{obs}}) \end{aligned}$$

where $\mathbf{G}_{\text{obs}} = \mathcal{F}^* \hat{\mathbf{G}}_{\text{obs}} \mathcal{F}$. That brings the forcing into the form $\mathbf{F}[\mathbf{q}] = -\Sigma^{-1}(\mathbf{q} - \boldsymbol{\mu})$ required by our analysis in section 3.2, which suggests a possibility of computing the acceleration matrices as $\mathbf{A}\mathbf{A}^* = \mathbf{G}_{\text{obs}}^{-1}$. Unfortunately this is not generally feasible, because the matrix \mathbf{G}_{obs} is singular unless the measurements are taken at every node of the domain. The singularity of \mathbf{G}_{obs} implies the singularity of the multivariate normal distribution of $\hat{\mathbf{q}}$ given by H_{obs} , which means the distribution is concentrated in a r -dimensional subspace of \mathbb{C}^N , $r < N$. Since \mathbf{q} results from a linear transformation of $\hat{\mathbf{q}}$, the multivariate normal distribution of \mathbf{q} is also degenerate. This implies that there is a linear subspace of configurations \mathbf{q} for which H_{obs} does not assign a probability, and therefore cannot be identified.

An empirical study of the SVD decomposition $\hat{\mathbf{G}}_{\text{obs}} = \mathbf{U}\mathbf{S}\mathbf{V}^*$ of the matrix $\hat{\mathbf{G}}_{\text{obs}}$ computed for the example of section 4.2.5 provides some insight into features of the degenerate distribution of \mathbf{q} defined by the observations. Specifically, the vectors forming a basis for $\ker \hat{\mathbf{G}}_{\text{obs}}$ have negligible terms associated with the lower Fourier modes of \mathbf{q} , i.e., $|V_{kj}| \approx 0$ for small $|k|$ and $\text{rank } \hat{\mathbf{G}}_{\text{obs}} < j$. This implies that the lower frequency components

of \mathbf{q} fall mostly on the subspace of identifiable configurations. In general, $|V_{k,j}| \neq 0$ for high $|k|$ and $\text{rank } \hat{\mathbf{G}}_{\text{obs}} < j$, which implies that in general high frequency features cannot be identified.

4.2.2. Regularization Hamiltonian

The regularization Hamiltonian extends the distributions of \mathbf{q} and $\hat{\mathbf{q}}$ in order to make them well-defined. After a real-space discretization, and accounting for the periodic boundary conditions (17), the ℓ_2 -norm regularization Hamiltonian takes the form

$$H_{\text{reg}} = \gamma \mathbf{q}^\top \mathbf{G}_{\text{reg}} \mathbf{q}, \quad (27)$$

where γ is a regularization hyperparameter and \mathbf{G}_{reg} is the circulant matrix

$$\mathbf{G}_{\text{reg}} = \frac{1}{\Delta x} \begin{bmatrix} 2 & -1 & & -1 \\ -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & 2 & -1 \\ -1 & & & -1 & 2 \end{bmatrix},$$

or $\mathbf{G}_{\text{reg}} = \text{circ}\{\mathbf{r}\}^\top$, where $\mathbf{r} = (2, -1, 0, \dots, 0, -1)^\top / \Delta x$ and $\Delta x = 2\pi/N$. The matrix \mathbf{G}_{reg} extends the probability distribution by assigning a high energy (low probability) to configurations with large high-frequency components. To demonstrate this, we rewrite \mathbf{G}_{reg} as

$$\mathbf{G}_{\text{reg}} = \text{circ}\{\mathbf{r}\}^\top = \mathcal{F}^* \text{diag}\{\hat{\mathbf{r}}\} \mathcal{F}, \quad \hat{r}_k = \frac{1}{\pi} \left[1 - \cos\left(\frac{2\pi k}{N}\right) \right],$$

where $\hat{\mathbf{r}}$ is the DFT of \mathbf{r} . The components \hat{r}_k of vector $\hat{\mathbf{r}}$ increase with frequency k , with the zeroth frequency giving rise to $\hat{r}_0 = 0$. The latter is to be expected since the regularization operator does not affect the observability of the zeroth frequency, which corresponds to the average of the initial release.

Note that a Fourier-space discretization of the regularization Hamiltonian leads to a similar bilinear form for \mathbf{G}_{reg} , with $\hat{r}_k = 2\pi k^2/N^2$. Indeed, these \hat{r}_k have a similar asymptotic behavior as $k \rightarrow 0$.

4.2.3. Acceleration Matrix

For the full Hamiltonian $H = H_{\text{obs}} + H_{\text{reg}}$, the forcing is given by

$$\mathbf{F}[\mathbf{q}] = -\hat{\mathbf{G}}_{\text{obs}}(\hat{\mathbf{q}} - \hat{\mathbf{q}}_{\text{obs}}) - \hat{\mathbf{G}}_{\text{reg}}\hat{\mathbf{q}} = -\mathbf{G}_{\text{obs}}(\mathbf{q} - \mathbf{q}_{\text{obs}}) - \mathbf{G}_{\text{reg}}\mathbf{q} \quad (28)$$

where $\mathbf{G}_{\text{obs}} = \mathcal{F}^* \hat{\mathbf{G}}_{\text{obs}} \mathcal{F}$ and $\mathbf{G}_{\text{reg}} = \mathcal{F}^* \hat{\mathbf{G}}_{\text{reg}} \mathcal{F}$. This suggests that choosing the acceleration matrix \mathbf{A} , such that $\mathbf{A}\mathbf{A}^*(\mathbf{G}_{\text{obs}} + \mathbf{G}_{\text{reg}}) = \mathbf{I}$, would reduce the correlation of the Markov chain. Since \mathbf{G}_{obs} and \mathbf{G}_{reg} are Hermitian (semi)positive definite, their sum $\mathbf{G} = \mathbf{G}_{\text{obs}} + \mathbf{G}_{\text{reg}}$ is at least Hermitian (semi)positive definite. In fact, \mathbf{G} is a full rank matrix and thus can be factorized via a Cholesky decomposition $\mathbf{G} = \mathbf{S}^\top \mathbf{S}$. The matrix \mathbf{A} , defined by $\mathbf{A}\mathbf{A}^*\mathbf{G} = \mathbf{I}$, is then given by

$$\mathbf{A} = \mathbf{A}_1 = \mathbf{S}^{-1}, \quad \mathbf{S}^\top \mathbf{S} = \mathbf{G}. \quad (29)$$

The added cost of computing the acceleration matrix \mathbf{A} is the Cholesky factorization cost, and the leapfrog scheme for gHMC

incurs four matrix-vector products. For dense matrices, these costs are $\mathcal{O}(N^3)$ and $\mathcal{O}(N^2)$, respectively.

An advantage of the Cholesky factorization is that the vector of momenta \mathbf{p} in (13) can be chosen as real and multivariate normal, with zero mean and identity covariance matrix. A drawback is its relatively high cost per step in the Markov chain. Moreover, the matrix \mathbf{G} becomes more poorly conditioned as $\gamma \rightarrow 0$, which affects the stability of the Cholesky decomposition.

A less computationally costly alternative for the construction of \mathbf{A} is to employ the following heuristic: Instead of using the full correlation matrices in Fourier space, $\hat{\mathbf{G}}_{\text{obs}}$ and $\hat{\mathbf{G}}_{\text{reg}}$, to define \mathbf{G} , we approximate it as $\mathbf{G} \approx \mathcal{F}^* \text{diag}\{\bar{\mathbf{g}}\} \mathcal{F}$, where $\bar{\mathbf{g}}$ is the vector with components $\bar{g}_i = \{\hat{\mathbf{G}}_{\text{obs}} + \hat{\mathbf{G}}_{\text{reg}}\}_{ii}$. This approximation allows one to factorize \mathbf{G} as $\mathbf{G} \approx \bar{\mathbf{G}} = \mathcal{F}^* \mathbf{D} \mathcal{F}$, where $\mathbf{D} = \text{diag}\{(\bar{\mathbf{g}})^{1/2}\}$ with the square root understood as element-wise. This argument suggests that the acceleration matrix \mathbf{A} can be constructed as

$$\begin{aligned} \mathbf{A} &= \mathbf{A}_2 = \frac{1}{N} \mathcal{F}^* \mathbf{D}^{-1}, \quad \mathbf{D} = \text{diag}\{(\bar{\mathbf{g}})^{1/2}\}, \\ \bar{\mathbf{g}} &= \text{diag}\{\hat{\mathbf{G}}_{\text{obs}} + \hat{\mathbf{G}}_{\text{reg}}\}, \end{aligned} \quad (30)$$

which gives

$$\mathbf{A}\mathbf{p} = \frac{1}{N} (\mathcal{F}^* \mathbf{D}^{-1} \mathbf{p}), \quad \mathbf{A}^* \mathbf{F} = \mathbf{D}^{-1} \left(\frac{1}{N} \mathcal{F} \mathbf{F} \right).$$

Note that we have replaced the transpose of \mathbf{A} with its Hermitian transpose due to the complex nature of the DFT. This implies that the transpose in (12) and (13) must be replaced with a Hermitian transpose, and that in order to guarantee that $\mathbf{q} \in \mathbb{R}^N$ we must generalize the momenta such that $\mathbf{p} \in \mathbb{C}^N$. Once the acceleration matrix \mathbf{A} in (30) is constructed, products of the form $\mathbf{A}\mathbf{p}$ and $\mathbf{A}^* \mathbf{F}$ can be computed using DFTs. The computational cost per leapfrog step is reduced from four matrix-vector products of cost $\sim \mathcal{O}(N^2)$ to four of cost $\sim \mathcal{O}(N \log N)$, and no Cholesky decomposition is necessary.

Since $\bar{\mathbf{g}} = \partial^2 H / \partial \hat{\mathbf{q}}^2$, the approximation (30) can be thought as building \mathbf{A} from the diagonal of the Hessian of H with respect to $\hat{\mathbf{q}}$ (a similar heuristic is employed in Neal, 1995 for Bayesian learning). This observation begs the following question: Why do we take $\bar{\mathbf{g}} = \partial^2 H / \partial \hat{\mathbf{q}}^2$ instead of $\bar{\mathbf{g}} = \partial^2 H / \partial \mathbf{q}^2$, which would produce a similar acceleration matrix \mathbf{A} without the Fourier transforms? The answer is that the matrix $\hat{\mathbf{G}}_{\text{obs}} + \hat{\mathbf{G}}_{\text{reg}}$ is more concentrated along its diagonal than \mathbf{G} is. Hence, more information about the observation operator is conserved by taking the diagonal of $\hat{\mathbf{G}}_{\text{obs}} + \hat{\mathbf{G}}_{\text{reg}}$ than the diagonal of \mathbf{G} .

4.2.4. Sampling of Momentum Vector

In order to retain the validity of the leapfrog method with generalized momenta, we require said momenta to be associated with a kinetic energy following a bilinear form. We achieve this by assuming that $\mathbf{p} \in \mathbb{C}^N$ has a general complex normal distribution $\mathcal{CN}(0, \mathbf{\Gamma}, \mathbf{C})$ with unit covariance $\mathbf{\Gamma}$ and

$$\mathbf{C} = \mathbf{A}^* \bar{\mathbf{G}} (\text{conj } \mathbf{A}) = \frac{1}{N} \mathbf{D} \mathcal{F} \mathcal{F}^\top \mathbf{D}^{-1}. \quad (31)$$

For \mathcal{F}_{pq} given by (23),

$$\begin{aligned} \frac{1}{N}(\mathcal{F}\mathcal{F}^\top)_{pq} &= \sum_{r=0}^{N-1} \omega^{r(p+q-N)} = P_{pq} \\ &= \begin{cases} 1 & \text{if } p+q = kN, k = \dots, -1, 0, 1, \dots, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

where P_{pq} are the components of the permutation matrix \mathbf{P} . Since \mathbf{D} is diagonal and \mathbf{P} is a permutation matrix (23), yields $\mathbf{C} = \mathbf{P}$.

Let $\mathbf{p} = \mathbf{X} + i\mathbf{Y}$ with $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^N$. The vector $\mathbf{V}^\top = [\mathbf{X}^\top \ \mathbf{Y}^\top]$ is multivariate normal with zero mean. Given $\mathbf{\Gamma}$ and \mathbf{C} , the cross-covariance matrix of this vector is

$$\mathbb{E}[\mathbf{X}\mathbf{Y}^\top] = \frac{1}{2} \text{Im}\{\mathbf{\Gamma} + \mathbf{C}\} = 0, \quad \mathbb{E}[\mathbf{Y}\mathbf{X}^\top] = \frac{1}{2} \text{Im}\{-\mathbf{\Gamma} + \mathbf{C}\} = 0,$$

since both $\mathbf{\Gamma}$ and \mathbf{C} are real. In other words, the real and imaginary parts of \mathbf{p} are mutually uncorrelated. The covariance

matrix of this vector is

$$\mathbb{E}[X_i X_j] = \frac{1}{2} \text{Re}\{\Gamma_{ij} + C_{ij}\} = \begin{cases} 1 & \text{if } i = j = 0, -N/2 \\ 1/2 & \text{if } i = -j, i, j \neq 0 \\ 0 & \text{otherwise,} \end{cases} \quad (32a)$$

$$\mathbb{E}[Y_i Y_j] = \frac{1}{2} \text{Re}\{\Gamma_{ij} - C_{ij}\} = \begin{cases} 1/2 & \text{if } i = -j, i, j \neq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (32b)$$

It follows from (32) that only the components $p_k = X_k + iY_k$ with $k = -k$ are correlated. Their covariances are $\mathbb{E}[X_k X_{-k}] = 0.5$, $\mathbb{E}[Y_k Y_{-k}] = -0.5$. Since \mathbf{p} must be complex-symmetric to guarantee that \mathbf{q} remains real, we generate \mathbf{p} as

$$\begin{aligned} X_{-N/2} &\sim \mathcal{N}(0, 1), \\ X_{-N/2+1} &\sim \mathcal{N}(0, 1/2), & Y_{-N/2+1} &\sim \mathcal{N}(0, 1/2) \\ &\vdots \\ X_0 &\sim \mathcal{N}(0, 1), \\ X_1 = X_{-1}, & & Y_1 = -Y_{-1}, \\ &\vdots \\ X_{N/2-1} = X_{-N/2+1}, & & Y_{N/2-1} = -Y_{-N/2+1}. \end{aligned}$$

Hence the vector \mathbf{p} is generated with N independent identically distributed normal random variables.

4.2.5. Computational Example

We apply the gHMC algorithm to the model problem (16) with parameters $D = 1.0$, $u = 10.0$, $N = 64$, and $\sigma_\epsilon = 0.02$. Measurements are taken at locations x_j , $j \in J = \{47, 63\}$ over the time period $(0, 6 \times 10^{-1})$. The measurements are shown in **Figure 3**. To infer the release profile from the observations, we employ $\mathbf{A} = \mathbf{A}_1$, $S = 5$ leapfrog steps, regularization parameter $\gamma = 1 \times 10^{-2}$, and hybrid timestep tuned to achieve a rejection rate of 30–35%. A total of 10 chains were generated, each with 2×10^4 samples retained after burn-in. These samples are employed to compute Monte Carlo estimates of the posterior mean and standard deviation, shown in **Figure 4**.

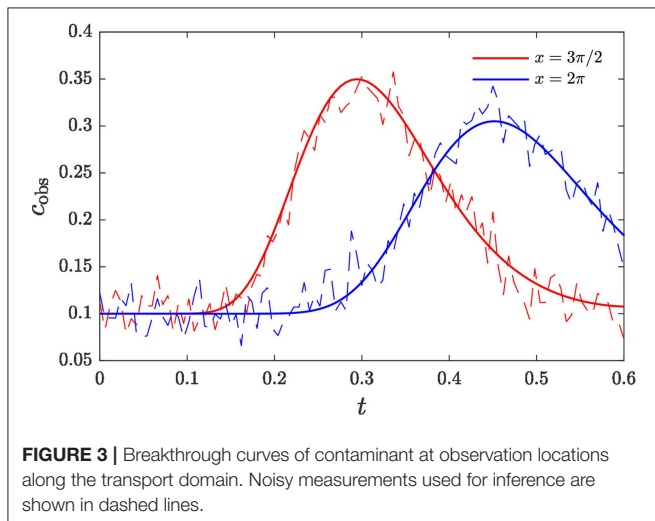


FIGURE 3 | Breakthrough curves of contaminant at observation locations along the transport domain. Noisy measurements used for inference are shown in dashed lines.

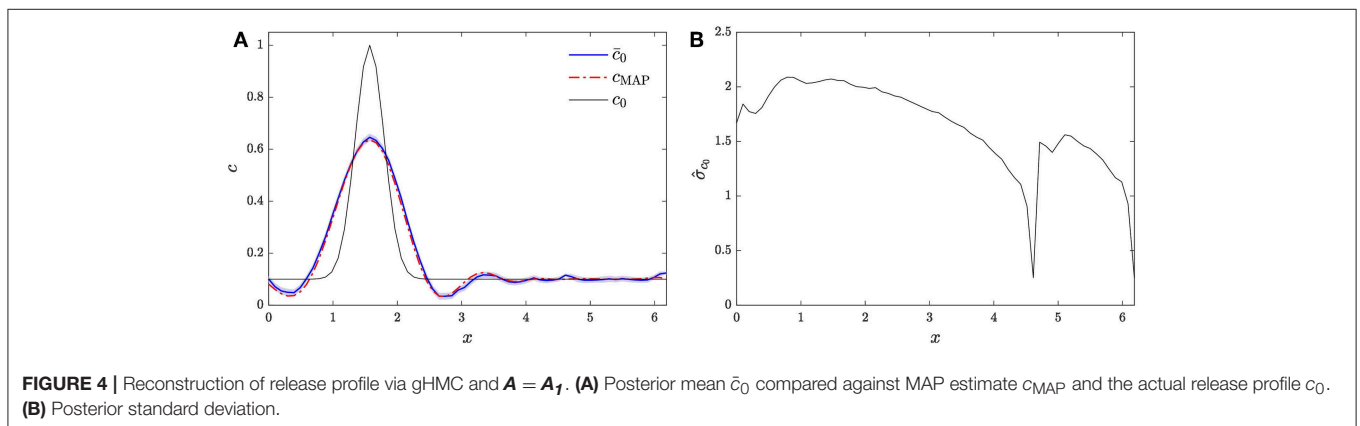


FIGURE 4 | Reconstruction of release profile via gHMC and $\mathbf{A} = \mathbf{A}_1$. **(A)** Posterior mean \bar{c}_0 compared against MAP estimate c_{MAP} and the actual release profile c_0 . **(B)** Posterior standard deviation.

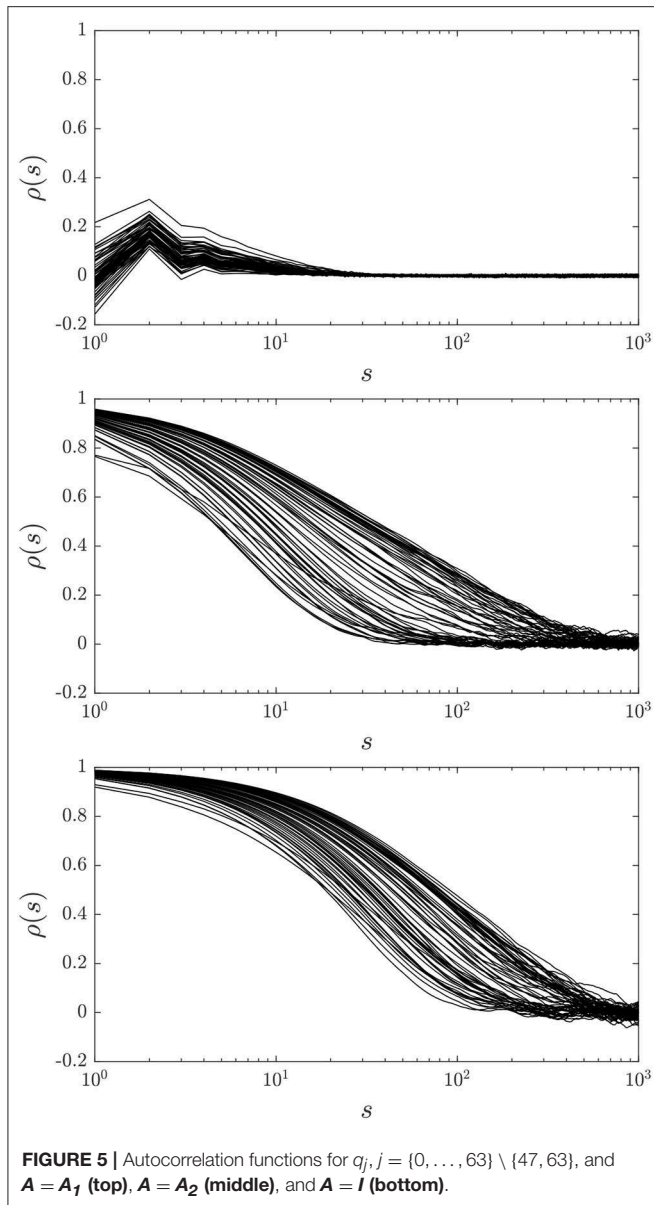


FIGURE 5 | Autocorrelation functions for $q_j, j = \{0, \dots, 63\} \setminus \{47, 63\}$, and $\mathbf{A} = \mathbf{A}_1$ (top), $\mathbf{A} = \mathbf{A}_2$ (middle), and $\mathbf{A} = \mathbf{I}$ (bottom).

Figure 4A compares the posterior mean estimate of the release profile, together with its 95% confidence interval, against the MAP estimate and the true release profile. It can be seen that the gHMC scheme is able to infer the main feature of the release profile. The gHMC estimate also compares favorably to the MAP estimate. **Figure 4B** shows the posterior standard deviation, which, as in section 4.1, is largely due to the relatively small number of observation locations, the relatively high measurement error, and the ill-posedness of the inverse problem.

Next, we study the effect of the choice of acceleration matrix \mathbf{A} on the Markov chains produced by the gHMC algorithm and the Monte Carlo standard error of the posterior mean estimate. Three alternatives for \mathbf{A} are considered: $\mathbf{A} = \mathbf{A}_1$ in (29), $\mathbf{A} = \mathbf{A}_2$ in (30), and $\mathbf{A} = \mathbf{I}$ (no acceleration). The Monte Carlo standard error is given by $\hat{\sigma}/\sqrt{n_{\text{eff}}}$, where $\hat{\sigma}$ is the posterior standard

TABLE 1 | Effective sample size reduction ratio η for various choices of regularization parameter γ and acceleration matrix \mathbf{A} .

γ	$\mathbf{A} = \mathbf{A}_1$	$\mathbf{A} = \mathbf{A}_2$	$\mathbf{A} = \mathbf{I}$
1×10^{-2}	3.4	20.7	22.3
1×10^{-3}	4.1	102.8	161.2

deviation of the inferred parameter, and n_{eff} denotes the MCMC effective sample size, given by $n_{\text{eff}} \equiv n/\eta$ (Kass et al., 1998), where n is the number of MCMC samples, and $\eta > 1$ is a reduction factor due to the correlation between MCMC samples. This reduction factor is given by

$$\eta = 1 + 2 \sum_{s=1}^{\infty} \rho(s),$$

where $\rho(s)$ is the autocorrelation of the MCMC chain at lag s . We note that, after convergence, gHMC chains for different choices of \mathbf{A} converge to the same posterior mean and standard deviation. The difference of performance between choices of \mathbf{A} is in terms of the reduction factor η : Better-performing choices of acceleration matrix result in smaller values of η , so that fewer samples are necessary to achieve a certain target error in the estimation of the posterior mean.

We compute 10 gHMC chains for each choice of \mathbf{A} and for two choices of regularization parameter γ , 1×10^{-2} and 1×10^{-3} , and employ the samples to compute the autocorrelations $\rho(s)$ and the effective sample size reduction ratios for each q_j , $j = 0, \dots, N-1$, except for $j = 47, 63$, which are included in the observations. **Figure 5** presents the autocorrelations for $\gamma = 1 \times 10^{-3}$ and each choice of \mathbf{A} . It can be seen that $\mathbf{A} = \mathbf{A}_1$ produces highly uncorrelated chains for each of the q_j studied, $\mathbf{A} = \mathbf{A}_2$ produces more correlated chains, and $\mathbf{A} = \mathbf{I}$ produces the most correlated chains. As expected, \mathbf{A}_2 provides a compromise between the low autocorrelation / high expense of the full Cholesky decomposition and the high autocorrelation / low cost of $\mathbf{A} = \mathbf{I}$.

The maximum effective sample size reduction ratio η for each choice of \mathbf{A} and γ is shown in **Table 1**. It can be seen that, consistent with **Figure 5**, $\mathbf{A} = \mathbf{A}_1$ produces highly uncorrelated chains, which leads to low values of η , and therefore to smaller standard errors. Similarly, $\mathbf{A} = \mathbf{I}$ leads to the most correlated chains, the highest values of η , and the highest standard errors, while the choice $\mathbf{A} = \mathbf{A}_2$ leads to values of η between those of $\mathbf{A} = \mathbf{A}_1$ and $\mathbf{A} = \mathbf{I}$. We also note that the values of η for $\mathbf{A} = \mathbf{A}_2$ and $\mathbf{A} = \mathbf{I}$ increase significantly by going from $\gamma = 1 \times 10^{-2}$ to $\gamma = 1 \times 10^{-3}$, which is to be expected as the latter is a more challenging case due to weaker regularization. On the other hand, the value of η for $\mathbf{A} = \mathbf{A}_1$ does not increase as dramatically when going from $\gamma = 1 \times 10^{-2}$ to $\gamma = 1 \times 10^{-3}$, which indicates that \mathbf{A}_1 is the best choice of acceleration matrix despite its higher computational cost per leapfrog step. We conclude that the gHMC scheme leads to a significant reduction in the estimation error over the use of HMC without an acceleration matrix.

For problems with reaction terms, the forcing $\mathbf{F} = -\nabla_{\mathbf{q}} H$ is not a linear function of \mathbf{q} as in (28). In such cases, the

selection of the acceleration matrix \mathbf{A} is not straightforward. The challenge is to find an approximation to the forcing that is linear in \mathbf{q} , i.e., preserves the form (28) with \mathbf{G}_{obs} and \mathbf{G}_{reg} independent of \mathbf{q} . This is required to guarantee the reversibility of the Hamiltonian dynamics.

Such an approximation can be obtained by disregarding the nonlinear reaction term and using \mathbf{G}_{obs} in (25) and \mathbf{G}_{reg} in (27), which are functions only of the temporal and spatial domain properties and the hyperparameters σ_ϵ^2 and γ . This selection is equivalent to taking $\mathbf{F} \approx -\mathbf{G}_{\text{lin}}(\mathbf{q} - \mathbf{q}_{\text{obs}})$, where \mathbf{G}_{lin} is the Hessian of the advection-diffusion (linear) portion of the Hamiltonian. It gives the acceleration matrix \mathbf{A}_1 of (29). This choice is justified for non-periodic boundary conditions if the contaminant plume does not reach the domain's boundaries during the simulation time. An alternative is to take only the diagonal portion of the Hessian of an advection-diffusion portion of the Hamiltonian. This would produce the acceleration matrix \mathbf{A}_2 in (30).

5. CONCLUSIONS AND FURTHER WORK

We presented a computationally efficient and accurate algorithm for identification of sources and release histories of (geo)chemically active solutes. The algorithm is based on a generalized hybrid Monte Carlo approach, in which MC sampling is accelerated by the use of discrete adjoint equations. Some of the salient features of our approach are: (1) its ability to handle nonlinear systems, since it requires no linearizations, and (2) its compatibility with various regularization strategies.

The introduction of an acceleration matrix to the gHMC scheme was tested for an advection-dispersion problem. While the example presented was limited to one-dimensional domains, periodic boundary conditions, and homogeneous porous media, our analysis demonstrated that the proposed acceleration matrices improve upon basic HMC; therefore, we consider the proposed acceleration strategy to be promising. The generalization of these constructions to problems with nonlinear reaction terms, two- and three-dimensional, heterogeneous media, and non-periodic boundary conditions, will be the subject of future work.

Finally, we note the importance of considering the heterogeneity of flow and transport parameters, such as the

hydraulic conductivity and dispersion coefficient tensors, for source identification tasks (Xu and Gómez-Hernández, 2018). Attempting to perform Bayesian inference when the values of these coefficients are assumed to be known but their values are erroneous may lead to model misspecification and consequently to posterior densities with little predictive value. Fortunately, the HMC and gHMC schemes presented in this work can accommodate the simultaneous identification of heterogeneous coefficients together with the release history by extending the state vector \mathbf{q} to include the discretized heterogeneous coefficients. The calculation of the gradient of the data misfit with respect to the extended state vector can be accomplished via discrete adjoint sensitivity analysis (Zhang et al., 2017) for complex dynamical systems. The extension of the presented framework to the identification of heterogeneous parameters of geophysical models will be considered in future work.

AUTHOR CONTRIBUTIONS

DB-S implemented the presented algorithms in MATLAB. All authors contributed equally to the preparation of this manuscript.

FUNDING

This work was supported in part by the Applied Mathematics Program within the U.S. Department of Energy Office of Advanced Scientific Computing Research Mathematical Multifaceted Integrated Capability Centers (MMICCS) project—AEOLUS—Advances in Experimental Design, Optimal Control, and Learning for Uncertain Complex Systems under award number DE-SC0019393, by the U.S. National Science Foundation under award number DMS-1802189, and by the U.S. Department of Energy Office of Basic Energy Research under award number DE-SC0019130. Pacific Northwest National Laboratory is operated by Battelle for the DOE under Contract DE-AC05-76RL01830.

ACKNOWLEDGMENTS

We thank N. Gulbahce and T. Bewley for fruitful discussions.

REFERENCES

- Alexander, F. J., Eyink, G. L., and Restrepo, J. M. (2005). Accelerated Monte Carlo for optimal estimation of time series. *J. Stat. Phys.* 119, 1331–1345. doi: 10.1007/s10955-005-3770-1
- Archambeau, C., Oppen, M., Shen, Y., Cornford, D., and Shawe-Taylor, J. (2007). "Variational inference for diffusion processes," in *Advances in Neural Information Processing Systems*, Vol. 20, eds J. C. Platt, D. Koller, Y. Singer, and S. Roweis (Boston, MA: MIT Press), 1–8.
- Atmadja, J., and Bagtzoglou, A. C. (2001a). Pollution source identification in heterogeneous porous media. *Water Resour. Res.* 37, 2113–2125. doi: 10.1029/2001WR000223
- Atmadja, J., and Bagtzoglou, A. C. (2001b). State of the art report on mathematical methods for groundwater pollution source identification. *Environ. Forensics* 2, 205–214. doi: 10.1080/15275920127949

- Bagtzoglou, A. C., and Atmadja, J. (2003). Marching-jury backward beam equation and quasi-reversibility methods for hydrologic inversion: application to contaminant plume spatial distribution recovery. *Water Resour. Res.* 39. doi: 10.1029/2001WR001021
- Bagtzoglou, A. C., Dougherty, D. E., and Tompson, A. F. B. (1992). Application of particle methods to reliable identification of groundwater pollution sources. *Water Resour. Manage.* 6, 15–23. doi: 10.1007/BF00872184
- Butcher, J. B., and Gauthier, T. D. (1994). Estimation of residual dense NAPL mass by inverse modeling. *Ground Water* 32, 71–78. doi: 10.1111/j.1745-6584.1994.tb00613.x
- Cao, Y., Li, S., Petzold, L., and Serban, R. (2003). Adjoint sensitivity analysis for differential-algebraic equations: the adjoint DAE system and its numerical solution. *SIAM J. Sci. Comput.* 24, 1076–1089. doi: 10.1137/S1064827501380630

- Daescu, D., Carmichael, G. R., and Sandu, A. (2000). Adjoint implementation of rosenbrock methods applied to variational data assimilation problems. *J. Comput. Phys.* 165, 496–510. doi: 10.1006/jcph.2000.6622
- Gugat, M. (2012). Contamination source determination in water distribution networks. *SIAM J. Appl. Math.* 72, 1772–1791. doi: 10.1137/110859269
- Herrera, G. S., and Pinder, G. F. (2005). Space-time optimization of groundwater quality sampling networks. *Water Resour. Res.* 41. doi: 10.1029/2004WR003626
- Hutchinson, M., Eyink, G. L., and Restrepo, J. M. (2017). A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors. *Inf. Fusion* 36, 130–148. doi: 10.1016/j.inffus.2016.11.010
- Ito, K., and Jin, B. (2015). *Inverse Problems: Tikhonov Theory and Algorithms*, Vol. 22. Series in Applied Mathematics. Singapore: World Scientific.
- Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998). Markov chain Monte Carlo in practice: a roundtable discussion. *Am. Stat.* 52, 93–100. doi: 10.1080/00031305.1998.10480547
- Landau, D. P., and Binder, K. (2009). *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge, UK: Cambridge University Press.
- Li, S., and Petzold, L. (2004). Adjoint sensitivity analysis for time-dependent partial differential equations with adaptive mesh refinement. *J. Comput. Phys.* 198, 310–325. doi: 10.1016/j.jcp.2003.01.001
- Lichtner, P. C., and Tartakovsky, D. M. (2003). Upscaled effective rate constant for heterogeneous reactions. *Stochas. Environ. Res. Risk Assess.* 17, 419–429. doi: 10.1007/s00477-003-0163-3
- Liu, C., and Ball, W. P. (1999). Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware. *Water Resour. Res.* 35, 1975–1985. doi: 10.1029/1999WR900092
- Michalak, A. M., and Kitanidis, P. K. (2003). A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification. *Water Resour. Res.* 39. doi: 10.1029/2002WR001480
- Michalak, A. M., and Kitanidis, P. K. (2004a). Application of geostatistical inverse modeling to contaminant source identification at Dover AFB, Delaware. *J. Hydraul. Res.* 42, 9–18. doi: 10.1080/00221680409500042
- Michalak, A. M., and Kitanidis, P. K. (2004b). Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling. *Water Resour. Res.* 40. doi: 10.1029/2004WR003214
- Neal, R. M. (1993). *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Technical Report CRG-TR-93-1, University of Toronto.
- Neal, R. M. (1995). *Bayesian learning for neural networks* (Ph.D. thesis). University of Toronto, Toronto, ON, Canada.
- Neupauer, R. M., Borchers, B., and Wilson, J. L. (2000). Comparison of inverse methods for reconstructing the release history of a groundwater contamination source. *Water Resour. Res.* 36, 2469–2475. doi: 10.1029/2000WR900176
- Neupauer, R. M., and Wilson, J. L. (1999). Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant. *Water Resour. Res.* 35, 3389–3398. doi: 10.1029/1999WR900190
- Neupauer, R. M. and Wilson, J. L. (2003). Backward location and travel time probabilities for a decaying contaminant in an aquifer. *J. Contam. Hydrol.* 66, 39–58. doi: 10.1016/S0169-7722(03)00024-X
- Neupauer, R. M., and Wilson, J. L. (2004). Forward and backward location probabilities for sorbing solutes in groundwater. *Adv. Water Resour.* 27, 689–705. doi: 10.1016/j.advwatres.2004.05.003
- Neupauer, R. M., and Wilson, J. L. (2005). Backward probability model using multiple observations of contamination to identify groundwater contamination sources at the Massachusetts Military Reservation. *Water Resour. Res.* 41. doi: 10.1029/2003WR002974
- Pirot, G., Krityakierne, T., Ginsbourger, D., and Renard, P. (2019). Contaminant source localization via Bayesian global optimization. *Hydrol. Earth Sys. Sci.* 23, 351–369. doi: 10.5194/hess-23-351-2019
- Shlomi, S., and Michalak, A. M. (2007). A geostatistical framework for incorporating transport information in estimating the distribution of a groundwater contaminant plume. *Water Resour. Res.* 43. doi: 10.1029/2006WR005121
- Skaggs, T. H., and Kabala, Z. J. (1994). Recovering the release history of a groundwater contaminant. *Water Resour. Res.* 30, 71–79. doi: 10.1029/93WR02656
- Skaggs, T. H., and Kabala, Z. J. (1995). Recovering the history of a groundwater contaminant plume: method of quasi-reversibility. *Water Resour. Res.* 31, 2669–2673. doi: 10.1029/95WR02383
- Snodgrass, M. F., and Kitanidis, P. K. (1997). A geostatistical approach to contaminant source identification. *Water Resour. Res.* 33, 537–546. doi: 10.1029/96WR03753
- Stanev, V. G., Iliev, F. L., Hansen, S., Vesselinov, V. V., and Alexandrov, B. S. (2018). Identification of release sources in advection-diffusion system by machine learning combined with Green's function inverse method. *Appl. Math. Model.* 60, 64–76. doi: 10.1016/j.apm.2018.03.006
- Toral, R., and Ferreira, A. (1994). “A general class of hybrid Monte Carlo methods,” in *Proceedings of Physics Computing*, Vol. 94, eds R. Gruber and M. Tomasi (New York, NY: EPS), 265–268.
- Verwer, J. G., Spee, E. J., Blom, J. G., and Hundsdorfer, W. (1999). A second-order rosenbrock method applied to photochemical dispersion problems. *SIAM J. Sci. Comput.* 20, 1456–1480. doi: 10.1137/S1064827597326651
- Vesselinov, V. V., Alexandrov, B. S., and O'Malley, D. (2018). Contaminant source identification using semi-supervised machine learning. *J. Contam. Hydrol.* 212, 134–142. doi: 10.1016/j.jconhyd.2017.11.002
- Vesselinov, V. V., Alexandrov, B. S., and O'Malley, D. (2019). Nonnegative tensor factorization for contaminant source identification. *J. Contam. Hydrol.* 220, 66–97. doi: 10.1016/j.jconhyd.2018.11.010
- Wahba, G. (1990). *Spline Models for Observational Data*, Vol. 59. Philadelphia, PA: Society for Industrial Mathematics (SIAM).
- Woodbury, A. D., and Urych, T. J. (1996). Minimum relative entropy inversion: theory and application to recovering the release history of a groundwater contaminant. *Water Resour. Res.* 32, 2671–2681. doi: 10.1029/95WR03818
- Xu, T., and Gómez-Hernández, J. J. (2016). Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble kalman filtering. *Water Resour. Res.* 52, 6587–6595. doi: 10.1002/2016WR019111
- Xu, T. and Gómez-Hernández, J. J. (2018). Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble kalman filter. *Adv. Water Resour.* 112, 106–123. doi: 10.1016/j.advwatres.2017.12.011
- Zhang, H., Abhyankar, S., Constantinescu, E., and Anitescu, M. (2017). Discrete adjoint sensitivity analysis of hybrid dynamical systems with switching. *IEEE Trans. Circ. Syst. I* 64, 1247–1259. doi: 10.1109/TCSI.2017.2651683

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor is currently co-organizing a Research Topic with one of the authors DT, and confirms the absence of any other collaboration.

Copyright © 2019 Barajas-Solano, Alexander, Anghel and Tartakovsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Discrete Sensitivity Analysis

For the problem in section 4.1 we use the linearized Runge-Kutta (Rosenbrock) method ROS2 of Verwer et al. (1999) for time stepping of the forward ODE problem. The advantage of using this method is that it allows for a linear implicit treatment of the dispersion operator and a linearization of the reaction operator, while the advection operator is treated explicitly.

We assume that the advection-dispersion-reaction equation can be discretized into an autonomous system of ODEs

$$\mathbf{c}_t = \mathbf{f}(\mathbf{c}) = (\mathbf{A}_D + \mathbf{A}_A)\mathbf{c} - \mathbf{R}(\mathbf{c}),$$

where \mathbf{c} is the state vector, \mathbf{A}_D is the discretized linear dispersion operator, \mathbf{A}_A is the discretized linear advection operator, and $\mathbf{R}(\mathbf{c})$ is the reaction vector. Time stepping is performed via a scheme

$$\mathbf{c}^{n+1} = \mathbf{c}^n + (2 - b)\Delta t \mathbf{k}_1 + b\Delta t \mathbf{k}_2, \quad (\text{A1})$$

$$(\mathbf{I} - \theta \Delta t \mathbf{J})\mathbf{k}_1 = \mathbf{f}(\mathbf{c}^n), \quad (\text{A2})$$

$$(\mathbf{I} - \theta \Delta t \mathbf{J})\mathbf{k}_2 = \mathbf{f}\left(\mathbf{c}^n + \frac{1}{2b}\Delta t \mathbf{k}_1\right) - \frac{1}{b}\mathbf{k}_1, \quad (\text{A3})$$

where $\mathbf{J} = \mathbf{f}_c(\mathbf{c}^n)$ is the Jacobian of \mathbf{f} with respect to the state. The coefficients θ and b are taken for this application as $\theta = 1 - \sqrt{2}/2$ and $b = 1/2$, respectively. The left-hand side operators of (A2, A3) are approximated via approximate matrix factorization (AMF) to obtain the split form

$$(\mathbf{I} - \theta \Delta t \mathbf{J}) \approx (\mathbf{I} - \theta \Delta t \mathbf{A}_D)(\mathbf{I} + \theta \Delta t \mathbf{R}_c(\mathbf{c}^n)).$$

The discussion in section 4.1 led us to conclude that it is necessary to compute products of the form $(d\mathbf{c}^{i+1}/d\mathbf{c}^j)^\top \mathbf{u}$ in order to apply

the discrete sensitivity technique of Daescu et al. (2000). The formulae for the computation of these products are derived from the time-stepping scheme (A1-A3). In particular, differentiating (A1) with respect to the state and multiplying by a test vector \mathbf{u} gives the single-step sensitivity product as

$$\left(\frac{d\mathbf{c}^{n+1}}{d\mathbf{c}^n}\right)^\top \mathbf{u} = \mathbf{u} + \frac{3}{2}\Delta t \left(\frac{d\mathbf{k}_1^n}{d\mathbf{c}^n}\right)^\top \mathbf{u} + \frac{1}{2}\Delta t \left(\frac{d\mathbf{k}_2^n}{d\mathbf{c}^n}\right)^\top \mathbf{u}.$$

The next task is to derive formulas for the Jacobians of the stage derivatives \mathbf{k}_1 and \mathbf{k}_2 . Let \mathbf{M} be the AMF-ed left-hand-side matrix of (A2, A3). Differentiating (A2, A3) with respect to the state and multiplying by the test vector \mathbf{u} gives the formulae

$$\left(\frac{d\mathbf{k}_1^n}{d\mathbf{c}^n}\right)^\top \mathbf{u} = \left[\mathbf{J}_0 - \left(\frac{d\mathbf{M}}{d\mathbf{c}^n}\mathbf{k}_1^n\right)\right]^\top \mathbf{v}, \quad \mathbf{M}^\top \mathbf{v} = \mathbf{u}$$

and

$$\left(\frac{d\mathbf{k}_2^n}{d\mathbf{c}^n}\right)^\top \mathbf{u} = \mathbf{J}_1^\top \mathbf{v} + \left(\frac{d\mathbf{k}_1^{n+1}}{d\mathbf{c}^n}\right)^\top (\Delta t \mathbf{J}_1 - 2\mathbf{I}) \mathbf{v} - \left(\frac{d\mathbf{M}}{d\mathbf{c}^n}\mathbf{k}_2^n\right)^\top \mathbf{v},$$

with $\mathbf{J}_0 = \mathbf{f}_c(\mathbf{c}^n)$, $\mathbf{J}_1 = \mathbf{f}_c(\mathbf{c}^n + \Delta t \mathbf{k}_1^n)$.

The computation of the products $(d\mathbf{M}/d\mathbf{c}^n)\mathbf{k}_i^n$, $i = 1, 2$ is highly problem-specific. It depends on the structure of the second-order derivatives of the reaction vector with respect to the state. For the reaction model (15) and a method-of-lines discretization, the Jacobian \mathbf{R}_c is diagonal, and so the computation of these products is straightforward. For different reaction models and more sophisticated discretization schemes the computation might be more involved.



A Machine Learning Based Hybrid Multi-Fidelity Multi-Level Monte Carlo Method for Uncertainty Quantification

Nagoor Kani Jabarullah Khan* and Ahmed H. Elsheikh

School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, United Kingdom

OPEN ACCESS

Edited by:

Philippe Renard,
Université de Neuchâtel, Switzerland

Reviewed by:

Maruti Kumar Mudunuru,
Los Alamos National Laboratory
(DOE), United States
Niklas Linde,
Université de Lausanne, Switzerland

*Correspondence:

Nagoor Kani Jabarullah Khan
nj7@hw.ac.uk

Specialty section:

This article was submitted to
Freshwater Science,
a section of the journal
Frontiers in Environmental Science

Received: 27 November 2018

Accepted: 20 June 2019

Published: 27 August 2019

Citation:

Jabarullah Khan NK and Elsheikh AH
(2019) A Machine Learning Based
Hybrid Multi-Fidelity Multi-Level Monte
Carlo Method for Uncertainty
Quantification.
Front. Environ. Sci. 7:105.
doi: 10.3389/fenvs.2019.00105

This paper focuses on reducing the computational cost of the Monte Carlo method for uncertainty propagation. Recently, Multi-Fidelity Monte Carlo (MFMC) method (Ng, 2013; Peherstorfer et al., 2016) and Multi-Level Monte Carlo (MLMC) method (Müller et al., 2013; Giles, 2015) were introduced to reduce the computational cost of Monte Carlo method by making use of low-fidelity models that are cheap to an evaluation in addition to the high-fidelity models. In this paper, we use machine learning techniques to combine the features of both the MFMC method and the MLMC method into a single framework called Multi-Fidelity-Multi-Level Monte Carlo (MFML-MC) method. In MFML-MC method, we use a hierarchy of proper orthogonal decomposition (POD) based approximations of high-fidelity outputs to formulate a MLMC framework. Next, we utilize Gradient Boosted Tree Regressor (GBTR) to evolve the dynamics of POD based reduced order model (ROM) (Xiao et al., 2017) on every level of the MLMC framework. Finally, we incorporate MFMC method in order to exploit the POD ROM as a level specific low-fidelity model in the MFML-MC method. We compare the performance of MFML-MC method with the Monte Carlo method that uses either a high-fidelity model or a single low-fidelity model on two subsurface flow problems with random permeability field. Numerical results suggest that MFML-MC method provides an unbiased estimator with speedups by orders of magnitude in comparison to Monte Carlo method that uses high-fidelity model only.

Keywords: uncertainty quantification, POD, multi-fidelity Monte Carlo method, multi-level Monte Carlo method, machine learning

1. INTRODUCTION

Effective propagation of uncertainties through nonlinear dynamical systems has become an essential task for model based engineering applications (e.g., water resources management, petroleum reservoir management) (Elsheikh et al., 2013; Petvipusit et al., 2014; Kani and Elsheikh, 2018). There are many possible sources of uncertainties in the input of multi-phase porous media flow models such as material properties (e.g., permeability, and porosity), boundary conditions, and geometrical information of the simulated domain. In this work, we focus on the canonical problem of uncertainty propagation in subsurface flow models due to the stochastic model inputs mainly the spatially distributed hydraulic conductivity field. In this setting, the high-fidelity model outputs [quantities of interest (QoI)] are usually defined as a time series of transport variables at selected grid blocks (e.g., well locations) in the porous media domain. The propagation

of uncertainties through multi-phase porous media flow models remains challenging because of high dimensionality of input parameter space (e.g., heterogeneous permeability) and the non-polynomial model nonlinearities (Elsheikh et al., 2012, 2013). For this class of problems, probabilistic techniques, including stochastic Galerkin (Ghanem and Spanos, 1991; Stefanou, 2009), and stochastic collocation methods (Babuška et al., 2007; Doostan and Owhadi, 2011) have limited applicability despite they are computationally very effective for quasi-linear flow models with the small number of random variables (Li and Zhang, 2007; Lin and Tartakovsky, 2009).

One viable option to handle such situations is the Monte Carlo method (MC) where repeated evaluations of the high-fidelity flow models using different instantiations of the random input are performed. The output of these simulations is post-processed for estimates of the desired statistics such as the mean and the variance of the QoI. Generally, the estimators of the MC method are unbiased. However, since the accuracy of the MC method is measured in terms of the estimator variance (Giles, 2013), the convergence rate of MC estimators toward the desired statistics scales as \sqrt{N} , where N is the number of random samples. Given this slow convergence rate of MC methods, the MC method is computationally expensive since a large number of high fidelity simulation have to be performed to obtain a reasonably accurate statistical estimate for the QoI. One notable advantage of MC methods in-comparison to other techniques (Li and Zhang, 2007; Lin and Tartakovsky, 2009) is the ease of implementation using black-box simulators. Also, the rate of convergence is independent of the dimensionality of the random model inputs.

In this work, in order to make use of the aforementioned advantages of the MC method and to alleviate the slow convergence rate, we employ a variant of control variate method (Ng, 2013; Giles, 2015) called Multi-Level Monte Carlo method (Giles, 2013, 2015) which makes use of the correlation between the high-fidelity model output and a multi-level hierarchy of low-fidelity model outputs. The key aspect of MLMC method is the repartition of the computational cost between different hierarchical levels of models based on the number of samples required to decrease the variance at each level. More precisely, the MLMC method relies on the fact that increasing the number of samples reduces the variance at low levels and at high levels, the level variances are expected to be typically small and thus MLMC method incurs few expensive high-fidelity simulations (Giles, 2013; Müller et al., 2013).

Similar to MLMC method, Multi-Fidelity Monte Carlo method (Ng, 2013; Peherstorfer et al., 2016) is another control variate method which combines the outputs from an arbitrary number of low-fidelity models with the high-fidelity model in order to speedup the statistical estimation of the QoI. The key aspect of MFMC approach is the initial selection of low-fidelity models and the corresponding number of model runs for each model (Ng, 2013; Peherstorfer et al., 2016). Ng (2013) proposed a multifidelity approach to reduce the cost of expensive objective functions in stochastic optimization problems by making use of inexpensive, low-fidelity models. Peherstorfer et al. (2016) extended the MFMC method introduced in Ng (2013) to accelerate uncertainty quantification (UQ) tasks by making use

of many number of low-fidelity models. Furthermore, MFMC method introduced in Ng (2013) can utilize low-fidelity models of any type, for example, up-scaled models (Durlafsky and Chen, 2012), POD reduced order models (Berkooz et al., 1993; Antoulas et al., 2001; Lassila et al., 2014) and response surface based models (Frangos et al., 2010) could be combined in the MFMC framework. We refer the readers to read the paper by Peherstorfer et al. (2018) for a complete review of MFMC method.

We now present a brief literature review of MLMC method as applied to uncertainty quantification (UQ) tasks. It appears Heinrich (2001) was the one to first apply MLMC in the context of parametric integration. Kebaier (2005) then used similar ideas for a two-level Monte Carlo method to approximate weak solutions to stochastic differential equations in mathematical finance. Giles (2008) extended the MLMC method to solve stochastic ordinary differential equations of Ito type. Barth et al. (2011) and Cliffe et al. (2011) introduced MLMC method for elliptic partial differential equations (PDEs) with stochastic coefficients. Abdulle et al. (2013) applied MLMC method to solve elliptic PDEs in divergence form, where the coefficients are random with multiple scales. Mishra et al. (2012) generalized MLMC method to nonlinear, scalar hyperbolic conservation laws with random initial data. Mishra et al. (2016) extended the work of Mishra et al. (2012) for systems of nonlinear, hyperbolic conservation laws in several space dimensions. Geraci et al. (2015) proposed a Multi-Level Multi-Fidelity method in which the MLMC estimator is modified at each level to benefit from a level specific low-fidelity model.

In the context of fluid flow in porous media, Müller et al. (2013) applied MLMC method for two-phase transport simulations of an oil reservoir with uncertain heterogeneous permeability. Efendiev et al. (2013) used mixed multi-scale finite element methods within the MLMC framework to speed up the computations involving multiphase flow and transport simulations. Efendiev et al. (2015) coupled the generalized multi-scale finite element method with the Multi-Level Markov chain Monte Carlo method (MLMCMC), which sequentially screens the proposal with different levels of approximations and combines the samples at different levels to arrive at an accurate estimate. Elsakout et al. (2015) demonstrated the performance of MLMCMC for uncertainty quantification tasks involving reservoir simulation with less computational cost in comparison to the standard Markov Chain Monte Carlo method. Fagerlund et al. (2016) combined selective refinement technique with the MLMC for estimating the sweep efficiency in a two-phase flow scenario where an absolute accuracy of failure probability in a magnitude 5 to 10 percent is required. Lu et al. (2016) applied MLMC method for estimating cumulative distribution functions of QoI obtained from the numerical approximation of large-scale stochastic subsurface simulations. For a complete review of MLMC method, we refer the readers to the following papers by Giles (2013) and Giles (2015).

Historically, MLMC method constructs a hierarchy of coarse spatial and/or time discretization models as low-fidelity models. However, it is also possible to formulate a sequence of low-fidelity models utilizing projection based reduced order models (Wang et al., 2017; Xiao et al., 2017) of different dimensions. For

example, Codina et al. (2015) employed different reduced basis ROMs in the MLMC framework to estimate the statistical outputs of stochastic elliptic PDEs. In that work, the authors proposed an algorithm for optimally choosing both the dimensions of the reduced basis ROMs and the number of Monte Carlo samples at each level to achieve a given error tolerance.

In this manuscript, we propose a Multi-Fidelity-Multi-Level Monte Carlo (MFML-MC) method to address some of the limitations of standard MLMC method with Galerkin projection based ROMs (Antoulas et al., 2001; Lassila et al., 2014; Codina et al., 2015) as low-fidelity models, in particular for large scale nonlinear UQ problems. We first note that the variance, and hence the mean square error of the standard MLMC estimator depends on the correlation between every two consecutive level ROMs. This requires a large number of levels with a small difference in the number of dimensions between every two consecutive ROMs. Therefore, the standard MLMC estimator not only requires many levels of ROMs but also requires ROMs of high dimensions until high correlation with the high-fidelity model is achieved. Hence, the MLMC method involving ROMs obtained directly from high-fidelity model solution data like the one mentioned in Codina et al. (2015) can significantly limit the performance of MLMC method. Second, Galerkin-projection ROMs like POD ROMs obtained from the nonlinear high-fidelity model are subject to severe convergence and stability issues especially when the dimensions of the ROMs are much smaller than the dimensions of the high-fidelity model (Bui-Thanh et al., 2007; He, 2010; Wang et al., 2012). This severely limits the use of POD ROMs with low dimensions in MLMC framework, and therefore we cannot expect the reduction in computational complexity by orders of magnitude as a result of state variable's dimension reduction (Kani and Elsheikh, 2018). Third, MLMC method based on ROMs requires reconstruction of the high-fidelity model state variable for every sample at each level for nonlinear problems. Such reconstruction of the high-fidelity model state variable involves a high dimensional matrix-vector multiplication, and therefore employing ROMs in the MLMC method can easily cause computational overheads, in particular for UQ problems with nonpolynomial nonlinearity. However, we note that this limitation about reconstructing the high-fidelity model solution to predict outputs of interest does not apply to linear ROMs with linear outputs. Fourth, finding the optimal dimensions of the ROMs is not guaranteed despite the additional computational complexity in the nonlinear integer optimization problem formulated in Codina et al. (2016).

The proposed MFML-MC method utilizes a number of ideas that are detailed as follows. The first idea of the MFML-MC approach is to obtain a sequence of POD based approximations of the QoI and use these sequence of POD based approximations as low-fidelity models in MLMC framework. More precisely, we compute the optimal POD bases from the singular value decomposition of the snapshot matrix built directly from the training samples of the QoI. We then employ the computed POD bases in the least-squares reconstruction method to obtain a sequence of POD based approximations of the QoI (see section 4 for more details). Since the dimension of the QoI is much smaller than the state variable's dimension, the dimension of the

basis vector utilized to approximate the QoI is much smaller than the basis vector utilized to build a standard POD ROM. Therefore, building QoI POD instead of a full state variable POD enables the efficient extraction of high-level PODs at a limited computational cost. The second idea is to employ the MFMC method at each level of the MLMC method so that the high-fidelity model is utilized to provide an unbiased estimator, while the low computational cost of low-fidelity models are exploited to run a very large number of realizations in order to obtain a low variance estimator. The third idea in the MFML-MC approach is to represent the difference between every two consecutive level models of the MLMC framework in a reduced dimension. We utilize principal component analysis (PCA) to perform this dimensionality reduction. The main reason to utilize PCA for dimensionality reduction is to exploit the linearity of the expected value operator. The fourth idea is to use a data-driven approach to construct a non-intrusive ROM (Wang et al., 2017; Xiao et al., 2017) in order to compute the reduced representation mentioned in the third step of MFML-MC method. We use Gradient Boosted Tree Regressor (GBTR) (Friedman, 2001) to formulate such level specific low-fidelity non-intrusive ROM in the MFMC setup. We then utilize the constructed non-intrusive ROM as a low-fidelity model in the MFMC setup on every level of the MFML-MC method. To the best of our knowledge, this paper presents the first attempt to combine the features of MFMC method and MLMC method using machine learning techniques for UQ analysis of nonlinear dynamical systems representing multi-phase porous media flow with uncertainty in the permeability field. In addition, this paper presents the first attempt to use the MFMC method to estimate the statistics of the vector-valued time series QoI while the standard MFMC method is mainly used for estimating the statistics of scalar QoI (Ng, 2013; Peherstorfer et al., 2016).

The remaining of this manuscript is organized as follows. In section 2, multi-phase porous media flow problem is formulated. In section 3, MC, MFMC, and MLMC methods are briefly explained. In section 4, MFML-MC method is introduced. In section 5, Numerical results for two subsurface multi-phase porous media flow problems showing the performance of MFML-MC method are reported. We note that building reduced order models for these porous media flow problems is quite challenging where standard POD-Galerkin reduced order models produce inaccurate and unstable results even for the cases where a large number of POD basis vectors is utilized (He et al., 2011; Kani and Elsheikh, 2018). Hence, in the two numerical test cases, standard MLMC with POD-Galerkin ROM had all the four limitations as mentioned earlier in this section. Finally, in section 6, conclusions and perspectives are drawn.

2. PROBLEM FORMULATION

We consider an immiscible two-phase (oil and water) flow in an incompressible porous media domain. The flow behavior of oil and water in a porous media domain can be described by conservation of mass and Darcy's law for each phase (Bastian, 1999; Chen et al., 2006; Aarnes et al., 2007). Neglecting the effects

of gravitation, capillary, and compressibility, and assuming the density ratio to be equal to one, Darcy's law for each phase can be described as

$$\mathbf{v}_\alpha = -\mathbf{K} \frac{k_{r\alpha}}{\mu_\alpha} \nabla p \quad (1)$$

where the subscript $\alpha = w$ denotes the water phase, the subscript $\alpha = o$ denotes the oil phase, \mathbf{v}_α is the phase velocity, p is the global pressure, \mathbf{K} is the absolute permeability tensor, $k_{r\alpha}$ is the relative permeability of phase α , μ_α is the viscosity of phase α (Bastian, 1999; Chen et al., 2006; Aarnes et al., 2007). The phase relative permeabilities $k_{r\alpha}$ models the interactions between the two phases and usually, $k_{r\alpha}$ is described as a function of phase saturation (volume of phase α in a given pore space of the porous media domain) (Aarnes et al., 2007).

The total conservation of mass can be expressed in terms of incompressibility condition that takes the form

$$\nabla \cdot \mathbf{v} = q \quad (2)$$

$\mathbf{v} = \mathbf{v}_o + \mathbf{v}_w$ is the total velocity vector and q is the total source and sink term. We can combine the equation of Darcy's law for each phase (Equation 1) and the conservation of mass (Equation 2) to derive equations for global pressure and water saturation:

$$\begin{aligned} \nabla \cdot \mathbf{K} \lambda \nabla p &= q \\ \phi \frac{\partial s_w}{\partial t} - \nabla \cdot (f_w \mathbf{v}) + q_w &= 0 \end{aligned} \quad (3)$$

where $\lambda = \lambda_w + \lambda_o$ is the total mobility, $\lambda_\alpha = k_{r\alpha}/\mu_\alpha$ is the phase mobility, $f_w = \lambda_w/(\lambda_w + \lambda_o)$ is termed as the fractional flow function for the water phase and with the constraint $s_w + s_o = 1$. In the rest of the manuscript, we use s in place of s_w to denote water saturation.

In this problem, we consider Equation (3) as the high-fidelity model and we solve Equation (3) for pressure and saturation using sequential formulation where we solve for pressure first and then solve for the water saturation. We use finite volume method to discretize the spatial derivatives of Equation (3) in a spatial domain of n grid blocks. We use implicit time stepping method to solve Equation (3) for the high-fidelity model state variable $\mathbf{y}_s \in \mathbb{R}^n$, where each component of \mathbf{y}_s is the water saturation value at the i th grid block.

The QoI is defined as $\mathbf{u}(t) \in \mathbb{R}^m$, where $u_i = y_s(x_i, y_i)$, $i = 1 \dots m \ll n$ at specific time steps (say $t = 10, 20, \dots, 200$). In the following, we use \mathbf{u} in place of $\mathbf{u}(t)$ to simplify the notation and we are interested in the first moment estimate (i.e., mean) of \mathbf{u} . The grid points of interest (x_i, y_i) $i = 1 \dots m$ can be a set of arbitrary user specific spatial locations. For example, a set of grid points where injectors and producers are located.

3. MULTI-FIDELITY MONTE CARLO AND MULTI-LEVEL MONTE-CARLO METHOD

Let \mathbf{x} be a realization of the input random vector $\mathbf{X}(\omega)$, $\omega \in \Omega$ where Ω is the sample space and the quantity of interest be the expectation of the random variable \mathbf{u} . The standard Monte

Carlo method estimates the expectation $\mathbb{E}[\mathbf{u}]$ of the random variable \mathbf{u} as

$$\hat{\mathbf{u}} = \frac{1}{N} \sum_{i=1}^N \mathbf{u}^i \quad (4)$$

where $\hat{\mathbf{u}}$ is the estimator of $\mathbb{E}[\mathbf{u}]$, $\mathbf{u}^i = \mathbf{u}(\mathbf{x}_i)$, and N is the number of realizations of the model output. As per the law of large numbers (Central Limit Theorem) (Giles, 2015), a sample based estimate of the expectation $\mathbb{E}[\mathbf{u}]$ introduces sampling error (mean square error) defined as

$$\epsilon = \text{Var}(\hat{\mathbf{u}}) = \frac{1}{N} \text{Var}(\mathbf{u}) \quad (5)$$

where $\text{Var}(\mathbf{u})$ is the variance of \mathbf{u} . As $\sqrt{\epsilon}$ known as standard error scales with $\frac{1}{\sqrt{N}}$ for a constant $\text{Var}(\mathbf{u})$, MC simulations are computationally prohibitive because of the slow convergence rate. One way to achieve a lower ϵ is to reduce the numerator in Equation (5) (Ng, 2013).

Control variate is a variance reduction technique which uses alternative estimator for $\mathbb{E}[\mathbf{u}]$, $\mathbf{u} \in \mathbb{R}$ that takes the form

$$\hat{\mathbf{u}}^{cv} = \hat{\mathbf{u}} + \beta (\hat{\mathbf{v}} - \mathbb{E}[\mathbf{v}]) \quad (6)$$

where $\mathbf{v}(\mathbf{x}) \in \mathbb{R}$ is an auxiliary random variable. The estimator $\hat{\mathbf{u}}^{cv}$ is an unbiased estimator of $\mathbb{E}[\mathbf{u}]$ with variance defined as Ng (2013)

$$\text{Var}(\hat{\mathbf{u}}^{cv}) = \text{Var}(\hat{\mathbf{u}}) (1 - \rho^2) \quad (7)$$

where ρ is the correlation between $\mathbf{u}(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$. Since ρ^2 lies between 0 and 1, $\text{Var}(\hat{\mathbf{u}}^{cv})$ is always less than $\text{Var}(\hat{\mathbf{u}})$. For UQ tasks, where the QoI is governed by partial differential equations, $\mathbf{u}(\mathbf{x})$ is obtained from a high-fidelity model output and $\mathbf{v}(\mathbf{x})$ is generally obtained from a low-fidelity model output. In general, we do not know exactly $\mathbb{E}[\mathbf{v}]$ and we have to use a more accurate estimate of $\mathbb{E}[\mathbf{v}]$. For example, Ng (2013) replaced $\mathbb{E}[\mathbf{v}]$ in Equation (6) by $\hat{\mathbf{v}} = \frac{1}{M} \sum_{i=1}^M \mathbf{v}(\mathbf{x}_i)$, where $M \gg M_{\text{HF}}$ and M_{HF} is the number of high-fidelity model samples. Furthermore, it was proved in Ng (2013) that for a fixed computational budget p , a perfectly correlated low-fidelity model is not the only condition for variance reduction over the standard MC estimator but the low-fidelity model must also be cheaper to evaluate than the high-fidelity model.

The potential limitation in the aforementioned multi-fidelity estimator (Ng, 2013) is that it repartitions the given computational budget p between the high-fidelity model and only a single low-fidelity model such that the mean square error of the estimator is minimized. In order to allow an arbitrary number of low-fidelity models into the control variate method, Peherstorfer et al. (2016) extended the multi-fidelity approach introduced in Ng (2013). Multi-Fidelity Monte Carlo method introduced in Peherstorfer et al. (2016) formulated an optimization problem that used an arbitrary number of low-fidelity models to derive an unbiased MFMC estimator of $\mathbb{E}[\mathbf{u}]$ that takes the form

$$\hat{\mathbf{u}}^{mf} = \hat{\mathbf{u}} + \beta^1 (\hat{\mathbf{v}}^1 - \hat{\mathbf{u}}) + \sum_{i=2}^I \beta^i (\hat{\mathbf{v}}^i - \hat{\mathbf{v}}^{i-1}) \quad (8)$$

where $\mathbf{v}^1 \dots \mathbf{v}^I \in \mathbb{R}$ are auxiliary random variables obtained from I number of different low-fidelity models, $\hat{\mathbf{v}}^i$ estimates the expectation $\mathbb{E}[\mathbf{v}^i]$ using M_i samples of low-fidelity model i , $\beta^1 \dots \beta^I \in \mathbb{R}$ are the coefficients. The low-fidelity model i uses $\mathbf{x}_1 \dots \mathbf{x}_{M_i}$ realizations of the input random vector $\mathbf{X}(\omega)$ to estimate $\hat{\mathbf{v}}^i$, whereas the low-fidelity model $i-1$ uses only the first M_{i-1} realizations of $\mathbf{X}(\omega)$ to estimate $\hat{\mathbf{v}}^{i-1}$. Therefore the two consecutive estimators $\hat{\mathbf{v}}^i$ and $\hat{\mathbf{v}}^{i-1}$ are dependent for all $i = 1 \dots I$. The cost of the MFMC estimator is $C(\hat{\mathbf{u}}^{mf}) = \sum_{i=1}^I C_i \cdot M_i + C_{\text{HF}} \cdot M_{\text{HF}}$, where C_{HF} is the cost of evaluating a high-fidelity model, and C_i is the cost of evaluating a low-fidelity model i for all $i = 1 \dots I$. In Peherstorfer et al. (2016), an optimization problem was formulated to select optimal values for the number of samples $\{M_{\text{HF}}^*, M_1^* \dots M_I^*\}$, and for the coefficients $\{\beta^{1*} \dots \beta^{I*}\}$ such that the mean square error of the MFMC estimator is lower than the Monte Carlo estimator for a fixed computational budget.

The multi-level idea is an another extension of the control variate approach in which a sequence of low-fidelity models at different levels ($\mathbf{v}_i \in \mathbb{R}^m$ with $i = 1 \dots I$) is used to evaluate an approximate statistics of \mathbf{u} . First, let the index i encodes the accuracy of \mathbf{v}_i with respect to the true solution $\mathbf{u} \in \mathbb{R}^m$. This means, as i is increased, the accuracy of \mathbf{v}_i is refined to approximate \mathbf{u} . Consequently, \mathbf{u} can be written as a telescopic sum in terms of \mathbf{v}_i with $i = 1 \dots I$, that takes the form (Müller et al., 2013)

$$\mathbf{u} = \mathbf{v}_1 - \mathbf{v}_0 + \mathbf{v}_2 - \mathbf{v}_1 + \dots + \mathbf{v}_I - \mathbf{v}_{I-1} + \mathbf{u} - \mathbf{v}_I = \sum_{i=0}^I \mathbf{Y}_i \quad (9)$$

where $\mathbf{Y}_i = \mathbf{v}_{i+1} - \mathbf{v}_i$ with $i = 0 \dots I-1$, $\mathbf{Y}_I = \mathbf{u} - \mathbf{v}_I$, and we set $\mathbf{v}_0 = 0$. Exploiting the linearity of the expected value operator \mathbb{E} , the expected value $\mathbb{E}[\mathbf{u}]$ defined in Equation (9) can be written as

$$\mathbb{E}[\mathbf{u}] = \sum_{i=0}^I \mathbb{E}[\mathbf{Y}_i] \quad (10)$$

The MLMC estimator for the expected value of \mathbf{u} is obtained by replacing the expected values on the right hand side of Equation (10) by ensemble averages and is defined as

$$\hat{\mathbf{u}}^{ml} = \sum_{i=0}^I \hat{\mathbf{Y}}_i = \sum_{i=0}^I \frac{1}{M_i} \sum_{j=1}^{M_i} \mathbf{Y}_i^j \quad (11)$$

The mean square error (mse) of MLMC estimator $\hat{\mathbf{u}}^{ml}$ is derived as

$$\epsilon^{ml} = \sum_{i=0}^I \text{Var}(\hat{\mathbf{Y}}_i) = \sum_{i=0}^I \frac{1}{M_i} \text{Var}(\mathbf{Y}_i) \quad (12)$$

It is evident from Equation (12) that the mse (ϵ^{ml}) of MLMC estimator is sum of several smaller contributions $\frac{1}{M_i} \text{Var}(\mathbf{Y}_i)$ with $i = 0 \dots I$.

The MLMC method is mainly based on the fact that $\frac{1}{M_i} \text{Var}(\mathbf{Y}_i)$ at low levels are reduced by increasing number

of samples (M_i) as low level samples are computed at low computational cost. At high levels, the level variances $\text{Var}(\mathbf{Y}_i)$ are expected to be typically small, thus M_i can be small and hence MLMC method incurs few expensive high-fidelity model simulations. In summary, MLMC method relies on the following variance hierarchy:

$$\text{Var}(\mathbf{Y}_0) > \text{Var}(\mathbf{Y}_1) > \text{Var}(\mathbf{Y}_2) > \dots > \text{Var}(\mathbf{Y}_I) \quad (13)$$

and also expects $C_0 < C_1 < C_2 < \dots < C_I$, where C_i is the computational cost to compute one sample of \mathbf{Y}_i . In MLMC method, the optimal values for the number of samples $\{M_0^* \dots M_I^*\}$ are computed by solving a constrained minimization problem where the cost function to be minimized is the total computational cost ($\sum_{i=0}^I C_i \cdot M_i$) of the MLMC method and constraint is set by fixing ϵ^{ml} to a specific value (say $\frac{\epsilon^2}{2}$) (Müller et al., 2013; Geraci et al., 2015). The optimal values for the number of samples are expressed as

$$M_i^* = \frac{2}{\epsilon^2} \left[\sum_{j=0}^I \sqrt{C_j \cdot \text{Var}(\mathbf{Y}_j)} \right] \sqrt{\frac{\text{Var}(\mathbf{Y}_i)}{C_i}} \quad i = 0, \dots, I \quad (14)$$

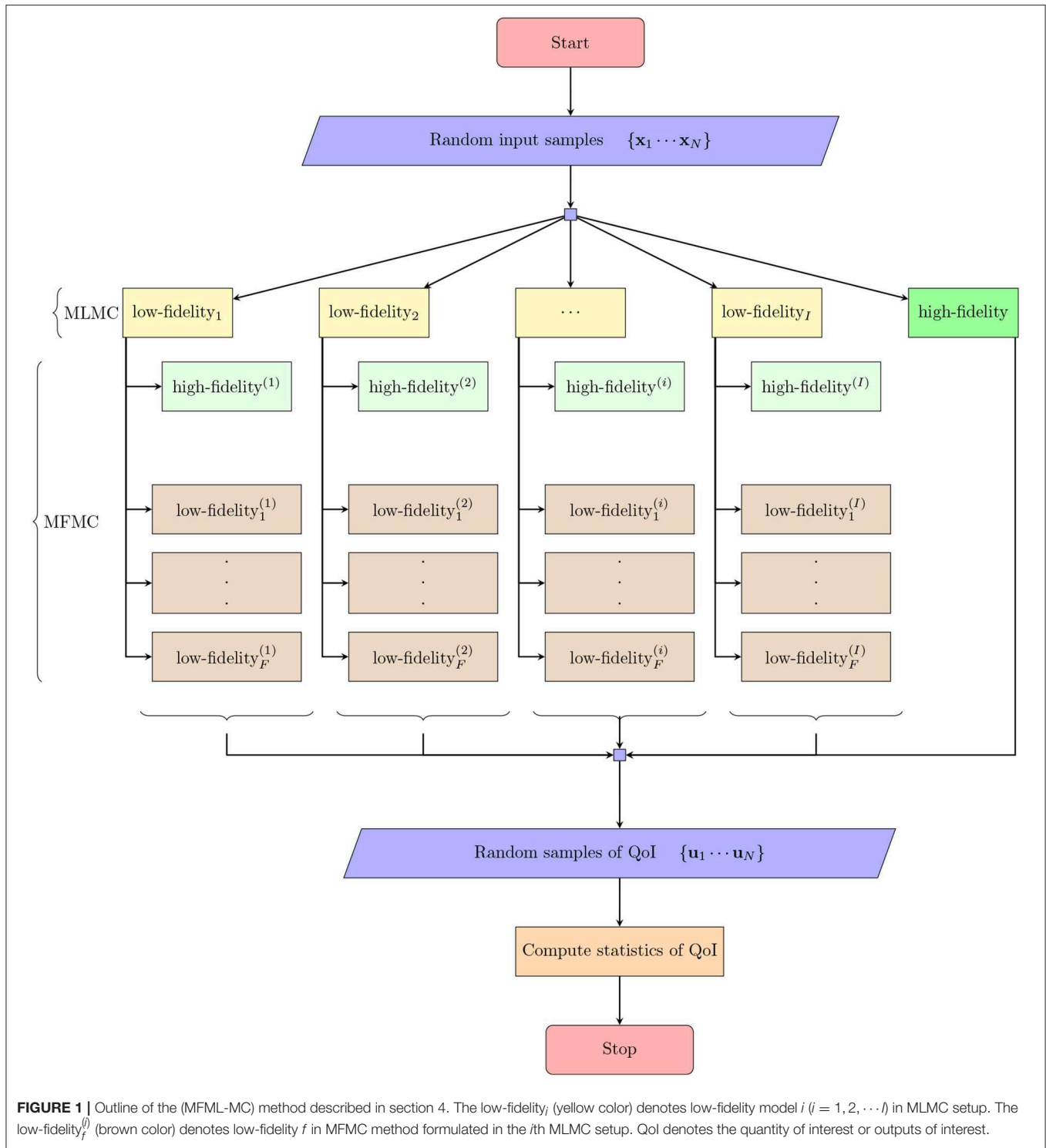
Although MLMC in general refer to control variate method with a sequence of I geometrical levels (mesh discretization levels), it can also be utilized with a sequence of I reduced basis models (Codina et al., 2016) or POD basis models. More specifically, a sequence of POD basis models can be employed as sequence of low-fidelity models $\mathbf{v}_1 \dots \mathbf{v}_I$.

A practical implementation of the MLMC algorithm is the following (Müller et al., 2013)

1. Fix a sequence of levels based on grid resolutions or POD basis $i = 1 \dots I$.
2. Fix a number of offline samples M_0 and fix a threshold for the estimated standard error.
3. Perform M_0 samples of high fidelity simulations.
4. If POD basis models, Derive I number of POD basis models.
5. Compute M_0 samples of \mathbf{Y}_i on every level.
6. Solve the optimization (Müller et al., 2013) problem to estimate M_i samples of \mathbf{Y}_i with $i = 0 \dots I$.
7. Update the estimates for $\mathbb{E}[\mathbf{Y}_i]$, $\text{Var}(\mathbf{Y}_i)$, and C_i on every level.
8. Compute and update the required number of samples M_i on each level.
9. On every level, if the updated M_i is more than the number of samples already computed, then add an additional sample of \mathbf{Y}_i and continue with step 6. If no level requires an additional sample, then quit.

4. MULTI-FIDELITY-MULTI-LEVEL MONTE CARLO METHOD

In this section, we present a novel variance reduction method called Multi-Fidelity-Multi-Level Monte Carlo (MFML-MC) method addressing the limiting facts observed in the standard MLMC method with Galerkin projection based ROMs (see section 1 for more details). In MFML-MC method, we formulate



a MLMC framework with *I* levels and then apply the techniques of MFMC method on every level of MLMC framework. **Figure 1** displays the outline of the MFML-MC method and its detailed formulation is described as five steps in the rest of this section.

The first step of MFML-MC method is to formulate a sequence of POD approximations of the QoI **u** and utilize these sequence

as low-fidelity models [**v**₁, ..., **v**_{*I*}] in MLMC framework. More precisely, in this approach, **v**_{*i*} is *i*th level POD approximation of **u** and is computed from least-squares reconstruction method defined as

$$\mathbf{u} \approx \mathbf{v}_i = \mathbf{U}_u^{r_i} \tilde{\mathbf{u}} = \mathbf{U}_u^{r_i} (\mathbf{U}_u^{r_i \top} \mathbf{u}) \quad (15)$$

where $\tilde{\mathbf{u}} \in \mathbb{R}^{r_i}$ is the reduced representation of \mathbf{u} , $\mathbf{U}_u^{r_i} \in \mathbb{R}^{m \times r_i}$ is the orthonormal matrix containing r_i orthonormal basis vectors in its columns. The optimal orthonormal basis vectors are computed from the singular value decomposition (SVD) of the snapshot matrix $\mathbf{X}_u = ((\mathbf{u}_1 \dots \mathbf{u}_T)^1 \dots (\mathbf{u}_1 \dots \mathbf{u}_T)^L)$, where T denotes the number of time steps and L denotes the number of training samples corresponding to different realizations of the stochastic input parameters. The SVD of \mathbf{X}_u is expressed as (Kani and Elsheikh, 2018)

$$\mathbf{X}_u = \mathbf{U}_u \Sigma_u \mathbf{W}_u \quad (16)$$

where $\mathbf{U}_u \in \mathbb{R}^{m \times m}$ is the left singular matrix, $(\sigma_1 > \sigma_2 > \sigma_3 > \dots \sigma_m \geq 0)$ are the singular values of the snapshot matrix \mathbf{X}_u . The associated error termed as least-squares errors in approximating \mathbf{u} by \mathbf{v}_i using only r_i basis vectors is given by (Berkooz et al., 1993; Lucia et al., 2004)

$$\varepsilon_i = \|\mathbf{u} - \mathbf{v}_i\|_2 = \sum_{j=r_i+1}^m \sigma_j \quad (17)$$

Please note that the dimension m of the basis vector in \mathbf{U}_u is much smaller than n (the number of grid points). Hence, for a large scale UQ problems where $m \ll n$, we can easily form many levels with smaller ε_i in this MLMC framework in comparison to standard MLMC method. Moreover, \mathbf{v}_i can be obtained by using less number of basis vectors ($r_i \approx m$ with $\varepsilon_i \approx 0$) in comparison to standard MLMC method with a sequence of Galerkin projection based ROMs.

The second step of MFML-MC method is to compute the reduced representation of \mathbf{Y}_i over all levels in MLMC framework (see Equation 9). The reduced representation of \mathbf{Y}_i is expressed as

$$\tilde{\mathbf{Y}}_i = \mathbf{U}_{Y_i}^{q_i \top} \mathbf{Y}_i \quad (18)$$

where $\tilde{\mathbf{Y}}_i \in \mathbb{R}^{q_i}$ is the reduced representation of \mathbf{Y}_i , $\mathbf{U}_{Y_i}^{q_i} \in \mathbb{R}^{m \times q_i}$ is the orthonormal matrix containing q_i orthonormal basis vectors in its columns. The optimal orthonormal basis matrix $\mathbf{U}_{Y_i}^{q_i}$ is computed from the singular value decomposition (SVD) of the snapshot matrix $\mathbf{X}_{Y_i} = ((\mathbf{Y}_{i1} \dots \mathbf{Y}_{iT})^1 \dots (\mathbf{Y}_{i1} \dots \mathbf{Y}_{iT})^L)$, where $\mathbf{Y}_{ij} = \mathbf{v}_{i+1j} - \mathbf{v}_{ij}$ ($i = 0 \dots I-1$) and $\mathbf{Y}_{ij} = \mathbf{u}_j - \mathbf{v}_{ij}$ for all $j = 1, \dots, T$. Since, \mathbf{Y}_i is computed from the difference between two consecutive levels of POD based approximations \mathbf{v}_{i+1} and \mathbf{v}_i , i.e., $\mathbf{Y}_i = \mathbf{v}_{i+1} - \mathbf{v}_i$, the least-squares error in approximating \mathbf{Y}_i by $(\mathbf{U}_{Y_i}^{q_i} \tilde{\mathbf{Y}}_i)$ is equivalent to the difference of two consecutive level ε (see Equation 17) which is expressed as $\Delta \varepsilon_i = \varepsilon_i - \varepsilon_{i+1} = \sum_{j=r_i+1}^{r_{i+1}} \sigma_j$.

Now the MLMC estimator (see Equation 11) for the expected value of \mathbf{u} is expressed as

$$\mathbb{E}[\mathbf{u}] = \sum_{i=0}^I \mathbb{E}[\mathbf{Y}_i] \approx \sum_{i=0}^I \mathbf{U}_{Y_i}^{q_i} \mathbb{E}[\tilde{\mathbf{Y}}_i] = \hat{\mathbf{u}}^{ml} = \sum_{i=0}^I \hat{\mathbf{Y}}_i = \sum_{i=0}^I \mathbf{U}_{Y_i}^{q_i} \hat{\tilde{\mathbf{Y}}}_i \quad (19)$$

The third step of MFML-MC method is to set r_i for all $i = 1 \dots I$. In this framework, we set $r_i = i$. Now, $\Delta \varepsilon_i = \sigma_i$ and therefore, we

expect \mathbf{Y}_i to be attracted to a certain low dimensional subspace of dimension $q_i = \Delta r_i = (r_{i+1} - r_i) = 1$ over all the levels.

In the fourth step, we extend the Multi-Level Multi-Fidelity method introduced in Geraci et al. (2015) by adopting multi-fidelity approach (see Equation 8) on every level of MLMC framework to derive an unbiased estimator of $\mathbb{E}[\tilde{\mathbf{Y}}_i]$ in Equation (19) that takes the form

$$\hat{\tilde{\mathbf{Y}}}_i^{mf} = \hat{\tilde{\mathbf{Y}}}_i + \beta_i^1 (\hat{\tilde{\mathbf{Y}}}_i^1 - \hat{\tilde{\mathbf{Y}}}_i) + \sum_{f=2}^{F_i} \beta_i^f (\hat{\tilde{\mathbf{Y}}}_i^f - \hat{\tilde{\mathbf{Y}}}_i^{f-1}) \quad (20)$$

where $\tilde{\mathbf{Y}}_i^1 \dots \tilde{\mathbf{Y}}_i^{F_i}$ are auxiliary random variables obtained from F_i number of level specific low-fidelity models of $\tilde{\mathbf{Y}}_i$, $\hat{\tilde{\mathbf{Y}}}_i^f$ estimates the expectation $\mathbb{E}[\tilde{\mathbf{Y}}_i^f]$ using M_i^f samples of $\tilde{\mathbf{Y}}_i^f$ for all $f = 1 \dots F_i$, and $\beta_i^1 \dots \beta_i^{F_i} \in \mathbb{R}$ are the coefficients on level i , $i = 0 \dots I$. In this paper, we set $F_i = 1$ for all $i = 0 \dots I$. Next, we use the optimization problem formulated in Peherstorfer et al. (2016) to select optimal values M_{HF}^* , M_i^{1*} such that the mean square error of the MFML-MC estimator $\hat{\tilde{\mathbf{Y}}}_i^{mf}$ on every level is lower than the Monte Carlo estimator $\hat{\tilde{\mathbf{Y}}}_i$ for the same computational budget.

Now, the MFML-MC estimator for the expected value of \mathbf{u} (see Equation 19) is expressed as

$$\mathbb{E}[\mathbf{u}] \approx \hat{\mathbf{u}}^{ml} = \sum_{i=0}^I \hat{\mathbf{Y}}_i \approx \sum_{i=0}^I \mathbf{U}_{Y_i}^{q_i} \hat{\tilde{\mathbf{Y}}}_i^{mf} \quad (21)$$

In the fifth step, we utilize a data-driven approach to derive a level specific low-fidelity model $\tilde{\mathbf{Y}}_i^1$ in the MFMC setup. In this data-driven approach, we first consider a discrete nonlinear dynamical system on every level ($i = 0 \dots I$) that takes the form

$$\tilde{\mathbf{Y}}_i^1(t+1) = \tilde{\mathbf{Y}}_i^1(t) + \mathbf{F}_i(\mathbf{x}, \tilde{\mathbf{Y}}_i^1(t)), \quad (22)$$

where $\mathbf{F}_i(\mathbf{x}, \tilde{\mathbf{Y}}_i^1(t))$ is the nonlinear term utilized to update $\tilde{\mathbf{Y}}_i^1(t+1)$ on level i for all $i = 0 \dots I$ (Nagoor Kani and Elsheikh, 2017). Next, we use GBTR (Friedman, 2001) on every level to approximate $\mathbf{F}_i(\mathbf{x}, \tilde{\mathbf{Y}}_i^1(t))$. We use $(\mathbf{x}, \tilde{\mathbf{Y}}_i^1(t))$ as an input to GBTR and compute $\mathbf{F}_i(\mathbf{x}, \tilde{\mathbf{Y}}_i^1(t))$ as an output. We fit GBTRs using the same training samples $((\mathbf{Y}_{i1} \dots \mathbf{Y}_{iT})^1 \dots (\mathbf{Y}_{i1} \dots \mathbf{Y}_{iT})^L)$ utilized in the second step.

5. NUMERICAL EXPERIMENTS

In this section, we present numerical results to evaluate the performance of MFML-MC method. The numerical results are based on two UQ tasks involving two-phase flow in the heterogeneous porous media domain. The two test cases are quarter five spot problem and the uniform flow problem with the uncertainties in the permeability field (Kani and Elsheikh, 2018). In section 5.1, we describe high-fidelity model setup, in section 5.2, we describe low-fidelity model setup in order to formulate MFML-MC framework, in section 5.3, we describe MLMC with Galerkin-POD ROMs setup, and in section 5.4, we define a set of error metrics that we utilize

to compare MFML-MC method with standard MC method that uses either high-fidelity model or low-fidelity model. In section 5.5, we provide the results for quarter five spot problem and in section 5.6, we provide the results for uniform flow problem.

5.1. High-Fidelity Model Setup

We consider two-phase flow of oil and water in a two dimensional porous media domain $[0, 1] \times [0, 1]$ where water is injected to displace the residual oil. We consider Equation (3) as a high-fidelity model to describe the flow behavior of oil and water. We define the relative permeability based on Corey's model $k_{rw}(s) = s^{*2}$, $k_{ro} = (1 - s^{*})^2$, where $s^{*} = (s - s_{wc}) / (1 - s_{or} - s_{wc})$, s_{wc} is the irreducible water saturation and s_{or} is the residual oil saturation (Aarnes et al., 2007). We set $s_{wc} = 0.2$, $s_{or} = 0.2$, and initial water saturation to $s_{wc}(0.2)$. We set the porosity field in the porous media domain to a constant value of 0.2. We set viscosity ratio of water and oil to 0.2. We consider uncertainties from the permeability field and assumed to be modeled as a log-normal distribution function with zero mean and exponential covariance that takes the form

$$\text{Cov} = \sigma_k^2 \exp \left[-\frac{|x_1 - x_2|}{\iota_k} \right] \quad (23)$$

where σ_k^2 is the variance, ι_k is the correlation length. We set σ_k^2 to 1 and ι_k to 0.1. Sample realizations of log-permeability values are displayed in **Figure 2**.

As mentioned in section 2, we use sequential formulation to solve Equation (3) for pressure and water saturation (Aarnes et al., 2007). We first generate a uniform mesh of 96×96 blocks in a spatial domain. We use finite volume method with two point flux approximation (Aarnes et al., 2007) to solve for pressure and an upwind finite-volume method with an implicit backward Euler method combined with Newton-Raphson iterative method to solve for saturation. We set time step size to 0.015 and we solve Equation (3) for 200 time steps. We solve pressure and update

velocity field at every 8th time step as pressure field changes much slower than saturation field over time. Time is measured by a non dimensional unit called pore volumes injected (PVI) (Ibrahima, 2016).

As defined in section 2, QoI is $\mathbf{u} \in \mathbb{R}^m$, where $u_i = y_s(x_i, y_i)$, $i = 1 \cdots m \ll n$ at specific time steps. The first moment estimate of $\mathbf{u}(t)$ at specific time steps are the desired statistic. The interested grid points $((x_i, y_i) \ i = 1 \cdots m)$ are 6×6 grid points ($m = 36$) uniformly selected from the 96×96 spatial domain. The interested time steps are $t = 10, 20, \dots, 200$. We solve Equation (3) for 25,000 random realizations of the permeability field and use Monte Carlo method to estimate the statistics of \mathbf{u} (Ibrahima, 2016).

5.2. Low-Fidelity Model Setup

We first compute the optimal POD bases matrices \mathbf{U}_u and \mathbf{U}_{Y_i} for all $i = 0 \cdots I$. We compute the POD matrices from the SVD of the snapshot matrices $\mathbf{X}_u, \mathbf{X}_{Y_i}$, $i = 0 \cdots I$. We built the snapshot matrices from 10 random samples of high-fidelity model solution data. In order to select the 10 random high-fidelity models to build snapshot matrices, we use K-means clustering algorithm to cluster 25,000 random permeability realizations into 10 clusters (Ghasemi, 2015). Then, we solve the high-fidelity model for a single permeability realization from each cluster.

Following that, the obtained matrix \mathbf{U}_u is utilized to build a sequence of POD approximations of \mathbf{u} (as detailed in the section 4) from the collected training data. Then the matrix $\mathbf{U}_{Y_i}^{q_i}$ is utilized to compute training samples of $\tilde{\mathbf{Y}}_i$ (the reduced representation of \mathbf{Y}_i) for all $i = 0 \cdots I$. We set $I = 18$, $r_i = i$, and therefore $q_i = 1$ as already mentioned in section 4.

Next, we build a level specific GBTR on every level ($i = 0 \cdots I$) to estimate \mathbf{F}_i and utilize the estimated \mathbf{F}_i in Equation (22) to compute $\tilde{\mathbf{Y}}_i^1(t+1)$. We use Scikit-learn (Pedregosa et al., 2011) a machine learning python package to implement the GBTRs. We use the training samples of $\tilde{\mathbf{Y}}_i$ to fit the level specific GBTR.

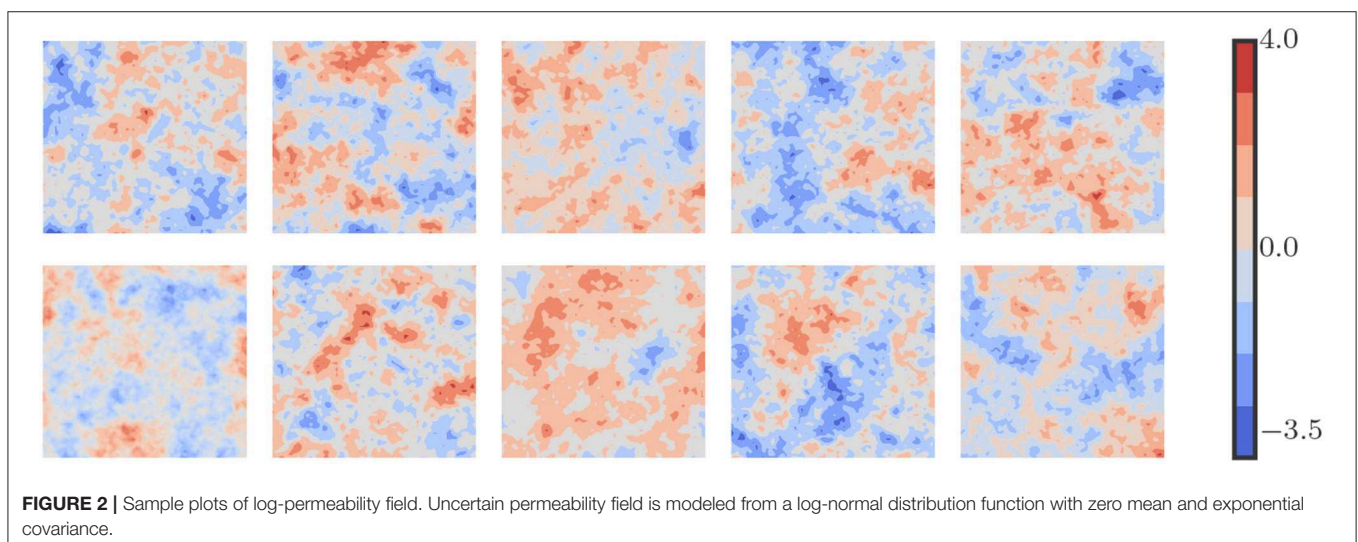


FIGURE 2 | Sample plots of log-permeability field. Uncertain permeability field is modeled from a log-normal distribution function with zero mean and exponential covariance.

5.3. Standard MLMC With POD-Galerkin ROMs Setup

We first compute optimal POD basis vectors for the pressure and saturation solution vectors from the SVD of the corresponding snapshot matrices. We built the snapshot matrices from the solution vectors (pressure and saturation) collected from the solutions of the high-fidelity model for 45 random realizations of the permeability field. We then built low-fidelity ROM models of different dimensions via Galerkin projection of the discretized system of the high-fidelity model Equation (3) on to the POD space spanned by the POD basis vectors.

Following that one can obtain MLMC framework using Galerkin projection POD ROMs as low-fidelity models. However, in the two numerical test cases namely, the quarter five spot problem (5.5), and the uniform flow problem (5.6), we obtained accurate and stable POD results only when the dimensions of the POD-Galerkin ROMs were on the order of magnitude nearly equivalent to the dimension of the high-fidelity state variable (Xiao et al., 2017; Kani and Elsheikh, 2018). Hence, the computational cost to obtain the QoI from the POD-Galerkin ROM is more than the computational cost to obtain the QoI from the high-fidelity model for a single realization. Therefore, it was infeasible to derive an effective MLMC framework with a hierarchy of low-fidelity models based on standard POD ROMs. This is expected because the governing equations of the flow problem Equation (3) has nonpolynomial nonlinearity and is well known issue in reduced order modeling for multi-phase subsurface flow problems (Chaturantabut and Sorensen, 2011; He et al., 2011; Jansen and Durlofsky, 2017; Kani and Elsheikh, 2018). We also note that we conducted extensive study on

reduced order modeling for these two problems in Kani and Elsheikh (2018) and we obtained inaccurate and unstable results when using POD ROMs. At this point, we request the readers to refer figures included in the numerical results section of Kani and Elsheikh (2018), where some of the standard POD ROM unstable results are displayed. Hence, we have not included the comparison of MFML-MC method with MLMC method based on standard POD ROMs as low-fidelity models.

5.4. Evaluation Metrics

We evaluate the performance of MFML-MC method using two time specific error metrics defined by

$$\begin{aligned}\hat{e}_t^{\text{bias}} &= \frac{1}{N_e} \sum_{j=1}^{N_e} \|\hat{\mathbf{u}}_t^{\text{ref}} - \hat{\mathbf{u}}_t^{(j)}\|_2^2 \\ \hat{e}_t^{\epsilon} &= \frac{1}{N_e} \sum_{j=1}^{N_e} \text{Var}(\hat{\mathbf{u}}_t^{(j)})\end{aligned}\quad (24)$$

where N_e is the number of runs utilized to estimate the errors, $\hat{\mathbf{u}}_t^{\text{ref}}$ is the reference result of $\mathbb{E}[\mathbf{u}_t]$ obtained from Monte Carlo estimate $\hat{\mathbf{u}}_t^{(\text{MC})}$ computed with $N = 25,000$ high-fidelity model samples. $\hat{\mathbf{u}}_t^{(j)}$ is the approximation of $\mathbb{E}[\mathbf{u}_t]$ that can be obtained from various estimators including Monte Carlo estimate that uses only high-fidelity model, Monte Carlo estimate that uses only low-fidelity model, and the MFML-MC estimator. We note that, $\hat{\mathbf{u}}_t^{(j)}$ is obtained for a fixed computational budget p . The computational budget is measured in terms of the cost required to run p number of MC realizations that uses only high-fidelity

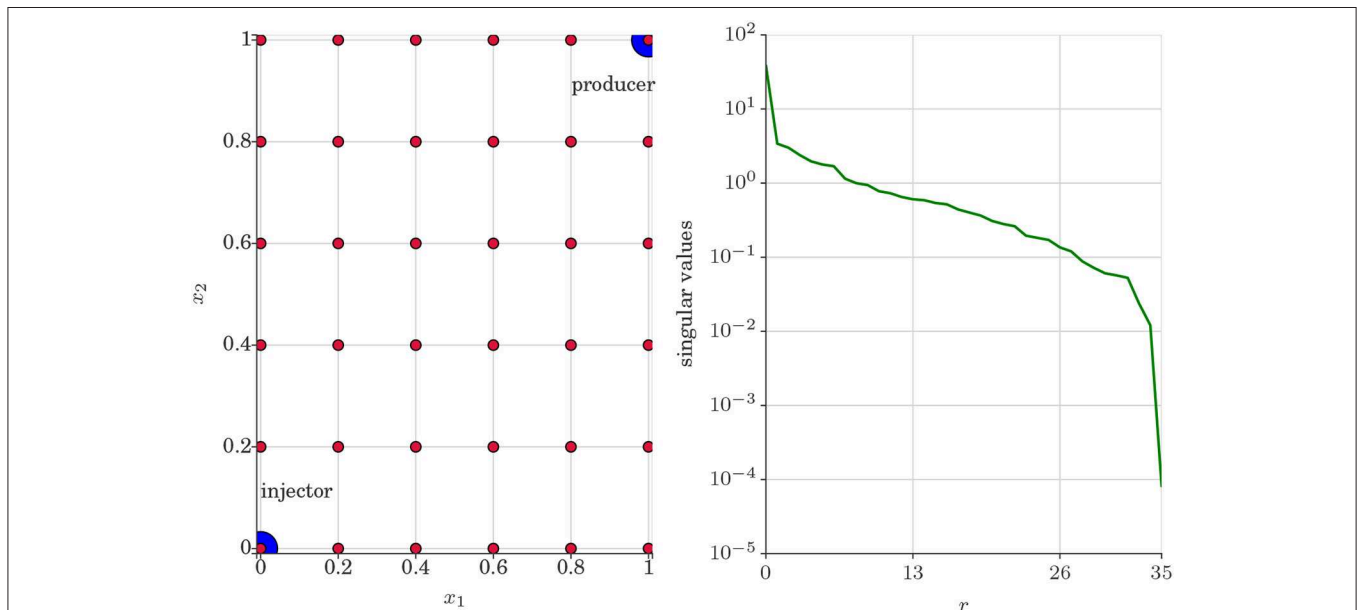
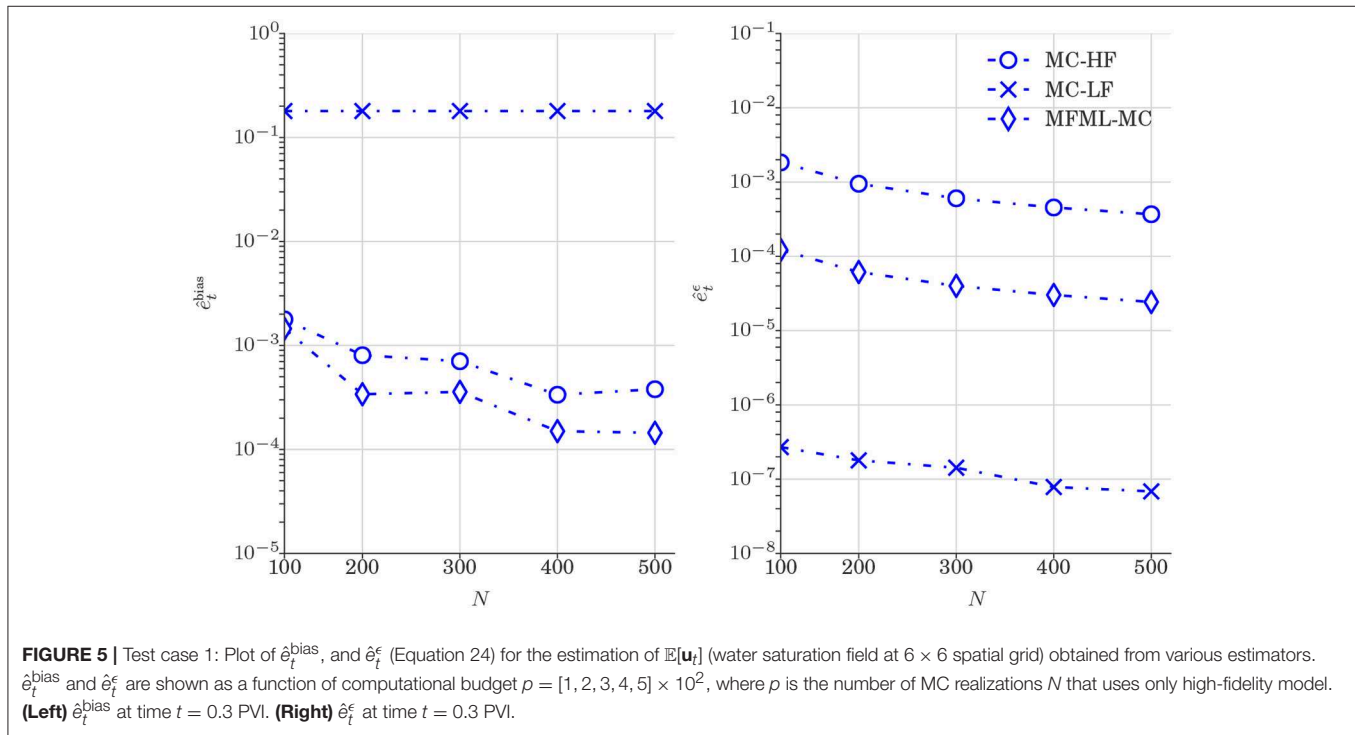
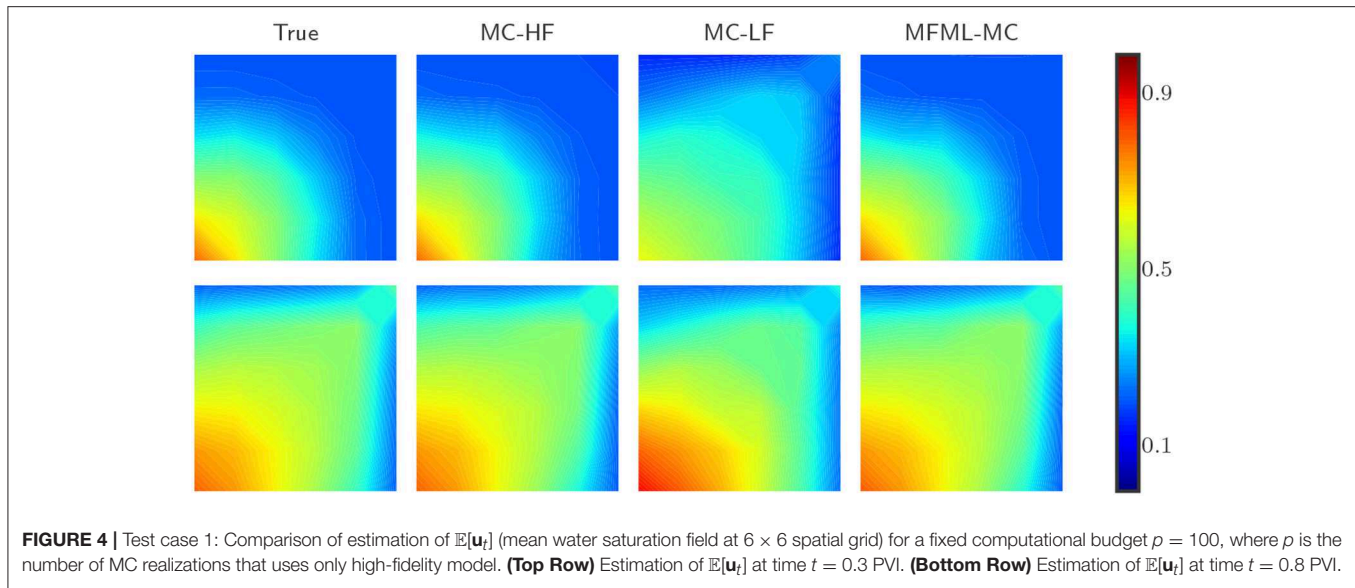


FIGURE 3 | Test case 1. (Left) Two dimensional quarter-five spot problem set-up where water is injected in the lower left corner (blue dot). Oil is displaced and produced with water in the upper right corner (blue dot). The red dots denotes spatial locations in the porous media domain where the statistics of the QoI \mathbf{u} are investigated. (Right) Decay of singular values of the snapshot matrix \mathbf{X}_U .



model. We also note that, $\hat{\mathbf{u}}_t^{(j)}$ is evaluated from a different set of independent samples for set $j = 1 \dots N_e$.

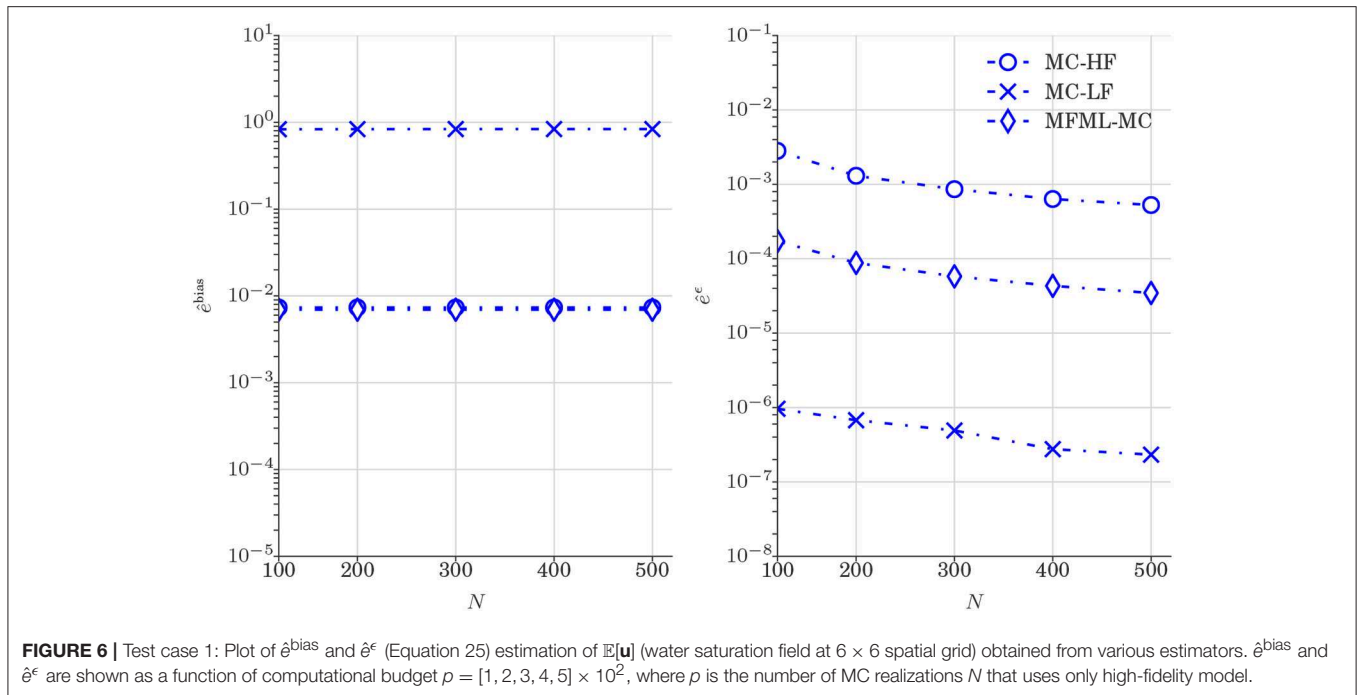
Additionally, we utilize two global error metrics defined as

$$\begin{aligned} \hat{e}_t^{\text{bias}} &= \frac{1}{N_e} \sum_{j=1}^{N_e} \max_{t=1}^T \|\hat{\mathbf{u}}_t^{\text{ref}} - \hat{\mathbf{u}}_t^{(j)}\|_2^2 \\ \hat{e}_t^\epsilon &= \frac{1}{N_e} \sum_{j=1}^{N_e} \max_{t=1}^T \text{Var}(\hat{\mathbf{u}}_t^{(j)}) \end{aligned} \quad (25)$$

where all the time snapshots of \mathbf{u} are used. We set $N_e = 15$ to evaluate the two time specific error metrics and the two global error metrics.

5.5. Numerical Test Case 1

Test case 1 is two dimensional quart-five spot problem where water is injected in the lower left corner $(0,0)$ of the porous media domain to produce oil and water in the top right corner $(1,1)$ (Kani and Elsheikh, 2018). We set q defined in the saturation equation (Equation 3) to 0.05 at $(0,0)$ and -0.05



at (1,1). We set no flux boundary condition in all the four sides of the porous media domain. The left panel of **Figure 3** displays the quart-five spot problem set up and the right panel of **Figure 3** displays the decay of the singular values of the snapshot matrix \mathbf{X}_u .

Figure 4 shows the results for the estimation of $\mathbb{E}[\mathbf{u}_t]$ (first moment of \mathbf{u}) obtained from the reference result (MC estimate with 25,000 samples) and from various MC estimators. In **Figure 4**, MC estimator that uses only high-fidelity model is denoted as MC-HF and that uses only low-fidelity model is denoted as MC-LF. In **Figure 4**, results shown in the top row are obtained at time = 0.3 PVI and results shown in the bottom row are obtained at time = 0.8 PVI. As shown in **Figure 4**, the estimation of $\mathbb{E}[\mathbf{u}_t]$ obtained from MC-LF deviates significantly from the reference result. This clearly shows that utilizing only low-fidelity model in MC framework resultant in biased estimation with respect to the reference result. Furthermore, **Figure 4** shows that the estimation of $\mathbb{E}[\mathbf{u}_t]$ obtained from MFML-MC estimator is almost indistinguishable from the reference result. This result confirm that combining higher number of low-fidelity model realizations with the high-fidelity model in MFML-MC framework can improve the estimator of the first moment of the saturation field.

Figure 5 reports the comparison of $\hat{\epsilon}_t^{\text{bias}}$ and $\hat{\epsilon}_t^{\epsilon}$ (see Equation 24) obtained from various estimators. The left of **Figure 5** reports $\hat{\epsilon}_t^{\text{bias}}$ and the right of **Figure 5** reports $\hat{\epsilon}_t^{\epsilon}$ as a function of computational budget $p = [1, 2, 3, 4, 5] \times 10^2$, where p is the number of MC-HF realizations. The results of $\hat{\epsilon}_t^{\text{bias}}$ from **Figure 5** shows that Monte Carlo estimator that uses MC-LF is a biased estimator of the mean QoI value. The results of MFML-MC estimator displayed in left of **Figure 5** confirm that the MFML-MC estimator is an unbiased estimator of the

TABLE 1 | Performance chart of MFML-MC estimator for test case 1.

ϵ	p	CPU Time (min)		Speedup
		MC-HF	MFML-MC	
10^{-4}	5×10^2	125	15	8.3
10^{-5}	9×10^3	2,250	210	10.5
10^{-6}	25×10^3	6,250	490	13.4

ϵ defined in Equation (5) is shown as a function of computational budget p , where p is the number of MC realizations that uses the high-fidelity model only. ϵ is estimated at time = 0.3 PVI.

expectation. This shows that despite the low-fidelity model is a poor approximation of the high-fidelity model, the error of the MFML-MC estimator can be significantly reduced if the low-fidelity model is combined with the high-fidelity. The right of **Figure 5** shows that the variance of the MFML-MC and MC-LF estimators are at least an order of magnitude less when compared to MC-HF. Nevertheless, while MC-LF is a biased estimator as shown in left of **Figure 5**, MFML-MC estimator that uses the low-fidelity model in combination with the high-fidelity model is an unbiased estimator of the expectation.

Figure 6 reports the comparison of $\hat{\epsilon}^{\text{bias}}$ and $\hat{\epsilon}^{\epsilon}$ (see Equation 25) obtained from various estimators. We can clearly observe the trend of $\hat{\epsilon}^{\text{bias}}$, and $\hat{\epsilon}^{\epsilon}$ in **Figure 6** are similar to the one observed in **Figure 5** which confirms that MFML-MC method leads to variance reduction with unbiased estimation at all time steps.

Table 1 compare the speedup factors of MFML-MC method with respect to the Monte Carlo estimator that uses the high-fidelity model only. In **Table 1**, MFML-MC achieves a speedup with respect to MC-HF that range from 8 up to 15 for the same specific ϵ .

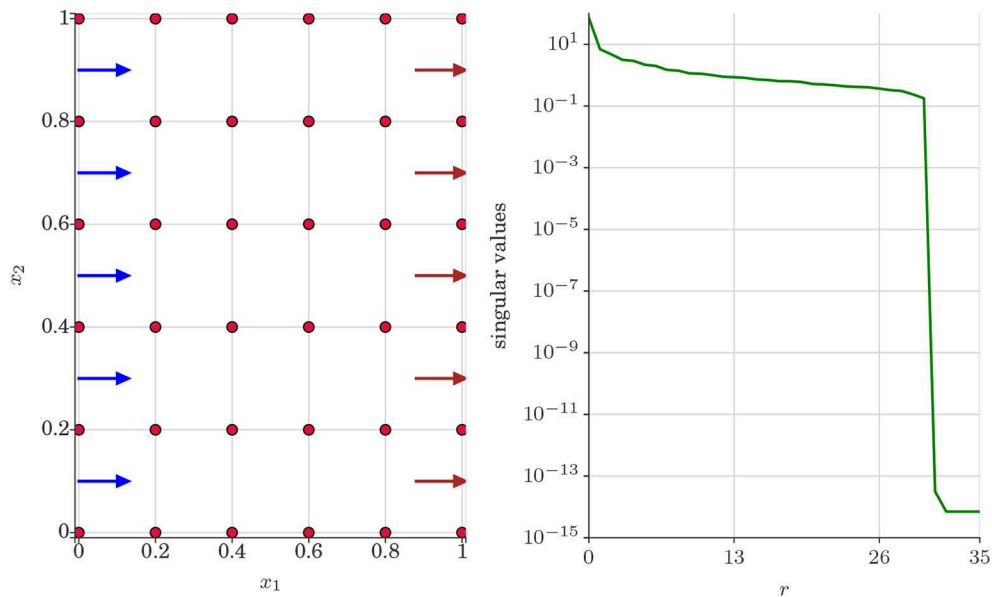


FIGURE 7 | Test case 2: **(Left)** Uniform flow problem set up where water is injected from the left side denoted by blue arrows. Oil and water are produced from the right side denoted by brown arrows. The red dots denotes the spatial locations of the porous media domain where the statistics of the QoI \mathbf{u} are investigated. **(Right)** Decay of singular values of the snapshot matrix \mathbf{X}_U .

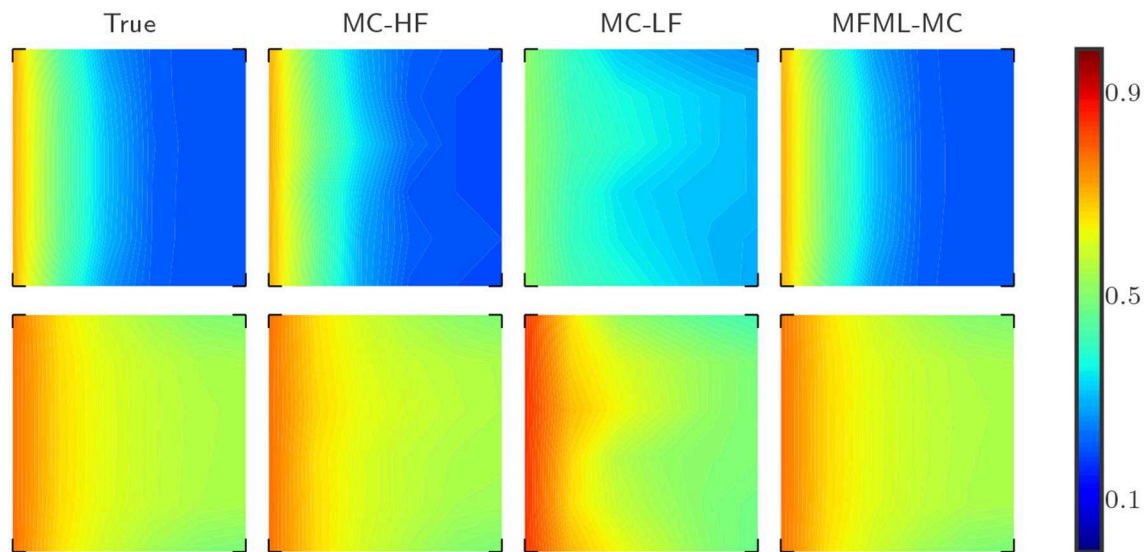


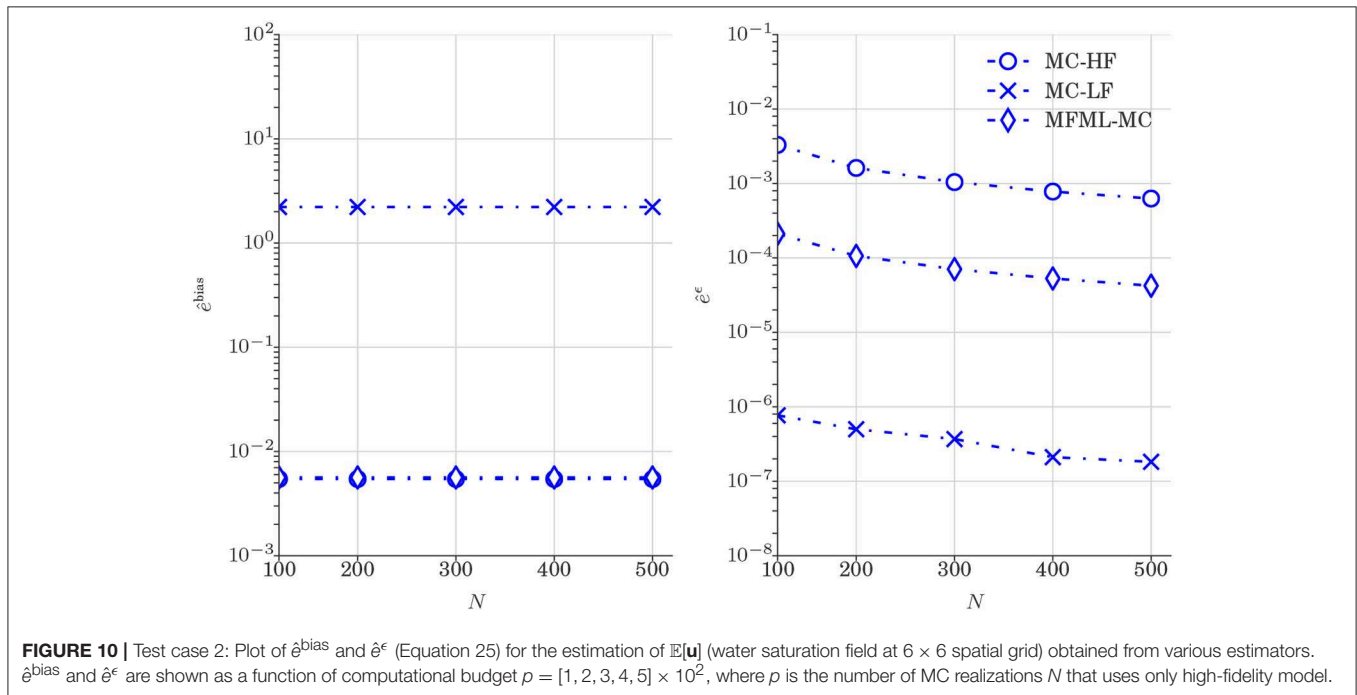
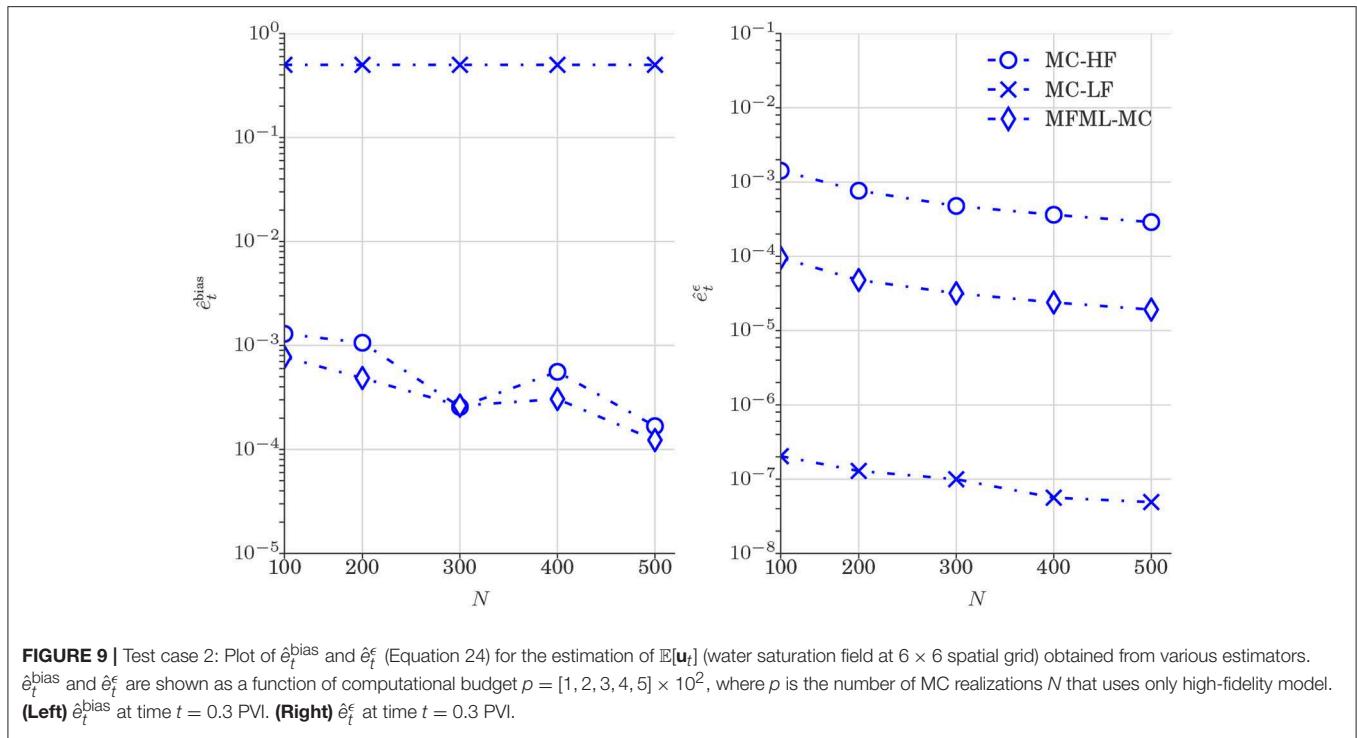
FIGURE 8 | Test case 2: Comparison of estimation of $\mathbb{E}[\mathbf{u}_t]$ (mean water saturation field at 6×6 spatial grid) for a fixed computational budget $p = 100$, where p is the number of MC realizations that uses only high-fidelity model. **(Top Row)** Estimation of $\mathbb{E}[\mathbf{u}_t]$ at time $t = 0.3$ PVI. **(Bottom Row)** Estimation of $\mathbb{E}[\mathbf{u}_t]$ at time $t = 0.8$ PVI.

5.6. Numerical Test Case 2

Test case 2 is a two dimensional uniform flow problem where water is injected from the left side of the porous media domain to produce oil and water from the right side. We set no flow boundary conditions in the remaining two sides (top and bottom) of the domain. We set inflow rate to 0.08 and outflow rate to 0.08 due to incompressibility constraint set in the problem (Kani and Elsheikh, 2018). The left panel of **Figure 7** displays uniform flow

problem set up and the right panel of **Figure 7** displays the decay of the singular values of the snapshot matrix \mathbf{X}_U .

Figure 8 shows the results for the first moment of the saturation field (\mathbf{u}) obtained from the reference result (MC estimate with 25000 samples) and from various MC estimators. The display settings defined in **Figure 8** are the same as the one defined in **Figure 4**. In **Figure 8**, we can see that the results obtained from MFML-MC method are almost indistinguishable



from the reference results whereas MC-LF yields extremely inaccurate results.

Figure 9 reports the comparison of \hat{e}_t^{bias} and \hat{e}_t^ϵ (see Equation 24) obtained from various estimators. The variance reduction can be clearly observed in **Figure 9** and the trend of **Figure 9** is similar to the one observed in **Figure 5** (Test case 1).

The results of **Figure 9** again confirm that combining the high-fidelity model with the low-fidelity model leads to a variance reduction. Please note that a similar confirmation was observed in **Figure 5**.

Figure 10 reports the comparison of \hat{e}_t^{bias} and \hat{e}_t^ϵ (see Equation 25) obtained from various estimators. As observed in

TABLE 2 | Performance chart of MFML-MC estimator for test case 2.

ϵ	p	CPU Time (min)		Speedup
		MC-HF	MFML-MC	
10^{-4}	5×10^2	148	14	10.8
10^{-5}	9×10^3	2,850	197	14.5
10^{-6}	25×10^3	7,950	410	19.4

ϵ defined in Equation (5) is shown as a function of computational budget p , where p is the number of MC realizations that uses the high-fidelity model only. ϵ is estimated at time = 0.3 PVI.

Figure 6, the results displayed in **Figure 9** shows that MFML-MC method leads to variance reduction with unbiased estimation.

Table 2 compare the speedup factors of MFML-MC method with respect to the MC method that uses the high-fidelity model only. In **Table 2**, MFML-MC achieves a speedup with respect to MC-HF that range from 10 up to 19 at a specific ϵ .

6. CONCLUSION

In this paper, we proposed a MFML-MC method combining the features of both the MFMC method and the MLMC method. In MFML-MC method, we formulated MLMC framework with a sequence of POD approximations of high-fidelity model outputs. Furthermore, in MFML-MC method, we formulated a MFMC setup on every level of MLMC framework in order to compute an unbiased statistical estimation. Finally, we utilized GBTR in the MFMC setup to formulate a level specific low-fidelity model.

We applied MFML-MC method on two uncertainty quantification problems involving two-phase flows in random

heterogeneous porous media where standard MLMC method with POD-Galerkin ROMs is ineffective. The uncertain permeability field is modeled from log-normal distribution function with exponential covariance function. Estimate of the first statistical moments of the water saturation at uniformly selected spatial grid points over a specific instant in time are calculated by MFML-MC, MC-HF, and MC-LF methods. Comparisons between MFML-MC and MC-LF suggested that MC-LF as a biased estimator and MFML-MC estimator as an unbiased estimator of the expectation. Comparisons between the MFML-MC and MC-HF computing times showed speedups of MFML-MC with respect to MC-HF that ranged from 8 up to 19 at equivalent accuracy.

Future work should consider the extension of MFML-MC method by utilizing two or more level specific low-fidelity models in the MFMC setup. In addition, it will also be interest to use MFML-MC method for history matching (Elsheikh et al., 2012, 2013), where we aim to minimize the mismatch between field observation data and the one computed from the high-fidelity model simulations by adjusting the geological model parameters. Future work should also verify the applicability of MFML-MC method for large-scale realistic problems with many wells and time varying injection rates by which the potential of MFML-MC method in speeding up a realistic Monte Carlo simulation can be magnified.

AUTHOR CONTRIBUTIONS

NJ developed the algorithm, coded the algorithm in python and obtained the results, and wrote the manuscript. AE is the Ph.D. supervisor of NJ. NJ did this paper under the guidance of AE.

REFERENCES

- Aarnes, J. E., Gimse, T., and Lie, K. A. (2007). "An introduction to the numerics of flow in porous media using Matlab," in *Geometric Modelling, Numerical Simulation, and Optimization* (Oslo: Springer), 265–306.
- Abdulle, A., Barth, A., and Schwab, C. (2013). Multilevel monte carlo methods for stochastic elliptic multiscale pdes. *Multisc. Model. Simulat.* 11, 1033–1070. doi: 10.1137/120894725
- Antoulas, A. C., Sorensen, D. C., and Gugercin, S. (2001). A survey of model reduction methods for large-scale systems. *Contemp. Math.* 280, 193–220. doi: 10.1090/conm/280/04630
- Babuška, I., Nobile, F., and Tempone, R. (2007). A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.* 45, 1005–1034. doi: 10.1137/050645142
- Barth, A., Schwab, C., and Zollinger, N. (2011). Multi-level monte carlo finite element method for elliptic pdes with stochastic coefficients. *Numerische Math.* 119, 123–161. doi: 10.1007/s00211-011-0377-0
- Bastian, P. (1999). *Numerical computation of multiphase flow in porous media*. (Ph.D. thesis). Habilitationsschrift Universität Kiel, Berlin, Germany.
- Berkooz, G., Holmes, P., and Lumley, J. L. (1993). The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.* 25, 539–575. doi: 10.1146/annurev.fl.25.010193.002543
- Bui-Thanh, T., Willcox, K., Ghattas, O., and Waanders, B. V. (2007). Goal-oriented, model-constrained optimization for reduction of large-scale systems. *J. Comput. Phys.* 224, 880–896. doi: 10.1016/j.jcp.2006.10.026
- Chaturantabut, S., and Sorensen, D. C. (2011). Application of POD and DEIM on dimension reduction of non-linear miscible viscous fingering in porous media. *Math. Comput. Modell. Dyn. Syst.* 17, 337–353. doi: 10.1080/13873954.2011.547660
- Chen, Z., Huan, G., and Ma, Y. (2006). *Computational Methods for Multiphase Flows in Porous Media*. Dallas, TX: SIAM.
- Cliffe, K. A., Giles, M. B., Scheichl, R., and Teckentrup, A. L. (2011). Multilevel monte carlo methods and applications to elliptic pdes with random coefficients. *Comput. Visualizat. Sci.* 14:3. doi: 10.1007/s00791-011-0160-x
- Codina, F. V., Ngoc, N. C., Giles, M. B., and Peraire, J. (2015). A model and variance reduction method for computing statistical outputs of stochastic elliptic partial differential equations. *J. Comput. Phys.* 297, 700–720. doi: 10.1016/j.jcp.2015.05.041
- Codina, F. V., Ngoc, N. C., Giles, M. B., and Peraire, J. (2016). An empirical interpolation and model-variance reduction method for computing statistical outputs of parametrized stochastic partial differential equations. *SIAM/ASA J. Uncertain. Quantificat.* 4, 244–265. doi: 10.1137/15M1016783
- Doostan, A., and Owhadi, H. (2011). A non-adapted sparse approximation of pdes with stochastic inputs. *J. Comput. Phys.* 230, 3015–3034. doi: 10.1016/j.jcp.2011.01.002
- Durlofsky, L. J., and Chen, Y. (2012). "Uncertainty quantification for subsurface flow problems using coarse-scale models," in *Numerical Analysis of Multiscale Problems*, eds I. Graham, T. Hou, O. Lakkis, and R. Scheichl (Berlin: Springer), 163–202.
- Efendiev, Y., Iliev, O., and Kronsbein, C. (2013). Multilevel monte carlo methods using ensemble level mixed msfm for two-phase flow and transport simulations. *Compu. Geosci.* 17, 833–850. doi: 10.1007/s10596-013-9358-y
- Efendiev, Y., Jin, B., Michael, P., and Tan, X. (2015). Multilevel markov chain monte carlo method for high-contrast single-phase flow problems.

- Commun. Comput. Phys.* 17, 259–286. doi: 10.4208/cicp.021013.260614a
- Elsakout, D., Christie, M., Lord, G., et al. (2015). “Multilevel markov chain monte carlo (mlmcmc) for uncertainty quantification,” in *SPE North Africa Technical Conference and Exhibition* (Cairo: Society of Petroleum Engineers).
- Elsheikh, A. H., Jackson, M., and Laforce, T. (2012). Bayesian reservoir history matching considering model and parameter uncertainties. *Math. Geosci.* 44, 515–543. doi: 10.1007/s11004-012-9397-2
- Elsheikh, A. H., Wheeler, M. F., and Hoteit, I. (2013). Nested sampling algorithm for subsurface flow model selection, uncertainty quantification, and nonlinear calibration. *Water Resour. Res.* 49, 8383–8399. doi: 10.1002/2012WR013406
- Fagerlund, F., Hellman, F., Målqvist, A., and Niemi, A. (2016). Multilevel monte carlo methods for computing failure probability of porous media flow systems. *Adv. Water Resour.* 94, 498–509. doi: 10.1016/j.advwatres.2016.06.007
- Frangos, M., Marzouk, Y., Willcox, K., and Waanders, B. V. (2010). *Surrogate and Reduced-Order Modeling: A Comparison of Approaches for Large-Scale Statistical Inverse Problems*. Alburquerque: John Wiley & Sons, Ltd.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232. doi: 10.1214/aos/1013203451
- Geraci, G., Eldred, M., and Iaccarino, G. (2015). “A multifidelity control variate approach for the multilevel monte carlo technique,” in *Center for Turbulence Research Annual Research Briefs* (Alburquerque).
- Ghanem, R. G., and Spanos, P. D. (1991). “Stochastic finite element method: response statistics,” in *Stochastic Finite Elements: A Spectral Approach* (New York, NY: Springer), 101–119.
- Ghasemi, M. (2015). *Model order reduction in porous media flow simulation and optimization*. (Ph.D. thesis). Texas AM Univeristy, Austin, TX.
- Giles, M. B. (2008). Multilevel monte carlo path simulation. *Operat. Res.* 56, 607–617. doi: 10.1287/opre.1070.0496
- Giles, M. B. (2013). “Multilevel Monte Carlo methods,” in *Monte Carlo and Quasi-Monte Carlo Methods 2012*, eds J. Dick, F. Kuo, G. Peters, and I. Sloan (Berlin: Springer), 83–103.
- Giles, M. B. (2015). Multilevel monte carlo methods. *Acta Numer.* 24, 259–328. doi: 10.1017/S096249291500001X
- He, J. (2010). *Enhanced linearized reduced-order models for subsurface flow simulation*. (M.S. thesis). Stanford Univeristy, Stanford, CA.
- He, J., Sætrom, J., and Durlöfsky, L. J. (2011). Enhanced linearized reduced-order models for subsurface flow simulation. *J. Comput. Phys.* 230, 8313–8341. doi: 10.1016/j.jcp.2011.06.007
- Heinrich, S. (2001). “Multilevel monte carlo methods,” in *Large-Scale Scientific Computing*, eds S. Margenov, J. Waśniewski, and P. Yalamov (Berlin: Heidelberg: Springer), 58–67.
- Ibrahim, F. (2016). *Probability distribution methods for nonlinear transport in heterogenous porous media*. (Ph.D. thesis). Stanford Univeristy, Stanford, CA.
- Jansen, J. D., and Durlöfsky, L. J. (2017). Use of reduced-order models in well control optimization. *Optimizat. Eng.* 18, 105–132. doi: 10.1007/s11081-016-9313-6
- Kani, J. N., and Elsheikh, A. H. (2018). “Reduced-order modeling of subsurface multi-phase flow models using deep residual recurrent neural networks,” in *Transport in Porous Media*, eds Z. Zhang, J. Chi, S. Sun, and Z. Pan (Springer), 1–29.
- Kebaier, A. (2005). Statistical romberg extrapolation: a new variance reduction method and applications to option pricing. *Ann. Appl. Probabil.* 15, 2681–2705. doi: 10.1214/105051605000000511
- Lassila, T., Manzoni, A., Quarteroni, A., and Rozza, G. (2014). “Model order reduction in fluid dynamics: challenges and perspectives,” in *Reduced Order Methods for Modeling and Computational Reduction*, eds A. Quarteroni and G. Rozza (Cham: Springer), 235–273.
- Li, H., and Zhang, D. (2007). Probabilistic collocation method for flow in porous media: comparisons with other stochastic methods. *Water Resour. Res.* 43. doi: 10.1029/2006WR005673
- Lin, G., and Tartakovsky, A. M. (2009). An efficient, high-order probabilistic collocation method on sparse grids for three-dimensional flow and solute transport in randomly heterogeneous porous media. *Adv. Water Resour.* 32, 712–722. doi: 10.1016/j.advwatres.2008.09.003
- Lu, D., Zhang, G., Webster, C., and Barbier, C. (2016). An improved multilevel monte carlo method for estimating probability distribution functions in stochastic oil reservoir simulations. *Water Resour. Res.* 52, 9642–9660. doi: 10.1002/2016WR019475
- Lucia, D. J., Beran, P. S., and Silva, W. A. (2004). Reduced-order modeling: new approaches for computational physics. *Prog. Aerospace Sci.* 40, 51–117. doi: 10.1016/j.paerosci.2003.12.001
- Mishra, S., Schwab, C., and Šukys, J. (2012). Multi-level monte carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions. *J. Comput. Phys.* 231, 3365–3388. doi: 10.1016/j.jcp.2012.01.011
- Mishra, S., Schwab, C., and Šukys, J. (2016). Multi-level monte carlo finite volume methods for uncertainty quantification of acoustic wave propagation in random heterogeneous layered medium. *J. Comput. Phys.* 312, 192–217. doi: 10.1016/j.jcp.2016.02.014
- Müller, F., Jenny, P., and Meyer, D. W. (2013). Multilevel monte carlo for two phase flow and buckley-leverett transport in random heterogeneous porous media. *J. Comput. Phys.* 250, 685–702. doi: 10.1016/j.jcp.2013.03.023
- Nagoor Kani, J., and Elsheikh, A. H. (2017). DR-RNN: a deep residual recurrent neural network for model reduction. *arxiv 1709.00939*.
- Ng, L. W.-T. (2013). *Multifidelity approaches for design under uncertainty*. (Ph.D. thesis). Massachusetts: Massachusetts Institute of Technology.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Peherstorfer, B., Willcox, K., and Gunzburger, M. (2016). Optimal model management for multifidelity monte carlo estimation. *SIAM J. Sci. Comput.* 38, A3163–A3194. doi: 10.1137/15M1046472
- Peherstorfer, B., Willcox, K., and Gunzburger, M. (2018). Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Rev.* 60, 550–591. doi: 10.1137/16M1082469
- Petvipisit, K. R., Elsheikh, A. H., Laforce, T. C., King, P. R., and Blunt, M. J. (2014). Robust optimisation of CO₂ sequestration strategies under geological uncertainty using adaptive sparse grid surrogates. *Comput. Geosci.* 18, 763–778. doi: 10.1007/s10596-014-9425-z
- Stefanou, G. (2009). The stochastic finite element method: past, present and future. *Comput. Methods Appl. Mech. Eng.* 198, 1031–1051. doi: 10.1016/j.cma.2008.11.007
- Wang, Z., Akhtar, I., Borggaard, J., and Iliescu, T. (2012). Proper orthogonal decomposition closure models for turbulent flows: a numerical comparison. *Comput. Methods Appl. Mech. Eng.* 237, 10–26. doi: 10.1016/j.cma.2012.04.015
- Wang, Z., Xiao, D., Fang, F., Govindan, R., Pain, C. C., and Guo, Y. (2017). Model identification of reduced order fluid dynamics systems using deep learning. *Int. J. Numer. Methods Fluids* 86, 255–268. doi: 10.1002/fld.4416
- Xiao, D., Lin, Z., Fang, F., Pain, C. C., Navon, I. M., Salinas, P., et al. (2017). Non-intrusive reduced-order modeling for multiphase porous media flows using smolyak sparse grids. *Int. J. Numer. Methods Fluids* 83, 205–219. doi: 10.1002/fld.4263

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Jabarullah Khan and Elsheikh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Using Meta-Models for Tsunami Hazard Analysis: An Example of Application for the French Atlantic Coast

Vito Bacchi^{1*}, Hervé Jomard¹, Oona Scotti¹, Ekaterina Antoshchenkova¹, Lise Bardet¹, Claire-Marie Duluc¹ and Hélène Hebert²

¹ Institute for Radiological Protection and Nuclear Safety, Fontenay-aux-Roses, France, ² CEA, DAM, DIF, Arpajon, France

OPEN ACCESS

Edited by:

Philippe Renard,
Université de Neuchâtel, Switzerland

Reviewed by:

Jacopo Selva,
Istituto Nazionale di Geofisica e
Vulcanologia (INGV), Italy
Tipaluck Krityakierne,
Mahidol University, Thailand

*Correspondence:

Vito Bacchi
vito.bacchi@gmail.com

Specialty section:

This article was submitted to
Solid Earth Geophysics,
a section of the journal
Frontiers in Earth Science

Received: 08 May 2019

Accepted: 04 February 2020

Published: 03 March 2020

Citation:

Bacchi V, Jomard H, Scotti O,
Antoshchenkova E, Bardet L,
Duluc C-M and Hebert H (2020)
Using Meta-Models for Tsunami
Hazard Analysis: An Example
of Application for the French Atlantic
Coast. *Front. Earth Sci.* 8:41.
doi: 10.3389/feart.2020.00041

This paper illustrates how an emulator (or meta-model) of a tsunami code can be a useful tool to evaluate or qualify tsunami hazard levels associated with both specific and unknown tsunamigenic seismic sources. The meta-models are statistical tools permitting to drastically reduce the computational time necessary for tsunami simulations. As a consequence they can be used to explore the tsunamigenic potential of a seismic zone, by taking into account an extended set of tsunami scenarios. We illustrate these concepts by studying the tsunamis generated by the Azores-Gibraltar Plate Boundary (AGPB) and potentially impacting the French Atlantic Coast. We first analyze the impact of two realistic scenarios corresponding to potential sources of the 1755-Lisbon tsunami (when uncertainty on seismic parameters is considered). We then show how meta-models could permit to qualify the tsunamis generated by this seismic area. All the results are finally discussed in light of tsunami hazard issued by the TSUMAPS-NEAM research project available online (<http://ai2lab.org/tsumapsneam/interactive-hazard-curve-tool/>). From this methodological study, it appears that tsunami hazard issued by TSUMAPS-NEAM research project is envelop, even when compared to all the likely and unlikely tsunami scenarios generated in the AGPB area.

Keywords: kriging surrogate, uncertainty quantification, tsunami modeling, hazard analysis, sensitivity analysis

INTRODUCTION

The evaluation of tsunami impact requires accurate simulation results for planning and risk assessment purposes because of the severe consequences which could be associated to this kind of event. Considering that tsunami phenomena involve a large span of parameters at different spatial and temporal scales (Behrens and Dias, 2015), even a single run of a tsunami numerical model can

Abbreviations: μ , shear modulus; $[N(m(x), s^2(x))]$, Gaussian process of mean " $m(x)$ " and variance " $s^2(x)$ "; $[M(x)]$, kriging surrogate; $\{X, Y\}$, are the coordinates of the design simulations used for kriging parameters evaluation; $C(\cdot)$, covariance kernel; CEA, commissariat à l'énergie atomique et aux énergies alternatives; D [m], average slip along the rupture surface; DTHA, deterministic tsunami hazard assessment; GSA, global sensitivity analysis; L [m], length of the rupture surface; MCS, maximum credible scenario; MCS_h, tsunami hazard level issued by an exploration of a very wide range of tsunamigenic scenarios; M_0 , seismic moment; M_w , seismic moment magnitude; MSE, mean squared error; PTHA, probabilistic tsunami hazard assessment; R^2 , squared correlation coefficient; RMSE, root mean squared error; UQ, uncertainty quantifications; W [m], width of the rupture surface.

be prohibitively long, in the order of minutes to days, according to the study area characteristics and to the resolution of the numerical model. Hence, a common practice when the computational code is time-consuming is the use of meta-models (also denoted surrogate-models or emulators). A meta-model is a mathematical model of approximation of the numerical model, built on a learning basis (Razavi et al., 2012a). Meta-models have been applied, for example, in hydraulic fields to model physical variables such as flows (Wolfs et al., 2015; Machac et al., 2016), flood damages (Yazdi and Salehi Neyshabouri, 2014), or in the field of the design for civil flood defenses (Richet and Bacchi, 2019). A comprehensive review of the use of meta-models in environmental research was proposed by Razavi et al. (2012a) for the interested reader.

Meta-models have also already been used in the field of tsunamis. For instance, Sraj et al. (2014) investigate the uncertainties in the resulting wave elevation predictions due to the uncertainty in the Manning's friction coefficient, using polynomial chaos expansion to build a surrogate model that is a computationally cheap approximation of the computer model. Sarri et al. (2012) used in a similar way a statistical emulator of the analytical landslide-generated tsunami model developed by Sammarco and Renzi (2008). More recently, Rohmer et al. (2018) studied the uncertainty related to the source parameters through a Bayesian procedure to infer (i.e., learn) the probability distribution of the source parameters of the earthquake. However, to our knowledge, meta-modeling has never been used for tsunami hazard analysis.

In this paper, we propose to apply meta-modeling techniques in the framework of deterministic tsunami hazard assessment (DTHA) and evaluate how it can be useful in seismic areas with no (or poor) seismotectonic knowledge. In such cases, when seismotectonic parameters are uncertain, it may be of interest to provide a first order idea of the tsunami hazard potential through DTHA, the implementation of DTHA being simpler than the probabilistic method (PTHA). The scenario-based (or DTHA) approach classically relies on the study of "maximum credible scenarios" (MCS). In particular, DTHA tries to explore the potential of the largest scenarios, by selecting some of the extreme ones (i.e., a recorded/reconstructed historical event) and simulating them for the target area through numerical modeling (JSCE, 2002; Lynett et al., 2004), without addressing the likelihood of occurrence of such a big event (Omira et al., 2016). With this approach, MCS is assessed through an expert opinion. The outputs of the deterministic analysis are, in general, tsunami travel time, wave height, flow depth, run-up, and current velocity maps corresponding to the chosen scenario (Omira et al., 2016).

As mentioned above, DTHA relies on a refined knowledge of the seismic sources generating the tsunamis. As a consequence, it could be hampered by the use of specific values of input parameters which may be subjective depending on the person or group carrying out the analysis (Roshan et al., 2016). A good example is the 1755 Lisbon tsunami, generated by an earthquake in the Azores Gibraltar Plate Boundary (AGPB). The Great 1755 Lisbon earthquake generated the most historically destructive tsunami near the Portugal coasts (Santos and Koshimura, 2015). Source location and contemporary effects of such tsunami are not

precisely identified and several earthquake scenarios have already been published in the literature in the last decades (Johnston, 1996; Baptista et al., 1998, 2003; Zitellini et al., 1999; Gracia et al., 2003; Terrinha et al., 2003; Gutscher et al., 2006; Grandin et al., 2007; Horsburgh et al., 2008; Barkan et al., 2009; Cunha et al., 2010). All of these studies show how variable the parameters of the seismic source can be and the importance to take into account their uncertainty.

In DTHA, the classical approach to deal with uncertainties consists in performing a limited number of deterministic simulations with conservative values of the seismic sources (e.g., JSCE, 2002; Lynett et al., 2004; Allgeyer et al., 2013). However, the great M 9.0 Tohoku-Oki subduction earthquake of 2011, the largest ever recorded in Japan (Saito et al., 2011), has clearly shown the limitations of the classical approach focused on the identification of known maximum tsunamigenic sources (MCS approach). Recently, Roshan et al. (2016) improved the DTHA procedure detailed in Yanagisawa et al. (2007) in order to better evaluate the effects of the seismic source uncertainties through Monte-Carlo simulations of a limited number of seismic source parameters (the dip angle, the strike and the source location), leading to around 300 tsunami scenarios. The authors presented an improvement of the classical MCS approach by introducing uncertainties on seismic source parameters.

In this context, the objective of this work is to propose a new methodology to evaluate or qualify tsunami hazard levels associated to both specific and unknown tsunamigenic seismic sources by integrating the uncertainty related to the seismic parameters in the DTHA procedure. The main idea is to develop a meta-model, or emulator, of the tsunami numerical model, that makes it possible to perform a large number of tsunami scenarios with reduced computational time and, consequently, to intensively explore a tsunamigenic area for which geological and geophysical datasets may be limited. The constructed meta-models, when exploited with the statistical criteria classically employed in uncertainty quantification (UQ) studies (Saltelli et al., 2000, 2008; Saltelli, 2002; Iooss and Lemaitre, 2015), can permit to go beyond the classical approach for DTHA and perform a better quantification of uncertainties of a large set of seismic source parameters.

In the next sections we will first present the methodology used in this study (see section "Methodology"), and then develop and validate the meta-models for AGPB related tsunamis impacting the French Atlantic Coast (see section "Application of the methodology to the French Coast"). In section "Potential application of meta-models for tsunami hazard analysis," we (1) evaluate the impact of two realistic scenarios corresponding to potential sources of the 1755-Lisbon tsunami (when uncertainty on seismic parameters is considered), and (2) present the analysis of the tsunamigenic potential of the AGPB zone, considered for the purpose of the exercise as a poorly known tsunamigenic area. Finally, we show how the numerical results obtained with this method can be discussed in light of tsunami hazard issued by the TSUMAPS-NEAM research project available on line¹.

¹<http://ai2lab.org/tsumapsneam/interactive-hazard-curve-tool/>

METHODOLOGY

The proposed approach relies on three main steps, as reported in **Figure 1**. STEP 1 consists in the construction of a numerical model able to reproduce the tsunami heights generated by a given seismic area and impacting a target area. In STEP 2, the numerical model simulates a regular set of physical tsunamigenic scenarios (so called design data-base) that are used for the construction and the validation of an emulator (the meta-model) able to reproduce the original model results in the target zone. In STEP 3, the validated meta-models may be used for DTHA assessment and/or qualification of other THA results. The UQ performed using meta-models instead of the original model permits to assess the uncertainty related to a given tsunami scenario (see section “In the case of expert opinion: uncertainty quantification in DTHA”) and to explore intensively the tsunamigenic area with nearly zero computational time (see section “Without expert opinion: exploration of a large tsunamigenic area”).

Step 1: Tsunami Simulations

The first step consists in the construction of a tsunami numerical model of the area to explore. In this study, the tsunami numerical simulations were performed by using the tsunami code reported in Allgeyer et al. (2013) (or CEA-code), which exploits two models, one for tsunami initialization and the other one for tsunami propagation. The initial seabed deformation caused by an earthquake is generated with the Okada model (Okada, 1985) and is transmitted instantaneously to the surface of the water.

This analytical model uses simplistic planar fault parameters with uniform slip, satisfying the expression of the seismic moment M_0 :

$$M_0 = \mu \cdot D \cdot L \cdot W \quad (1)$$

where μ denotes the shear modulus, D [m] the average slip along the rupture of length L [m] and width W [m]. Then, the seismic moment magnitude “ M_w ” is directly computed through equation 2 (Hanks and Kanamori, 1979), as follow:

$$M_w = \frac{2}{3} \cdot \log_{10}(M_0) - 6, 07 \quad (2)$$

The following parameters are also required for tsunami initiation: *longitude*, *latitude*, and *depth* [km] of the center of the source, *strike* [degrees], *dip* [degrees], and *rake* [degrees]. A conceptual scheme of the input parameters for the tsunami-code and the numerical domain used in this study are reported in **Figure 2**. Then, the computation of the tsunami propagation is based on hydrodynamic equations, under the non-linear shallow water approximation (the Boussinesq equations as reported in Allgeyer et al., 2013). Shallow water equations are discretized using a finite-differences method in space and time (FDTD). Pressure and velocity fields are evaluated on uniform separate grids according to Arakawa’s C-grid (Arakawa, 1972). Partial derivatives are approximated using upwind finite-differences (Mader, 2004). Time integration is performed using the iterated Crank-Nicholson scheme. No viscosity terms are taken into account in our simulations. The only parameters of this model are the bathymetry (space step and depth resolution) and the time step.

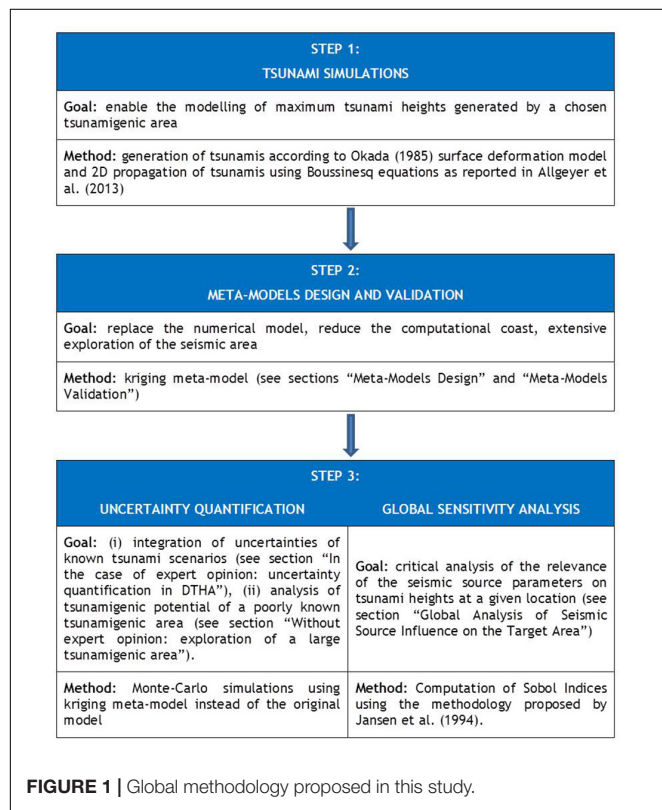
Step 2: Meta-Model Design and Validation

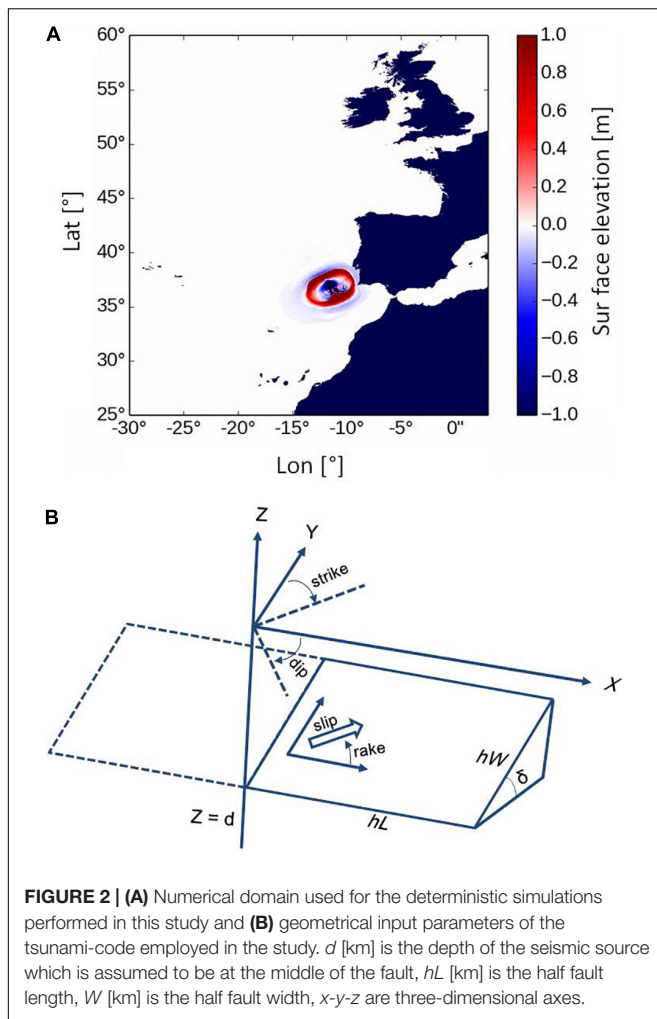
Meta-Models Design

A variety of metamodels have been applied in the water resources literature (Santana-Quintero et al., 2010; Razavi et al., 2012b). Moreover, some examples of applications have already been published in the context of flood management (e.g., Yazdi and Salehi Neyshabouri, 2014; Löwe et al., 2018) and in the field of tsunamis (e.g., Sarri et al., 2012; Sraj et al., 2014; Rohmer et al., 2018). The classical steps for meta-models construction and validation are reported in various studies (i.e., Saltelli, 2002; Saltelli et al., 2008; Faivre et al., 2013), and are shortly summarized in **Table 1**.

For the study presented here, we rely on conditional Gaussian processes (also known as kriging (Roustant et al., 2012), derived from Danie Krige’s pioneering work in mining (Krige, 1951), later formalized within the geostatistical framework by Matheron (1963). Kriging meta-model has already shown good predictive capacities in many practical applications (see Marrel et al., 2008, for example), it became a standard meta-modeling method in operational research (Santner et al., 2003; Kleijnen, 2005) and it has performed robustly in previous water resource applications (Razavi et al., 2012b; Villa-Vialaneix et al., 2012; Löwe et al., 2018).

A general kriging model “ $M(x)$ ” (which later provides an estimation of the maximum tsunami height in a given location)





can be defined for $x = (x_1, \dots, x_d) \in D \in \mathbb{R}^d$ as the following Gaussian process “ $N(\cdot)$ ”:

$$M(x) = N(m(x), s^2(x))$$

Where, for simple kriging:

- $m(x) = C(x)^T C(X)^{-1} Y$ is the conditional mean;

- $s^2(x) = c(x) - C(x)^T C(X)^{-1} C(x)$ is the conditional variance;
- $C : (u, v) \in D^2 \rightarrow C(u, v) \in \mathbb{R}$ is the covariance Kernel;
- $c(x)$ is the vector of covariance between the kriging predictions and the original model evaluations.

It must be noted that $\{X, Y\}$ are the coordinates of the design simulations used for kriging parameters evaluation:

- X is a matrix containing for each design simulation (for a total of N simulations) the value of the seismic source model parameters defined in section “Step 1: tsunami simulations” (for a total of 9 inputs parameters):

$$X = \begin{pmatrix} x_{1,1} & \dots & x_{1,9} \\ \vdots & & \vdots \\ x_{N,9} & \dots & x_{N,9} \end{pmatrix}$$

- Y is a vector containing, for each gauge, the maximum water depth associated to a simulation:

$$Y = \begin{pmatrix} h_{\max,1} \\ \vdots \\ h_{\max,N} \end{pmatrix}$$

More than a commonplace deterministic interpolation method (like splines of any order) this model is much more informative owing to its predicted expectation and uncertainty. The fitting procedure of this model includes the choice of a covariance model [here a tensor product of the “Matern5₂” function (Roustant et al., 2012)], and then the covariance parameters (e.g., range of covariance for each input variable, variance of the random process, nugget effect), could be estimated using Maximum Likelihood Estimation (standard choice we made) or Leave-One-Out minimization [known to mitigate the arbitrary covariance function choice (see Bachoc, 2013)].

It must be noted that the kriging interpolation technique requires computing and inverting the $n \times n$ covariance matrix $C(X, X)$ between the observed values $Y(X)$, which leads to a $O(n^2)$ complexity in space and $O(n^3)$ in time (Rullière et al., 2018). In practice, this computational burden makes Gaussian process regression difficult to use when the number of observation points is in the range $[10^3, 10^4]$ or greater, as in this study. As a consequence, we used in this article the procedure for estimating the parameters of kriging reported in Rullière et al. (2018), by using an adapted R-tool available on line². The full details of this methodology are reported in the abovementioned paper. This approach is proven to have better theoretical properties than other aggregation methods that can be found in the literature, and permitted us to drastically reduce the computational time necessary for meta-models construction and validation.

²<https://github.com/drulliere/nestedKriging>

TABLE 1 | Steps for meta-models construction, validation and evaluation of the uncertainty.

Step for meta-models design	Description
Initial design database	Monte-Carlo sampling in the distribution of the model input parameters.
Construction	Estimation of meta-model parameters according to equations reported in section “Meta-models design” using training set data base
Accuracy/Optimization	k-fold cross validation (Breiman and Spector, 1992; Kohavi, 1995; Hastie et al., 2009).
Meta-model uncertainty	RMSE and R^2 issued from the Accuracy/Optimization step.

Meta-Models Validation

A general method for meta-models validation is the K-fold cross-validation method (Friedman et al., 2001). The principle of cross-validation is to split the data into K folds of approximately equal size A_1, A_2, \dots, A_K . For $k = 1$ to K , a model $\hat{Y}^{(-k)}$ is fitted from the data $U_{j \neq k} A_k$ (all the data except the A_k fold) and this model is validated on the fold A_k . Given a criterion of quality L as the Mean Square Error:

$$L = MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3)$$

the quantity used for the “evaluation” of the model is computed as follow:

$$L_k = \frac{1}{n/K} \sum_{i \in A_k} L(y_i Y^{(-k)}(x_i)), \quad (4)$$

where \hat{y}_i and y_i are, respectively, the meta-model and the model response and n is the number of simulations in the k^{th} sample. When K is equal to the number of simulations of the training set, the cross-validation method corresponds to the leave-one-out technique not performed in this study. The methodology employed is described in the DiceEval R-package reference-manual (Dupuy et al., 2015). In our application case, we considered $K = 10$.

In this study, the accuracy of the meta-model is evaluated through several statistical metrics permitting to quantify the overall quality of regression models. This includes:

- **R-squared (R^2)**, representing the squared correlation between the observed outcome values and the values predicted by the model. The higher the adjusted R^2 is, the better the model is;
- **Root Mean Squared Error (RMSE)**, which measures the average prediction error made by the model in predicting the outcome for an observation. It corresponds to the average difference between the observed known outcomes and the values predicted by the model. The lower the RMSE is, the better the model is.

Step 3: Uncertainty Quantification and Global Sensitivity Analysis

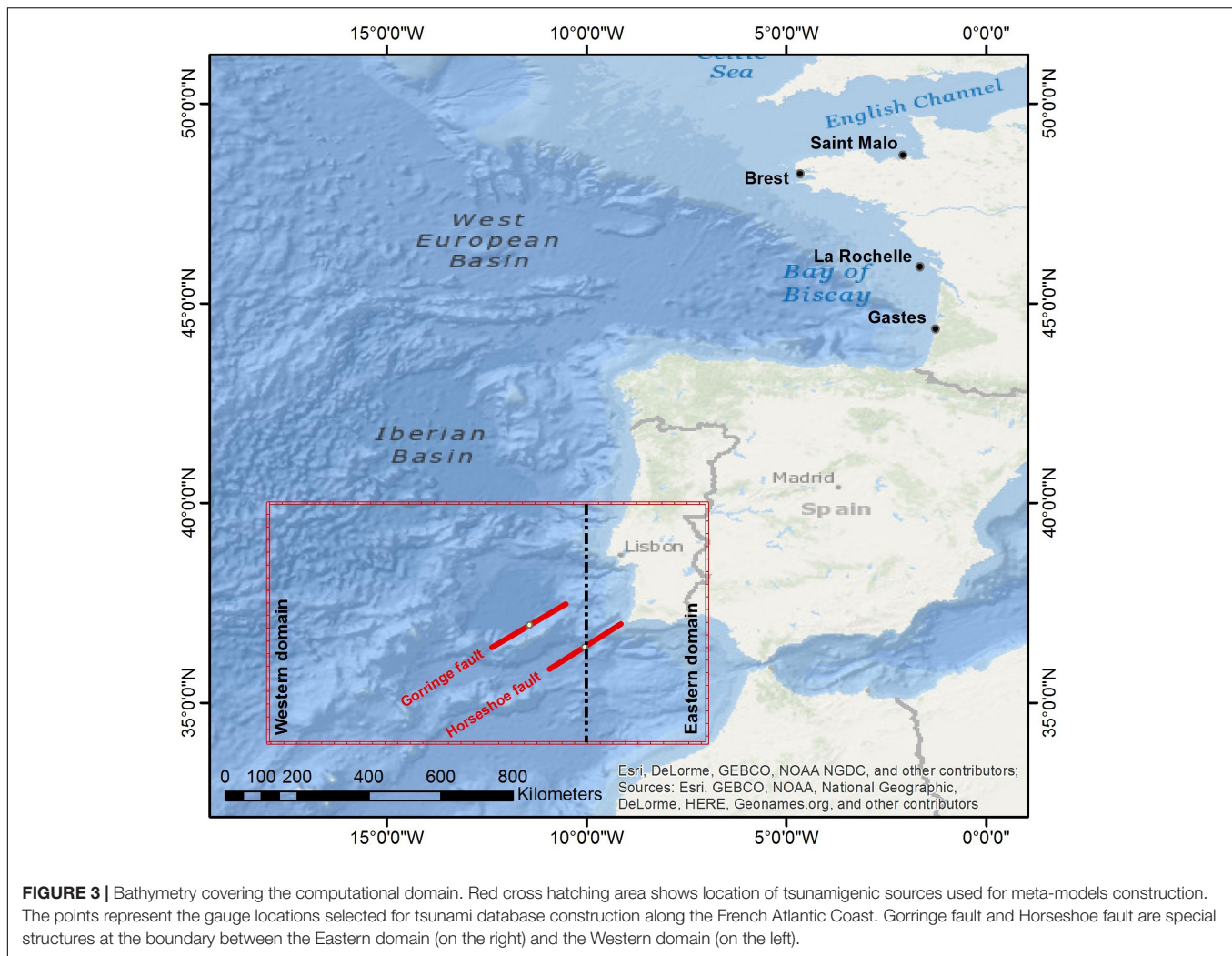
Considering the variety and the complexity of the geophysical mechanisms involved in tsunami generation, tsunami hazard assessment is generally associated with strong uncertainties (aleatory and epistemic). In PTHA, uncertainties are classically integrated in a rigorous way (Sørensen et al., 2012; Horspool et al., 2014; Selva et al., 2016) and quantified using the logic-tree approach (Horspool et al., 2014) and/or random simulations performed using the Monte-Carlo sampling of probability density functions of geological parameters (Sørensen et al., 2012; Horspool et al., 2014). An alternative and interesting approach was recently proposed by Selva et al. (2016), consisting in the use of an event tree approach and ensemble modeling (Marzocchi et al., 2015). Moreover, a new procedure was recently proposed by Molinari et al. (2016) for the quantification of uncertainties related to the construction of a tsunami data-base based on the quantification of elementary effects.

In this work we propose a classical methodology that could also be adapted to analyze tsunamigenic regions with poor (or no) information on crustal characteristics and based on the classical uncertainty study steps (Saltelli et al., 2004, 2008; Faivre et al., 2013; Iooss and Lemaître, 2015). This methodology, which was already tested in other hydraulic context in recent years (Nguyen et al., 2015; Abily et al., 2016), relies on Monte Carlo simulations for UQ steps and on GSA (Global Sensitivity Analysis) approaches for the analysis of the AGPB tsunamigenic potential, by computing Sobol indices (Sobol, 1993, 2001). These methods rely on sampling based strategies for uncertainty propagation, willing to fully map the space of possible model predictions from the various model uncertain input parameters and then, allowing to rank the significance of the input parameter uncertainty contribution to the model output variability (Baroni and Tarantola, 2014). The objectives with this approach are mostly to identify the parameter or set of parameters which significantly impact model outputs (Iooss et al., 2008; Volkova et al., 2008). GSA approaches are robust, have a wide range of applicability, and provide accurate sensitivity information for most models (Adetula and Bokov, 2012). Moreover, even if they are theoretically defined for linear mathematical systems, it was demonstrated that they are well suited to be applied with models having non-linear behavior and when interactions among parameters occur (Saint-Geours, 2012), as in the present study. For these reasons, these indices were already adopted for the analysis of bi-dimensional hydrodynamic simulations in urban areas (Abily et al., 2016) or of complex coastal models including interactions between waves, current and vegetation (Kalra et al., 2018) and they seem well suited for the present work.

For the computation of Sobol' indices, a large variety of methodologies are available, as the so-called “extended-FAST” method (Saltelli et al., 1999), already used in previous studies by IRSN (Nguyen et al., 2015). In this study, we used the methodology proposed by Jansen et al. (1994) already implemented in the open source sensitivity-package R (Pujol et al., 2017). This method estimates first order and total Sobol' indices for all the factors “ v ” at higher total cost of “ $v \times (p + 2)$ ” simulations (Faivre et al., 2013).

APPLICATION OF THE METHODOLOGY TO THE FRENCH COAST

The French Atlantic coast is subjected to two main seismogenic sources that could generate tsunamis, one in the lesser Indies, and a second one from the AGPB. In this application we only consider the AGPB and we only compute water heights offshore for four locations (**Figure 3**), ignoring the necessary refinements for propagation to the coast. Because in this study case the sources are far from the considered gauges, we propose a very simplified approach to characterize the source region. In the following we first present how the meta-model was constructed and validated through a series of statistical tests in comparison with published data from Allgeyer et al. (2013).



Numerical Tool and Design Data-Base

All the simulations were performed on the same bathymetric grid with a space resolution of $2'$ (~ 3.6 km). The numerical model was not directly validated by the comparison with similar simulations from literature. In fact, considering the rough bathymetrical grid resolution, the developed numerical model is not adapted to the estimation of the tsunamis run-up and the inundation areas and it can't be used for a real assessment of the tsunami hazard along the French Atlantic Coast. However, this work being methodological, we consider that the numerical results are consistent with the objectives of the study. Moreover, the tsunami-code was largely validated through extensive benchmarks in the framework of the TANDEM research project (Violeau et al., 2016) by ensuring its ability to reproduce tsunamis generation and propagation. As a consequence, the order of magnitude of the tsunami heights computed in this study should be realistic and adapted to the test of the methodology.

In order to perform the numerical simulations needed for the meta-model construction and validation (see section "Step 2: meta-model design and validation"), the CEA-code

was coupled with the IRSN Promethee bench. Promethee is an environment for parametric computation that allows carrying out UQ studies, when coupled (or warped) to a code. This software is freely distributed by IRSN³ and allows the parameterization with any numerical code and is optimized for intensive computing. Promethee was first linked to the numerical code by means of a set of software links (similar to bash scripts). In this way, numerical simulations were directly launched by the IRSN environment. Then, the statistical analysis, such as the Monte-Carlo simulations used for the meta-model construction (see section "Meta-models design") and UQ (see section "Analysis of the Impact of Tsunamis on a Target Area through UQ") or the Sobol indices computation (see section "Global Analysis of Seismic Source Influence on the Target Area") were also driven by Promethee, which integrates R statistical computing environment by permitting this kind of analysis (R Core Team, 2016).

In this methodological work, we considered a widened AGPB tsunamigenic area and we chose to explore as largely as possible

³<http://promethee.irsn.org/doku.php>

TABLE 2 | Summary of the variation range of the seismic source input parameters for the design, the western, the eastern database and for the tsunami scenarios associated to the Gorrige bank and the Horseshoe bank (hypothesis from Duarte et al., 2013; Grevenmeyer et al., 2017).

Seismic source parameter*	Design database	Western database	Eastern database	Gorrige bank	Horseshoe bank	Global database**
Strike [degrees]	0 – 360	0 – 360	0 – 360	40 – 70	40 – 70	0 – 360
Length [km]	40 – 600	40 – 227	40 – 600	180 – 200	160 – 200	40 – 400
Dip [degrees]	1 – 90	60 – 90	10 – 90	10 – 40	10 – 40	10 – 90
Rake [degrees]	–180 : +180	–180 : 0	0 : +180	70 : 110	70 : 110	–180 : +180
Width [km]	10 – 310	10 – 23	10 – 310	16 – 224	16 – 228	20 – 200
Slip [m]	1 – 25	1 – 25	1 – 25	5 – 25	5 – 25	1 – 25
Longitude [degrees]	–18 : –7	–18 : –10	–10 : –7	–12 : –11	–10.5 : –9.5	–18 : –7
Latitude [degrees]	34 – 40	34 – 40	34 – 40	36.5 – 37.5	35.8 – 36.5	34 – 40
Depth*** [km]	2 – 60	5 – 10	5 – 30	10 – 40	10 – 40	5 – 30
Sismogenic depth [km]	4 – 120	10 – 20	10 – 60	20 – 80	20 – 80	10 – 60
Magnitude range M_w [-]	6.7 – 9.3	6.7 – 8.3	6.8 – 9.3	7.7 – 8.9	7.7 – 8.9	6.7 – 9.3

*Seismic source parameters are assumed uniformly distributed and are randomly sampled for the construction of the Global database. **The global database is composed of the scenarios simulated with the meta-models (more than 50,000 tsunamis scenarios) for the construction of the western and the eastern database. ***Note that depth is considered to be the depth of the seismic source which is assumed to be in the middle of the fault (see Figure 2).

the potential tsunami height along the French Atlantic Coast generated by earthquakes from 34° to 40°N and from 18 to 7°W, encompassing to the East the southern part of Portugal down to Morocco, and reaching the oceanic sea-floor west of the Madeira-Tore rise (as reported in Figure 3). Because the design database is a learning base for meta-modeling, the range of variation of the input parameters (column “Design database” in Table 2) need to be large in order to cover a wide range of earthquake scenarios. Thus, if correctly estimated, meta-models will be able to reproduce the model behavior for a large range of variations of the seismic inputs parameters, including physical scenarios from geological studies of the zone.

In order to build the design database, fault parameters as defined in section “Step 1: tsunami simulations” and Table 2 were sampled randomly and independently with the Monte-Carlo method and supposing uniform distributions. The uniform distribution was chosen in order to build meta-models able to reproduce tsunami heights generated by various tsunamigenic sources with the same accuracy. The resulting earthquake magnitudes are computed using the sampled parameters with equation 2. The shear modulus chosen for the magnitude estimation is a constant value assumed to be equal to 30 GPA. This design database is a matrix which associates to a given combination of fault parameters estimates the maximum simulated water height at each point of the numerical grid and also at four selected locations along the French Atlantic Coast called gauges (Table 3 and Figure 3), namely, from North to South, “Saint-Malo,” “Brest,” “La Rochelle” and “Gastes.” The

maximum tsunami water height is the relevant parameter when estimating tsunami hazard.

Meta-Models Construction and Validation

Meta-models were constructed using the *NestedKriging* procedure described in Rulhière et al. (2018), as reported in section “Step 2: meta-model design and validation.” The design database contains 5839 scenarios used for the meta-model construction and validation. The water height characteristics associated to these scenarios are reported in Table 4. Each meta-model is a function able to compute the maximum tsunami water height at the gauge location for a given set of seismic source parameters (strike, length, dip, rake, width, slip, longitude, latitude and depth). Obviously, the input parameters should be included in the parameter range used for the meta-model construction and reported in Table 2.

For meta-model validation, the design data-base is split into K folds ($K = 10$ in this study, for a total of 584 simulations) of approximately equal size and a model is fitted from the data and validated on the fold A_k . K -fold cross validation is used for two main purposes: (i) to tune hyper parameters of the meta-model and (ii) to better evaluate the prediction accuracy of the meta-model. In both of these cases, the choice of k should permit to ensure that the training and testing sets are drawn from the same distribution. Especially, both sets should contain sufficient variation such that the underlining distribution is represented.

TABLE 3 | Location and water depth of the French Gauges chosen for meta-model construction.

Gauge	Longitude [degrees]	Latitude [degrees]	Depth [m]
Saint-Malo	–2.08	48.7	8
Brest	–4.65	48.26	28
La Rochelle	–1.65	45.93	40
Gastes	–1.27	44.35s	15

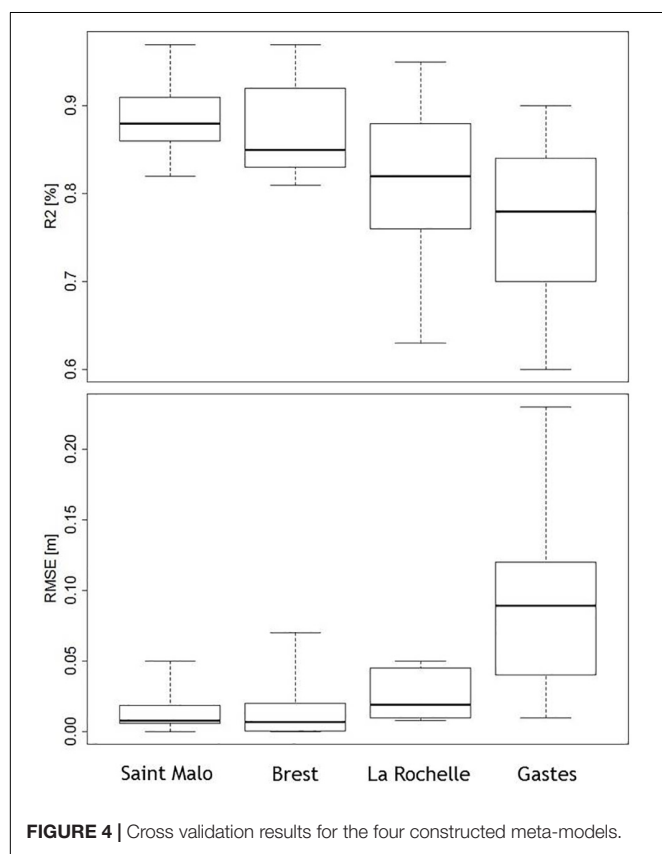
TABLE 4 | Maximum water height associated with the design database; μ , σ and Max correspond to the mean, standard-deviation and maximum modeled values.

	Saint Malo	Brest	La Rochelle	Gastes
Min [m]	0.00	0.00	0.00	0.00
μ [m]	0.21	0.47	0.37	0.30
$\mu + \sigma$ [m]	0.39	0.93	0.71	0.60
$\mu + 2\sigma$ [m]	0.57	1.38	1.05	0.90
Max [m]	1.60	5.66	3.12	3.07

TABLE 5 | Summary of meta-model evaluations using cross-validation technique (mean values).

Gauges	L_RMSE [–]	L_R2 [%]
Saint-Malo	0.0142	88.66%
Brest	0.0093	87.33%
La Rochelle	0.0262	81.55%
Gastes	0.10	76.33%

The statistical parameters for cross validation are defined in section “Meta-models validation.”

**FIGURE 4 |** Cross validation results for the four constructed meta-models.

From a practical point of view, the value of K is typically chosen as a good compromise between the computational times needed for the analysis and its reliability (Hastie et al., 2009). Indeed, there is not, to our knowledge, a well-established methodology allowing identifying the optimum number of folds necessary for cross-validation. In this methodological work, we considered $K = 10$ as a robust value with regards to the objectives of the study.

Results in terms of the criteria of quality L are reported in Table 5 and Figure 4. It appears that the mean computed values from cross validation are satisfying when considering the large range of parameters variations of the design data-base and the methodological purpose of the study. Indeed, except for the “Gastes” gauge, the mean R^2 is higher than 80%, which is satisfying according to Marrel et al. (2009) and Storlie et al. (2009), and the mean $RMSE$ of few centimeters, indicating that the kriging meta-model is a good emulator choice for reproducing the CEA-tsunami code behavior.

However, it must be underlined that, out of four gauges, results obtained for “Gastes” gauge are not satisfying, at least in terms of statistical performance. In fact, the large variation of the $RMSE$ parameter (from 0.03 to 0.25 m) and the low values of R^2 (varying from 0.6 to 0.9) reported in Figure 4 suggest that further numerical runs should be necessary to improve the accuracy of kriging.

It must be noted that the methodology used for the meta-model validation is a “state of the art” methodology permitting to focus on the ability of the meta-model to reproduce the mean model response and to estimate the model variability (represented by the variance). Even if this is common in literature, for hazard studies it would also be of interest to focus in the future on other criteria that account for the behavior in the tails of the distributions of the simulated values (extreme values).

Validation With Results From Allgeyer et al. (2013)

We perform an additional test in order to evaluate the ability of our meta-models to reproduce state of the art tsunami scenarios generated by the AGPB and impacting the French coast. With this aim, we compare our meta-models results with tsunami height simulated at the same location by Allgeyer et al. (2013). This comparison is of interest in order to confirm the ability of the constructed meta-models to reproduce the order of magnitude of the modeled tsunami height at a given location.

In this study, the authors analyzed the impact of a Lisbon-like tsunami on the French Atlantic Coast through numerical modeling. The authors focused on the simulated maximum tsunami water height in the North Atlantic associated to three different sources for the 1755 events derived from Johnston (1996), Baptista et al. (2003), and Gutscher et al. (2006), for a total of five tsunami scenarios (see Table 6). The same scenarios were simulated with the constructed meta-models for the four French Atlantic Gauges and with the CEA-tsunami model with a more refined grid, spacing of 1'. Even if meta-models prediction slightly overestimate the modeled tsunami height (see Figure 5 and Table 7), these results indicate a good agreement between the meta-modeled and the modeled water height, for the “Saint Malo,” “Brest” and “La Rochelle” gauges. On the contrary, these results confirm that further numerical runs should be necessary to improve the accuracy of kriging for “Gastes” gauge, which largely overestimate the tsunami heights of the original model (Figure 5).

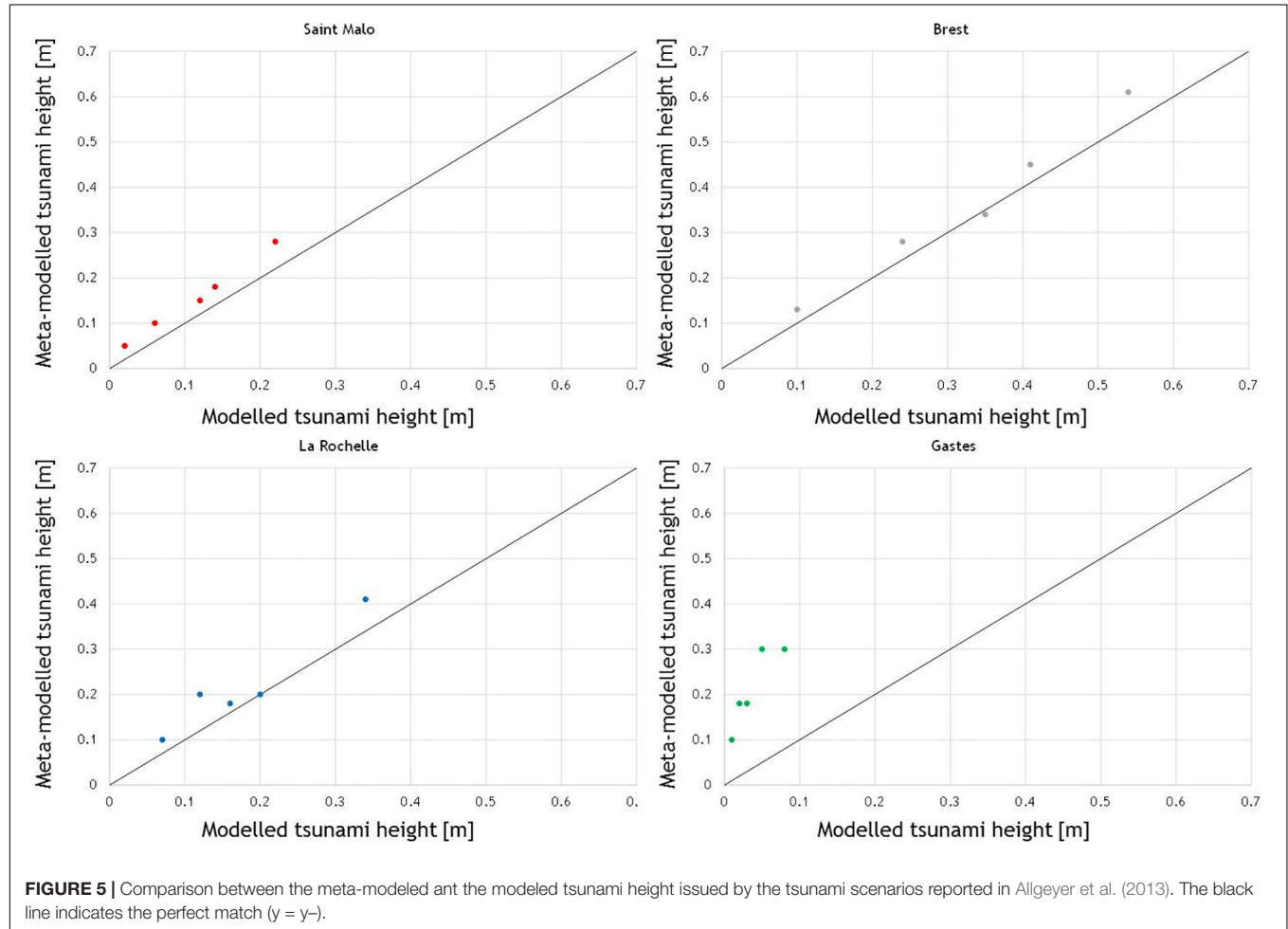
Considering the methodological purpose of the study, these results are satisfying. However, for a practical application, a more extended set of physical scenarios should be of interest for the validation of the tsunami height predicted by meta-models.

POTENTIAL APPLICATION OF META-MODELS FOR TSUNAMI HAZARD ANALYSIS

The objective of this section is to present how meta-models can be employed for (i) the integration of uncertainties of “known”

TABLE 6 | Seismic sources simulated in Allgeyer et al. (2013) for the 1755 Lisbon-tsunami.

Hypothesis reported in Allgeyer et al. (2013)									
Source	Strike [°]	Length [km]	Dip [°]	Rake [°]	Width [km]	Slip [m]	Lon [°]	Lat [°]	Depth [km]
Johnston, 1996	60	200	40	90	80	13.1	11.45	36.95	27
Baptista et al., 2003	250	155	45	90	55	20	−8.7	36.1	20.5
Gutscher et al., 2006	21.7	96	24	90	55	20	−10	36.8	20.5
	346	180	5	90	197	20	−7.5	35.5	15.1
	346	180	30	90	12	20	−8.6	35.3	3.2

**TABLE 7** | Maximum tsunamis height simulated in Allgeyer et al. (2013) for the 1755 Lisbon-tsunami and computed with meta-models.

Source	Numerical simulations from Allgeyer et al. (2013)				Results from meta-models constructed in this study			
	Saint Malo [m]	Brest [m]	La Rochelle [m]	Gastes [m]	Saint Malo [m]	Brest [m]	La Rochelle [m]	Gastes [m]
Johnston, 1996	0.22	0.54	0.34	0.08	0.28	0.61	0.41	0.3
Baptista et al., 2003	0.06	0.24	0.12	0.02	0.1	0.28	0.2	0.18
Gutscher et al., 2006	0.14	0.35	0.16	0.03	0.18	0.34	0.18	0.18
	0.12	0.41	0.2	0.05	0.15	0.45	0.2	0.3
	0.02	0.1	0.07	0.01	0.05	0.13	0.1	0.1

tsunami scenarios (see section “In the case of expert opinion: uncertainty quantification in DTHA”) and (ii) the analysis of the tsunamigenic potential of a poorly known tsunamigenic area (see section “Without expert opinion: exploration of a large tsunamigenic area”). Finally, the obtained results are discussed in light of the tsunami hazard issued by a probabilistic analysis (see section “Qualifying these approaches with respect to Probabilistic Tsunami Hazard Assessment”).

We recall here that this is not an operational tsunami hazard assessment and that all the presented results are purely methodological.

In the Case of Expert Opinion: Uncertainty Quantification in DTHA

DTHA is classically assessed by means of considering particular source scenarios (usually maximum credible scenario) and the associated maximum tsunami height is generally retained as hazard level (MCS). A more sophisticated method was recently proposed by Roshan et al. (2016). The authors tested around 300 tsunamis scenarios in a [8 – 9.5] magnitude range associated with various faults potentially impacting the Indian coast. Finally, the authors suggested that an appropriate water level for hazard assessment (e.g., mean value or mean plus sigma value) should be retained. They proposed the mean value of the simulated water heights, as test, by considering that this value may need to be revisited in the future.

In the case of the tsunamis hazard associated with the AGPB, the 1755 Lisbon tsunami is the classical reference scenario. In order to illustrate how to integrate the evaluation of uncertainties in a MCS approach, we focus on two specific and nearly deterministic scenarios considered as likely sources generating the Lisbon 1755 tsunami, namely the Gorringe and Horseshoe structures (Bufo et al., 1988; Stich et al., 2007; Cunha et al., 2012; Duarte et al., 2013; Grevenmeyer et al., 2017). Both structures were modeled taking into account available maps (Cunha et al., 2012; Duarte et al., 2013) and fault parameters (Stich et al., 2007; Grevenmeyer et al., 2017) summarized in **Table 2**. For the computation of the tsunami heights associated with these scenarios, fault parameters are considered uniformly distributed and are randomly sampled in their range of variation (**Table 2**), for a total of nearly 10,000 tsunami scenarios. The magnitude range associated with the sources of these tsunami scenarios varies from 7.7 to 8.9, which is coherent with the range of estimated magnitudes for the 1755 earthquake (Johnston, 1996; Gutscher et al., 2006). The convergence of statistics (defined as the evolution of the mean modeled tsunami height and of the mobile standard deviation) is largely achieved, indicating that the number of tsunami scenarios is sufficient to represent the expected variability of tsunamis height.

The distribution of tsunami heights resulting from the two Gorringe and Horseshoe sources are reported in **Figure 6** for each gauge and a summary of the numerical values in **Table 8**. It can be observed that the tsunami heights generated by the Horseshoe sources are globally lower than those generated by the Gorringe sources. At first glance, this result does not seem surprising considering the closer proximity of the French gauges

to the Gorringe bank. It can also be observed that both scenarios are affected by strong uncertainties. Indeed, the ratio between the maximum and the mean tsunami height is very large and it can vary from 2 to 5, according to the chosen gauges. Moreover, the variability of the modeled values around the mean value can be higher than 1.0 m for the Gorringe scenarios and it is always higher than 0.24 m (for Saint Malo gauge).

If we assume that the MCS approach is to take into account the worst possible scenarios, a hazard level corresponding to the maximum modeled water height could be retained for each gauge. However, if we consider the excursion of our numerical results this hazard level could be too high. This is probably why, as previously mentioned, other authors proposed to set a mean level as representative of the tsunami hazard (Roshan et al., 2016). In general, a first methodological conclusion is that the impact of the uncertainty on the source parameters on water height can be high and it should be taken into account by decision makers.

Without Expert Opinion: Exploration of a Large Tsunamigenic Area

Let's now assume, for a methodological purpose, that AGPB is a poorly known seismic area and that the objective of the study are to (i) evaluate the “possible” impact of tsunamis generated by this area on a target location (in this example, the French gauges) and (ii) to better understand the relative influence of the source parameters on these tsunamis. This latter analysis should permit to better guide the geological investigations in the area.

With this objective, UQ and GSA appear as two useful and complementary tools, respectively to answer to objectives (i) and (ii).

Analysis of the Impact of Tsunamis on a Target Area Through UQ

We propose to evaluate the tsunami hazard level, called here MCS_h, by exploring meta-models results based on a very wide range of tsunamigenic scenarios, beyond those proposed so far in the literature as we suppose that this tsunamigenic area is not well known. As a consequence, we accept that through this approach we explore the effects of both likely and very unlikely scenarios relative to the MCS (the “MCS_h” scenarios) that could potentially arise in a context of poor seismotectonic knowledge. As in the previous case, depending on the specific hazard target (civil or industrial facilities), and its location with respect to the source zone, the end-user of this methodology needs to decide which level of water height to choose from the obtained distribution. In this exercise, we will consider the maximum simulated water height for MCS_h.

We build a database (called *global database*) of tsunami scenario generated by the considered AGPB area at four French Atlantic Gauges, with the aim to cover a wider range of tsunamigenic scenarios (**Table 2**). For the purpose of this methodological paper, the modeled area, which encompasses different seismotectonic domains, was split in a very simplistic way into two main seismic source zones (**Figure 3**): a western domain where normal to transtensive earthquakes occur within a thin crust and an eastern domain where reverse to transpressive earthquakes mainly occur on a thicker crust. We considered the

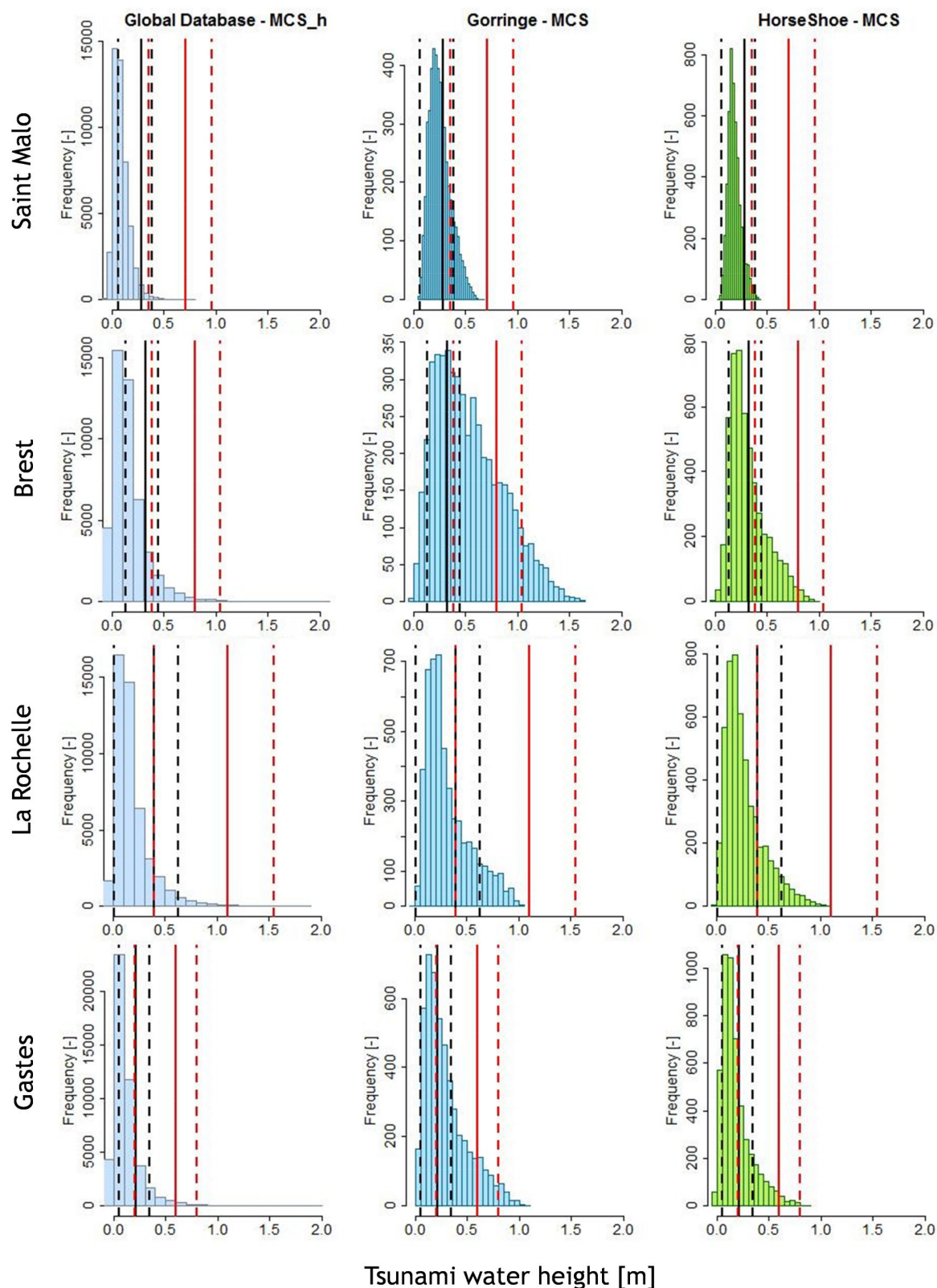


FIGURE 6 | Tsunami height frequency distribution associated to the global data base (gray-blue), the Gorrige bank (blue) and the Horseshoe bank (green). The black and red lines correspond, respectively, to the mean tsunami height associated to a return period of 1 000 and 10 000 years issued by TSUMAPS-NEAM research project. Dotted lines represent the 2th and 98th percentile of the hazard curves.

TABLE 8 | Maximum tsunami height distributions associated with the global database (GD), HorseShoe (HS), and Gorringe (GR) scenarios; μ_m , σ_m and Max_m correspond to the mean, standard-deviation and maximum meta-modeled values.

Scenarios	Saint Malo			Brest			La Rochelle			Gastes		
	GD	HS	GR	GD	HS	GR	GD	HS	GR	GD	HS	GR
Min_m [m]	0.00	0.03	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
μ_m [m]	0.09	0.18	0.25	0.18	0.31	0.53	0.19	0.26	0.32	0.13	0.18	0.29
$\mu_m + \sigma_m$ [m]	0.17	0.24	0.35	0.34	0.49	0.85	0.36	0.43	0.53	0.27	0.33	0.5
$\mu_m + 2\sigma_m$ [m]	0.24	0.3	0.46	0.51	0.66	1.17	0.54	0.63	0.74	0.41	0.47	0.7
Max_m [m]	0.77	0.42	0.66	2.01	0.96	1.60	1.81	1.05	1.04	1.92	0.87	1.06

Western domain west of the 10°W meridian and the *Eastern domain* east of this meridian, coinciding roughly with the base of the continental slope facing the Portuguese coastline. The main fault characteristics considered to build the tsunami data-base in both eastern and western domains are reported in **Table 2**, following data contained in Buforn et al. (1988), Molinari and Morelli (2011), Cunha et al. (2012).

The considered seismogenic thickness takes into account the depth of the observed seismicity as well as the fact that part of the upper mantle can potentially be mobilized during major earthquakes: the western domain seismogenic crust is considered to be up to 20 km deep (after Baptista et al., 2017), and up to 60 km for the eastern domain (after Silva et al., 2017). The fault parameters are considered uniformly distributed and are randomly sampled in their range of variation. We finally filtered the resulting database according to an aspect-ratio criterion, allowing the ratio between the length and the width of the faults not to exceed the value of 10, which corresponds to an upper bound of what is observed in nature (Mc Calpin, 2009). The final global database contains nearly 50,000 tsunami scenarios, resulting in earthquake magnitudes varying from 6.7 to 9.3 (**Table 2**), depending on the explored earthquake source characteristics and calculated from Eq 1. This range of magnitudes is consistent with the magnitude range of the design database.

The tsunami water heights distributions associated to the four gauges are reported in **Table 8** and **Figure 6**. As for the previous paragraph, these results show a very large variability in tsunamis height, which is not surprising considering the large range of variation of the source parameters. However, if we compare these results with those obtained in the previous section, we can also observe that Gorringe and Horseshoe banks are among the major contributors to tsunami hazard along the French Atlantic Coast, generated by seismic sources in the AGPB. Indeed, **Figure 6** clearly shows that even if some isolated tsunami scenarios can generate a hazard level higher than the Gorringe Scenarios, globally, these sources are representative of the higher tsunamis from the global data-base.

Global Analysis of Seismic Source Influence on the Target Area

A sensitivity analysis is hereafter performed in order to decipher the relative influence of the seismic source parameters. Homma and Saltelli (1996) introduced the total sensitivity index which measures the influence of a variable jointly with all its

interactions. If the total sensitivity index of a variable is zero, this variable can be removed because neither the variable nor its interactions at any order have an influence on the results. This statistical index (called Sobol index “ST” in this paper) is here of particular interest in order to highlight the earthquake source parameters that mostly control the tsunamis height at each tested gauge. In **Figure 7**, we reported the total Sobol index for the four meta-models of the French Atlantic Gauges computed with the methodology proposed by Jansen et al. (1994) using the sensitivity-package R (Pujol et al., 2017). The accuracy of Sobol indices performed with Jansen’s method depends on the number of model evaluations. For instance, in this study, we performed nearly 20 000 simulations using meta-models. Results show that the slip parameter is globally the most-influencing parameter for all the French Atlantic Gauges meta-models. Concretely, nearly 50% of the variance of the tsunami water height (the uncertainty) could be reduced by a better knowledge of this parameter. This result is quite obvious considering that the fault-slip directly conditions the ocean floor deformation and hence the tsunami amplitude.

However, this analysis also suggests that the most influencing parameters for the four gauges are slightly different, depending on their location. One can differentiate results obtained for the southern gauges (i.e., La Rochelle and Gastes) from those obtained at northern gauges (i.e., Saint Malo and Brest):

- For the southern gauges, width is the second most relevant parameter, especially at La Rochelle where it is almost as important as the slip parameter. Other important parameters are length and rake, suggesting that for these gauges, fault source parameters in terms of magnitudes (depending on width, slip and length according to equations 1 and 2) and kinematics are the most important in generating hazard;
- For the northern gauges, apart from slip, strike, width and rake are also important, but slightly less than the longitude parameter. This means that the location of the source is here of major importance. A possible physical reason could be associated to the lack of the natural barrier composed by the north of Spain and Portugal which protects southern gauges from AGPB related tsunamis compared to northern gauges. As a consequence, the northern gauges should be more exposed to hazard in comparison to southern gauges, located in the shadow of Portugal and Spain.

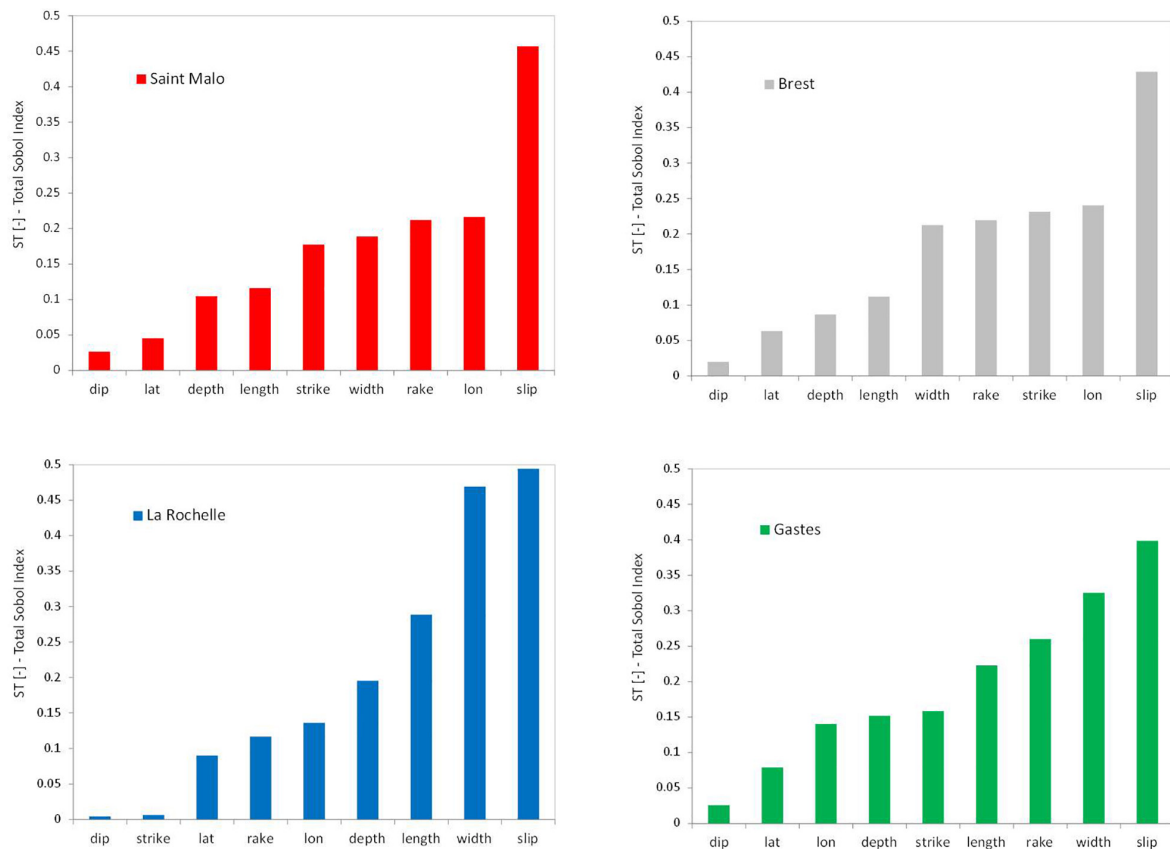


FIGURE 7 | Total order Sobol Index computed for the French Atlantic Gauges.

Qualifying These Approaches With Respect to Probabilistic Tsunami Hazard Assessment

Let's imagine now that for the target area a reference hazard level is provided through a probabilistic study and that we aim to qualify the robustness of this hazard level with regards to possible tsunamis height issued by expert scenarios (MCS, see section "In the case of expert opinion: uncertainty quantification in DTHA") or by the exploration of poorly known tsunamigenic area (MCS_h, see section "Without expert opinion: exploration of a large tsunamigenic area"). The idea is here to compare the results from a probabilistic analysis with the deterministic approach proposed in this study, which should permit to completely cover the uncertainty related to seismic source parameters.

With this aim, we compare our results to the tsunami hazard level (Figure 6) issued from TSUMAPS-NEAM research project⁴ in four points located near to the gauges used in this study. TSUMAPS-NEAM results are provided in terms of Maximum Inundation Height (MIH), which is the estimated maximum flow depth from the envelope of the tsunami wave at all times, as reported in the NEAMTHM18 documentation (Basili et al., 2018, 2019). TSUMAPS-NEAM has developed a long-term PTHA for

earthquake-induced tsunamis for the coastlines of the NEAM region (NE Atlantic, the Mediterranean, and connected seas). TSUMAPS-NEAM results largely relied on inputs from the EU FP7 project ASTARTE, of the GAR15 (global risk quantification under the HFA), and national PTHAs like those of USA and Italy. One of the major results of the project is the Interactive Hazard Curve Tool⁵ which represents online hazard maps for different hazard probability/average return periods (mean, median, 2nd, 16th, 84th, and 98th percentiles hazard curves). In TSUMAPS-NEAM, tsunamis are computed using an approach which relies on the use of unit sources to reproduce tsunami scenarios (e.g., Molinari et al., 2016; Baptista et al., 2017). The main differences with our methodology are that these methods rely on a decomposition of the tsunami waves form recorded in a target area with simplified methods (e.g., the Green's law), and on the main assumption that the non-linear terms of tsunami propagation are negligible. The scope of these studies is to develop a fast emulator, permitting to replace a tsunami model at some selected locations, along the same philosophy of the meta-models developed here. As a consequence, according to the author's purpose, MIH is suitable for a regional, initial screening assessment type such as the objective of TSUMAPS-NEAM. These results must be considered only as input reference study for

⁴<http://www.tsumaps-neam.eu/>

⁵<http://ai2lab.org/tsumapsneam/interactive-hazard-curve-tool/>

further site-specific assessment. Considering the methodological objective of this work, it is of interest to compare our results with this previous work.

For this exercise, we selected two targets return periods of 1,000 and 10,000 years from the hazard curves of TSUMAPS-NEAM. To take into account the uncertainty related to these results, we consider the mean, the 2th and the 98th percentile hazard curves. These targets, and the associated uncertainty, are typical of classical risk analysis and permit to compare the distributions issued by PTHA with our results. For instance, after the Fukushima accident, a return period of 10,000 years is considered as the target for hazard assessment in the field of nuclear safety (WENRA, 2015).

Figure 6 illustrates the comparison of our results (MCS and MCS_h) with those selected from the TSUMAPS-NEAM project. One can notice that:

- The tsunami intensity associated to a target mean return period of 1,000 years does not cover the tsunamis intensities related to expert scenarios (Gorringe and Horseshoe banks) or to MCS_h scenarios (all the likely and unlikely scenarios from AGPB);
- The tsunami intensity associated to a target mean return period of 10,000 years covers all the tsunami intensities related to seismic sources from the AGPB, including very unlikely tsunami scenarios (MCS_h);
- A target hazard level of 10,000 years covers the tsunamis related to Horseshoe banks;
- A target hazard level of 10,000 years covers the tsunamis related to Gorringe bank, with the exception of “Brest” gauge and to some extent that of “Gastes”.

Even if these results are purely methodological, it appears that a target hazard level of 10,000 years issued from TSUMAPS-NEAM project covers all the variability related to seismic sources from the AGPB, even by exploring very unlikely tsunami scenarios (MCS_h) and most of the variability related to Gorringe and Horseshoe scenarios. However, concerning the “Brest” gauge, the methodology indicates a high sensitivity of this coastal area to the characteristics of the Gorringe seismic sources.

DISCUSSION AND CONCLUSION

The research work presented in this paper was performed in order to test the interest of UQ for the analysis and the qualification of the DTHA generated by earthquakes. We propose a new methodology, permitting the assessment of the uncertainty related to tsunami hazard through the analysis of a wide range of tsunami scenarios at a given location. This concept should permit to define a hazard level which goes beyond the definition of the Maximum Credible Scenario (MCS) classically reported in the literature (JSCE, 2002; Lynett et al., 2004) and employed for DTHA and permit to integrate uncertainty in hazard quantification. Moreover, tsunami hazard evaluated through UQ can also permit the exploration of tsunamigenic potential of a poorly known seismic zone, as well as a qualification of PTHA.

From a methodological point of view, meta-models appear as a very efficient and viable solution to the problem of generating many computationally expensive tsunamis simulations. As reported in Behrens and Dias (2015), the statistical emulator gives perfect predictions at the input points that are used in its generation process (it interpolates). Statistical emulation does not accelerate the model itself. The significant advantage of using the emulator is that it is much less computationally demanding to be evaluated and, therefore, it can be employed to carry out fast predictions and inexpensive analyses, such as sensitivity and uncertainty analyses reported in this study. Even if it is the first time, to our knowledge, that meta-models are proposed for THA, they have been already employed for tsunami modeling (Sarri et al., 2012; Sraj et al., 2014; Rohmer et al., 2018). As reported in section “Qualifying these approaches with respect to Probabilistic Tsunami Hazard Assessment,” in the field of tsunami hazard, an alternative approach relies on the use of unit sources to reproduce tsunami scenarios (e.g., Molinari et al., 2016; Baptista et al., 2017). Results from these studies are satisfactory for most of the practical applications such as probabilistic tsunami hazard analysis, tsunami source inversion and tsunami warning systems. However, we consider that our methodology can be proposed as an alternative to these studies as it does not rely on any assumption on tsunami propagation and can be applied everywhere in the model domain, without any limitation. For instance, meta-models could be constructed and validated from detailed simulations in a given target area, including complex physical phenomena as overtopping, run-up, and breaking. However, we want to stress that meta-model construction is time consuming and needs an appropriate design data-base, which requires a good compromise between the range of variations of the inputs and the number of simulations.

Concerning our application at four selected location, with the exception of “Gastes” gauge, both the statistical tests we performed and the comparison with result from Allgeyer et al. (2013) suggest that the meta-models are able to reproduce the tsunamis generated by the AGPB. Thus, the constructed meta-models could be employed in a further study to roughly evaluate the impact of other seismic scenarios from the AGPB and impacting the French Atlantic Coast offshore. In order to increase the accuracy of “Gastes” meta-model, the design data-base should be however enriched.

Results from GSA suggests that beyond earthquake magnitudes, the position and the orientation of the faults are influent parameters, at least for the sites considered along the northern French Atlantic Coast. Indeed, sensitivity analysis can be a useful tool not only to parametrize the design data-base but also to orient future geological surveys in a specific area.

In conclusion, MCS_h in the AGPB study region as defined in our study could be implicitly associated to a mean return period of 10,000, when considering the strong hypothesis we did on source characteristics. In this sense, results from UQ shows that a hazard level of 10,000 years issued from TSUMAPS-NEAM project covers a very wide range of uncertainties related to characterization of seismic sources from the AGPB, even by exploring very unlikely tsunami scenarios (MCS_h) and

most of the uncertainties related to Gorringer and Horseshoe banks expert opinion scenarios. However, concerning the “Brest” gauge, even a target hazard level of 10,000 years does not seem appropriate to cover all the uncertainties related to the Gorringer source, indicating a high sensitivity of this coastal area to the characteristics and the kinematics of the Gorringer seismic source.

PERSPECTIVES

For an operational far-field DTHA, it should be necessary, at first, to improve the actual numerical model in order to better represent the tsunamis run-up and the inundated areas, with a more accurate bathymetric grid. For an operational application to locations closer to the source zones (i.e., Portugal or Spain, Morocco), it may be challenging to gather the necessary details for a proper establishment of both models and meta-models. However this point deserves further attention because uncertainties will always remain. Thus even in such cases, where more refined databases need to be established, the proposed approach should be of interest to at least efficiently explore in a more exhaustive way the uncertainties (e.g., different probability distribution of inputs parameters, non-uniform slip distribution, fault geometries) in the source parameters linked to the MCS approach. Moreover, it must be noted that for this methodological study we chose a very simple but widely used source description (planar faults with homogeneous slip). However, more robust simulations should take into account a more complex source representation (e.g., 3D geometry, heterogeneous slip distribution), as recently suggested by Davies and Griffin (2019). Indeed, meta-models may also be useful to account for these parameters, allowing for many simulations.

From a methodological point of view, it could be of interest to compare our methodology with the alternative approaches using unit sources to reproduce tsunami scenarios (e.g., Molinari et al., 2016; Baptista et al., 2017), in terms of simulations needed, computational time, accuracy, for instance.

Finally, from a numerical point of view, it would be of interest to (1) introduce recurrence models for each tsunamigenic source to go toward PTHA calculations, and (2) introduce meta-models

in systems developed for tsunami early warning, considering the low computational time inherent to this statistical tool.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the Institute for Radiological Protection and Nuclear Safety (IRSN).

AUTHOR CONTRIBUTIONS

VB: development of the methodology, meta-models construction and validation, and sensitivity and hazard analysis. HJ and OS: seismic sources definition, and seismological and hazard analysis. EA: tsunamis modeling (model construction and simulations of the design data-base). LB and C-MD: technical discussions. HH: tsunamis modeling and technical discussions. VB, HJ, and OS wrote the manuscript.

FUNDING

This research was funded by the French National Agency for Research (ANR).

ACKNOWLEDGMENTS

This study was performed through a cooperative research between the Institute for Radiological Protection and Nuclear Safety (IRSN) and the CEA, in the framework of the French ANR-TANDEM research project. We want to thank ENS Paris and particularly the Geology Laboratory for their technical support on this project. We acknowledge especially Eric Calais for the availability of a part of the team's cluster and Pierpaolo Dubernet for the help he offered to install Promethee. The present work also benefited from the inputs of Msc. Hend Jebali and Dr. Audrey Gailler and from the valuable technical assistance of Dr. Maria Lancieri, Dr. Yann Richet, and Miss Ludmila Provost.

REFERENCES

- Abily, M., Bertrand, N., Delestre, O., Gourbesville, P., and Duluc, C.-M. (2016). Spatial Global sensitivity analysis of high resolution classified topographic data use in 2d urban flood modelling. *Environ. Model. Softw.* 77, 183–195. doi: 10.1016/j.envsoft.2015.12.002
- Adetula, B. A., and Bokov, P. M. (2012). Computational method for global sensitivity analysis of reactor neutronic parameters. *Sci. Technol. Nucl. Installations* 2012:11. doi: 10.1155/2012/109614
- Allgeyer, S., Daubord, C., Hébert, H., Loevenbruck, A., Schindelé, F., and Madariaga, R. (2013). Could a 1755-like tsunami reach the french atlantic coastline? Constraints from twentieth century observations and numerical modeling. *Pure Appl. Geophys.* 170, 1415–1431. doi: 10.1007/s00024-012-0513-5
- Arakawa, A. (1972). *Design of the UCLA General Circulation model. In Numerical Simulation of Weather and Climate*. Tech. Rep. No. 7. Los Angeles, CA.
- Bachoc, F. (2013). Cross validation and maximum likelihood estimations of hyper-parameters of gaussian processes with model misspecification. *Computa. Stati. Data Anal.* 66, 55–69. doi: 10.1016/j.csda.2013.03.016
- Baptista, M., Miranda, P., Miranda, J., and Victor, L. M. (1998). Constrains on the source of the 1755 Lisbon tsunami inferred from numerical modelling of historical data on the source of the 1755 Lisbon tsunami. *J. Geodyn.* 25, 159–174. doi: 10.1016/S0264-3707(97)00020-3
- Baptista, M. A., Miranda, J. M., Chierici, F., and Zitellini, N. (2003). New study of the 1755 earthquake source based on multi-channel seismic survey data and tsunami modeling. *Nat. Hazards Earth Syst. Sci.* 3, 333–340. doi: 10.5194/nhess-3-333-2003
- Baptista, M. A., Miranda, J. M., Matias, L., and Omira, R. (2017). Synthetic tsunami waveform catalogs with kinematic constraints. *Nat. Hazards Earth Syst. Sci.* 17, 1253–1265. doi: 10.5194/nhess-17-1253-2017
- Barkan, R., Uri, S., and Lin, J. (2009). Far field tsunami simulations of the 1755 Lisbon earthquake: Implications for tsunami hazard to the US East Coast and the Caribbean. *Mar. Geol.* 264, 109–122. doi: 10.1016/j.margeo.2008.10.010
- Baroni, G., and Tarantola, S. (2014). A General probabilistic framework for uncertainty and global sensitivity analysis of deterministic models A

- hydrological case study. *Environ. Model. Softw.* 51, 26–34. doi: 10.1016/j.envsoft.2013.09.022
- Basili, R., Brizuela, B., Herrero, A., Iqbal, S., Lorito, S., Maesano, F. E., et al. (2018). *NEAM Tsunami Hazard Model 2018 (NEAMTHM18) Online Data of the Probabilistic Tsunami Hazard Model for the NEAM Region from the TSUMAPS-NEAM project*, Chap. Rome. Istituto Nazionale di Geofisica e Vulcanologia (INGV).
- Basili, R., Brizuela, B., Herrero, A., Iqbal, S., Lorito, S., Maesano, F. E., et al. (2019). *NEAMTHM18 Documentation the Making of the TSUMAPS-NEAM Tsunami Hazard Model 2018*. Rome: Istituto Nazionale di Geofisica e Vulcanologia (INGV).
- Behrens, J., and Dias, F. (2015). New computational methods in tsunami science, philosophical transactions of the royal society a mathematical. *Phys. Eng. Sci.* 373:20140382. doi: 10.1098/rsta.2014.0382
- Breiman, L., and Spector, P. (1992). Submodel selection and evaluation in regression. The X-random case. *Int. Stat. Rev. Int. Stat.* 60, 291–319. doi: 10.2307/1403680
- Bufo, E., Udias, A., and Mezcu, J. (1988). Seismicity and focal mechanisms in south Spain. *Bull. Seismol. Soc. Am.* 78:2024.
- Cunha, T., Matias, L., Terrinha, P., Negredo, A., Rosas, F., Fernandes, R., et al. (2012). Neotectonics of the SW Iberia margin, gulf of cadiz and alboran sea reassessment including recent structural, seismic and geodetic data. *Geophys. J. Int.* 188, 850–872. doi: 10.1111/j.1365-246x.2011.05328.x
- Cunha, T., Watts, A., Pinheiro, L., and Myklebust, R. (2010). Seismic and gravity anomaly evidence of large-scale compressional deformation off SW Portugal. *Earth Planet. Sci. Lett.* 293, 171–179. doi: 10.1016/j.epsl.2010.01.047
- Davies, G., and Griffin, J. (2019). Sensitivity of probabilistic tsunami hazard assessment to far-field earthquake slip complexity and rigidity depth-dependence Case study of australia. *Pure Appl. Geophys.* 1–28. doi: 10.1007/s00024-019-02299-w
- Duarte, J. C., Rosas, F. M., Terrinha, P., Schellart, W. P., Boutelier, D., Gutscher, M.-A., et al. (2013). Are subduction zones invading the Atlantic? Evidence from the southwest Iberia margin. *Geology* 41, 839–842. doi: 10.1130/g34100.1
- Dupuy, D., Helbert, C., and Franco, J. (2015). Dice design and dicee v2.0 R packages for design and analysis of computer experiments. *J. Stat. Softw.* 65, 1–38.
- Faivre, R., Iooss, B., Mahévas, S., Makowski, D., and Monod, É. (2013). *Analyse de sensibilité et Exploration de Modèles. Application aux Sciences de la Nature et de l'Environnement*. Bengaluru: HAL.
- Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The Elements of Statistical Learning*, Vol. 1. New York, NY: Springer series in statistics.
- Gracia, E., Danobeitia, J., Vergés, J., and Parsifal Team. (2003). Mapping active faults offshore Portugal (36 N–38 N) implications for seismic hazard assessment along the southwest Iberian margin. *Geology* 31, 83–86.
- Grandin, R., Borges, J. F., Bezzeghoud, M., Caldeira, B., and Carrilho, F. (2007). Simulations of strong ground motion in SW Iberia for the 1969 February 28 (M = 8.0) and the 1755 November 1 (M ~ 8.5) earthquakes-II. Strong ground motion simulations. *Geophys. J. Int.* 171, 807–822. doi: 10.1111/j.1365-246x.2007.03571.x
- Grevenmeyer, L., Lange, D., Villinger, H., Custodio, S., and Matias, L. (2017). Seismotectonics of the horseshoe abyssal plain and gorringe bank, eastern atlantic ocean constraints from ocean bottom seismometer data. *J. Geophys. Res. Solid Earth*. 122, 63–78. doi: 10.1002/2016jb013586
- Gutscher, M.-A., Baptista, M., and Miranda, J. (2006). The Gibraltar Arc seismogenic zone (part 2) constraints on a shallow east dipping fault plane source for the 1755 Lisbon earthquake provided by tsunami modeling and seismic intensity. *Tectonophysics* 426, 153–166. doi: 10.1016/j.tecto.2006.02.025
- Hanks, T. C., and Kanamori, H. (1979). A moment magnitude scale. *J. Geophys. Res. Solid Earth* 84, 2348–2350.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media. doi: 10.1007/b94608
- Homma, T., and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.* 52, 1–17. doi: 10.1016/0951-8320(96)00002-6
- Horsburgh, K., Wilson, C., Baptie, B., Cooper, A., Cresswell, D., Musson, R., et al. (2008). Impact of a Lisbon-type tsunami on the UK coastline and the implications for tsunami propagation over broad continental shelves. *J. Geophys. Res. Oceans* 113. doi: 10.1029/2007JC004425
- Horspool, N., Pranantyo, I., Griffin, J., Latief, H., Natawidjaja, D., Kongko, W., et al. (2014). A probabilistic tsunami hazard assessment for Indonesia. *Nat. Hazards Earth Syst. Sci.* 14, 3105–3122.
- Iooss, B., and Lemaître, P. (2015). *A Review on Global Sensitivity Analysis Methods, in Uncertainty Management in Simulation-Optimization of Complex Systems*. Springer, 101–122.
- Iooss, B., Ribatet, M., and Marrel, A. (2008). Global sensitivity analysis of stochastic computer models with joint metamodels. *arXiv* [preprint],
- Jansen, M. J., Rossing, W. A., and Daamen, R. A. (1994). Monte carlo estimation of uncertainty contributions from several independent multivariate sources. In: *Predictability and nonlinear modelling in natural sciences and economics* / Grasman, ed J. G. van Straaten (Dordrecht: Kluwer Academic Publishers), 334–343. doi: 10.1007/978-94-011-0962-8_28
- Johnston, A. C. (1996). Seismic moment assessment of earthquakes in stable continental regions—III. New Madrid 1811–1812, Charleston 1886 and Lisbon 1755. *Geophys. J. Int.* 126, 314–344. doi: 10.1111/j.1365-246x.1996.tb05294.x
- JSCE (2002). *Tsunami Assessment Method for Nuclear Power Plants in Japan*. Tokyo: JSCE.
- Kalra, T. S., Aretxabaleta, A., Seshadri, P., Ganju, N. K., and Beudin, A. (2018). Sensitivity analysis of a coupled hydrodynamic-vegetation model using the effectively subsampled quadratures method. *Geosci. Model. Dev.* 10, 4511–4523.
- Kleijnen, J. P. (2005). An overview of the design and analysis of simulation experiments for sensitivity analysis. *Eur. J. Operat. Res.* 164, 287–300. doi: 10.1016/j.ejor.2004.02.005
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai* 1137–1143. Available online at: https://www.researchgate.net/profile/Ron_Kohavi/publication/2352264_A_Study_of_Cross-Validation_and_Bootstrap_for_Accuracy_Estimation_and_Model_Selection/links/02e751bccc14c5e91c000000.pdf
- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. South. Afr. Inst. Min. Metall.* 52, 119–139.
- Löwe, R., Urich, C., Kulahci, M., Radhakrishnan, M., Deletic, A., and Arnbjerg-Nielsen, K. (2018). Simulating flood risk under non-stationary climate and urban development conditions—experimental setup for multiple hazards and a variety of scenarios. *Environ. Model. Softw.* 102, 155–171. doi: 10.1016/j.envsoft.2018.01.008
- Lynett, P., Wei, Y., and Arcas, D. (2004). “Tsunami hazard assessment Best modelling practices and state-of-the-art technology (nureg/cr-7223),” in *Numerical Modeling of Water Waves*. ed C. L. Mader. (Boca Raton, FL: CRC press).
- Machac, D., Reichert, P., Rieckermann, J., and Albert, C. (2016). Fast mechanism-based emulator of a slow urban hydrodynamic drainage simulator. *Environ. Model. Softw.* 78, e54–e67. doi: 10.1016/j.envsoft.2015.12.007
- Mader, C. L. (2004). *Numerical Modeling of Water Waves*. Boca Raton, FL: CRC press.
- Marrel, A., Iooss, B., Laurent, B., and Roustant, O. (2009). Calculations of the sobol indices for the gaussian processes metamodel. *Reliab. Eng. Syst. Saf.* 94, 742–751. doi: 10.1016/j.res.2008.07.008
- Marrel, A., Iooss, B., Van Dorpe, F., and Volkova, E. (2008). An efficient methodology for modeling complex computer codes with Gaussian processes. *Comput. Stat. Data Anal.* 52, 4731–4744. doi: 10.1016/j.csda.2008.03.026
- Marzocchi, W., Taroni, M., and Selva, J. (2015). Accounting for epistemic uncertainty in PSHALogic tree and ensemble modeling. *Bull. Seismol. Soc. Am.* 105, 2151–2159. doi: 10.1785/0120140131
- Matheron, G. (1963). Principles of geostatistics. *Econ. Geol.* 58, 1246–1266. doi: 10.2113/gsecongeo.58.8.1246
- Mc Calpin, J. (2009). *Paleoseismology*. San Diego, CA: Academic, 629.
- Molinari, I., and Morelli, A. (2011). EPcrusta reference crustal model for the European Plate. *Geophys. J. Int.* 185, 352–364. doi: 10.1111/j.1365-246x.2011.04940.x
- Molinari, I., Tonini, R., Lorito, S., Piatanesi, A., Romano, F., Melini, D., et al. (2016). Fast evaluation of tsunami scenarios uncertainty assessment for a Mediterranean Sea database. *Nat. Hazards Earth Syst. Sci.* 16, 2593–2602. doi: 10.5194/nhess-16-2593-2016
- Nguyen, T.-M., Richet, Y., Balayn, P., and Bardet, L. (2015). Propagation des incertitudes dans les modèles hydrauliques 1D. *La Houille Blanche* 55–62. doi: 10.1051/lhb/20150055
- Okada, Y. (1985). Surface deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Am.* 75, 1135–1154.

- Omira, R., Matias, L., and Baptista, M. A. (2016). Developing an event-tree probabilistic tsunami inundation model for NE atlantic coasts application to a case study. *Pure Appl. Geophys.* 173, 3775–3794. doi: 10.1007/s00024-016-1367-z doi: 10.1007/978-3-319-55480-8_7
- Pujol, G., Iooss, B., and Janon, A. (2017). *Sensitivity: Global Sensitivity Analysis of Model Outputs*. Available online at: <https://CRAN.R-project.org/package=sensitivity>
- R Core Team (2016). *A Language and Environment for Statistical Computing*, in: *R Foundation for Statistical Computing*. Vienne: R core Team.
- Razavi, S., Tolson, B. A., and Burn, D. H. (2012a). Review of surrogate modeling in water resources. *Water Resour. Res.* 48, doi: 10.1029/2011WR011527
- Razavi, S., Tolson, B. A., and Burn, D. H. (2012b). Numerical assessment of metamodeling strategies in computationally intensive optimization. *Environ. Model. Softw.* 34, 67–86. doi: 10.1016/j.envsoft.2011.09.010
- Richet, Y., and Bacchi, V. (2019). Inversion algorithm for civil flood defense optimization Application to two-dimensional numerical model of the garonne river in france. *Front. Environ. Sci.* 7:160. doi: 10.3389/fenvs.2019.00160
- Rohmer, J., Rousseau, M., Lemoine, A., Pedreros, R., Lambert, J., and Benki, A. (2018). Source characterisation by mixing long-running tsunami wave numerical simulations and historical observations within a metamodel-aided ABC setting. *Stoch. Environ. Res. Risk Assess.* 32, 967–984. doi: 10.1007/s00477-017-1423-y
- Roshan, A. D., Basu, P. C., and Jangid, R. S. (2016). Tsunami hazard assessment of Indian coast. *Nat. Hazards* 82, 733–762. doi: 10.1007/s10661-018-7048-x doi: 10.1007/s11069-016-2216-1
- Roustant, O., Ginsbourger, D., and Deville, Y. (2012). DiceKriging, dice optim two R packages for(the)analysis of computer experiments by kriging-based metamodeling and optimization. *J. Stat. Softw.* 51, 1–55.
- Rullière, D., Durrande, N., Bachoc, F., and Chevalier, C. (2018). Nested kriging predictions for datasets with a large number of observations. *Stat. Comput.* 28, 849–867. doi: 10.1007/s11222-017-9766-2
- Saint-Geours, N. (2012). *Sensitivity Analysis of Spatial Models application to Cost-Benefit Analysis of (flood)Risk Management Plans*. Montpellier: Université Montpellier II-Sciences et Techniques du Languedoc.
- Saito, T., Ito, Y., Inazu, D., and Hino, R. (2011). Tsunami source of the 2011 Tohoku–Oki earthquake, Japan Inversion analysis based on dispersive tsunami simulations. *Geophys. Res. Lett.* 38, doi: 10.1029/2011GL049089
- Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.* 145, 280–297. doi: 10.1016/s0010-4655(02)00280-1
- Saltelli, A., Chan, K., and Scott, E. M. (2000). *Sensitivity Analysis*. New York, NY: Wiley.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2008). *Global Sensitivity Analysis the Primer*. New York, NY: John Wiley & Sons.
- Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity Analysis in Practice a Guide to Assessing Scientific Models*. New York, NY: John Wiley & Sons.
- Saltelli, A., Tarantola, S., and Chan, K.-S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 41, 39–56. doi: 10.1080/00401706.1999.10485594
- Sammarco, P., and Renzi, E. (2008). Landslide tsunamis propagating along a plane beach. *J. Fluid Mech.* 598, 107–119. doi: 10.1017/S0022112007009731 doi: 10.1017/s0022112007009731
- Santana-Quintero, L. V., Montano, A. A., and Coello, C. A. (2010). “A review of techniques for handling expensive functions in evolutionary multi-objective optimization,” in *Computational Intelligence in Expensive Optimization Problems*, eds M. H. Lim, and Y. S. Ong (Springer Berlin Heidelberg), 29e59.
- Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. 1st edn. New York, NY: Springer-Verlag.
- Santos, A., and Koshimura, S. (2015). The historical review of the 1755 Lisbon Tsunami. *J. Geodesy Geomat. Eng.* 1, 38–52.
- Sarri, A., Guillas, S., and Dias, F. (2012). Statistical emulation of a tsunami model for sensitivity analysis and uncertainty quantification. *arXiv* 205, 1780–1803.
- Selva, J., Tonini, R., Molinari, I., Tiberti, M. M., Romano, F., Grezio, A., et al. (2016). Quantification of source uncertainties in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA). *Geophys. J. Int.* 205, 1780–1803. doi: 10.1093/gji/ggw107
- Silva, S., Terrinha, P., Matias, L., Duarte, J. C., Roque, C., Ranero, C. R., et al. (2017). Micro-seismicity in the Gulf of Cadiz Is there a link between micro-seismicity, high magnitude earthquakes and active faults? *Tectonophysics*. 717, 226–241. doi: 10.1016/j.tecto.2017.07.026
- Sobol, I. M. (1993). Sensitivity estimates for non linear mathematical models. *Math. Model. Comput. Exp.* 1, 407–414.
- Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* 55, 271–280. doi: 10.1016/s0378-4754(00)00270-6
- Sørensen, M. B., Spada, M., Babeyko, A., Wiemer, S., and Grünthal, G. (2012). Probabilistic tsunami hazard in the mediterranean Sea. *J. Geophys. Res. Solid Earth* 117, doi: 10.1029/2010JB008169
- Sraj, I., Mandli, K. T., Knio, O. M., Dawson, C. N., and Hoteit, I. (2014). Uncertainty quantification and inference of Manning’s friction coefficients using dart buoy data during the Tōhoku Tsunami. *Ocean Model.* 83, 82–97. doi: 10.1016/j.ocemod.2014.09.001
- Stich, D., Mancilla, F. D. L., Pondrelli, S., and Morales, J. (2007). Source analysis of the February 12th 2007, Mw 6.0 Horseshoe earthquake Implications for the 1755 Lisbon earthquake. *Geophys. Res. Lett.* 34, doi: 10.1029/2007GL030012
- Storlie, C. B., Swiler, L. P., Helton, J. C., and Sallaberry, C. J. (2009). Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliab. Eng. Syst. Saf.* 94, 1735–1763. doi: 10.1016/j.res.2009.05.007
- Terrinha, P., Pinheiro, L. M., Henriot, J.-P., Matias, L., Ivanov, M., Monteiro, J. H., et al. (2003). Tsunamigenic-seismogenic structures, neotectonics, sedimentary processes and slope instability on the southwest Portuguese Margin. *Mar. Geol.* 195, 55–73. doi: 10.1016/s0025-3227(02)00682-5
- Villa-Vialaneix, N., Follador, M., Ratto, M., and Leip, A. (2012). A comparison of eight metamodeling techniques for the simulation of N 2O fluxes and N leaching from corn crops. *Environ. Model. Softw.* 34, e51–e66. doi: 10.1016/j.envsoft.2011.05.003
- Violeau, D., Ata, R., Benoit, M., Joly, A., Abadie, S., Clous, L., et al. (2016). “A database of validation cases for tsunami numerical modelling, Sustainable Hydraulics in the Era of Global Change,” in *Proceedings of the 4th IAHR Europe Congress* (Liege), 377.
- Volkova, E., Iooss, B., and Van Dorpe, F. (2008). Global sensitivity analysis for a numerical model of radionuclide migration from the RRC “Kurchatov Institute” radwaste disposal site. *Stoch. Environ. Res. Risk Assess.* 22, 17–31. doi: 10.1007/s00477-006-0093-y
- WENRA (2015). Guidance document issue tNatural hazards head document, Guideline for the WENRARHWG Safety Reference Levels for Natural hazards introduced as lesson learned from TEPCO Fukushima Daiichi accident. Available online at: http://www.wenra.org/media/filer_public/2016/11/04/wenra_guidance_on_extreme_weather_conditions_-_2016-10-11.pdf
- Wolfs, V., Meert, P., and Willems, P. (2015). Modular conceptual modelling approach and software for river hydraulic simulations. *Environ. Model. Softw.* 71, 60e77. doi: 10.1016/j.envsoft.2015.05.010
- Yanagisawa, K., Imamura, F., Sakakiyama, T., Annaka, T., Takeda, T., and Shuto, N. (2007). Tsunami assessment for risk management at nuclear power facilities in Japan. *Pure Appl. Geophys.* 164, 565–576. doi: 10.1007/978-3-7643-8364-0_16
- Yazdi, J., and Salehi Neyshabouri, S. A. A. (2014). Adaptive surrogate modeling for optimization of flood control detention dams. *Environ. Model. Softw.* 61, 106–120. doi: 10.1016/j.envsoft.2014.07.007
- Zitellini, N., Chierici, F., Sartori, R., and Torelli, L. (1999). The tectonic source of the 1755 Lisbon earthquake and tsunami. *Annali di Geofisica* 42, 49–55. doi: 10.4401/ag-3699

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Bacchi, Jomard, Scotti, Antoshchenkova, Bardet, Duluc and Hebert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Inversion Algorithm for Civil Flood Defense Optimization: Application to Two-Dimensional Numerical Model of the Garonne River in France

Yann Richet* and Vito Bacchi

Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, France

OPEN ACCESS

Edited by:

Philippe Renard,
Université de Neuchâtel, Switzerland

Reviewed by:

Jeremy Rohmer,
Bureau de Recherches Géologiques
et Minières, France
Rachid Ababou,
UMR5502 Institut de Mécanique des
Fluides de Toulouse (IMFT), France

Roland Löwe,
Technical University of Denmark,
Denmark

*Correspondence:

Yann Richet
yann.richet@irsn.fr

Specialty section:

This article was submitted to
Freshwater Science,
a section of the journal
Frontiers in Environmental Science

Received: 18 December 2018

Accepted: 30 September 2019

Published: 06 November 2019

Citation:

Richet Y and Bacchi V (2019)
Inversion Algorithm for Civil Flood
Defense Optimization: Application to
Two-Dimensional Numerical Model of
the Garonne River in France.
Front. Environ. Sci. 7:160.
doi: 10.3389/fenvs.2019.00160

The objective of this study is to investigate the “inversion approach” for flood defense optimization in an inundated area. This new methodology within this engineering field consists in defining a “safety criterion” (for instance, “the water level in a given location must be lower than a given value”) and the combined analysis of all the uncertain controlled parameters (i.e., flood defense geometry, location, etc.) that ensure the safety objective for all the possible combinations of uncontrolled parameters (i.e., the flow hydrograph parameters) representing the natural phenomenon is not exceeded. To estimate this safety set, a metamodeling approach will be used which significantly reduces the number of model evaluations required. This algorithm relies on a kriging surrogate built from a few model evaluations, sequentially enriched with new numerical model evaluations as long as the remaining uncertainty of the entire safety set remains too high. Also known as “Stepwise Uncertainty Reduction,” this algorithm is embedded in the “Funz” engine (<https://github.com/Funz>) tasked with bridging the numerical model and any design of experiments algorithm. We applied this algorithm to a real two-dimensional numerical model of the Garonne river (France), constructed using the open-source TELEMAC-2D model. We focused our attention mainly on the maximum water depth in a given area (the “safety criterion”) when considering the influence of a simplified flood defense during a flooding event. We consider the two safety control parameters describing the slab and dyke elevations of the flood defense system, to design against the full operating range of the river in terms of possible watershed flooding. For this application case, it appears that less than 200 simulations are needed to properly evaluate the restricted zone of the design parameters (the “safety zone”) where the safety criterion is always met. This provides highly valuable data for full risk-informed management of the area requiring protection.

Keywords: kriging surrogate, Bayesian optimization, inversion, level set, uncertainty, hydraulic modeling

1. INTRODUCTION

It is well-known that the world’s major lowland rivers (the Rhine, the Po, the Elbe River, and the Loire River) are protected against flooding by embankments or other flood defenses (Ciullo et al., 2019). The embankments and so-called primary flood defenses such as flood walls and dams (Kind, 2014) are aimed primarily at reducing the likelihood of flooding in the protected area, and

historically they have been the most commonly-adopted flood risk reduction measure (Ciullo et al., 2019). The design of these flood protection measures is therefore of major importance society-wide, and can have a considerable impact on the economic and demographic development of the alluvial plains (White, 1945).

The classical approach for flood defense systems design was developed in the Netherlands in the wake of the 1953 disaster (see Vrijling, 2001). Since, reliability-based flood defense design strategies have been developed all over the world (see Vrijling, 2001; van Gelder and Vrijling, 2004; Ciullo et al., 2019). These strategies mainly involve the statistical quantification of the hazard (see Vrijling, 2001; Apel et al., 2004; van Gelder and Vrijling, 2004; Polanco and Rice, 2014) and the “economic” optimization of the flood defense systems (see Vrijling, 2001; van Gelder and Vrijling, 2004; Ciullo et al., 2019). The optimum design is considered to be the value at which the total cost of investment (which increases with the height of the flood defense) and the present value of the risk (which diminishes with the increasing height) takes its minimum (see Vrijling, 2001; van Gelder and Vrijling, 2004; Ciullo et al., 2019). Another good example is the optimization model introduced by van Dantzig (1956) for the embankment height, which was further developed by other authors, as reported by Eijgenraam et al. (2016).

These statistical studies are well-suited to the definition of a global flood protection strategy (at “country” scale), as they are not supported by intensive numerical modeling. Although the value of these models is indisputable, the flooding probabilities of the protected areas are assumed to be independent of one another, disregarding the change in hydraulic load along the river stretch as a consequence of the state (e.g., failure, increase in safety) of the embankments elsewhere (Ciullo et al., 2019).

At a local scale, in practical engineering applications, the classical method for the design of flooding protections for urban and industrial facilities relies on the development of sophisticated and refined numerical model systems able to reproduce surface flow accurately for a given chosen “design scenario” (such as the 100-year return period flood event). The quantities of major interest for the design of a flood defense (wall, urban structures, drainage network, etc.) are often related to the parameters describing this design scenario, such as the water height at a given location, or the water velocity in the inundated area (Milanesi et al., 2015). Once these quantities are evaluated, the flood defense is accordingly designed.

Although robust, this “classical approach” can appear too simplified considering the variability of natural phenomena and the numerous uncertainties related to natural hazard modeling. Ideally, flood disaster mitigation strategies should be based on a comprehensive assessment of the flood risk, combined with a thorough investigation of the uncertainties associated with the risk assessment procedure (Apel et al., 2004). Specifically, numerous studies have demonstrated the influence of uncertainties on flood hazard assessment (see Apel et al., 2004; Alho and Mäkinen, 2010; Domeneghetti et al., 2013; Maurizio et al., 2014; Nguyen et al., 2015; Abily et al., 2016; Bacchi et al., 2018), sometimes underlining the hard-to-estimate damage caused by flooding (see Apel et al., 2004). These studies are

often based on the use of simplified numerical models which reduce the computational time (see Apel et al., 2004; Alho and Mäkinen, 2010; Domeneghetti et al., 2013; Maurizio et al., 2014; Nguyen et al., 2015), and they are an example of how uncertainty quantification techniques could be employed for the assessment of natural hazards. If applied to the design of flood defenses, these studies can be considered an “improvement” on the classical approach, since they make it possible to better evaluate the uncertainties related to the “design” parameters, which in this case are the target values derived from uncertainty quantification (i.e., a target water height).

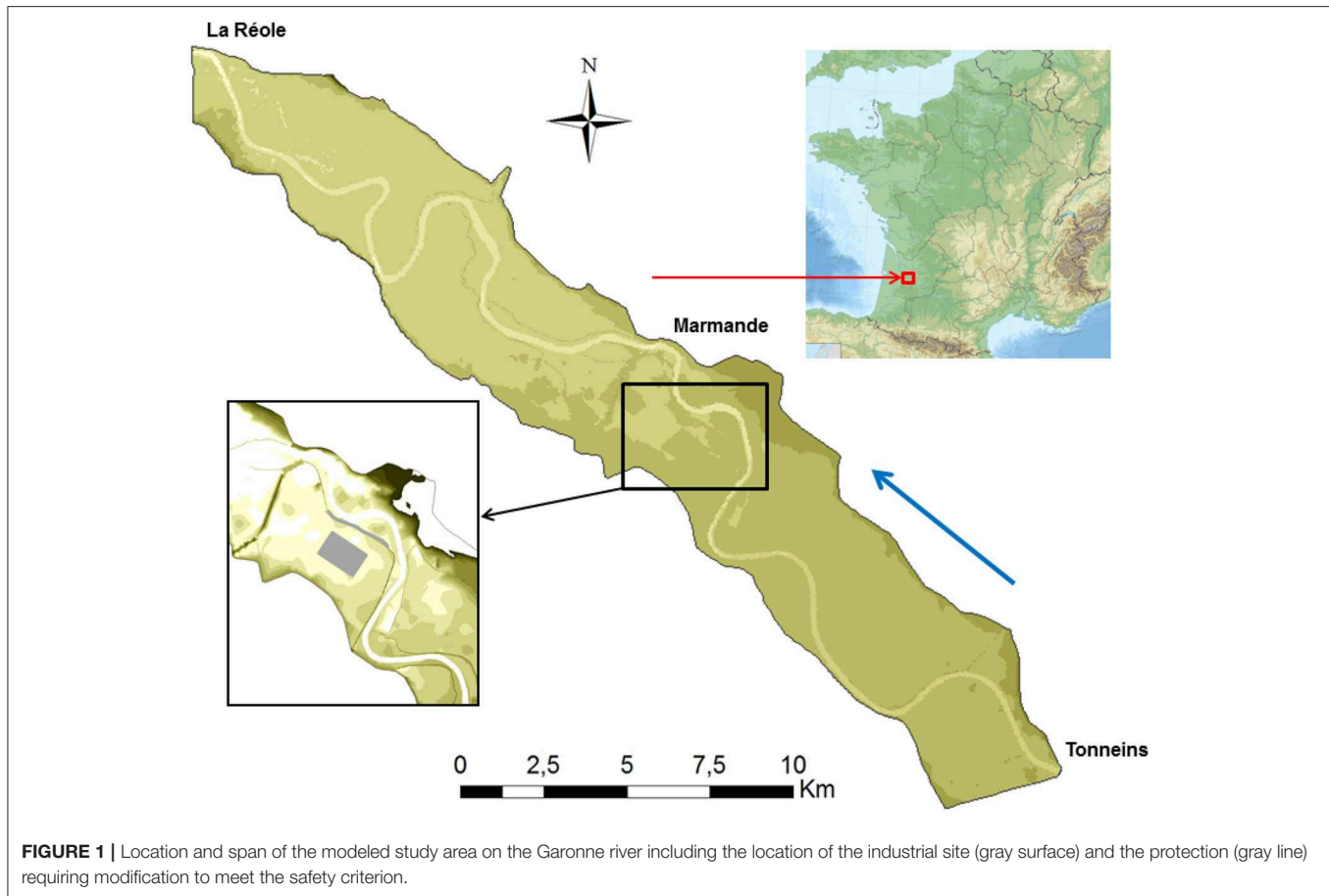
However, both the “classical” and “improved” approach for flood defense design suffer the same limitations, as they do not allow the end-user of the methodology to robustly evaluate the best flood defense configuration (the geometry) for the natural variability of the simulated phenomenon. Within this context, the objective of this work is therefore to investigate the “robust inversion approach,” subsequently referred to as RSUR, for the design of a flood defense in a two-dimensional inundated area. This method consists in defining a “safety criterion” (such as “the water level in the slab must remain lower than 25 cm”) and the analysis of suitable design parameters (for instance, the elevation of the flood defenses) that ensure the safety objective for all the possible combinations of uncertain input parameters (for instance, the flow hydrograph characteristics) describing the natural phenomenon to be met.

To estimate this safety set, a standard surrogate approach will be used (Jones et al., 1998), which significantly reduces the number of model evaluations needed. This algorithm relies on a kriging surrogate built from a few model evaluations, sequentially enriched with new numerical model evaluations as long as the remaining uncertainty of the entire safety set remains too high (Chevalier, 2013). This algorithm therefore appears well-suited to the rigorous study of the uncertainties of very refined numerical models traditionally used in engineering applications. Belonging to a more general class of “Stepwise Uncertainty Reduction,” this algorithm is embedded in the “Prométhée” workbench (using the “Funz” engine) tasked with bridging the numerical model and any design of experiments algorithm (further technical information is provided in an **Annex**).

In this research work, we first introduce the engineering problem we want to solve (section 2). Our proposed methodology for model resolution is then presented (section 3) with a focus on the numerical tools we develop for the methodology’s application to the chosen real case (section 4). Lastly, the results and main conclusions and perspectives are reported (section 5). The results will be introduced as the safety set and will be analyzed in terms of safety control within the river’s operating range. Once properly evaluated, this constrained zone provides highly valuable data for full risk-informed management of the river.

2. PROBLEM SPECIFICATION

The study area is a 50-km-long reach on the Garonne river between Tonneins and La Réole (**Figure 1**). The area was settled to protect flood plains by organizing flooding and flood storage



between 1760 and 1850, when many earthen levees were built to protect the harvest against spring floods (LPCB, 1983; SMEPAG, 1989). The river was canalized to protect residents from flooding after the historic 1875 flood event (SMEPAG, 1989).

In this section of the Garonne river, the flood defense is actually designed for a river flood of nearly $3,500 \text{ m}^3/\text{s}$. Specifically, successive storage areas give the Garonne profiles a particular configuration and allow flooding in the floodplain to be controlled. **Figure 2** shows three flow characteristics on a typical cross-section of the Garonne to illustrate the flooding sequence: (1) base flow $1,100 \text{ m}^3/\text{s}$; (2) bankfull flow $2,400 \text{ m}^3/\text{s}$; and (3) flow before overflow of the levees with the lowest protection level $3,500 \text{ m}^3/\text{s}$. Consequently, flooding of the less protected areas between Tonneins and La Réole occurs with a low return period, i.e., ~ 10 years, and only a few levees have a standard of protection higher than 30 years. Due to the flat topography and the presence of a steep floodplain lateral slope, the floodplains are largely inundated even for high-probability floods. The December 1981 flood was one of the largest floods occurring since the most severe flood on record (1875); this flood event was used as a reference for our study. During this 9-day event, the peak discharge measured at Tonneins reached $6,040 \text{ m}^3/\text{s}$, corresponding approximately to a 20-year flood, and the floodplains were fully inundated.

Within this context, it was decided to site an industrial area spanning nearly 1 km^2 in the vicinity of the left bank of the

Garonne river (see gray area in **Figure 1**). This zone is protected by a dyke that is nearly 2 km long, 20 m wide and at a constant elevation of nearly 25.8 m NGF¹ (see gray line in **Figure 1**). However, the local topography of this zone varies between 18 and 22 m NGF, and the area is fully inundated during a flooding event characterized by a peak discharge higher than $3,500 \text{ m}^3/\text{s}$ according to the current design of the dykes (see **Figure 2**). To protect the new industrial area against flooding, a decision was made to investigate the impact of modifying the current crest of the dyke and the basement of the future industrial area.

More practically, the objective of the study is to identify the actual dyke elevation and the platform elevation (the slab), to ensure the industrial area remains operational under all possible watershed flooding exceeding the defense (occurring at nearly $3,500 \text{ m}^3/\text{s}$). Specifically, the water height in the vicinity of the industrial area must not exceed 0.25 m above the slab, which is the safety criterion of this study. With this aim, we made the following assumptions for the uncontrolled input parameters (the flow hydrograph) and the controlled input parameters (crest of the flood defense, basement of the industrial area), both of which have a different role in the RSUR algorithm workflow presented in **Table 1** below:

¹Nivellement Général de la France (Above Ordnance Datum).

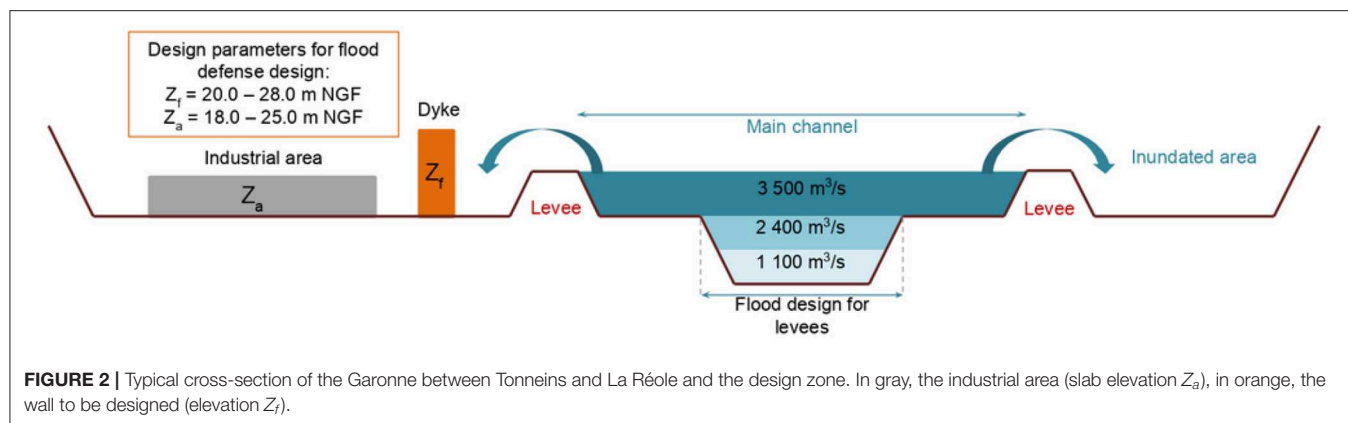


TABLE 1 | Qualified parameters of the study: time to peak discharge T_p , peak discharge rate Q_p , dyke elevation Z_f , slab elevation Z_a .

Uncontrolled parameters		Controlled parameters	
T_p [s]	Q_p [m^3/s]	Z_f [m NGF]	Z_a [m NGF]
3600.0–86400.0	2600.0–8000.0	20.0–28.0	18.0–25.0

The methodology used and the numerical chain developed for the study are presented in the next sections.

3. METHODOLOGY FOR ROBUST DESIGN OF PROTECTIONS

The intrinsic nature of many engineering problems like flooding protection is a combination of several “canonical” problems (a word used to express a natural orthogonality), mainly within the following mathematical classes:

- Optimization (e.g., search for the worst flood conditions).
- Inversion (e.g., set defense to not exceed 25 cm of flooding).
- Randomization (e.g., integrate a probabilistic flooding model according to a given model like the Gumbel law).

For instance, here we choose to describe the flood protection problem as a combination of designing protections to avoid exceeding a given water level (i.e., inversion part) when faced with the worst rain conditions for a given return period (i.e., max/optimization part). It should be noted that qualification of these aspects of the problem is somewhat arbitrary and may largely depend on the country’s regulation practice and the safety objective considered.

However, regardless of the choice made, the main nature of the problem considered (the identification/inversion of a flood level) is often “tainted” by secondary issues (worst flooding conditions), and all the parameters belong to one of these canonical roles. As an example, we could also mention other common engineering problems like robust optimization (optimizing some parameters and considering others as random) or constraint optimization (optimizing some parameters, keeping others verifying an [in]equality).

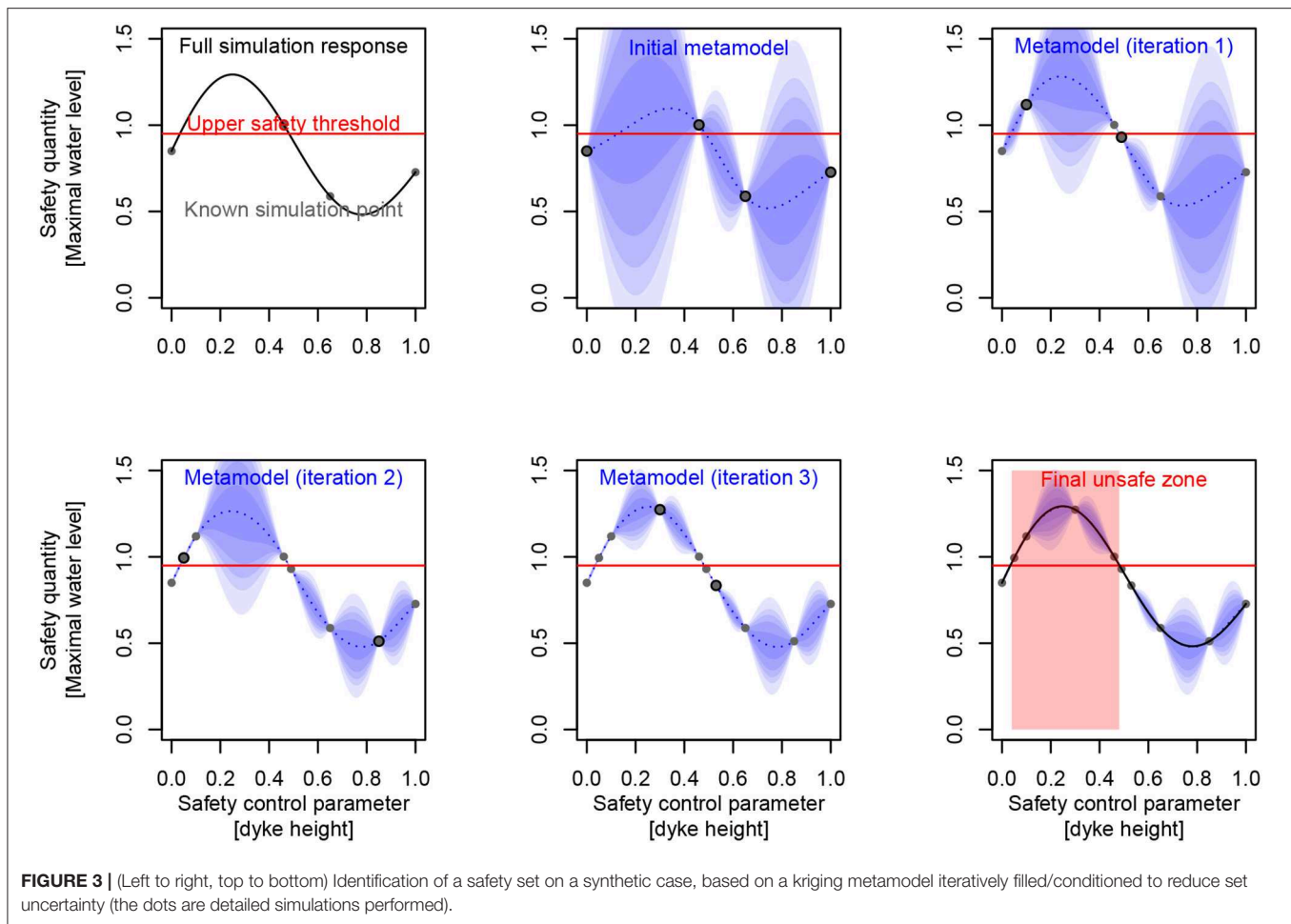
This “real-world” engineering practice also brings more complexity than canonical problems and thus requires dedicated algorithms in order to be solved efficiently (Chevalier, 2013). In the following, we will focus on the “robust inversion” problem, where the objective is to restore the civil engineering safety set identified for the worst flooding conditions without any probabilistic assumption.

3.1. Bayesian Metamodeling

The fairly common practice of Bayesian metamodeling is now a standard for solving any engineering task requiring numerous CPU-expensive simulations. Indeed, since the seminal paper describing the Efficient Global Optimization (EGO) algorithm (Jones et al., 1998; Roustant et al., 2012), many improvements have been proposed, investigating the algorithms’ efficiency (Picheny and Ginsbourger, 2013) or different problems to solve (Chevalier et al., 2014).

The rationale behind this approach consists in replacing most of the costly numerical simulations with an inexpensive surrogate function to investigate the properties that are relevant to our engineering purpose, like possible optimizers, excursion set or its main parameter effects. This surrogate function is designed to interpolate a few known “true” simulation points (Figure 3), taken as conditioning events of an initial uncertain/random function. Starting with a largely uncertain metamodel (Figure 3), this iterated process leads to a very precise metamodel around “true” simulated points, cleverly chosen in relation to our engineering objective (Figure 3):

A variety of metamodels have been applied in the water resources literature (Santana-Quintero et al., 2010; Razavi et al., 2012). Moreover, some examples of applications in the context of flood management have already been published (e.g., Yazdi and Salehi Neyshabouri, 2014; Löwe et al., 2018). However, like the previously mentioned EGO algorithm, for the study presented here, we will rely on conditional Gaussian processes (also known as kriging Roustant et al., 2012), derived from Danie Krige’s pioneering work in mining, later formalized within the geostatistical framework by Matheron (1973). Our choice is mainly motivated by this non-parametric metamodeling because although some properties of the considered response surface may be assumed (like continuity, derivability, etc.), its precise



shape cannot be assumed a priori. Thus, kriging became a standard metamodeling method in operational research (Santner et al., 2003; Kleijnen, 2015) and has performed robustly in previous water resource applications (Razavi et al., 2012; Villa-Vialaneix et al., 2012; Löwe et al., 2018). More practically, it is of considerable interest when investigating engineering objectives like max-minima or level sets of such random functions, inheriting convenient properties of Gaussian processes, and the method is fast provided the datasets exceed no more than a few hundred observations. The kriging model (which later provides an estimation of water height in an industrial area) is then defined for $x \in S$ (x will then represent study variables like the slab and dyke height or flow hydrograph parameters) as in the following Gaussian process:

$$M(x) = \mathcal{N}(m(x), s^2(x)) \quad (1)$$

where (for “simple” kriging):

- $\{X, Y\}$ are the coordinates of the “true” numerical simulations, taken as conditioning events of the statistical process:
- thus conditional mean is $m(x) = C(x)^T C(X)^{-1} Y$
- thus conditional variance is $s^2(x) = c(x) - C(x)^T C(X)^{-1} C(x)$
- C is the covariance kernel $C(.) = \text{Cov}(X, .)$, $c(.) = \text{Cov}(x, .)$

More than a commonplace deterministic interpolation method (like splines of any order), this model is much more informative owing to its predicted expectation and uncertainty. The fitting procedure of this model includes the choice of a covariance model [here a tensor product of the “Matern5₂” function, (Roustant et al., 2012)], and then the covariance parameters (e.g., range of covariance for each input variable, variance of the random process, nugget effect, etc.), could be estimated using Maximum Likelihood Estimation (standard choice we made) or Leave-One-Out minimization [known to mitigate the arbitrary covariance function choice (see Bachoc, 2013)], or even sampled within a full Bayesian framework (too far from our computational constraints at present). Once such a metamodel is specified, the following criteria will draw on such information to optimize the algorithm’s iterative sampling policy.

3.2. Design of (Numerical) Experiments

Once the Bayesian metamodeling framework is provided, the remaining issue is to define the sampling criterion used to fill the design of experiments: each batch of experiments will be proposed by the algorithm, then evaluated by the numerical simulator, and returned to the algorithm in order to propose the next batch (Algorithm 1). This general iterative process must be defined for each problem addressed: X input space, Y output

Algorithm 1: Sequential design of numerical experiments procedure based on J criterion of interest

Result: Alogrithm target (e.g., optimum, control set, sensitivity coefficients, ...)

Input : number of iterations, input variables space S , black-box function f , criterion maximizing algorithm target J

Output: sample of points in input space S , corresponding output values from f , metamodel M

```

1 Choose a preliminary uniform random sample  $X$  for some  $x \in S$ 
2 Evaluate  $Y = f(X)$ 
3 Fit metamodel  $M$  on  $\{X, Y\}$ 
4 while not reached end of iterations do
5   Compute  $X_{new} = \operatorname{argmax}_{x \in S} (J_M(x))$ 
6   Evaluate  $Y_{new} = f(X_{new})$ 
7   Append  $X = X \cup X_{new}$  and  $Y = Y \cup Y_{new}$ 
8   Fit metamodel  $M$  on  $\{X, Y\}$ 
9 end

```

target, and J criterion of interest (see later canonical or hybrid concrete instantiations):

Improvements to this fully sequential procedure have been proposed, like asynchronization (Le Riche et al., 2012) which may reduce the servers' sleep time between iterations (especially if the simulations have very different computing times). Nevertheless, it is often easier to consider synchronous batching (Ginsbourger et al., 2010) as an efficient turn-around to reduce user time, so the algorithm we will actually use becomes (Algorithm 2):

Note that this last algorithm may be greatly improved and likened to the previous one, when the criterion J is computable (in closed form) for a whole batch of points (see Chevalier and Ginsbourger, 2013 for such optimization criterion), which is sadly not often the case.

3.2.1. Canonical Optimization Problem

The EGO algorithm (Jones et al., 1998) is the common entry point for batch sequential kriging algorithms, as it proposes an efficient criterion (standing for criterion of interest J) called the "Expected Improvement" (EI):

$$J: x \longrightarrow EI(x) = E[(\min(Y) - M(x))^+] \quad (2)$$

- $\cdot \longrightarrow \cdot^+$ being "positive part" function,
- M is the kriging metamodel (see Equation 1),
- X is the optimization input space,
- Y are the black-box function evaluations on X (see Equation 1).

Computing this criteria (using the previous kriging formulas) is quite simple, so the main issue relates to the optimization of this criterion, whose maximum will define the most "promising" point for the next batch. This strategy will just propose the next point, and although a "multiple expected improvement" criterion (Chevalier and Ginsbourger, 2013) allows many points

Algorithm 2: Batch sequential design of numerical experiments procedure based on J criterion of interest

Result: Alogrithm target (e.g., optimum, control set, sensitivity coefficients, ...)

Input : number of iterations, input variables space S , black-box function f , criterion maximizing algorithm target J , batch size

Output: sample of points in input space S , corresponding output values from f , metamodel M

```

1 Choose a preliminary uniform random sample  $X$  for some  $x \in S$ 
2 Evaluate  $Y = f(X)$ 
3 Fit metamodel  $M$  on  $\{X, Y\}$ 
4 while not reached end of iterations do
5   Set  $X^* = X, Y^* = Y, M^* = M$ 
6   while not reached end of batch do
7     Compute  $x_{new}^* = \operatorname{argmax}_{x \in S} (J_{M^*}(x))$ 
8     Estimate  $y_{new}^* = E[M^*(x_{new}^*)]$  (known as "kriging believer" proxy)
9     Append  $X^* = X^* \cup x_{new}^*$  and  $Y^* = Y^* \cup y_{new}^*$ 
10    Fit metamodel  $M^*$  on  $\{X^*, Y^*\}$ 
11  end
12  Get  $X_{new} = \mathbb{C}_{X^*} X$  (ie. new points from  $X^*$  not yet in  $X$ )
13  Evaluate  $Y_{new} = f(X_{new})$ 
14  Append  $X = X \cup X_{new}$  and  $Y = Y \cup Y_{new}$ 
15  Fit metamodel  $M$  on  $\{X, Y\}$ 
16 end

```

to be proposed at a time, it remains practically more robust (numerically speaking) to use heuristics, which are usually preferred by practitioners: "Constant Liar," "Kriging Believer" (see Picheny and Ginsbourger, 2013).

3.2.2. Canonical Inversion Problem

Beyond such criteria dedicated to optimization problems, others are proposed to solve the (also) canonical problem of inversion, like Bichon's criterion (Bichon et al., 2008) which is similar to Expected Improvement, but focuses on proposing points closest to the inversion target:

$$J: x \longrightarrow EE(x) = E[(s(x) - |T - M(x)|)^+] \quad (3)$$

- $\cdot \longrightarrow \cdot^+$ being "positive part" function,
- M is the kriging metamodel and s its variance (see Equation 1),
- X belongs to the inversion input space S ,
- T being the target value of inversion output Y .

Other criteria for inversion have been proposed (Ranjan et al., 2008), but all of these so-called "punctual" criteria have the same intrinsic limitation of searching for a punctual solution to the inversion problem, while the answer should lie in a non-discrete space (usually a union of subsets of S).

Trying to solve the inversion problem more consistently will require defining an intermediate value like the uncertainty of

the excursion set above (or below) the inversion target (Bect et al., 2012). A suitable criterion will then focus on decreasing this value, and the final inversion set will be identified through the metamodel instead of its sampling points. This leads to the “Stepwise Uncertainty Reduction” (SUR) family of criteria (also abusively naming the following inversion algorithm) which takes a non-closed form, unlike the previous punctual ones:

$$\begin{aligned} J: x &\longrightarrow SUR(x) \\ &= -E \left[\int_S P[T < M_{n+\{x\}}(x')] \times P[T \geq M_{n+\{x\}}(x')] dx' \right] \\ &= -E \left[\int_S P[T < M_{n+\{x\}}(x')] \times (1 - P[T < M_{n+\{x\}}(x')]) dx' \right] \end{aligned} \quad (4)$$

- $M_{n+\{x\}}$ is the kriging metamodel conditioned by the n points $\{X, Y\}$ (just like “ M ” in Equation 1), **plus** the new point x ,
- X belongs to the inversion input space S ,
- T is the target value of inversion output Y ,
- so $P[T < M_{n+\{x\}}(x')]$ stands for the probability of exceedance in x' , then $P[T < M_{n+\{x\}}(x')] \times (1 - P[T < M_{n+\{x\}}(x')])$ gives a null contribution in the integral as far as $M_{n+\{x\}}(x')$ is certainly upper or lower than T .

At the cost of a significantly more computer-intensive task, the SUR criterion will propose more informative points and avoid the over-clustering defect often encountered with “punctual” criteria (even with the EGO algorithm whose exploration/exploitation trade-off is indeed due to heuristic tuning).

3.2.3. Robust Inversion Hybrid Problem

A by-product of the SUR algorithm is the extensive formulation used, which allows a flexible expression of the interest expectation. Therefore, it is now possible to use more complex interest values, for instance relying on the process properties of each subspace: S_c for controlled parameters and S_u for uncontrolled parameters. The “Robust Inversion” will then use the statistic of the marginal uncontrolled maximum (on S_u) of the process to integrate into the criterion (Chevalier, 2013):

$$\begin{aligned} J: x &\longrightarrow RSUR(x = \{x_c, x_u\}) \\ &= -E_n \left[\int_{S_c} P_{n+1}(x'_c) \times (1 - P_{n+1}(x'_c)) dx'_c \right] \end{aligned} \quad (5)$$

where:

- $x = \{x_c, x_u\}^{**}$
 - x_c stands for the coordinate of x in the input controlled subspace S_c ,
 - x_u stands for the coordinate of x in the input uncontrolled subspace S_u ,
- $P_{n+1}(x_c) = P[T < \max_{x_u \in S_u} (M_{n+\{x_c, x_u\}}(x_c, x_u))]$ is the marginal uncontrolled maximum probability:
 - $M_{n+\{x\}}$ is the kriging metamodel conditioned by the n points $\{X, Y\}$ (just like “ M ”), **plus** the new point x ,
 - T being the target value of inversion.

Using such an expression combines non-homogeneous subspaces, here an inversion subspace (for controlled variables) S_c , and an optimization subspace (for uncontrolled variables) S_u . Thus, this criterion leads to a hybrid algorithm of optimization “inside” inversion. More generally, this approach may be used to establish other criteria to solve hybrid problems. For instance, just replacing the “ \max_{x_u} ” statistic (used to define $P_{n+1}(x_c)$) by a mean or quantile on S_u may be useful for solving a probabilistic inversion problem, instead of the present worst-case inversion. Such hybrid algorithms are often much closer to solving real engineering concerns than purely canonical algorithms.

At this point of reasoning, it is very important to understand that a one-point marginal prediction of any kriging model is not sufficient to fully access the process behavior (and thus integrate it). Indeed, the correlation function behind process $C(.,.)$ is strongly related to the process sample functional properties, which may vary greatly depending on the kernel assumption (Figure 4). Such a process statistic (maximum here) is then impossible to compute as an independent sum of punctual evaluations of $x \in S$. A simple, but more costly approach consists in using a simulation of the random processes instead of its prediction statistical model.

Using this latter criterion (Equation 5) on our practical case study, we will now ask the algorithm (Algorithm 2) to sample the S space, trying to identify the safety set (dyke and slab height), irrespective of the flooding duration and discharge values.

In a self-supporting form, the batch sequential algorithm using this RSUR criterion on a TELEMAC-2D model of the Garonne river becomes:

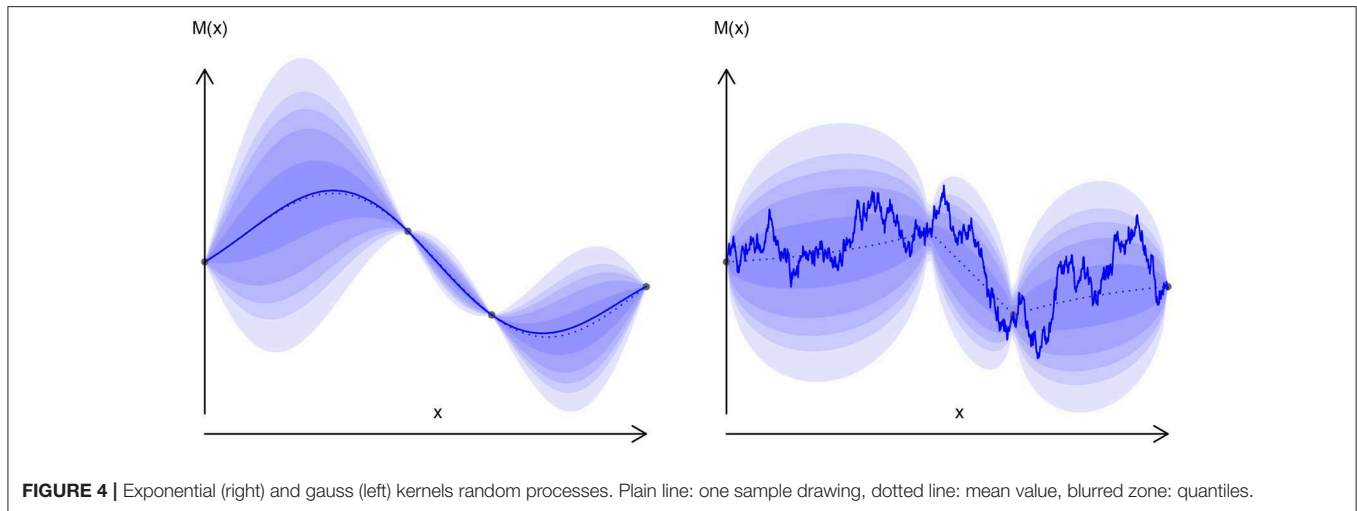
4. SOFTWARE AND NUMERICAL TOOLS

This work lies within a trend of applied research aimed at engineering enhancement. In practice, the previous (Algorithm 3) requires a heterogeneous set of hardware and software to be applied. The workbench which drives the simulations according to the algorithm: (Richet, 2019) [which may also be used through its Graphical User Interface (Richet, 2011)] is intended to fill the gap between:

- the flooding simulations: **TELEMAC-2D**, running in parallel on some Amazon Web Services “instances” (i.e., virtual machines);
- the robust inversion algorithm: **RSUR**, from the RobustInv R package (Chevalier et al., 2017). As a “reproducible research” target, these components are all available as free software and standard hardware (see **Annex** for details of how to implement them on a cloud platform).

4.1. Hydrodynamic Model of the Study Area

The Institut de radioprotection et de sûreté nucléaire (Institute for Radiological Protection and Nuclear Safety, IRSN) has been involved in the “Garonne Benchmark” project instigated by EDF. The aim of the project was to obtain an uncertainty quantification with hydraulic modeling on a stretch of the Garonne river. Earlier work carried out by Besnard and Goutal (2011) and Bozzi et al. (2015), has investigated discharge and roughness uncertainty



Algorithm 3: Batch sequential procedure to sample the safety frontier ($\max_t(H_{p1}) < 0.25$ m NFG) in control space $\{Z_a, Z_f\}$

Result: Safety frontier ($\max_t(H_{p1}) < 0.25$ m NFG) in control space $\{Z_a, Z_f\}$

Input : number of iterations, batch size

Output: Sampling of safety frontier ($\max_t(H_{p1}) < 0.25$ m NFG) in control space $\{Z_a, Z_f\}$, corresponding output values from TELEMAC-2D, kriging metamodel

```

1 Choose preliminary Latin Hypercube random sample  $X = \{T_p, Q_p, Z_a, Z_f\} \in \{[3600, 86400] \times [2600, 8000] \times [18, 25] \times [20, 28]\}$ 
2 Simulate with TELEMAC-2D  $H_{p1}(X)$  (one simulation for each point of  $X$ )
3 Evaluate  $Y = \max_t(H_{p1}(X))$ 
4 Fit Matern3/2 kriging  $M$ : maximize likelihood on  $\{X = \{T_p, Q_p, Z_a, Z_f\}, Y = \max_t(H_{p1})\}$  with hyper-parameters (variance and covariance range of  $T_p, Q_p, Z_a, Z_f$ )
5 while not reached end of iterations do
6   Set  $X^* = X, Y^* = Y, M^* = M$ 
7   while not reached end of batch do
8     Maximize on  $x_{new}^* = \{t_p, q_p, z_a, z_f\}$  criterion  $RSUR_{M^*}(x_{new}^*) = -E \left[ \int_{[18,25] \times [20,28]} P_{n+1}(z_a, z_f) \times (1 - P_{n+1}(z_a, z_f)) dz_a dz_f \right]$ , where  $P_{n+1}(z_a, z_f) = P[0.25 < \max_{t_p, q_p} (M_{n+x_{new}^*}^*(t_p, q_p, z_a, z_f))]$  stands for the probability of exceeding 0.25 for any possible  $\{t_p, q_p\}$  at a given coordinate  $(z_a, z_f)$ 
9     Estimate  $y_{new}^* = E[M^*(x_{new}^*)]$  (mean value of  $\max_t(H_{p1})$  predicted by kriging)
10    Append  $X^* = X^* \cup x_{new}^*$  and  $Y^* = Y^* \cup y_{new}^*$ 
11    Fit metamodel  $M^*$  on  $\{X^*, Y^*\}$ 
12  end
13  Get  $X_{new} = \mathbb{C}_{X^*} X$  (ie. new points from  $X^*$  not yet in  $X$ )
14  Simulate with TELEMAC-2D  $H_{p1}(X_{new})$  (one simulation for each point of  $X_{new}$ )
15  Set  $Y_{new} = \max_t(H_{p1}(X_{new}))$  (for each point of  $X_{new}$ )
16  Append  $X = X \cup X_{new}$  and  $Y = Y \cup Y_{new}$ 
17  Fit Matern3/2 kriging  $M$ : maximize likelihood on  $\{X = \{T_p, Q_p, Z_a, Z_f\}, Y = \max_t(H_{p1})\}$ 
18 end

```

in 1D and 2D hydraulic models. In order to contribute to the project, a version of the MASCARET 1D and TELEMAC-2D model, as well as hydraulic and hydrological data required to build these numerical models, were provided to the project's participants. In Besnard and Goutal (2011), the models' capacities to represent a major flood event were compared.

In this study, we use 2D model TELEMAC-2D from the open TELEMAC-MASCARET system (<http://www.opentelemac.org>). TELEMAC-2D solves 2D depth-averaged equations (i.e., shallow

water Equation 6). Disregarding the Coriolis, wind and viscous forces and assuming a vertically hydrostatic pressure distribution and incompressible flow, the 2D depth-averaged dynamic wave equations for open-channel flows can be written in conservative and vector form as:

$$\partial U / \partial t + \nabla \cdot F = S \quad (6)$$

where

- t = time (seconds)
- (x, y, z) = coordinate system with the x-axis longitudinal, y-axis transversal and the z-axis vertical upward (*meters*)
- $U = [h, hu, hv]^T$ = vector of conservative variables, with h being flow depth (*meters*), u and $v = x$ and y -components of the velocity vector (*meters/seconds*)
- $F = F(U) = [E(U), G(U)]^T$ = flux vector with $E = [hu, hu^2 + gh^2/2, huv]^T$ and $G = [hv, huv, hv^2 + gh^2/2]^T$ (g is gravitational acceleration in *meters/seconds²*)
- $S = S_0 + S_f$, with $S_0 = [0, -gh\partial z_b/\partial x, -gh\partial z_b/\partial y]^T$ = dimensionless bottom slope and $S_f = [0, -gn^2u(u^2 + v^2)^{1/2}/h^{1/3}, -gn^2v(u^2 + v^2)^{1/2}/h^{1/3}]^T$ = energy losses due to the bottom and wall shear stress, where z_b is bed elevation and n is Manning roughness (*seconds/meters^{1/3}*).

The roughness is flow and sediment dependent, but for simplicity it is assumed to be constant in each of the numerical runs. In this work, turbulence is modeled using a constant eddy viscosity value.

A two-dimensional model of the study area was constructed by EDF within the framework of the “Garonne Benchmark” (Besnard and Goutal, 2011). The model is composed of nearly 82,116 triangular elements with different lengths varying from 10 m (for the dyke crest, or the main channel of the Garonne River) to 300 m for the inundated areas. The floodplain topography and bathymetry are represented by the interpolation of the triangular mesh on the photogrammetry data (downstream part of the study area) and the national topographic map (upstream part). The model covers nearly 136 km² of the Garonne river basin. It is forced upstream (the “Tonneins” section) by a flow hydrograph and downstream by considering steady flow conditions (the “La Réole” section). The model calibration is reported in Besnard and Goutal (2011).

The presented model was modified slightly in this study to introduce the industrial area reported in **Figures 1, 2**. The industrial platform was inserted into the model by elevating the topography of the corresponding mesh at a constant mean level varying from 18.0 to 25.0 m NGF (**Table 1**).

For the purpose of the study, triangular flow hydrographs were used in the upstream section of the model. At the beginning of each simulation, a steady flow corresponding to a permanent discharge of 2,100 m³/s is imposed on the model. The flow discharge is then increased linearly until the maximum water discharge (Q_p) is reached at the time step corresponding to T_p . Once the peak discharge Q_p is reached, the discharge is decreased linearly until the permanent flow of 2,100 m³/s is reached at a time step of $2 * T_p$.

4.2. Robust Inversion Algorithm

The Robust Stepwise Uncertainty Reduction (RSUR) algorithm was implemented in a dedicated R package (Chevalier et al., 2017). It should be used in the same way as other R packages dedicated to Bayesian optimization or inversion like DiceOptim (Picheny and Ginsbourger, 2013) or KrigInv (Chevalier et al., 2012). However, for convenient and simple integration (in the Funz workbench), the standardized wrapper of the MASCOT-NUM research group (Monod et al., 2019) was applied to

provide a front API² (<http://www.gdr-mascotnum.fr/template.html>) with the following R functions (<https://github.com/Funz/algorithm-RSUR>):

- **RSUR(options)**: basic algorithm constructor, will return RSUR object to hold algorithm state
- **getInitialDesign(rsur,input,output)**: function to return a first random sample based on input and output information (dimension, bounds, etc.)
- **getNextDesign(rsur,X,Y)**: function to return iterative new samples which maximize the RSUR criterion, based on previous {X, Y} conditions
- **displayResults(rsur,X,Y)**: function to display the current state of the RSUR algorithm based on previous {X, Y} conditions.

4.3. Funz Workbench

Developed by the IRSN and distributed under the BSD³ license (<https://github.com/Funz>), Funz is a server client engine designed to support parametric scientific simulations. An overhanging graphical user interface design for practical engineering is also available at <http://promethee.irsln.fr>. Funz can be easily and quickly linked to any computer simulation code through a set of wrapping expressions (a set of regexp-like lines in the ASCII file). It uses the R programming language that is freely available and widely used by the scientific community working with applied mathematics. In addition to its use by the research community, this language is also used by many regulatory organizations. It is therefore simple to integrate algorithms developed and validated by the scientific community into Funz, which can then be applied to the previously-linked computer codes. Thus, to perform this study, Funz was used to link Telemac (plugin available at <https://github.com/Funz/plugin-Telemac>) and the RSUR algorithm (available at <https://github.com/Funz/algorithm-RSUR>).

Moreover, in order to perform a given set of computations, Funz can bring together independent servers, clusters (above their own queue manager if available), workstations, virtual servers (see **Annex** for Amazon Web Services / EC2 example), and even desktop computers running Windows, MacOS, Linux, Solaris, or other operating systems. While more focused on medium-sized performance computing (usually less than 100 concurrent connected instances), the high performance bottleneck is indeed delegated to each connected simulator instance, able to require dozens of CPUs independently (or even launch a Funz master itself).

5. RESULTS, ANALYSIS, AND CONCLUSIONS

5.1. Overview Results of Algorithm

As previously mentioned, the RSUR algorithm is parameterized with:

- Controlled parameters: $Z_a \in [18, 25]$ (slab elevation), $Z_f \in [20, 28]$ (dyke elevation),

²Application Program Interface.

³Berkeley Software Distribution.

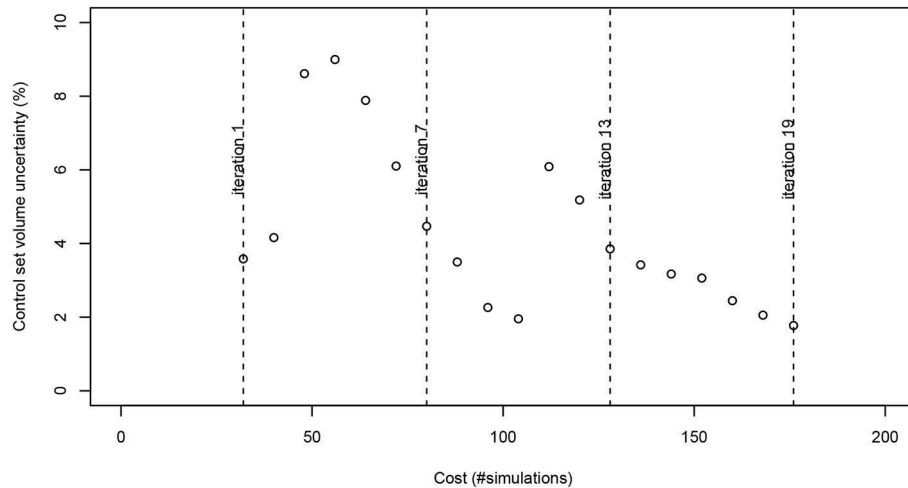


FIGURE 5 | Convergence of RSUR criterion during iterations.

- Uncontrolled parameters: $T_p \in [3600, 86400]$ (time to peak), $Q_p \in [2600, 8000]$ (peak discharge rate),
- Objective function: $\max_i(H_{p1}) < 0.25$ ($p1$ being one point in the center of the slab, this means we aim to not exceed 25 cm of water on this slab),
- Batches of 8 calculations at each iteration (plus 2^4 calculations of input parameter boundary combinations).
- 20 iterations of (algorithm 2).

Among these, the computing parameters (20 batches of 8 simulations) are defined arbitrarily, considering that:

- the number of iterations (20) has to be increased with the number of variables, so that the relative uncertainty of the control set reaches some percent in the end;
- the batch size (8) is mainly an opportunistic choice related to our computing resources, but it is also an empirical equilibrium to limit the batching effect of the 8 simultaneous points chosen [to mitigate the temporary hypothesis effect in M^* , see (Algorithm 2)].

The convergence of the algorithm is measured by the remaining uncertainty on the control set volume (numerically, this is the opposite of the RSUR criterion value (see Chevalier et al., 2014 for computing details) in **Figure 5**. This quantifies the “fuzzy zone” where it is still unclear whether the safety limit is exceeded or not, or in other words, where the limit is exceeded with a probability which is not 0 or 1 (visually the “gray zone”, while the “white zone” contains definite unsafe points, and the “black zone” definite safe points, **Figure 6**).

It should be noted that some intermediate raising of the RSUR criterion occurs (e.g., between iterations 10 and 11), when the kriging metamodel is changing abruptly because of the last data acquired, thus correcting the fitting of some kriging parameters. Between iterations 10 and 11, the range of covariance over the Z_f variable decreases (from 9.1 to 7.43), so the algorithm

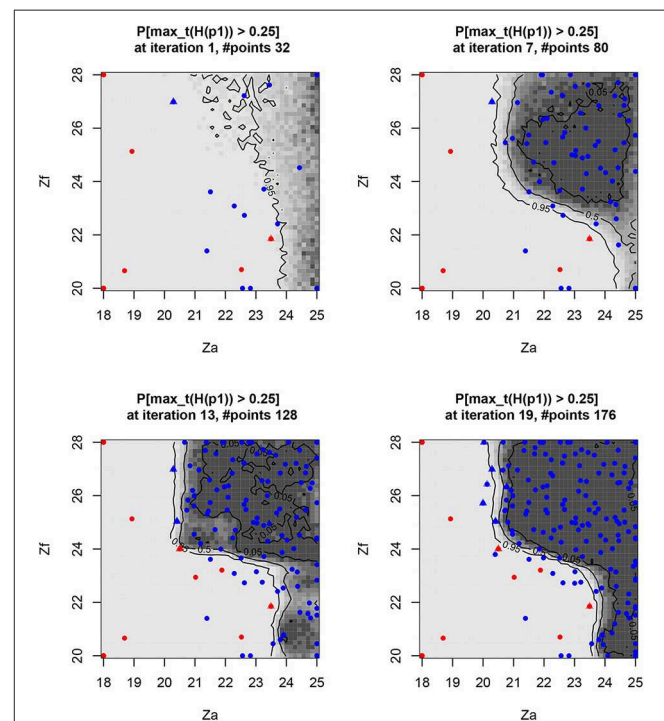


FIGURE 6 | Probability of exceeding the level of 25 cm on the slab, depending on the controlled Z_a and Z_f parameters, along some RSUR iterations (last added points as triangles, exceeding the points in red). The fairly dense sampling of the “safe” zone represents a practical safety guarantee.

expects a lower regularity on the Z_f dimension, which leads us to “discover” an unexpected safe zone for $Z_a > 24$ & $Z_f < 23$.

The safety-controlled set $S_c = \{Z_a, Z_c\}$ is iteratively identified, with increasing accuracy along the iterations (**Figure 6**):

It should be noted that the sampling density of the three zones is intrinsically different because of the asymmetry of the algorithm concerning the property of “unsafety”:

- The assumed “safe” zone needs to be sampled densely to assume its true safety⁴.
- The “unsafe” zone is discovered with sparse sampling, as just one exceeding (of level set) point is sufficient.
- The frontier between safe/unsafe is explored very densely, it being the critical information returned by the algorithm.

In addition to this raw result on the S_c subspace, it is also useful to consider some specific coordinates in their S_u projection. The response surface interpolated (by the kriging mean predictor) on such interest points, whose side of the frontier may lead to the safety criterion being exceeded (**Figures 7–9**) ($\max_t(H_{p1}) > 0.25$ are red points):

- Definitely “unsafe”: exceeding $\max_t(H_{p1}) > 0.25$, for (Z_a, Z_f) in the unsafe zone:
- On the frontier: $\max_t(H_{p1}) \sim 0.25$, for (Z_a, Z_f) in the undefined safety zone:
- Almost definitely “safe”: $\max_t(H_{p1}) < 0.25$, for (Z_a, Z_f) taken in the safe zone:

It should be noted that it was quite expected that maximum duration and discharge would be observed as the most penalizing configuration of this study, which is confirmed by these projections. When such an assumption may be proven prior to a study, it should be reached by using just an SUR algorithm, reducing the optimization space to a known value: $S_u = \{T_p, Q_p\} = \{\max(T_p), \max(Q_p)\}$.

5.2. Validation of Results

Validation of the previous results [i.e., the accuracy of the safe/unsafe limit where $\max_t(H_{p1}) \simeq 0.25$] should be performed considering the accuracy of the kriging metamodel obtained at the end of the algorithm iterations. Nevertheless, it is sufficient to verify that the metamodel is sufficiently accurate for $\max_t(H_{p1}) \leq 0.25$, as the inaccuracy of the metamodel for higher values where $\max_t(H_{p1}) > 0.25$ has no crippling impact on safety.

The best possible measurements of metamodel inaccuracy should be obtained using a fully-independent test basis of points, which is unattainable in most practical cases, where financial and resource constraints prevail. A cheaper alternative lies in cross-validation inaccuracy estimators, keeping in mind that this estimate is a “proxy” of the metamodel prediction error (see Bachoc, 2013). We will use the standard “Leave-One-Out” (LOO) estimator which computes the Gaussian prediction at each design point $x_i \in X$ when x_i is artificially removed to re-build the kriging metamodel (see Equation 7):

$$y_{-i} = M_{n-\{x_i\}} = \mathcal{N}(m_{-i}(x), s_{-i}^2(x)) \quad (7)$$

⁴This will remain an hypothesis, as long as this black-box algorithm does not consider any physical knowledge about the case study. A deeper physical study might have rigorously excluded any risk for $Z_a > 24$, for instance.

Where $M_{n-\{x_i\}}$ is the kriging metamodel conditioned by the n points $\{X, Y\}$ (just like “ M ” in (1)), **minus** the point x_i (previously belonging to X).

We compare this “blind” (somewhat, considering that x_i was not really chosen randomly) prediction y_{-i} (expectation and 95% confidence interval) to the true value y_i we already know (see **Figure 10**):

We observe that:

- the target value $y \simeq 0.25$ is sampled closely by several points (many of them where $y \simeq 0$);
- these “safety limit” points are quite well-predicted (small bias and confidence interval), but even at this last iteration few points are still underestimated in term of risk (i.e., $y_{\text{true value}} > y_{\text{predicted}}$), but it might be improved with some more iterations;
- although the unsafe zone is not well-predicted (often large confidence interval), this is not a major concern for the result (focusing at $\max_t(H_{p1}) < 0.25$).

Moreover, along the algorithm iterations the number and accuracy of the interest points increases (i.e., 95% confidence interval and bias decreasing together), so that the safety limit uncertainty decreases at the same time (see **Figure 11**):

5.3. Engineering Analysis

From an engineering perspective, the numerical tool developed in this study (sections 3.1, 3.2, and 3.3) allows us to identify the possible design parameter combinations (Z_a, Z_f) suitable to protect the industrial area against all the possible flooding events identified for the Garonne river (Q_p, T_p) (see all hydrographs simulated in **Figure 12**):

In particular, according to the parameters reported in **Table 1**, we study the combination of the elevation of the industrial basement Z_a and the current dyke height Z_f for all the possible flood events that could cause flooding of the area of interest, and characterized by a peak discharge varying from 2,600 to 8,000 m^3/s and a total duration varying from 2 to 48 h. This choice of uncontrolled parameters for the study seems robust considering that the historic flood event for this area does not exceed 6,040 m^3/s (SMEPAG, 1989).

The results presented in **Figure 6** show that the most relevant parameter for the industrial area’s defense against flooding is elevation of the industrial area Z_a . Specifically, the numerical results show that:

- If the basement of the industrial area is elevated above 24 m NGF, the safety criteria is always met even without modification of the current protection.
- On the contrary, if the basement of the industrial area is below 21 m NGF, any modification of the dyke elevation would be useless as the safety criteria is never met.
- An intermediary solution could be to set the basement of the industrial area between 21 and 24 m NGF by controlling the dyke elevation between 22 and 24 m NGF. The optimum point could be chosen according to other practical considerations, like the cost of the civil engineering work (Vrijling, 2001; van Gelder and Vrijling, 2004; Ciullo et al., 2019).

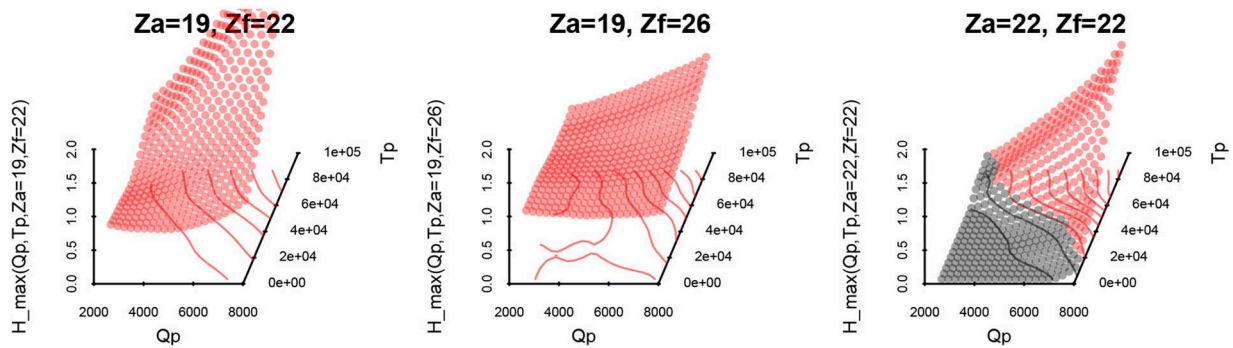


FIGURE 7 | Metamodel mean response surface on S_u for surely unsafe points.

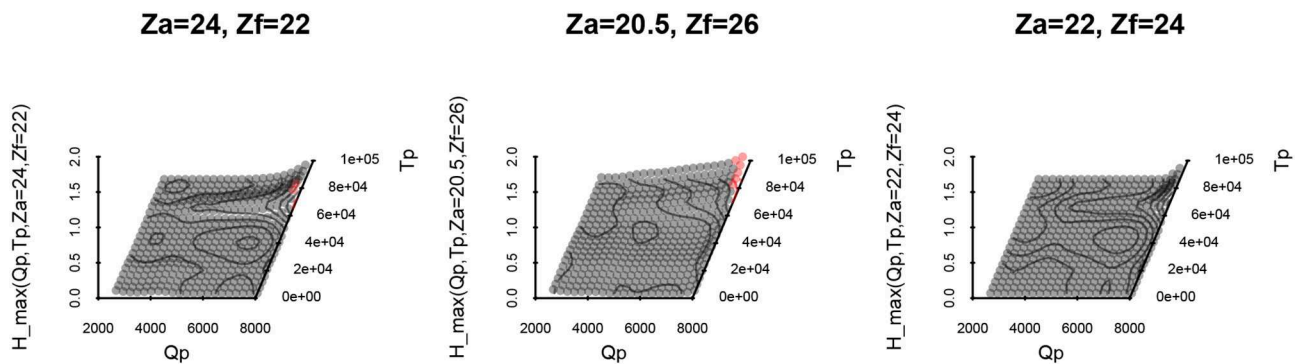


FIGURE 8 | Metamodel mean response surface on S_u for safety limit points.

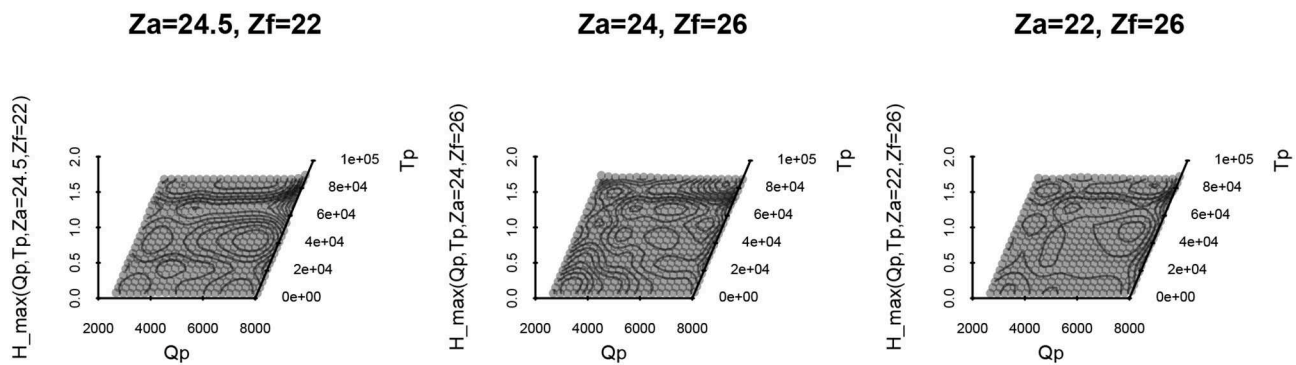


FIGURE 9 | Metamodel mean response surface on S_u for surely safe points.

5.4. Perspectives

The general methodology proposed in this study seems quite efficient for designing civil engineering for safety purposes. Although the true penalizing configuration of the worst flood event may have been assumed beforehand (which was at least confirmed by a blind metamodel), it might not be the general case for more complex models (in terms of protection degrees of freedom). It should be noted that the

rising number of variables (both controlled and uncontrolled) will have a cost in terms of metamodel predictability, and therefore in terms of the number of simulations needed to achieve enough accuracy for the control set limits. Tuning of such algorithm convergence parameters remains to be done for general cases.

Moreover, the developed numerical tool could be employed in several flooding hazard applications, for instance:

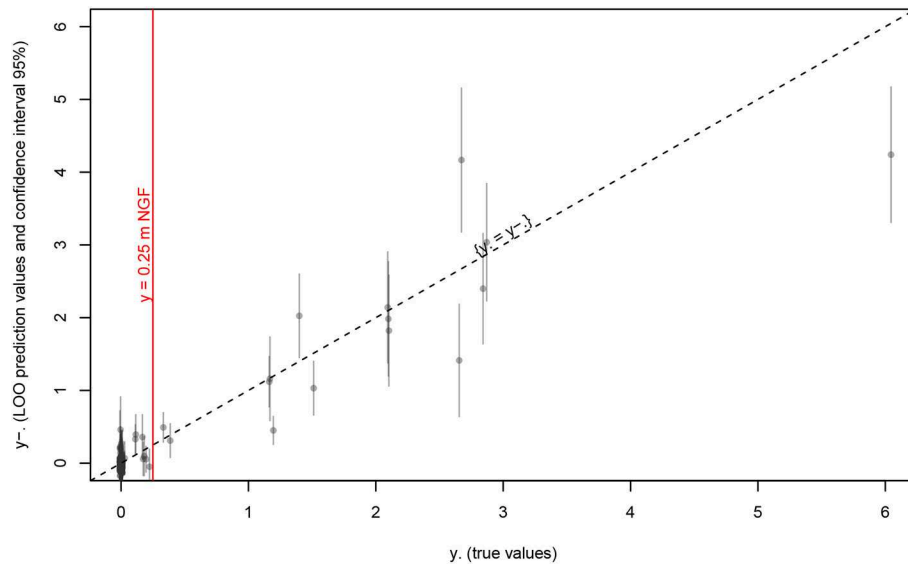


FIGURE 10 | Kriging prediction of $\max_t(H_{p1})$ vs. true values at the last iteration (19). Gray bars are issued from kriging variance, red line is the safety target, and dashed line is unbiased prediction.

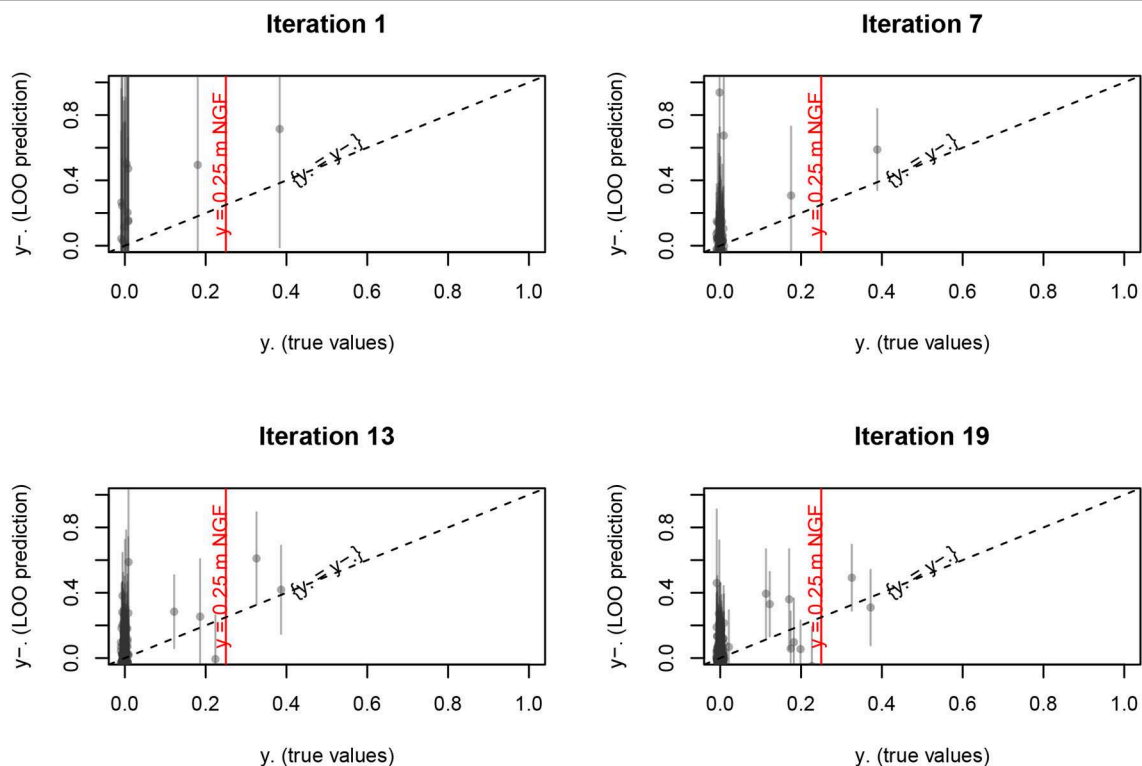


FIGURE 11 | Kriging prediction of $\max_t(H_{p1})$ vs. the true values at iteration 1, 7, 13, and 19.

- Quantification of the margins in the existing protections of the civil/industrial area (i.e., the algorithm ensures the protections are robust against any possible flood event).
- Reconstitution of a historical event (i.e., it could be possible to replace the safety criterion with historical knowledge and check for all the possible configurations of the

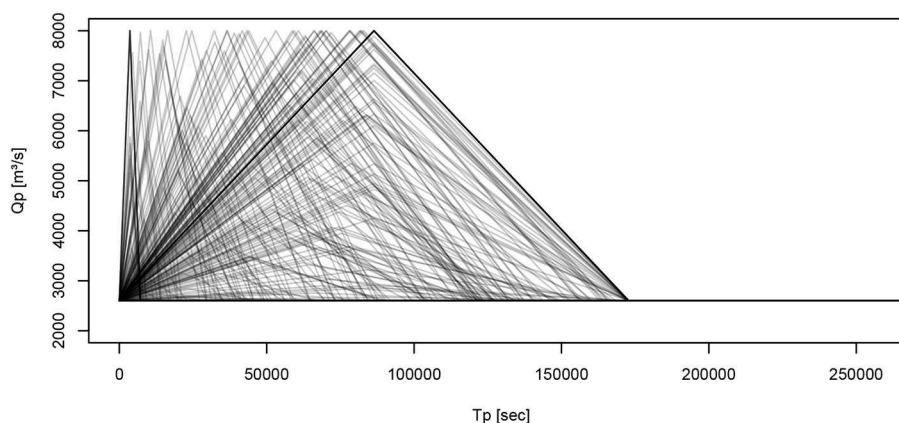


FIGURE 12 | Focused (expected penalizing) flow hydrographs (T_p , Q_p).

controlled and uncontrolled parameters reproducing the historical data).

Lastly, the primary question that remains is the connection of such an engineering practice to the probabilistic framework often adopted in regulation practice. The mainstream idea should be to weight the hydrograph by their probability of occurrence (thus inside the integral expression of the RSUR criterion), which will link the protection design to the flood return period as required.

In a probabilistic framework, results from this kind of analysis should indicate which return period the protections are

associated with and even which return period is hazardous for the civil or industrial installation.

AUTHOR CONTRIBUTIONS

YR: algorithmic inversion part and computing process. VB: Garonne model and case study definition.

ACKNOWLEDGMENTS

The authors want to thank Nicole Goutal and Cédric Goeury (EDF R&D) that provide the data necessary for this study.

REFERENCES

- Abily, M., Bertrand, N., Delestre, O., Gourbesville, P., and Duluc, C.-M. (2016). Spatial global sensitivity analysis of high resolution classified topographic data use in 2D urban flood modelling. *Environ. Model. Softw.* 77, 183–195. doi: 10.1016/j.envsoft.2015.12.002
- Alho, P., and Mäkinen, J. (2010). Hydraulic parameter estimations of a 2D model validated with sedimentological findings in the point bar environment. *Hydrol. Process.* 24, 2578–2593. doi: 10.1002/hyp.7671
- Apel, H., Thielen, A. H., Merz, B., and Blöschl, G. (2004). Flood risk assessment and associated uncertainty. *Nat. Haz. Earth Syst. Sci.* 4, 295–308. doi: 10.5194/nhess-4-295-2004
- Bacchi, V., Duluc, C.-M., Bardet, L., Bertrand, N., and Rebour, V. (2018). “Feedback from uncertainties propagation research projects conducted in different hydraulic fields: outcomes for engineering projects and nuclear safety assessment,” in *Advances in Hydroinformatics* (Springer), 221–241. doi: 10.1007/978-981-10-7218-5_15
- Bachoc, F. (2013). Cross validation and maximum likelihood estimations of hyperparameters of Gaussian processes with model misspecification. *Comput. Stat. Data Anal.* 66, 55–69. doi: 10.1016/j.csda.2013.03.016
- Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Stat. Comput.* 22, 773–793. doi: 10.1007/s11222-011-9241-4
- Besnard, A., and Goutal, N. (2011). Comparison between 1D and 2D models for hydraulic modeling of a floodplain: case of Garonne River. *La Houille Blanche*, 42–47. doi: 10.1051/lhb/2011031
- Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M. (2008). Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA J.* 46, 2459–2468. doi: 10.2514/1.34321
- Bozzi, S., Passoni, G., Bernardara, P., Goutal, N., and Arnaud, A. (2015). Roughness and discharge uncertainty in 1D water level calculations. *Environ. Model. Assess.* 20, 343–353. doi: 10.1007/s10666-014-9430-6
- Chevalier, C. (2013). *Fast Uncertainty Reduction Strategies Relying on Gaussian Process Models* (Ph.D. thesis). University of Bern, Bern, Switzerland.
- Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast Parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics* 56, 455–465. doi: 10.1080/00401706.2013.860918
- Chevalier, C., and Ginsbourger, D. (2013). “Fast computation of the multi-points expected improvement with applications in batch selection,” in *International Conference on Learning and Intelligent Optimization* (Berlin; Heidelberg: Springer), 59–69.
- Chevalier, C., Picheny, V. V., and Ginsbourger, D. (2012). *KrigInv: Kriging-Based Inversion for Deterministic and Noisy Computer Experiments*. Available online at: <http://CRAN.R-Project.org/package=KrigInv>
- Chevalier, C., Richet, Y., and Caplin, G. (2017). *RobustInv: Robust Inversion of Expensive Black-Box Functions*. French National Institute for Radio-logical Protection; Nuclear Safety (IRSN). Available online at: <https://github.com/IRSN/RobustInv>
- Ciullo, A., de Bruijn, K. M., Kwakkel, J. H., and Klijn, F. (2019). Accounting for the uncertain effects of hydraulic interactions in optimising embankments heights: proof of principle for the IJssel River. *J. Flood Risk Manage.* 12:e12532. doi: 10.1111/jfr3.12532

- Domeneghetti, A., Vorogushyn, S., Castellarin, A., Merz, B., and Brath, A. (2013). Probabilistic flood hazard mapping: effects of uncertain boundary conditions. *Hydrol. Earth Syst. Sci.* 17, 3127–3140. doi: 10.5194/hess-17-3127-2013
- Eijgenraam, C., Brekelmans, R., Den Hertog, D., and Roos, K. (2016). Optimal strategies for flood prevention. *Manage. Sci.* 63, 1644–1656. doi: 10.1287/mnsc.2015.2395
- Ginsbourger, D., Le Riche, R., and Carraro, L. (2010). “Kriging is well-suited to parallelize optimization. Adaptation learning and optimization,” in *Computational Intelligence in Expensive Optimization Problems*, eds C. K. Goh and Y. Tenne (Berlin; Heidelberg: Springer), 131–162.
- Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *J. Glob. Optimiz.* 13, 455–492.
- Kind, J. M. (2014). Economically efficient flood protection standards for the Netherlands. *J. Flood Risk Manage.* 7, 103–117. doi: 10.1111/jfr3.12026
- Kleijnen, J. P. C. (2015). *Design and Analysis of Simulation Experiments*, 2nd Edn. Tilburg: Springer.
- Le Riche, R., Girdziusas, R., and Janusevskis, J. (2012). “A study of asynchronous budgeted optimization,” in *NIPS Workshop on Bayesian Optimization and Decision Making* (Lake Tahoe).
- Löwe, R., Urich, C., Kulahci, M., Radhakrishnan, M., Deletic, A., and Arnbjerg-Nielsen, K. (2018). Simulating flood risk under non-stationary climate and urban development conditions – Experimental setup for multiple hazards and a variety of scenarios. *Environ. Model. Softw.* 102, 155–171. doi: 10.1016/j.envsoft.2018.01.008
- LPCB (1983). *Vallée de La Garonne - Enquête Sur Les Ruptures de Dignes de Décembre 1981 Entre Meilhan et Port Sainte Marie*. LPCB.
- Matheron, G. (1973). The intrinsic random functions, and their applications. *Adv. Appl. Probab.* 5, 439–368.
- Maurizio, M., Luigia, B., Stefano, B., and Roberto, R. (2014). “Effects of river reach discretization on the estimation of the probability of levee failure owing to piping,” in *EGU General Assembly Conference Abstracts*, Vol. 16. (Vienna).
- Milanesi, L., Pilotti, M., and Ranzi, R. (2015). A conceptual model of people’s vulnerability to floods. *Water Resour. Res.* 51, 182–197. doi: 10.1002/2014WR016172
- Monod, H., Iooss, B., Mangeant, F., and Prieur, C. (2019). *GdR Mascot-Num: Research Group Dealing with Stochastic Methods for the Analysis of Numerical Codes*. French National Center for Scientific Research. Available online at: <http://www.gdr-mascotnum.fr/>
- Nguyen, T., Richet, Y., Balayn, P., and Bardet, L. (2015). Propagation des incertitudes dans les modèles hydrauliques 1D. *La Houille Blanche* 5, 55–62. doi: 10.1051/lhb/20150055
- Picheny, V., and Ginsbourger, D. (2013). Noisy kriging-based optimization methods: a unified implementation within the Diceoptim Package. *Comput. Stat. Data Anal.* 71, 1035–1053. doi: 10.1016/j.csda.2013.03.018
- Polanco, L., and Rice, J. (2014). A reliability-based evaluation of the effects of geometry on levee underseepage potential. *Geotech. Geol. Eng.* 32, 807–820. doi: 10.1007/s10706-014-9759-2
- Ranjan, P., Bingham, D., and Michailidis, G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics* 50, 527–541. doi: 10.1198/004017008000000541
- Razavi, S., Tolson, B. A., and Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resour. Res.* 48, 7401. doi: 10.1029/2011WR011527
- Richet, Y. (2011). *Promethee*. French National Institute for Radio-logical Protection; Nuclear Safety (IRSN). Available online at: <http://promethee.irsn.org>
- Richet, Y. (2019). *Funz*. French National Institute for Radio-logical Protection; Nuclear Safety (IRSN). Available online at: <http://doc.funz.sh/>.
- Roustant, O., Ginsbourger, D., and Deville, Y. (2012). “DiceKriging, DiceOptim: two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *J. Stat. Softw.* 51, 1–55. doi: 10.18637/jss.v051.i01
- Santana-Quintero, L. V., Montano, A. A., and Coello, C. A. C. (2010). “A review of techniques for handling expensive functions in evolutionary multi-objective optimization,” in *Computational Intelligence in Expensive Optimization Problems*, eds Y. Tenne and C. K. Goh (Berlin; Heidelberg: Springer), 29–59.
- Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*, 1st Edn. New York, NY: Springer-Verlag.
- SMEPAG (1989). *Monographie Des Crues de La Garonne (Du Pont Du Roy Au Bec d’Ambès)*. SMEPAG.
- van Dantzig, D. (1956). Economic decision problems for flood prevention. *Econometrica* 24, 276–287.
- van Gelder, P. H. A. J. M., and Vrijling, J. K. (2004). “Reliability based design of flood defenses and River Dikes,” in *Life Time Management of Structures*. Available online at: https://www.researchgate.net/profile/Phajm_Gelder/publication/228943806_Reliability_based_design_of_flood_defenses_and_river_dikes/links/004635325941e6b8f7000000.pdf
- Villa-Vialaneix, N., Follador, M., Ratto, M., and Leip, A. (2012). A comparison of eight metamodeling techniques for the simulation of N 2O fluxes and N leaching from corn crops. *Environ. Model. Softw.* 34, 51–66. doi: 10.1016/j.envsoft.2011.05.003
- Vrijling, J. K. (2001). Probabilistic design of water defense systems in the Netherlands. *Reliabil. Eng. Syst. Safe.* 74, 337–344. doi: 10.1016/S0951-8320(01)00082-5
- White, G. F. (1945). *Human Adjustments to Floods*. Department of Geography Research, the University of Chicago.
- Yazdi, J., and Salehi Neyshabouri, S. A. A. (2014). Adaptive surrogate modeling for optimization of flood control detention dams. *Environ. Model. Softw.* 61, 106–120. doi: 10.1016/j.envsoft.2014.07.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Richet and Bacchi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ANNEX: SUPPLEMENTARY INFORMATION FOR IMPLEMENTATION

As already mentioned, this study relies on the Funz engine (Richet, 2019) which provides the computing back end to distribute calculations according to the RSUR algorithm. Considering that the reproducibility of this study is a standard deliverable, the following supplementary information details both the software and hardware implementation.

The “master” computer hosting Funz is a basic desktop computer (just running the Funz engine and RSUR algorithm), which will also remotely start the 8 EC2 instances (**Script 1**), create the SSH tunnels for protocol privacy and install Telemac on each instance.

```
for i in `seq 1 8`
do
  FunzDaemon-EC2.sh -d "lib/Funz/scripts/
  install_OpenTelemac.sh lib/*.slf" \
  -c "bash./scripts/install_OpenTelemac
  .sh" -o $i &
done
```

Script 1: Deploy 8 Funz services on EC2, including the Telemac installation.

All TELEMATC-2D calculations are then performed on the 8 servers (suited to the mesh used in this model) started on the Amazon Web Services EC2 cloud computing platform (**Screenshot 1**).

The main Funz script (**Script 2**) then starts the RSUR algorithm on the Telemac Garonne model, which is compiled (meaning study parameters are inserted in the template model files) and sent to the EC2 instances as required when the RSUR asks for a calculation point.

```
Funz.sh RunDesign \
-m Telemac \
-if t2d_garonne_hydro.cas t2d_garonne.cli
Qmax_Garonne_CMWR3 \
  princi_wall.f poi.txt loi_hq_Garonne \
-iv $Q_p$=[2600,8000] $T_p$=[3600,86400]
$Z_f$=[20,28] $Z_a$=[18,25] \
-oe "Numeric:max_t(H_p1)" \
-d RSUR -do xinv.index='3,4'
xopt.index='1,2' ytarget='<0.25' \
  initBatchSize='8'
initBatchBounds='true' batchSize='8' \
  iterations='20'
```

Script 2: The main Funz command which starts the RSUR algorithm on the Telemac Garonne river model (files t2d_garonne_hydro.cas t2d_garonne.cli Qmax_Garonne_CMWR3 princi_wall.f poi.txt loi_hq_Garonne).

Ultimately, this comprehensive study of 20 RSUR iterations requires about 10 h availability for all computing instances, equating to a cost of less than \$100 (at the current standard rates of the main cloud computing platforms).

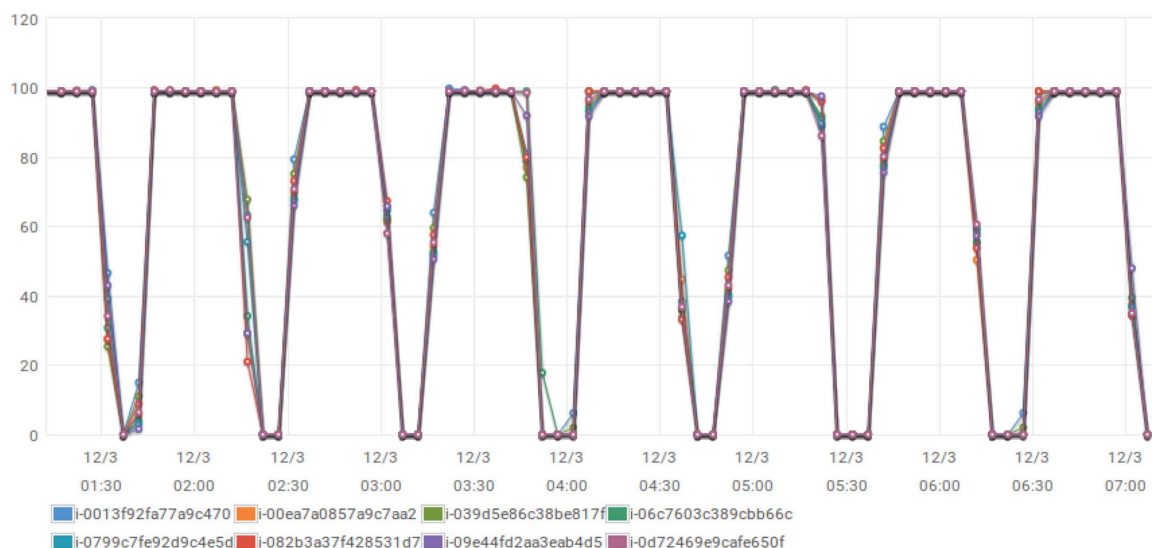
CloudWatch Monitoring Details

CPU Utilization (Percent)

Statistic: Average

Time Range: Last 6 Hours

Period: 5 Minutes



Screenshot 1 | 8 Amazon EC2 instances running the Telemac calculations on demand from the RSUR algorithm. Sleep times (non-negligible) are waiting for next RSUR batch.



Combining Clustering Methods With MPS to Estimate Structural Uncertainty for Hydrological Models

Troels Norvin Vilhelmsen^{1*}, Esben Auken¹, Anders Vest Christiansen¹,
Adrian Sanchez Barfod¹, Pernille Aabye Marker² and Peter Bauer-Gottwein³

¹ HydroGeophysics Group, Department of Geoscience, Aarhus University, Aarhus, Denmark, ² Sweco, Glostrup, Denmark,

³ DTU Environment, Department of Environmental Engineering, Technical University of Denmark, Lyngby, Denmark

OPEN ACCESS

Edited by:

Philippe Renard,
Université de Neuchâtel, Switzerland

Reviewed by:

Liangping Li,
South Dakota School of Mines
and Technology, United States
Frederic Nguyen,
University of Liège, Belgium
Ryan Martin,
University of Alberta, Canada

*Correspondence:

Troels Norvin Vilhelmsen
troels.norvin@geo.au.dk

Specialty section:

This article was submitted to
Hydrosphere,
a section of the journal
Frontiers in Earth Science

Received: 16 December 2018

Accepted: 24 June 2019

Published: 10 July 2019

Citation:

Vilhelmsen TN, Auken E,
Christiansen AV, Barfod AS,
Marker PA and Bauer-Gottwein P
(2019) Combining Clustering Methods
With MPS to Estimate Structural
Uncertainty for Hydrological Models.
Front. Earth Sci. 7:181.
doi: 10.3389/feart.2019.00181

This study presents a novel expansion of the clay-fraction (CF)/resistivity clustering method aiming at developing realizations of subsurface structures based on multiple point statistics (MPS). The CF-resistivity clustering method is used to define a data driven training image (TI) for MPS simulations. By combining this TI with uncertainty estimates obtained from correlation between the resistivity models and the unique categories in the TI, subsurface realizations are generated honoring geophysical and lithological data. The generated subsurface realizations were calibrated in a steady state groundwater model. Forecasts of well catchment zones were derived based on two wells located in areas with different levels of structural uncertainty. The catchment probability maps derived from the structural realizations were compared with the well catchment forecasted by a deterministic subsurface structure, and we are able to capture this catchment within the estimated uncertainties. We believe that this study is the first to combine MPS methods with a complete data driven workflow going directly from lithological and geophysical data to realizations of the subsurface structures. The main benefits of this is that it is data driven, fast, reproducible, and transparent.

HIGHLIGHTS

- Fast, transparent, and data driven uncertainty estimate for subsurface structures in groundwater models.
- The methodology uses multiple point statistics to capture the complexity of the subsurface structures better.
- Method supports generation of three-dimensional training images for practical application of multiple point statistics.

Keywords: groundwater modeling, uncertainty analysis, multiple point statistics, structural uncertainty, SkyTEM, SNESIM, MODFLOW

INTRODUCTION

Groundwater models are routinely used as tools for decision support in water resource related questions (e.g., Boronina et al., 2003; Almasri and Kaluarachchi, 2007; Mylopoulos et al., 2007; Saravanan et al., 2011; Sedki and Ouazar, 2011; Manghi et al., 2012; Enzenhoefer et al., 2014). However, despite being advocated often (e.g., Pappenberger and Beven, 2006;

Tartakovsky et al., 2012) uncertainty analysis is still not common practice in practical applications of groundwater modeling studies (Sanchez-Vila and Fernandez-Garcia, 2016; Delottier et al., 2017). Some arguments for the missing adaptation amongst professionals include limited access to software solutions (Renard, 2007; Tartakovsky et al., 2012), insufficient teaching and training of students, that can bring methods into practice (Renard, 2007; Sanchez-Vila and Fernandez-Garcia, 2016), reluctance among clients and decision makers to embrace model results presented with uncertainty (Freeze, 2004), and a further push for application of stochastic methods in practice (Renard, 2007), and finally limited access to regional scale, high resolution datasets. Most of the literature cited above, refers to uncertainty analysis in a broad context.

In the following, we will limit our focus to the sub-domain that deals with stochastic methods to represent uncertainty on subsurface structures in hydrological models. Traditionally, when performed, these stochastic simulations have been based on 2-point statistics such as sequential indicator simulation or sequential Gaussian simulation (Deutsch and Journel, 1998). However, the simplified assumptions of the 2-point statistics will often be insufficient to describe the complexity of the subsurface that governs groundwater flow (Zinn and Harvey, 2003). An alternative to the 2-point statistics is multiple-point statistics (MPS) (Hu and Chugunova, 2008; Mariethoz and Caers, 2015). In MPS, the statistical model consists of a training image (TI). This TI describes the spatial variability and patterns expected in the subsurface. As the name implies, MPS use multiple point information to estimate the correlation in structural patterns, and thus yields a more realistic representation of subsurface structure. A detailed review of the various MPS methods available will not be provided here, additional information can be found in e.g., Hu and Chugunova (2008) or Linde et al. (2015).

A common challenge for practical application of MPS methods is the lack of availability of a TI (Bastante et al., 2008; dell’Arciprete et al., 2012). TIs are often of conceptual nature, and encapsulates the experts knowledge of the system being analyzed (Caers, 2001). TIs can also be generated using parametric equations or simple templates (Maharaja, 2008). Using this approach a set of TI can easily be developed that encapsulates different conceptual interpretations. This approach was used by Hermans et al. (2015) for a local scale study of an alluvial aquifer. A limitation of this approach is that it is difficult to produce TIs that resemble the full complexities of the system, and the resulting TI will often be an over simplified version of the field conditions. Another common approach to generating TI is through manual interpretation of subsurface structures (Huysmans et al., 2008; Hoyer et al., 2017; Barfod et al., 2018). This approach integrates geological expert knowledge about the system, thus making it a more direct representation of it. For practical applications this approach has two limitations. First, it is a more time consuming methodology, resulting in a more expensive end product. Second, the derived TI may be highly non-stationary, therefore putting high constraints on additional auxiliary data to constrain the modeling (Strebelle, 2002). A TI can also be generated directly from the datasets collected in the field (Silva and Deutsch, 2014). Similar to

manual interpretations, this approach will often result in a non-stationary TI, and the simulations must therefore be made by including additional auxiliary variables. Another limitation to this approach is that the resolution of the TI is limited by the collected data (Linde et al., 2015). Processes resulting in structural variability below this scale of resolution can therefore not be encapsulated in the TI. It is therefore important, to acknowledge that model realizations cannot be produced with a higher resolution than the input dataset can provide. This methodology also requires input datasets with a sufficiently high spatial density and areal coverage rarely possible to obtain by e.g., drilling.

Large spatial coverage can be obtained using airborne electromagnetic (AEM) methods. The SkyTEM system (Sørensen and Auken, 2004) is one of the available AEM methods that have been used extensively for groundwater mapping (Møller et al., 2009b; Pryet et al., 2012; Korus et al., 2017; Knight et al., 2018). By inverting the EM dataset a model of the electrical resistivity of the subsurface can be produced. In sedimentary environments where the groundwater is uninfluenced by pore water salinity, structures with high resistivity will often be the water bearing units or aquifers, and structures with low resistivity will be the aquitards or confining units. This link between subsurface hydrological units and resistivity can, however, be both spatially variable and uncertain. To reduce this uncertainty, the interpretations can be assisted using lithological logs from boreholes. This process has been formalized in the clay fraction (CF) inversion method (Christiansen et al., 2014; Foged et al., 2014). Marker et al. (2015) further extended this methodology using k-means clustering to reduce the dimensionality of the CF and resistivity data to derive a set of hydrogeological units. In the following this methodology will be named CF-resistivity clustering.

In this study, we investigate how to utilize the CF-resistivity clustering method to create a 3D TI. A TI can thereby be created specifically for the investigated area using a fast and data driven approach. Subsequently, this TI is used together with an estimate of the subsurface uncertainty and hard data from boreholes to generate multiple realizations of the subsurface structures. Structural realizations are then incorporated into a groundwater flow model and calibrated to available hydrological data. Finally, the model realizations are used to estimate the uncertainty of a well catchment zone. By presenting an approach to deal with structural uncertainty analysis, we show how geostatistical methods can be applied to problems relevant to most groundwater resource professionals.

MATERIALS AND DATA

Study Area

The study area used to document the proposed methodology is located north-west of the town of Aarhus, Denmark (Figure 1). It is of principle interest to the local authorities and the public water supply company, due to its proximity to Aarhus, and its rich groundwater resources are important for the supply of drinking water. The area has been subject to several geophysical and geological mapping campaigns

(Kronborg et al., 1990; Danielsen et al., 2003; Sandersen and Jørgensen, 2003; Jørgensen and Sandersen, 2006; Møller et al., 2009a; Høyer et al., 2015). These studies have revealed a dense network of buried valley structures, which are incised into a thick sequence of Paleogene clays. In most parts of the area, these clays form the lower impermeable bed for the primary groundwater reservoirs located inside these valley structures. However, not all valleys are filled with sand and gravel (thus constituting aquifers); some are filled up with glacial deposits mainly consisting of clay till. These clay-filled valleys can act as hydraulic barriers and clay layers covering the primary aquifers. Besides the valley structures, plateaus between the valleys can hold till deposits consisting of either sand/gravel and clay deposited during the last glaciation. A more detailed description of the geological structure of the area can be found in Høyer et al. (2015).

Geophysical and Lithological Data

Lithological information from boreholes was extracted from the Danish national well database, Jupiter (Møller et al., 2009a). Within the test site, approximately 400 boreholes were available. The majority of these boreholes are shallow (below 50 m depth), and only a small subset (~3%) reaches depths of more than 100 m (**Figure 1**). The quality of the lithological information provided by these boreholes varied, based on their drilling method, purpose, age, etc. Therefore, to utilize these boreholes in a modeling context, their quality must be quantified. In the present study, this was done using the weighting scheme presented by He et al. (2015).

The geophysical dataset was collected using the SkyTEM304 system (Sørensen and Auken, 2004) in August 2013. The data consist of 24,600 SkyTEM soundings collected over 330 line kilometers flown with a line spacing of 100 m. Soundings were made with a spacing of 25 m along the flight lines. The geophysical dataset was processed using Aarhus Workbench following the methodology described by Auken et al. (2009). During processing, data biased by EM-noise are removed. The resulting data positions after processing are shown in **Figure 1**, where each position represents one 1D EM model. The inversion was carried out in a spatially constrained inversion (SCI) (Viezzoli et al., 2008) with a 1D sharp model formulation (Vignoli et al., 2015) as implemented in the AarhusInv inversion code (Auken et al., 2015). In an SCI, a stratified geological environment is mimicked through horizontal and vertical regularization constraints between proximate 1D resistivity models. The sharp inversion setup secures that smooth vertical and horizontal transitions in resistivity are penalized in the objective function, thus promoting more homogeneous resistivity models with fewer but sharper boundaries. To avoid over interpreting resistivity models with low sensitivity at large depths, all models have been individually blanked at their estimated depth of investigation (DOI) calculated as described by Christiansen and Auken (2012).

Hydrological Data and Forcing

Time series of stream base-flow estimates were determined for three sub-catchments within the regional model domain (outlined in **Figure 1C**), using the automatic filter approach presented by Arnold and Allen (1999) and Arnold et al. (1995).

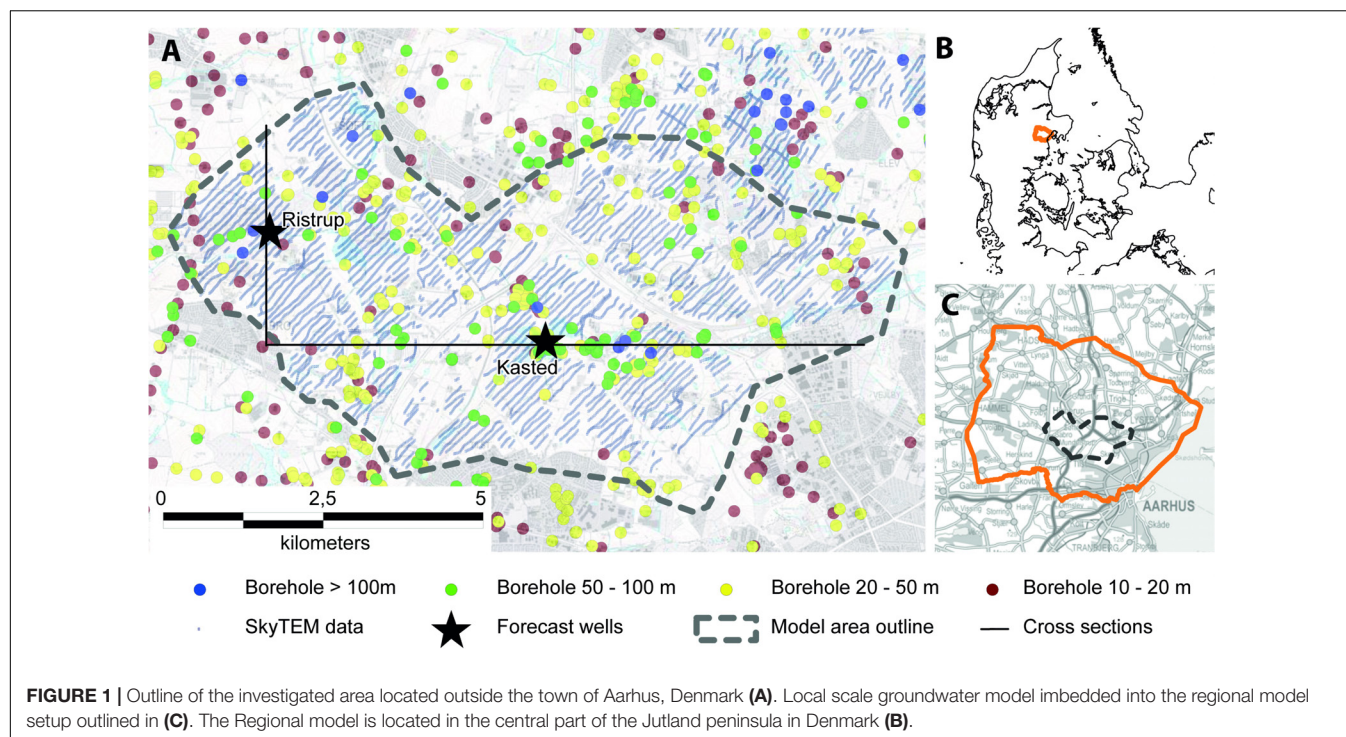
Estimates of steady state hydraulic heads were obtained for a total of 506 samples within the regional model area (**Figure 1C**). 106 of these samples were collected in the 62 wells that are screened in multiple aquifers, while 336 of the wells only have one screening interval, and thus a single steady state head estimate. Within the local groundwater model area, a total of 94 observations were available. 19 of the samples were collected from the 11 wells screened at multiple depths, while 75 of the wells only have one screening interval. Uncertainty of each head estimate was evaluated, taking into account factors such as the source of *xy* coordinates (GPS, read from map etc.); the method used to determine the reference level of a borehole; the length and temporal variation of time series used for estimating steady state heads; and the quality of the borehole (ranging from boreholes being part of the national groundwater monitoring program to old wells used for single household water supply).

Recharge rates were estimated using input from the national groundwater resource model of Denmark (Henriksen et al., 2003; Sonnenborg et al., 2003). First, the three dominating soil types near terrain were extracted from the model (sand, clay, and humus). On top of that, four land use types were defined (urban, farmland, forest, and water). By combining these, a total of 11 unique combinations were achieved. Subsequently, information on temperature, daily potential evapotranspiration, and precipitation was extracted from the national water resource model (assumed constant over the domain), and these were mapped onto the 11 unique zones. The recharge time series was then estimated using a simple linear root zone model (Jeppesen, 2001). Since the model used in the present study is steady state, these recharge estimates were averaged to a single value for each of the 11 zones.

Groundwater Model Setup

The groundwater model used in the study is set up in MODFLOW-USG. This version of MODFLOW was chosen such that the regional groundwater flow could be incorporated into the model setup while still having a fine numerical resolution within the local focus area. Details on the definition of the regional flow model can be found in Vilhelmsen et al. (2018), and the same model setup was used in Marker et al. (2017). The outline of the regional scale model can be seen from **Figure 1C** where it is marked with an orange line. The outline of the local scale model can be seen from **Figure 1A** where it is marked with a gray dotted line. Both the local model and the regional model have 11 model layers, and the local and the regional model have a horizontal discretization of 50 m and 100 m, respectively.

Two pumping wells have been chosen as focus for the present study. Both wells are located in existing well field areas. The location of the wells is marked with black stars on **Figure 1**. The Kasted well is part of the Kasted well field, and the Ristrup well is located within the Ristrup well field. These two locations were also chosen to represent a subarea with more complex geological settings (Kasted well), where noise in the geophysical data set resulted in areas with limited data coverage. The other well (Ristrup well) is located in a more homogeneous geological setting, where there was almost full coverage of the EM data.



METHODS AND THEORY

Clay Fraction-Resistivity Clustering

To obtain categorical information from lithological logs and resistivity models, we used the clay fraction (CF)-resistivity clustering methodology at borehole locations and sounding sites (Marker et al., 2017). The basic idea of the CF-resistivity clustering is to reduce the dimensionality of the input data (resistivity models and lithological logs), such that these are represented by a few (three or four) categorical variables. These categorical variables form the basis for the structural analysis of the subsurface.

The first step in this analysis is the clay-fraction inversion. Details on this methodology can be found in Christiansen et al. (2014) and Foged et al. (2014), and only a short resumé will be given here. In the CF inversion, boreholes are subdivided into equal discrete vertical intervals. Within each interval, the fraction of clay is determined, such that a CF of 1.0 is obtained if the discrete interval only contains clay and a CF of 0.0 if the interval does not have clay layers. The CF estimated at the boreholes, together with the resistivity models (averaged over the same elevation intervals), constitutes the dataset used in the CF inversion. The objective of the inversion is then to estimate a two-parameter error function that translates resistivities into CF such that the optimal translation occurs between CF estimated from boreholes and CF estimated from the resistivity models. Here, the two parameters are defined as the upper threshold, where all resistivities are translated to 0% clay, and the lower threshold is where resistivity values are translated to 100% clay. Since the correlation between resistivity and borehole lithology can vary spatially

within the survey area, the translation function parameters are defined on a 3D grid, where each grid node holds the two-parameter translation model. To constrain the inversion, the translation functions are regularized such that smooth translation parameters are preferred.

Once the CF inversion has been completed, each sounding site will have a set of resistivity and CF values with a predefined vertical resolution. To include the information at borehole locations, we also chose to interpolate a resistivity value using kriging from the geophysical soundings onto the boreholes. This provided us with both a resistivity and a CF at borehole locations.

For dimensionality reduction of the multivariable cloud to obtain a categorical dataset, we used K-means clustering (Wu, 2012), as implemented in Scikit-Learn in Python (Pedregosa et al., 2011). The clustering was performed on the principal components of the CF and the normalized resistivities (these were normalized to a scale of 0.0 to 1.0).

The outcome of the cluster analysis is a set of categorical values distributed in a three-dimensional space, which we can use to condition the generation of the subsurface realizations. The location of the categorical values is defined by the location and depth of the SkyTEM soundings and the locations and drilling depth of the boreholes. To fill the gaps between these data points, we use simple indicator kriging (IK) (Goovaerts, 1997, p. 293). Here, each cluster was treated as an indicator variable, which was each fitted with a variogram model. IK was then used to estimate the probability of each cluster distributed on a 3D grid. Based on these probabilities, an estimate of the most likely representation of the subsurface structure, given a predefined number of clusters, was created. This model constitutes the TI in the following simulations.

Multiple-Point Geostatistics

We could have used the estimated distribution of categorical variables together with the estimated variogram models to generate realizations of the subsurface structures with sequential indicator simulation (SIS) (Deutsch and Journel, 1998). This approach was used by Marker et al. (2017) for the same study area. However, SIS is limited by the amount of structural information that can be carried in the variogram model (Journel, 2005; Journel and Zhang, 2006). As the density of conditioning data increases, the dependency of the simulation result on the variogram model is reduced. However, to approximate the required data coverage the geophysical data had to be treated as hard data in the simulation. This constitutes a problem, since the probability of a given cluster will be dependent on the CF and the resistivity value of the pertaining data point. By treating the geophysical data as soft data, this information can be incorporated into the analysis, but it reduces the amount of hard data for the SIS, thus rendering it inapplicable.

To alleviate this problem, we have employed the MPS method single normal equation simulator (SNESIM) (Strebelle, 2002) in the present study. The higher order statistics incorporated in MPS methods allows them to capture and reproduce more complicated structures compared to variogram-based methods. Structural realizations will also create more coherent geological units, often making a better representation of actual field conditions. In SNESIM, the conditional information needed for the geostatistical simulation is obtained by scanning the TI. In SNESIM, this scanning is performed once, and the information is stored in a search tree structure. SNESIM is a sequential approach and it will visit each simulation node in the grid based on a random path. To incorporate hard data in the simulation, the data categories are assigned to the most proximate grid node prior to simulation. Soft data is incorporated into the analysis by updating the TI based probabilities, using the soft data probabilities. This update is done using the tau model (Journel, 2002). Conditional soft data must therefore be transformed into probabilities that can be used for conditioning. In the present study, this is done based on the conditional probabilities derived from cluster histograms. Alternatively, this information could be obtained from other statistical correlations between geophysics and lithology (Hermans et al., 2015; Barfod et al., 2016; Christensen et al., 2017b).

Single normal equation simulator is a complex algorithm, with many adjustable parameters. Fine tuning of the algorithm is therefore not always straight forward. In the present study, we derived our settings using a combination of the guidelines defined by Liu (2006) and He et al. (2014), and looping over parameter combinations using the Python interface implemented in SGeMS (Remy et al., 2009). The optimal parameters for the present study is defined in **Table 1**. These parameters were defined based on a compromise between simulation results and computational

burden. Parameters are only defined for the subset of parameters that deviate from default values (Remy et al., 2009).

In the present study we used vertical proportions and soft conditioning to define the target distributions. Both were defined based on the outcome of the cluster analysis. The entire workflow for generating structural realizations from MPS based on CF-resistivity clustering is outlined in **Figure 2**.

Groundwater Model, Particle Tracking and Inversion

The groundwater model was defined in MODFLOW-USG (Panday et al., 2015). MODFLOW-USG is a control-volume finite difference version of the well-known USGS groundwater flow code MODFLOW (Harbaugh et al., 2000; Harbaugh, 2005). Recharge was incorporated using the recharge package. Major stream segments were simulated using the River package. Trenches and minor stream segments were simulated using the drain package. Groundwater abstraction wells were simulated using the well package, and outer boundary conditions on the regional model edge were defined as either no-flow or constant head boundaries. This was determined based on the local hydrological conditions.

All model realizations, generated using SNESIM, were incorporated into the groundwater flow model and calibrated using PEST (Doherty, 2016). This was done by introducing the local groundwater model in a regional model used to simulate the potential interactions with the regional flow system. The parameters included in the inversion were the horizontal hydraulic conductivities of the determined clusters (except for the largest cluster, which was fixed based on sensitivity analysis in both the three- and the four-cluster model). Vertical hydraulic conductivities were tied to their horizontal counterparts at a ratio of 1:10. The two other parameter groups were the conductance terms of the drains and the streams within the local area, and the recharge multiplier. The objective function was defined as

$$\phi = \sum_{i=1}^{N_h} w_i ((h_{sim,i} - h_{obs,i}) / \sigma_{h,i})^2 + \sum_{i=1}^{N_q} ((q_{sim,i} - q_{obs,i}) / \sigma_{q,i})^2 \quad (1)$$

Where N_h is the number of steady state head observations, h_{sim} is the model simulated hydraulic heads, and h_{obs} are their observed counter parts. The residuals on the heads were weighted with the inverse of the observation standard deviation. Similar, q_{sim} , q_{obs} , and σ_q are the simulated, observed and uncertainty of the stream flow data. For the head data we also introduced an arbitrary weight (w_i) to the hydraulic heads. This was done to reduce the individual contribution to groups of observations located closely. All inversions were well posed, and we did not need to impose any regularization constrains.

To determine the catchment zones for the two pumping well scenarios we used mod-PATH3DU (Muffels et al., 2014).

TABLE 1 | Parameters used in the SNESIM simulations. Parameters are only defined when they deviate from the default values.

Cmin = 100	Max. conditioning = 100	Number of multigrids = 8	Subgrids = 1	Number of previously simulated nodes = 6	$\tau = 2$
------------	-------------------------	--------------------------	--------------	--	------------

Flow chart for performing MPS**simulations from CF-resistivity analysis**

1. Check ability for resistivity distribution to resolve major structural units (e.g. using a resistivity atlas Barfod et al. (2016))
2. Invert for CF distribution in the subsurface based on resistivity and boreholes (Foged et al. (2014))
3. Perform clustering on principal components of CF and normalized resistivity
4. Estimate indicator variograms for each cluster and estimate cluster probabilities using Indicator Kriging (IK)
5. Calculate maximum probability over all clusters for visualizing structural uncertainty
6. Calculate mode model (TI) from [5]
 - Visualize structural uncertainty on TI using maximum probability
7. Calculate vertical proportions from results obtained in [3]
8. Run MPS simulation using TI estimated from [6], vertical proportions from [7] and probability distribution from [4]. Additional hard data can be introduced at this step. In the present study we used screen intervals from pumping wells as hard data for cluster 0.

FIGURE 2 | Outline for generating structural realizations based on MPS.

Catchment zones were determined using backward tracking with 16,000 particles from the pumping well to terrain. This number of particles was the minimum required to make the catchment independent on the particle starting locations. Particle starting locations were randomly distributed within a cylinder. The height of this cylinder was equal to the pumping-well screening interval, and the diameter of the cylinder was determined such that the amount of recharge occurring within the well catchment zone balanced the amount of groundwater abstracted by the well. This estimation was done by trial and error based on the calibrated mode model. The catchment was discretized into 50 m by 50 m cells, and a cell was determined to be part of the catchment if at least one particle terminated in the cell. This tracking was performed for each of the stochastically generated subsurface structures. Once completed, the catchment probability maps were generated based on an averaging over all realizations.

RESULTS

Figure 3 shows the results of the k-means cluster analysis. The clustering was performed for three and four clusters. This was chosen based on two criteria. The three-cluster model was chosen since this is most comparable with the deterministic model available for the area, and the four-cluster model was chosen since this was deemed optimal in Marker et al. (2017), based on an optimal data fit criteria. The number of units are also limited

by the resolution of the input variables. In this case consisting of resistivity and CF values. The subdivision of clusters in the clouds spanned by the resistivities and the CF can be seen in the bottom subplots (E and F). Each point in these clouds thus represents a single pair of CF and resistivity values. These clusters are then mapped in histograms for the resistivities (subplot A and B). These histograms can also be used to estimate the conditional probabilities based on the kernel density functions (subplot C and D).

Each point in the data clouds represents a specific discrete interval in a resistivity sounding or a borehole. This means that the clusters can be mapped back as a point distribution in three dimensions. The fitted variogram models used for IK can be seen from **Figure 4**. Very importantly, the model obtained with IK (the mode model) is used as a TI in the MPS simulations. Cross sections through the mode model are shown in **Figure 5**. From the three-dimensional distribution of clusters, we can also calculate the vertical fractions of each cluster with depth. The result of this calculation can be seen from **Figure 5**. This information is used as additional conditional information in the SNESIM algorithm.

Figure 5 contains the information needed for the MPS simulations for the three-cluster model. A similar plot can be made for the four-cluster model, but this was considered redundant in the present case. The figure shows cross sections through the TI in the top left, and the vertical fractions in the top right. In the bottom, probability maps for each cluster are shown. These probability maps are determined from the conditional probabilities estimated using the histograms shown in **Figures 3A,B**. For each data point in space, we calculated the conditional probability of a given cluster based on the histograms. Subsequently, a variogram model was fitted to the probabilities for each cluster, and the probabilities were interpolated in three dimensions to fill out the model grid. Finally, the probabilities over all clusters were scaled such that they summed to one within each cell in the model grid. These probability grids are the ones shown in the bottom of **Figure 5**. The same methodology was used to estimate the probabilities in the four-cluster model. A similar methodology to estimate the probabilities was used by Hermans et al. (2015). The only conditional data that is not presented in **Figure 5** is the hard data used in the analysis. Hard data were only used for screen intervals of pumping wells in groundwater model. In the MPS simulations, these intervals were conditioned to be part of cluster 0, namely the cluster with the highest sand fraction. This is a reasonable assumption, since most of these wells have been active for several years and must therefore be screened in sandy deposits.

Besides using the estimated probabilities for conditioning the MPS simulations, they can also be used to illustrate the uncertainty of the underlying structures. Such plots have been generated in **Figures 6, 7**. **Figure 6** shows a cross section through the Kasted well field (the location of the cross section can be seen from **Figure 1**). **Figure 6A** shows the cross-section through the three-cluster model, **Figure 6C** shows the cross section through the four-cluster model, and **Figure 6D** show the cross section through the deterministic model structure. Each of these models have then been overlaid with a transparency filter based

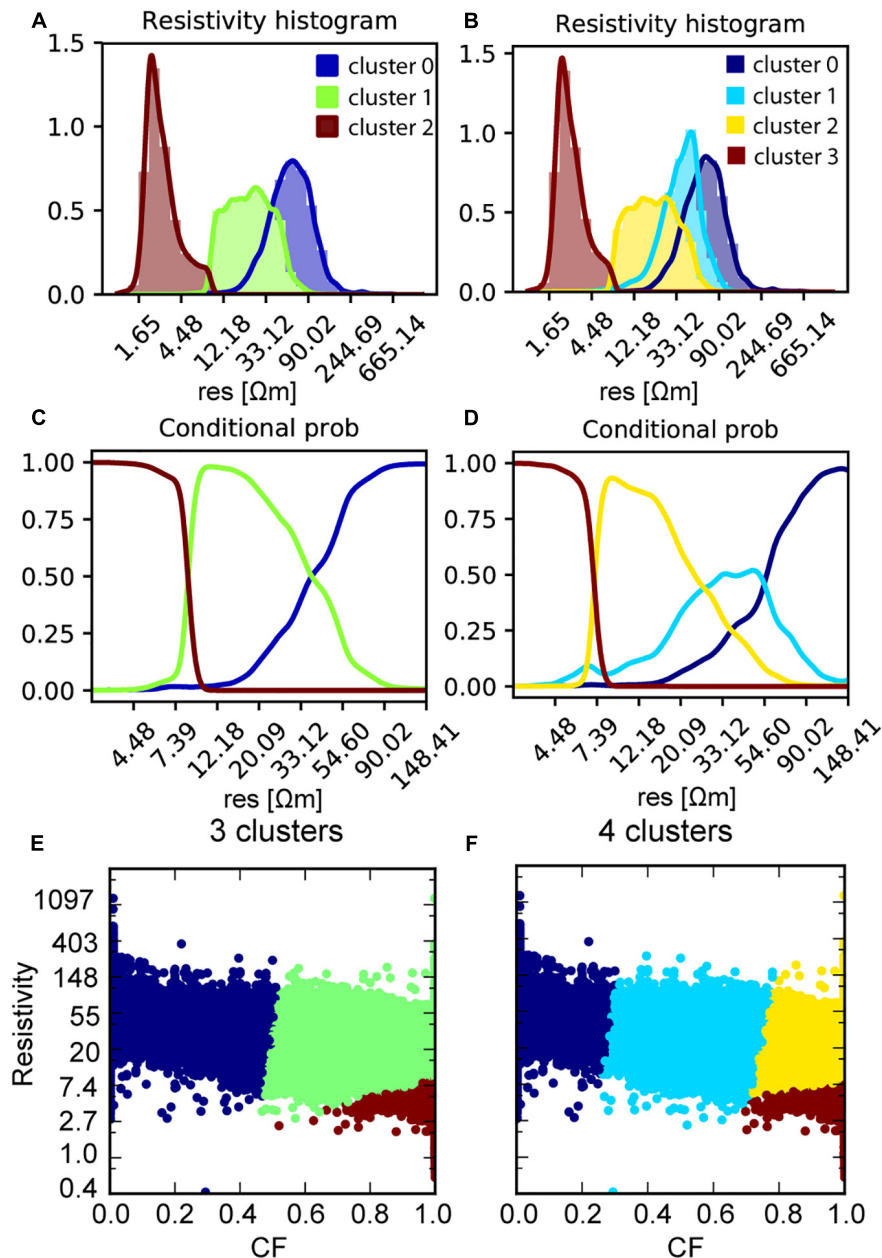
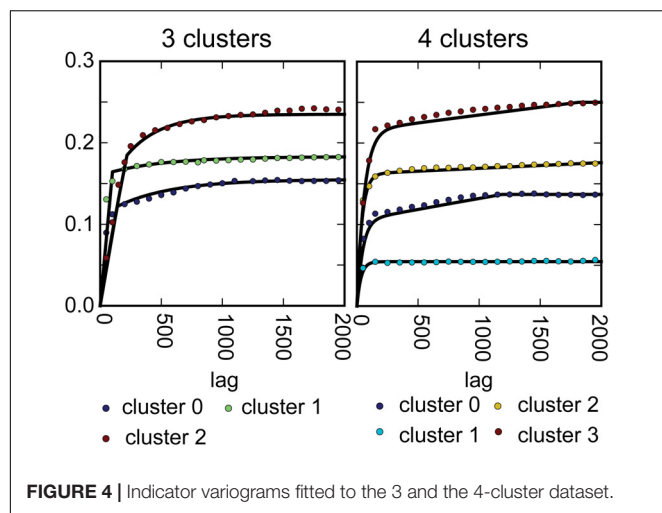


FIGURE 3 | Results of the cluster analysis on CF and resistivities. **(A)** shows the result of the 3-cluster analysis plotted as histograms in the resistivities. **(B)** shows the histograms for the 4-cluster analysis on resistivities. **(C)** shows the conditional probabilities of the 3-cluster analysis. **(D)** shows the conditional probabilities of the 4-cluster analysis. **(E)** show clustering of the cloud spanned by resistivities and CF into 3 clusters. Finally, **(F)** shows the clustering of the cloud spanned by resistivities and CF subdivided by 4 clusters.

on the uncertainties determined using the approach described above. Here transparent parts indicate high uncertainty. For the deterministic model we used the uncertainties derived from the three-cluster model. Similar plots can be seen for the cross section through the Ristrup well field in **Figure 7**. Based on the cross sections- some interesting aspects can be observed. First, many of the dissimilarities between the deterministic and the auto generated models are smeared out. Second, the plots conform with the background information from the area, namely that

the heterogeneity and the uncertainty in the structure appears larger in the Kasted well field compared to the Ristrup well field. Making this comparison, it should of course be taken into account that the Kasted profile is three times as long as the Ristrup profile.

Figure 8 shows a horizontal cross-section through the same depth interval of the deterministic model, the TI, three out of the 100 generated model realizations generated using SNESIM, and three model realizations generated using SISIM.



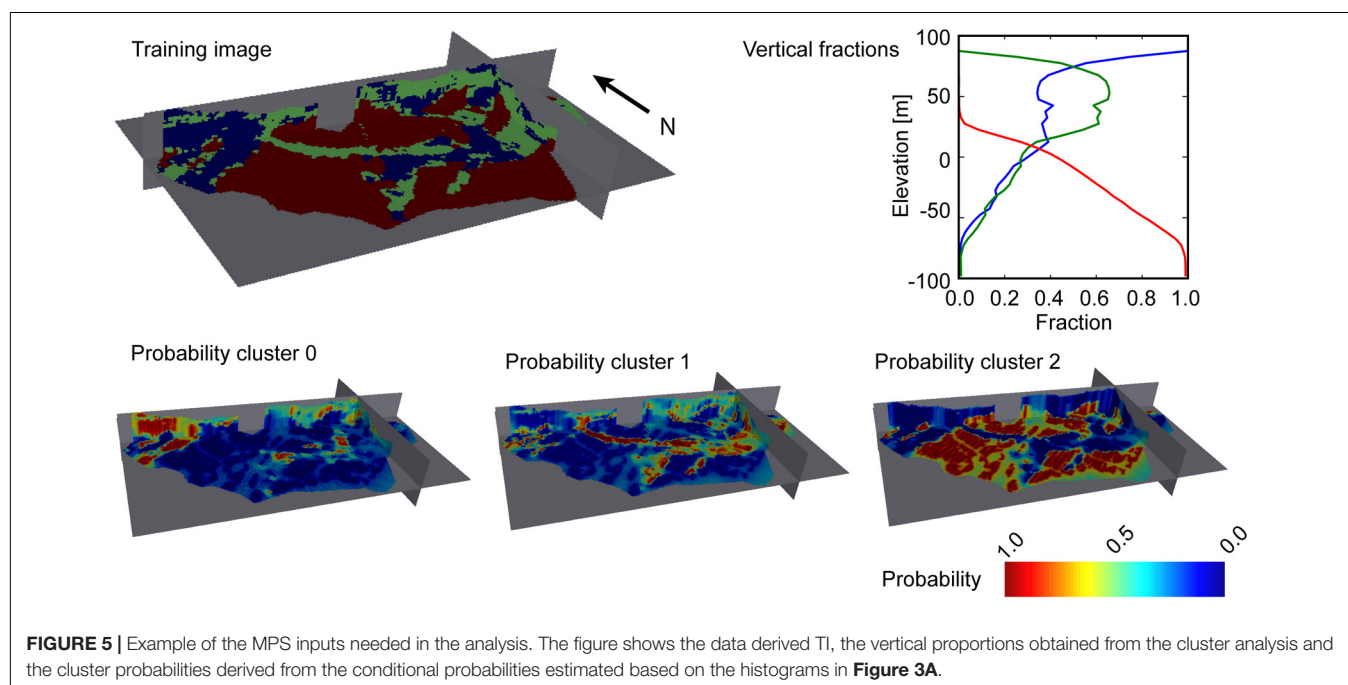
The model realizations generated with SNESIM have much more resemblance with both the manual interpretation and the TI. Also, the realizations generated with SISIM, show patterns with limited correlation to either of the remaining models presented on the figure. It should also be noted, that the three realizations generated with SISIM are quite similar. This is a feature caused by using the geophysical data as hard data in the simulations.

The generated model realizations from SNESIM were all incorporated into the groundwater model, and calibrated to hydraulic heads and stream flows. The result of this calibration with respect to simulated and observed hydraulic heads can be seen from **Figure 9**. By comparing the three plots, the performance of the models is similar. There is a slight scatter about the identity line, but it appears similar for the three setups,

both with respect to distance and outliers. Taking the structural uncertainty into account, most simulations estimated using the deterministic modeling approach fall within the simulated uncertainties of the structural realizations.

Analyzing the estimated parameter values, the log transformed hydraulic conductivity for the three-cluster model is -4.97 m/s for cluster 0 with a standard deviation of 0.14 and -5.21 m/s for cluster 1 with a standard deviation of 0.27. Cluster 2 was fixed in the inversion, due to the very low hydraulic conductivity of this unit. The value for cluster 1 may be a bit high compared to the expected value, but it is not unreasonable, taking the uncertainty into account. It is also in line with the expected result, that the uncertainty of the more clay dominated cluster 1 is larger than the sand dominated cluster 0. For the four-cluster model the parameters are (uncertainty in brackets) -5.00 m/s (0.14) for cluster 0, -5.02 m/s (0.75) for cluster 1, and -5.82 m/s (0.59) for cluster 2. Cluster 3 was fixed for the same reasons as outlined above. These parameter values are also in line with the expected, except maybe for cluster 1, which is a bit high, but its pertaining uncertainty is also high due to its limited presence in the model. This can also be seen from the cross sections on **Figures 6C, 7C**. The data fits obtained in this study are very similar to those obtained by Marker et al. (2017).

All the calibrated models were subsequently used to forecast the location of the catchment zones for the wells at the two different well fields. The result for the Kasted well using the three cluster models can be seen from **Figure 10**. The figure has been subdivided into two plots, one containing the estimated probability using the structural realizations (subplot A), and one containing the catchment probability overlaid with the particle endpoints obtained from the deterministic structure. Based on the model realizations, the majority of the deterministic catchment is contained within the uncertainty of the stochastic



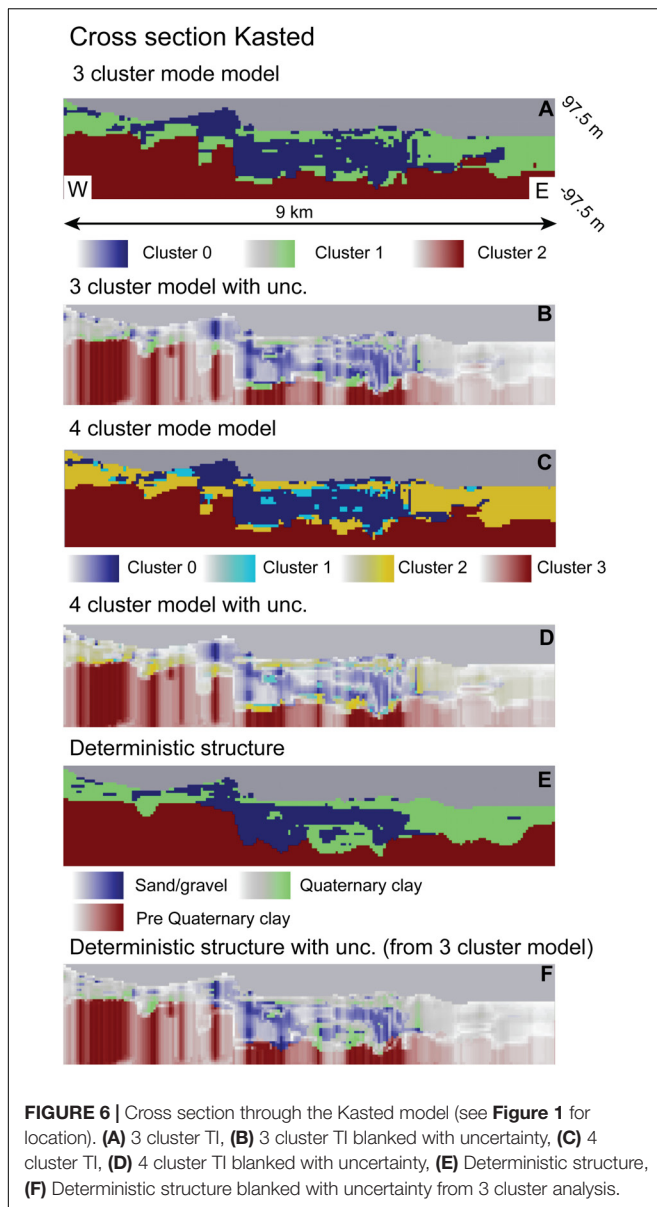


FIGURE 6 | Cross section through the Kasted model (see **Figure 1** for location). (A) 3 cluster TI, (B) 3 cluster TI blanked with uncertainty, (C) 4 cluster TI, (D) 4 cluster TI blanked with uncertainty, (E) Deterministic structure, (F) Deterministic structure blanked with uncertainty from 3 cluster analysis.

simulations. The only discrimination between the two is in the tail stretching toward the west. The tail of the stochastic realizations is offset toward the north. Subplot B also shows the outline of the catchment estimated by Marker et al. (2017). The general trends in the estimated catchments using the SISIM and the SNESIM approach are similar. However, the SNESIM approach is closer to the catchment of the model based on the deterministic structure.

The catchment probability for the Kasted well estimated using the four-cluster model is presented in **Figure 11**. Only small differences can be observed between the catchments estimated with the three and the four cluster models. These are all observed in the low probability region, but almost the entire catchment estimated by the deterministic model structure is contained by the stochastic models.

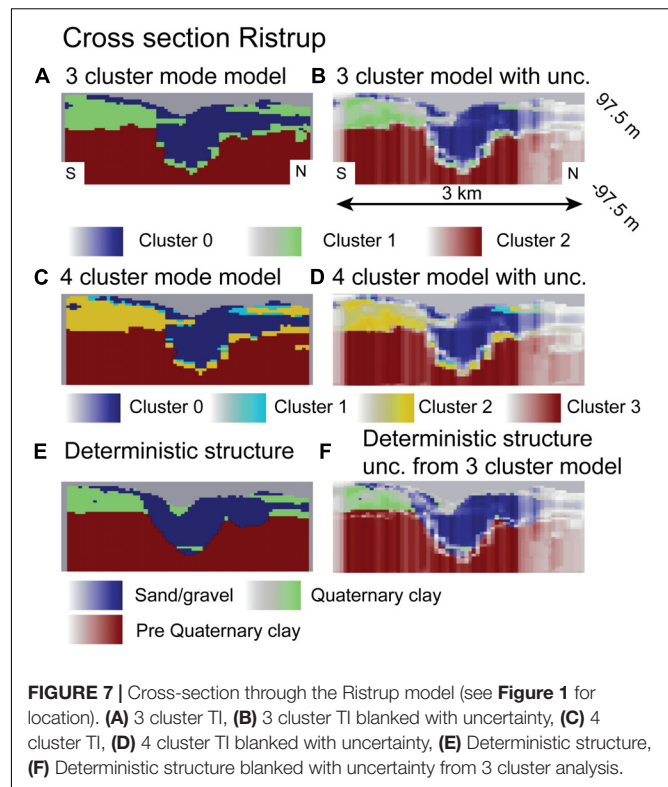


FIGURE 7 | Cross-section through the Ristrup model (see **Figure 1** for location). (A) 3 cluster TI, (B) 3 cluster TI blanked with uncertainty, (C) 4 cluster TI, (D) 4 cluster TI blanked with uncertainty, (E) Deterministic structure, (F) Deterministic structure blanked with uncertainty from 3 cluster analysis.

For the Ristrup well, we have only presented the well catchment using the four-cluster model. These results can be seen from **Figure 12**. Here, the solution is dominated by the effect of the better resolved geology. This is caused by a full coverage with geophysics and the higher resistivity contrasts. Combined, this results in an overall decrease in uncertainty which can be overed by comparing observed in **Figures 7B,D** compared to **Figures 6B,D**. In the plot this is also apparent from the estimated catchment probabilities, especially when these are compared to the catchments estimated using the deterministic structure. These are much better aligned compared to the Kasted Well (**Figure 11**).

DISCUSSION

In this study we have documented a methodology to automatically generate subsurface structures directly from geophysical data and borehole descriptions. The methodology is fast and reproducible, meaning that realistic subsurface structures for hydrological models can be generated in a few days, once the input data have been prepared. For the geophysical data this includes processing and inversion, and for the lithological data this includes a quality control. In the study, we have extended the CF-resistivity clustering method (Foged et al., 2014; Marker et al., 2015, 2017) to include MPS. This has a major benefit with respect to reproducing the structural patterns observed in the TI, which cannot be resolved using the two-point statistics contained in a variogram model. In the present study this is evident by comparing structural realizations generated using SISIM and SNESIM (**Figure 8**). Here, the

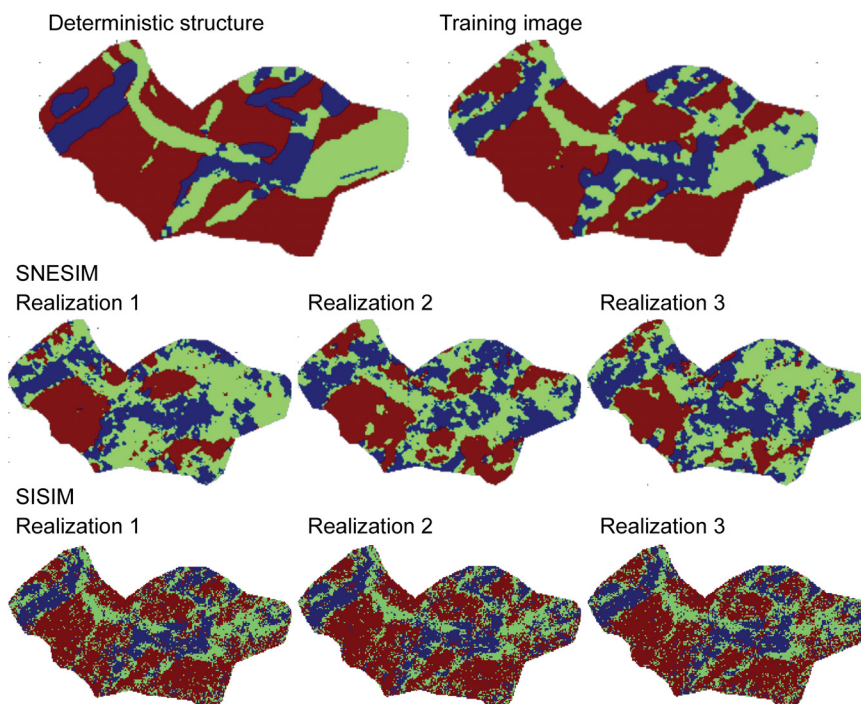


FIGURE 8 | Deterministic structure and TI together with three structural representations generated using SNESIM and SISIM, respectively.

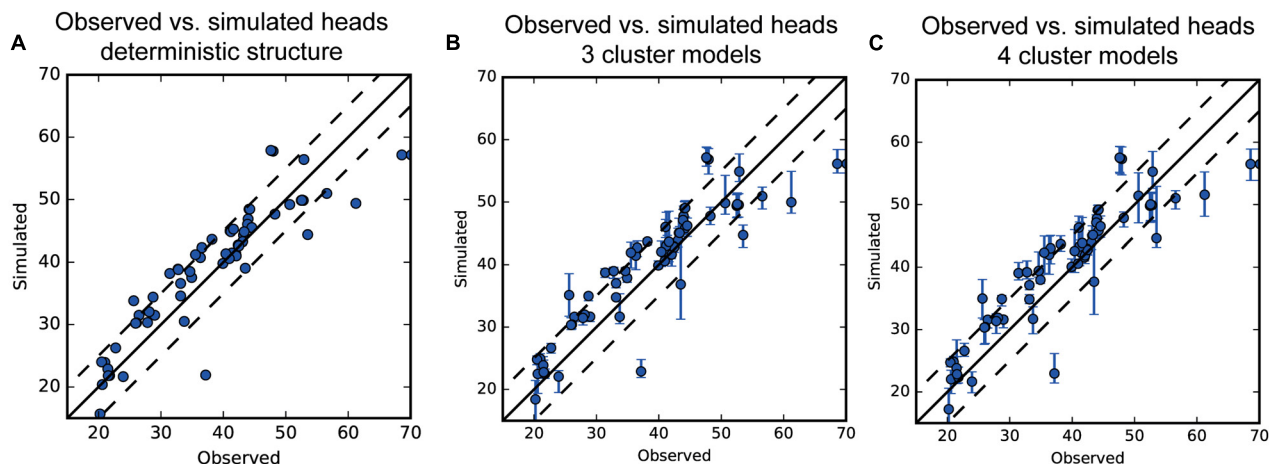
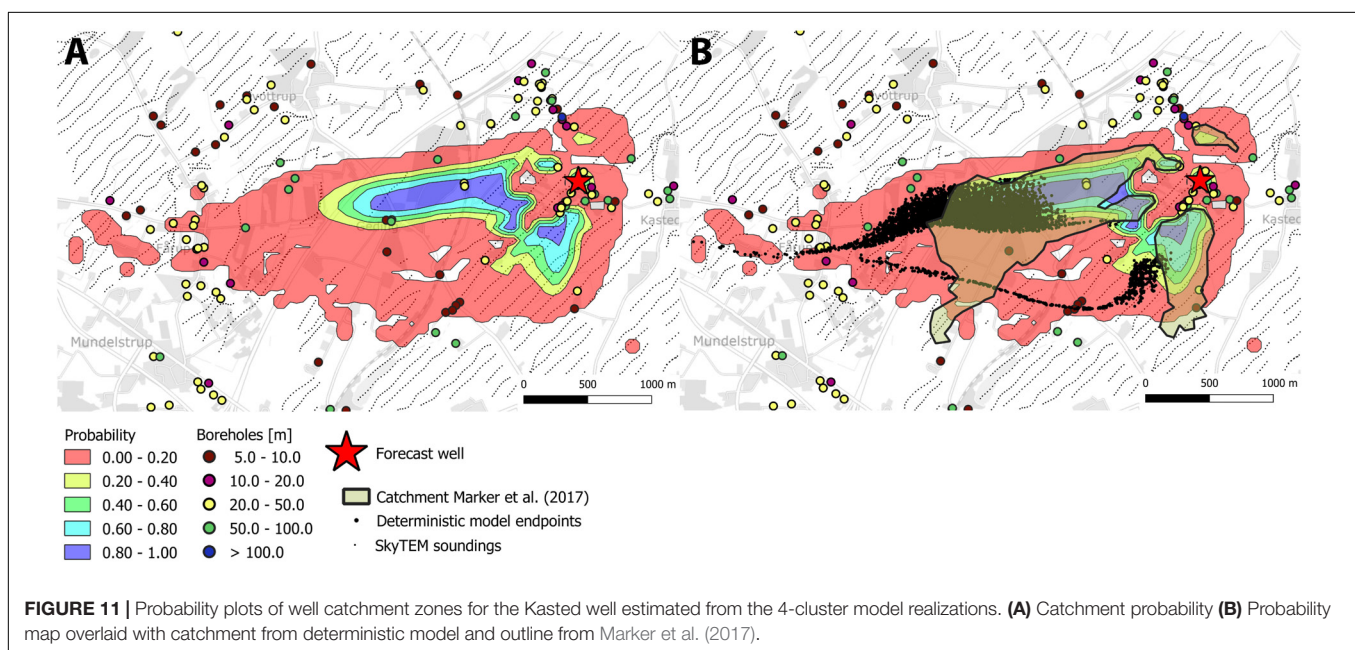
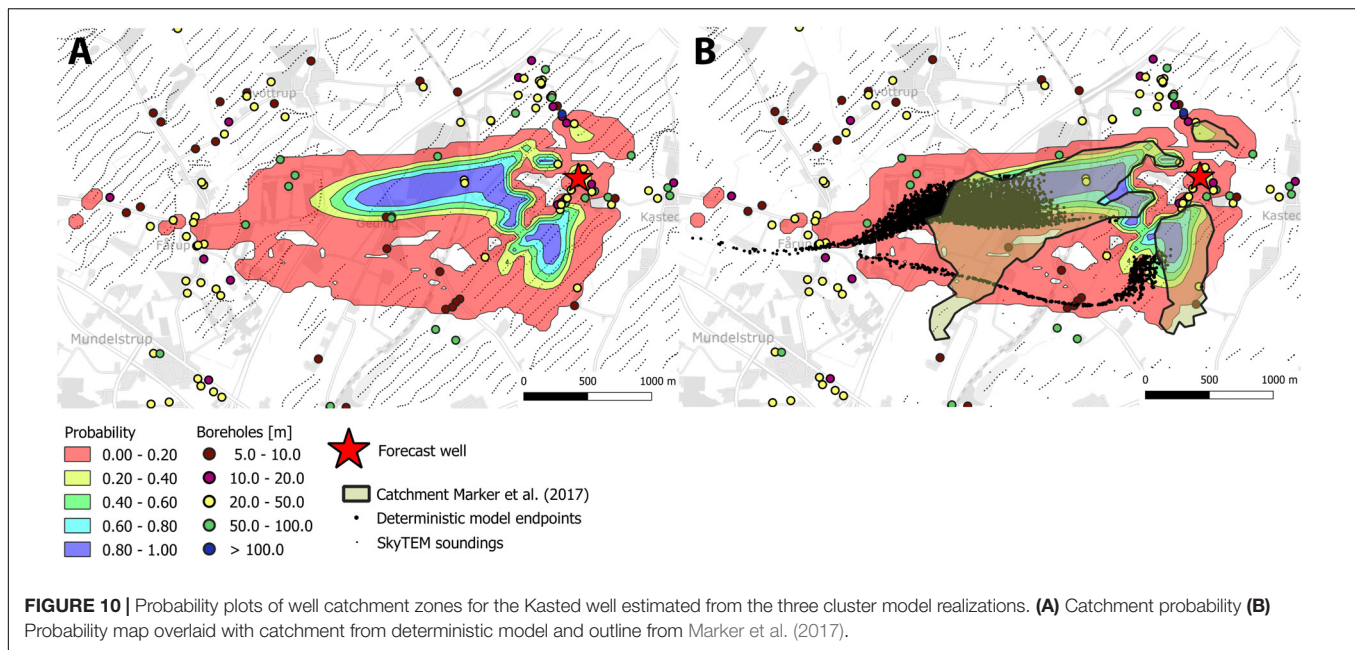


FIGURE 9 | Observed vs. simulated steady state heads. (A) Deterministic model structure, (B) 3 Cluster models, (C) 4 cluster models. Dotted lines show the plus and minus 5 m offsets from the identity line.

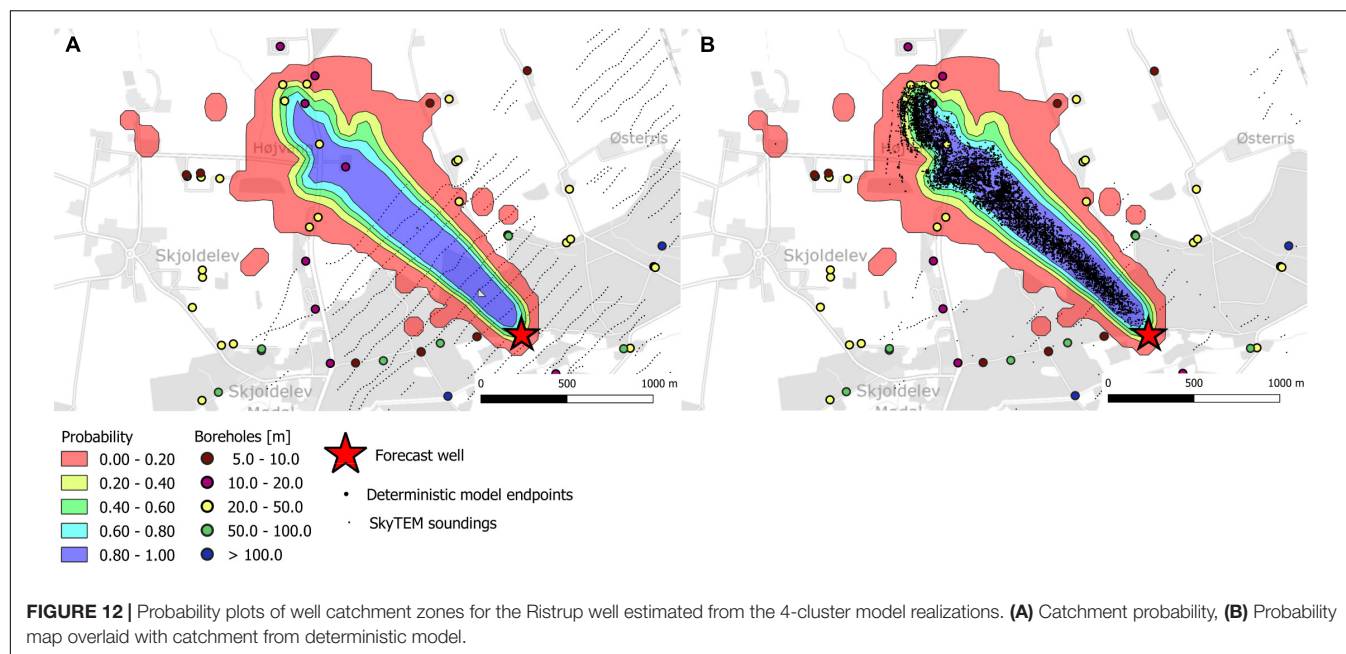
SISIM realization have limited resemblance with both the TI and the manual interpretations of the subsurface structures. Moreover, we also argue, that similarities between the SIS generated models are much larger than the models generated with MPS. This is an artifact of applying the geophysical data as hard constraints, and applying 2-point statistics that cannot preserve to expected connectivity of the subsurface structures. To avoid over conditioning the MPS simulations we have updated the framework such that CF estimates and resistivity data are not considered as hard data in the analysis.

This is an important contribution, since the uncertainty of the pertaining cluster is highly dependent on the estimated resistivity. This can easily be acknowledged from the conditional probabilities in **Figure 3B**. A resistivity value of approximately $40 \Omega\text{m}$ would result in cluster 1, but the probability that this is actually cluster 0 is approximately 30%. In the present study we have utilized the SNESIM algorithm, but the proposed methodology could be utilized with other MPS methods such as e.g., direct sampling Mariethoz et al. (2010) or image quilting Hoffmann et al. (2017).



One of the challenges faced when including geophysical data in automated structural modeling is that the smoothness constraints often imposed as regularization in the geophysical inversion lead to smooth transitions between different subsurface units. Most often, these results are not directly related to changes in lithology, since the smoothing effect caused by the regularization results in a smearing of layer boundaries (Jørgensen et al., 2013). To avoid this unwanted effect, we have adopted the sharp inversion formulation in the present study (Vignoli et al., 2015). This regularization imposes a penalty on gradients in resistivity, thus producing fewer, but sharper transitions between layers in the geophysical model. This

methodology has already proven advantageous in groundwater modeling studies (Christensen et al., 2017a) and structural interpretations (Barfod et al., 2018). Despite having sharp layer boundaries in individual 1D resistivity models, these have to be interpolated if a full 3D representation of the subsurface resistivity is required. Traditionally this is done using kriging, thus resulting in smooth transmissions in-between the sharp resistivity models. Similar to the limitations outlined above, these smooth transitions are not representations of the subsurface lithology, but an effect of the interpolation methodology (Jørgensen et al., 2013; Høyer et al., 2015). Utilizing such resistivity grids in automated structural modeling



will unavoidably lead to unwanted artifacts that cannot be related to structures. Such effects can be seen in Foged et al. (2014), Marker et al. (2015), and Jørgensen et al. (2013). Similar to Marker et al. (2017), we therefore choose to manage the geophysical data interpretations at the sounding sites instead of at the interpolated grids. This was done by transforming resistivity values into clusters directly at the resistivity and borehole data locations. Using this methodology, we have minimized the smoothing effects to the extent that was present in the resistivity models, while at the same time acknowledging the uncertainty of the underlying structure at the sounding sites.

Based on the results we argue, that the use of MPS methods compared to SIS as applied by Marker et al. (2017) has several advantages. First, we have not treated the geophysical data as hard data in the simulation. Second, we can easily incorporate supplementary information into the framework, such as the vertical proportions or other auxiliary variables, and third, we are not limited to the information that can be carried in the variogram model. The latter being particular important in areas where the data density is low. In a more general context, similar methodologies could be applied using SIS (Hadavand and Deutsch, 2017), however, this methodology would still be limited to the information content that can be carried in the variogram model.

Compared to previous studies employing CF resistivity clustering, the proposed methodology is not as limited by ability of the EM methods to resolve the major lithological units. This could be areas where clay units have high resistivities or where salinity reduces the resistivity of aquifers. In such areas this would result in none uniqueness visualized as overlapping probability density functions, thereby increasing the conditional uncertainties. The result would be a relatively increase in the uncertainty of the derived structural realizations.

In studies, where the majority of the structural modeling is performed based on EM data, this could potentially be an important supplement to the derived model or models. To make the modeler aware of such potential issues, it is recommended to perform an analysis of the EM resolution capabilities. This could either be done using a probability density function as the ones shown in this study, by utilizing a resistivity atlas (Barfod et al., 2016), or based on background geological knowledge of the area. The proposed method could also be restricted to a subdomain of the field site, where the assumption is valid, and other modeling strategies could be utilized elsewhere, similar to Høyer et al. (2016) or He et al. (2015).

The suggested methodology is different from most applied methods to perform MPS modeling, where the TI is derived from important expert knowledge (Linde et al., 2015). This study is not meant to neglect such background or process specific information. We do, however, argue, that the strength of the methodology is limited to the time it takes from data to model development, and it does not exclude incorporation of expert knowledge into the TI in cases where this is deemed necessary to obtain a satisfactory modeling result.

In the present study, we limited our analysis to three and four categorical units. This is a simplification of the true complexity in large scale aquifer systems. However, it is still common practice to parameterize subsurface structures in groundwater models with only a few distinct units. In the present study we set this limit based on two criteria, namely the comparison to the deterministic structure and the comparison to the study by Marker et al. (2017). However, the optimal number of clusters is also limited by the input data. Here these consists of clay fractions, which to a large extent is a binary input. 13% of the clay fractions have values below 0.05 and 73% of the clay fractions have values above 0.95. The geophysical data applied in the present case will have

the highest sensitivity to subsurface units with high electrical conductance. In this area this will mainly be Paleogene clays and clay tills. The method will also be able to distinguish between sand and clays, but due to the limited sensitivity to units with low electrical conductivity, it is more difficult to separate e.g., units of sand and gravel, which both have low electrical conductance in the present field site.

Non-stationarity in the TI is a challenge normally faced when performing MPS. In the present case we accounted for non-stationarity by conditioning structural simulations to a probability distribution of the different units. In some areas, this may not be sufficient to address non-stationarity. Under such circumstances the proposed methodology may not be applicable, and other methodologies may have to be applied. Alternatively if the area can be subdivided into more stationary subparts, a TI can be estimated for each subdomain. These subareas can then either be managed individually using the proposed methodology, or the can be simulated using multiple TI. Such capabilities are not available in the implementation of SNESIM applied here, but they are available in the Direct Sampling algorithm (Mariethoz et al., 2010).

Finally, the present study is limited to an analysis of a method to quantify contribution from structural uncertainty to forecast of a well catchment zone. Several other sources of uncertainty exists in groundwater models, including, hydraulic conductivities, heterogeneity in the distinct units, boundary conditions, forcing data etc. We did not include these sources, as we wanted to single out the contribution from the uncertain structures. The proposed methodology could, however, easily be combined with these other contributions to encapsulate a more complete estimate of the forecast uncertainty.

CONCLUSION AND PERSPECTIVES

This study documents a methodology, where the recently developed CF-resistivity clustering method is combined with MPS to estimate the contribution from large scale subsurface structural uncertainty. The derived structures were used to estimate the contribution to groundwater model forecasts uncertainty of two different well catchment zones; one screened in a highly resolved and simpler geological setting and one in a noisy EM environment with a more complex geology.

Compared to the previous studies using the CF-resistivity clustering method, the present combination with MPS methods allows for an improved framework for handling uncertainties, and it addresses the limitations by the two-point variogram models.

REFERENCES

- Almasri, M., and Kaluarachchi, J. (2007). Modeling nitrate contamination of groundwater in agricultural watersheds. *J. Hydrol.* 343, 211–229. doi: 10.1016/j.jhydrol.2007.06.016
- Arnold, J., and Allen, P. (1999). Automated methods for estimating baseflow and ground water recharge from streamflow records. *J. Am. Water Resour. Assoc.* 35, 411–424. doi: 10.1111/j.1752-1688.1999.tb03599.x

Similar to previous uses, the main advantage of the methodology lies in the fact that it is fast, data driven, reproducible, and transparent. We therefore believe that the potential for the method is large with respect to model screening and analysis, either as a direct input to groundwater flow models, or as a support tool for structural modeling tasks.

The methodology is limited by the direct resolution capabilities of the EM methods. Effectively this restricts the resolution to a few distinct units. In the present study, this resulted in modeling of three and four categorical variables. We found limited differences in the performance of the respective groups of models, both with respect to their ability to fit the hydraulic data and their forecast uncertainty. It is therefore not expected that the proposed methodology can provide higher meaningful resolutions than this, except if the model domain is subdivided into separate regimes, which have distinct different geophysical responses or hydrogeological structures. To introduce more units or variability in the models, this must be done in subsequent steps, e.g., by using MPS methods to simulate small scale heterogeneity in the large scale structures (Strebelle, 2002) or using pilot points in the groundwater model calibration (Doherty, 2003). The SkyTEM dataset and the borehole descriptions used in the study can be downloaded from the national Danish databases (Møller et al., 2009a). Borehole and lithological information is stored in the Jupiter database, and the geophysical data is stored in the Gerda database.

AUTHOR CONTRIBUTIONS

TV: main contributor to the manuscript and primary developer of the presented content. EA: expert on airborne geophysical methods, leader in the interpretations of geophysical data, and codeveloper of the methods. AC: developer of the clay fraction inversion technique. AB: introduced TV to MPS methods, and assisted in the development of the structural models. PM: assisted in the development of the groundwater flow model. PB-G: original supervisor in the development of the clustering methods used to derive the structures.

FUNDING

This manuscript was supported by the HyGEM, Integrating geophysics, geology, hydrology for improved groundwater and environmental management, Project No. 11-116763 and rOPEN, Open landscape nitrate retention mapping, Project No. 6150-00006B. Both project are funded by Innovation Fund Denmark.

- Arnold, J. G., Allen, P. M., Muttiah, R., and Bernhardt, G. (1995). Automated base-flow separation and recession analysis techniques. *Ground Water* 33, 1010–1018. doi: 10.1111/j.1745-6584.1995.tb00046.x
- Auken, E., Christiansen, A. V., Fiandaca, G., Schamper, C., Behroozmand, A. A., Binley, A., et al. (2015). An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data. *Exploration Geophys.* 2015, 223–235. doi: 10.1071/eg13097

- Auken, E., Christiansen, A. V., Westergaard, J. H., Kirkegaard, C., Foged, N., and Viezzoli, A. (2009). An integrated processing scheme for high-resolution airborne electromagnetic surveys, the SkyTEM system. *Explor. Geophys.* 40, 184–192. doi: 10.1071/EG08128
- Barfod, A., Vilhelmsen, T., Jørgensen, F., Christiansen, A., Hoyer, A., Straubhaar, J., et al. (2018). Contributions to uncertainty related to hydrostratigraphic modeling using multiple-point statistics. *Hydrol. Earth Syst. Sci.* 22, 5485–5508. doi: 10.5194/hess-22-5485-2018
- Barfod, A. A. S., Møller, I., and Christiansen, A. V. (2016). Compiling a national resistivity atlas of Denmark based on airborne and ground-based transient electromagnetic data. *J. Appl. Geophys.* 134, 199–209. doi: 10.1016/j.jappgeo.2016.09.017
- Bastante, F., Ordonez, C., Taboada, J., and Matias, J. (2008). Comparison of indicator kriging, conditional indicator simulation and multiple-point statistics used to model slate deposits. *Eng. Geol.* 98, 50–59. doi: 10.1016/j.enggeo.2008.01.006
- Boronina, A., Renard, P., Balderer, W., and Christodoulides, A. (2003). Groundwater resources in the Kouris catchment (Cyprus): data analysis and numerical modelling. *J. Hydrol.* 271, 130–149. doi: 10.1016/S0022-1694(02)00322-0
- Caers, J. (2001). Geostatistical reservoir modelling using statistical pattern recognition. *J. Pet. Sci. Eng.* 29, 177–188. doi: 10.1016/S0920-4105(01)00088-2
- Christensen, N. K., Ferre, T. P. A., Fiandaca, G., and Christensen, S. (2017a). Voxel inversion of airborne electromagnetic data for improved groundwater model construction and prediction accuracy. *Hydrol. Earth Syst. Sci.* 21, 1321–1337. doi: 10.5194/hess-21-1321-2017
- Christensen, N. K., Minsley, B. J., and Christensen, S. (2017b). Generation of 3-D hydrostratigraphic zones from dense airborne electromagnetic data to assess groundwater model prediction error. *Water Resour. Res.* 53, 1019–1038. doi: 10.1002/2016wr019141
- Christiansen, A. V., and Auken, E. (2012). A global measure for depth of investigation. *Geophysics* 77, WB171–WB177. doi: 10.1190/geo2011-0393.1
- Christiansen, A. V., Foged, N., and Auken, E. (2014). A concept for calculating accumulated clay thickness from borehole lithological logs and resistivity models for nitrate vulnerability assessment. *J. Appl. Geophys.* 108, 69–77. doi: 10.1016/j.jappgeo.2014.06.010
- Danielsen, J. E., Auken, E., Jørgensen, F., Søndergaard, V. H., and Sørensen, K. I. (2003). The application of the transient electromagnetic method in hydrogeophysical surveys. *J. Appl. Geophys.* 53, 181–198. doi: 10.1016/j.jappgeo.2003.08.004
- dell'Arciprete, D., Bersezio, R., Felletti, F., Giudici, M., Comunian, A., and Renard, P. (2012). Comparison of three geostatistical methods for hydrofacies simulation: a test on alluvial sediments. *Hydrogeol. J.* 20, 299–311. doi: 10.1007/s10040-011-0808-0
- Delottier, H., Pryet, A., and Dupuy, A. (2017). Why should practitioners be concerned about predictive uncertainty of groundwater management models? *Water Resour. Manag.* 31, 61–73. doi: 10.1007/s11269-016-1508-2
- Deutsch, C. V., and Journel, A. G. (1998). *GSLIB: Geostatistical Software Library and User's Guide. Second Edition*. Oxford: Oxford University Press.
- Doherty, J. (2003). Ground water model calibration using pilot points and regularization. *Ground Water* 41, 170–177. doi: 10.1111/j.1745-6584.2003.tb02580.x
- Doherty, J. (2016). *PEST, Model-Independent Parameter Estimation, User Manual Part I: PEST, SENSAN and Global Optimisers*. Brisbane, QLD: Watermark Numerical Computing.
- Enzenhoefer, R., Bunk, T., and Nowak, W. (2014). Nine steps to risk-informed wellhead protection and management: a case study. *Groundwater* 52, 161–174. doi: 10.1111/gwat.12161
- Foged, N., Marker, P. A., Christiansen, A. V., Bauer-Gottwein, P., Jørgensen, F., Høyer, A.-S., et al. (2014). Large scale 3D-modeling by integration of resistivity models and borehole data through inversion. *Hydrol. Earth Syst. Sci.* 18, 4349–4362. doi: 10.5194/hess-18-4349-2014
- Freeze, R. (2004). The role of stochastic hydrogeological modeling in real-world engineering applications. *Stoch. Environ. Res. Risk Assess.* 18, 286–289.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation. Applied Geostatistics Series*. New York, NY: Oxford University Press, 483.
- Hadavand, M., and Deutsch, C. (2017). Facies proportion uncertainty in presence of a trend. *J. Pet. Sci. Eng.* 153, 59–69. doi: 10.1016/j.petrol.2017.03.036
- Harbaugh, A., Banta, E. R., Hill, M. C., and McDonald, M. G. (2000). *MODFLOW-2000, the U.S. Geological Survey Modular Ground-Water Model – User Guide to Modularization Concepts and the Ground-Water Flow Process*. U.S. Geological Survey Open-File Report 00–92. Reston: U.S. Geological Survey.
- Harbaugh, A. W. (2005). *MODFLOW-2005 - The U.S. Geological Survey Modular Ground-Water Model - the Ground-Water Flow Process*. Reston: U.S. Geological Survey.
- He, X., Højberg, A. L., Jørgensen, F., and Refsgaard, J. C. (2015). Assessing hydrological model predictive uncertainty using stochastically generated geological models. *Hydrol. Process.* 29, 4293–4311. doi: 10.1002/hyp.10488
- He, X. L., Sonnenborg, T. O., Jørgensen, F., and Jensen, K. H. (2014). The effect of training image and secondary data integration with multiple-point geostatistics in groundwater modelling. *Hydrol. Earth Syst. Sci.* 18, 2943–2954. doi: 10.5194/hess-18-2943-2014
- Henriksen, H., Trolborg, L., Nyegaard, P., Sonnenborg, T., Refsgaard, J., and Madsen, B. (2003). Methodology for construction, calibration and validation of a national hydrological model for Denmark. *J. Hydrol.* 280, 52–71. doi: 10.1016/S0022-1694(03)00186-0
- Hermans, T., Nguyen, F., and Caers, J. (2015). Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: workflow and case study. *Water Resour. Res.* 51, 5332–5352. doi: 10.1002/2014wr016460
- Hoffmann, J., Scheidt, C., Barfod, A., and Caers, J. (2017). Stochastic simulation by image quilting of process-based geological models. *Comput. Geosci.* 106, 18–32. doi: 10.1016/j.cageo.2017.05.012
- Hoyer, A., Vignoli, G., Hansen, T., Vu, L., Keefer, D., and Jørgensen, F. (2017). Multiple-point statistical simulation for hydrogeological models: 3-D training image development and conditioning strategies. *Hydrol. Earth Syst. Sci.* 21, 6069–6089. doi: 10.5194/hess-21-6069-2017
- Høyer, A.-S., Jørgensen, F., Sandersen, P. B. E., Viezzoli, A., and Møller, I. (2015). 3D geological modelling of a complex buried-valley network delineated from Borehole and AEM Data. *J. Appl. Geophys.* 2015, 94–102. doi: 10.1016/j.jappgeo.2015.09.004
- Høyer, A.-S., Vignoli, G., Hansen, T. M., Vu, L. T., Keefer, D. A., and Jørgensen, F. (2016). Multiple-point statistical simulation for hydrogeological models: 3D training image development and conditioning strategies. *Hydrol. Earth Syst. Sci. Discuss* 21, 1–29. doi: 10.5194/hess-2016-567
- Hu, L. Y., and Chugunova, T. (2008). Multiple-point geostatistics for modeling subsurface heterogeneity: a comprehensive review. *Water Resour. Res.* 44, W11413. doi: 10.1029/2008WR006993
- Huysmans, M., Peeters, L., Moermans, G., and Dassargues, A. (2008). Relating small-scale sedimentary structures and permeability in a cross-bedded aquifer. *J. Hydrol.* 361, 41–51. doi: 10.1016/j.jhydrol.2008.07.047
- Jeppesen, J. (2001). *Vandbalancen for Rodzonen på Als*. Aarhus: Aarhus University.
- Jørgensen, F., Møller, R. R., Nebel, L., Jensen, N., Christiansen, A. V., and Sandersen, P. (2013). A method for cognitive 3D geological voxel modelling of AEM data. *Bull. Eng. Geol. Environ.* 72, 421–432. doi: 10.1007/s10064-013-0487-2
- Jørgensen, F., and Sandersen, P. B. E. (2006). Buried and open tunnel valleys in Denmark-erosion beneath multiple ice sheets. *Q. Sci. Rev.* 25, 1339–1363. doi: 10.1016/j.quascirev.2005.11.006
- Journel, A., and Zhang, T. F. (2006). The necessity of a multiple-point prior model. *Math. Geol.* 38, 591–610. doi: 10.1007/s11004-006-9031-2
- Journel, A. G. (2002). Combining knowledge from diverse sources: an alternative to traditional data independence hypotheses. *Math. Geol.* 34, 573–596.
- Journel, A. G. (2005). “Beyond covariance: the advent of multiple-point geostatistics,” in *Proceedings of the Seventh International Geostatistics Congress*, eds O. Leuangthong and C. V. Deutsch (Banff, AB: Springer), 225–233. doi: 10.1007/978-1-4020-3610-1_23
- Knight, R., Smith, R., Asch, T., Abraham, J., Cannia, J., Viezzoli, A., et al. (2018). Mapping aquifer systems with airborne electromagnetics in the central valley of California. *Groundwater* 56, 893–908. doi: 10.1111/gwat.12656
- Korus, J., Joeckel, R., Divine, D., and Abraham, J. (2017). Three-dimensional architecture and hydrostratigraphy of cross-cutting buried valleys using airborne electromagnetics, glaciated Central Lowlands, Nebraska, USA. *Sedimentology* 64, 553–581. doi: 10.1111/sed.12314
- Kronborg, C., Bender, H., Bjerre, R., Friberg, R., Jacobsen, H., Kristiansen, L., et al. (1990). Glacial stratigraphy of east and central Jutland. *Boreas* 19, 273–287. doi: 10.1111/j.1502-3885.1990.tb00451.x

- Linde, N., Renard, P., Mukerji, T., and Caers, J. K. (2015). Geological realism in hydrogeological and geophysical inverse modeling: a review. *Adv. Water Resour.* 86, 86–101. doi: 10.1016/j.advwatres.2015.09.019
- Liu, Y. H. (2006). Using the Snesim program for multiple-point statistical simulation. *Comput. Geosci.* 32, 1544–1563. doi: 10.1016/j.cageo.2006.02.008
- Maharaja, A. (2008). TiGenerator: object-based training image generator. *Comput. Geosci.* 34, 1753–1761. doi: 10.1016/j.cageo.2007.08.012
- Manghi, F., Williams, D., Safely, J., and Hamdi, M. R. (2012). Groundwater flow modeling of the arlington basin to evaluate management strategies for expansion of the Arlington desalter water production. *Water Resour. Manag.* 26, 21–41. doi: 10.1007/s11269-011-9899-6
- Mariethoz, G., and Caers, J. (2015). *Multiple-Point Geostatistics : Stochastic Modeling with Training Images*. Chichester: John Wiley & Sons Inc, 364.
- Mariethoz, G., Renard, P., and Straubhaar, J. (2010). The direct sampling method to perform multiple-point geostatistical simulations. *Water Resour. Res.* 46, W11536. doi: 10.1029/2008WR007621
- Marker, P. A., Foged, N., He, X., Christiansen, A. V., Refsgaard, A., Auken, E., et al. (2015). Performance evaluation of groundwater model hydrostratigraphy from airborne electromagnetic data and lithological borehole logs. *Hydrol. Earth Syst. Sci.* 19, 3875–3890. doi: 10.5194/hess-19-3875-2015
- Marker, P. M., Vilhelmsen, T. N., Foged, N., Wernberg, T., Auken, E., and Bauer-Gottwein, P. (2017). Probabilistic predictions using a groundwater model informed with airborne EM data. *Adv. Water Resour.* 2017:13.
- Møller, I., Søndergaard, V. H., and Jørgensen, F. (2009a). Geophysical methods and data administration in Danish groundwater mapping. *Geological Survey of Denmark and Greenland Bulletin* 17, 41–44.
- Møller, I., Verner, H., Søndergaard, V. H., Flemming, J., Auken, E., and Christiansen, A. V. (2009b). Integrated management and utilization of hydrogeophysical data on a national scale. *Near Surface Geophys.* 7, 647–659. doi: 10.3997/1873-0604.2009031
- Muffels, C., Wang, X., Tonkin, M., and Neville, C. (2014). *User's Guide for mod-PATH3DU A groundwater path and travel-time simulator*. Bethesda, MD: S.S. Papadopoulos & Associates, Inc.
- Mylopoulos, N., Mylopoulos, Y., Veranis, N., and Tolikas, D. (2007). Groundwater modeling and management in a complex lake-aquifer system. *Water Resour. Manag.* 21, 469–494. doi: 10.1007/s11269-006-9025-3
- Panday, S., Langevine, C. D., Niswonger, R. G., Ibaraki, M., and Hughes, J. D. (2015). *MODFLOW-USG version 1.3.00: An Unstructured Grid Version of MODFLOW for Simulating Groundwater Flow and Tightly Coupled Processes Using a Control Volume Finite-Difference Formulation: U.S. Geological Survey Software Release*. Reston: U.S. Geological Survey.
- Pappenberger, F., and Beven, K. (2006). Ignorance is bliss: or seven reasons not to use uncertainty analysis. *Water Resour. Res.* 42, W05302. doi: 10.1029/2005WR004820
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pryet, A., d'Ozouville, N., Violette, S., Deffontaines, B., and Auken, E. (2012). Hydrogeological settings of a volcanic island (San Cristóbal, Galapagos) from joint interpretation of airborne electromagnetics and geomorphological observations. *Hydrol. Earth Syst. Sci. Discuss* 16, 4571–4579. doi: 10.5194/hess-16-4571-2012
- Remy, N., Boucher, A., and Wu, J. (2009). *Applied Geostatistics with SGeMS : A User's Guide*. Cambridge: Cambridge University Press, 264.
- Renard, P. (2007). Stochastic hydrogeology: what professionals really need? *Ground Water* 45, 531–541. doi: 10.1111/j.1745-6584.2007.00340.x
- Sanchez-Vila, X., and Fernandez-Garcia, D. (2016). Debates-Stochastic subsurface hydrology from theory to practice: why stochastic modeling has not yet permeated into practitioners? *Water Res. Res.* 52, 9246–9258. doi: 10.1002/2016wr019302
- Sandersen, P., and Jørgensen, F. (2003). Buried Quaternary valleys in western Denmark—occurrence and inferred implications for groundwater resources and vulnerability. *J. Appl. Geophys.* 53, 229–248. doi: 10.1016/j.jappgeo.2003.08.006
- Saravanan, R., Balamurugan, R., Karthikeyan, M. S., Rajkumar, R., Anuthaman, N. G., and Gopalakrishnan, A. N. (2011). Groundwater modeling and demarcation of groundwater protection zones for Tirupur Basin - A case study. *J. Hydro Environ. Res.* 5, 197–212. doi: 10.1016/j.jher.2011.02.003
- Sedki, A., and Ouazar, D. (2011). Simulation-optimization modeling for sustainable groundwater development: a Moroccan coastal aquifer case study. *Water Resour. Manag.* 25, 2855–2875. doi: 10.1007/s11269-011-9843-9
- Silva, D., and Deutsch, C. (2014). A multiple training image approach for spatial modeling of geologic domains. *Math. Geosci.* 46, 815–840. doi: 10.1007/s11004-014-9543-0
- Sonnenborg, T., Christensen, B., Nyegaard, P., Henriksen, H., and Refsgaard, J. (2003). Transient modeling of regional groundwater flow using parameter estimates from steady-state automatic calibration. *J. Hydrol.* 273, 188–204. doi: 10.1016/S0022-1694(02)00389-X
- Sørensen, K. I., and Auken, E. (2004). SkyTEM - A new high-resolution helicopter transient electromagnetic system. *Exploration Geophys.* 35, 191–199.
- Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol.* 34, 1–21.
- Tartakovsky, D., Nowak, W., and Bolster, D. (2012). Introduction to the special issue on uncertainty quantification and risk assessment. *Adv. Water Resour.* 36, 1–2. doi: 10.1016/j.advwatres.2011.12.010
- Viezzoli, A., Christiansen, A. V., Auken, E., and Sørensen, K. I. (2008). Quasi-3D modeling of airborne TEM data by Spatially constrained inversion. *Geophysics* 73, F105–F113.
- Vignoli, G., Fiandaca, G., Christiansen, A. V., Kirkegaard, C., and Auken, E. (2015). Sharp spatially constrained inversion with applications to transient electromagnetic data. *Geophys. Prospecting* 63, 243–255. doi: 10.1111/1365-2478.12185
- Vilhelmsen, T. N., Marker, P. A., Foged, N., Wernberg, T., Auken, E., Christiansen, A. V., et al. (2018). A regional scale hydrostratigraphy generated from geophysical data of varying age, type and quality. *Water Resour. Manag.* 33, 539–553. doi: 10.1007/s11269-018-2115-1
- Wu, J. (2012). *Advances in K-means Clustering: A Data Mining Thinking*. Berlin: Springer.
- Zinn, B., and Harvey, C. (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 Resour. Res.* 39:1051. doi: 10.1029/2001WR001146

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Vilhelmsen, Auken, Christiansen, Barfod, Marker and Bauer-Gottwein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Hydrogeophysical Parameter Estimation Using Iterative Ensemble Smoothing and Approximate Forward Solvers

Corinna Köpke^{1*}, Ahmed H. Elsheikh² and James Irving¹

¹ Institute of Earth Sciences, University of Lausanne, Lausanne, Switzerland, ² School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, United Kingdom

OPEN ACCESS

Edited by:

Philippe Renard,
University of Neuchâtel, Switzerland

Reviewed by:

Jaime Gomez-Hernandez,
Universitat Politècnica de València,
Spain

Thomas Mejer Hansen,
Aarhus University, Denmark

*Correspondence:

Corinna Köpke
corinna.koepke@dlr.de

† Present Address:

Corinna Köpke,
Deutsches Zentrum für Luft- und
Raumfahrt, Institut für den Schutz
Maritimer Infrastrukturen,
Bremerhaven, Germany

Specialty section:

This article was submitted to
Freshwater Science,
a section of the journal
Frontiers in Environmental Science

Received: 14 November 2018

Accepted: 27 February 2019

Published: 20 March 2019

Citation:

Köpke C, Elsheikh AH and Irving J
(2019) Hydrogeophysical Parameter
Estimation Using Iterative Ensemble
Smoothing and Approximate Forward
Solvers. *Front. Environ. Sci.* 7:34.
doi: 10.3389/fenvs.2019.00034

In iterative ensemble smoother approaches and ensemble methods in general, the ensemble size governs the accuracy of the parameter estimates obtained. However, employing large ensembles may be computationally infeasible in applications with expensive forward solvers. Here, we reduce the computational cost of using large ensembles in iterative ensemble smoothing through the use of a proxy solver. To correct the proxy response for the corresponding model error, the latter of which can bias posterior parameter estimates if left untreated, we propose a local basis approach. With this approach, the discrepancy between the detailed and proxy solvers is learned for a subset of the ensemble and collected in a dictionary that grows with each iteration. For each ensemble member, the K -nearest neighbors in the dictionary are employed to build an orthonormal basis which is used to identify the model-error component of the residual by projection. The proposed methodology reduces the effects of overfitting the data with the proxy solver, but may lead to underfitting of the data in the absence of a sufficient number of dictionary entries, meaning that the number of ensemble members relative to the number of detailed-solver runs cannot be inflated arbitrarily. We present our approach in the context of the ensemble smoother with multiple data assimilations (ES-MDA) algorithm, and show its successful application to a high-dimensional synthetic example that involves crosshole ground-penetrating radar (GPR) travel-time tomography.

Keywords: ensemble methods, ES-MDA, proxy model, model error, inversion, uncertainty quantification

1. INTRODUCTION

Inverse problems commonly involve computationally expensive forward solvers and large numbers of unknown parameters that are spatially distributed. For risk assessment and effective environmental decision making, parameter uncertainties are required. These can be obtained through, for example, Bayesian stochastic inversion whereby the corresponding posterior distributions are typically sampled using Markov-chain-Monte-Carlo (MCMC) methods. The Bayesian-MCMC framework offers the advantages of providing a natural quantification of parameter uncertainties, as well as the flexibility to incorporate probabilistic information about priors and measurement errors into the inverse problem (Robert and Casella, 2004). However, depending on the forward solver and the dimensionality of the model-parameter space involved, it can be extremely computationally expensive. In many real-world applications, for example,

millions of forward model executions may be required to obtain meaningful posterior statistics with Bayesian-MCMC methods (Ruggeri et al., 2015). Although several recent modifications to the standard Metropolis-Hastings algorithm have significantly improved the computational efficiency of MCMC (e.g., Haario et al., 2001; Hansen et al., 2012; Cotter et al., 2013; Chen et al., 2016; Vrugt, 2016; Beskos et al., 2017), these modifications are often still not enough to make such methods practically feasible for many inverse problems.

One way to significantly reduce the computational cost of stochastic parameter estimation is to employ ensemble-based methods. With such methods, an initial ensemble of model parameter sets, drawn from the Bayesian prior distribution, is updated into posterior samples taking into account the available data. The most popular ensemble-based method is the ensemble Kalman filter (EnKF) (Evensen, 1994, 2007), which was developed as a robust sequential data-assimilation technique. A modification of the EnKF for solving parameter-estimation problems is the ensemble smoother (ES), whereby all available data are assimilated in one global update step rather than sequentially. The underlying equations for both the EnKF and ES may be derived from Bayesian statistics (e.g., van Leeuwen, 2001; Evensen, 2007). To deal with non-linear problems, iterative ensemble techniques have been proposed (e.g., Reynolds et al., 2006; Emerick and Reynolds, 2012b; Elsheimh et al., 2013; Stordal and Elsheimh, 2015). The ensemble smoother with multiple data assimilation (ES-MDA) is one of such techniques, in which the single update step of ES is replaced with a number of smaller updates (Emerick and Reynolds, 2012a). The large advantage of ES methods over MCMC for stochastic parameter estimation is that the executions of the forward solver can be parallelized in a straightforward manner.

Despite the computational advantages of ensemble methods over Bayesian-MCMC approaches, it is well known that large ensembles are required for the most accurate parameter estimates and predictions (e.g., Buizza and Palmer, 1998; Chen and Zhang, 2006; Evensen, 2007). As a result, we still have with ensemble methods the possibility that, for high-dimensional inverse problems involving expensive forward solvers, accurately sampling from the posterior distribution will remain computationally prohibitive. In such cases, the only solution is to employ an approximate forward solver or proxy. Generating such a proxy can be achieved by simplifying the physics of the problem (e.g., Scholer et al., 2012; Josset et al., 2015a,b), by coarsening the forward model discretization (e.g., Arridge et al., 2006; Calvetti et al., 2014), or by constructing a surrogate model based on, for example, polynomial chaos expansion, Gaussian processes, or neural network techniques (e.g., Khu and Werner, 2003; Rasmussen and Williams, 2006; Marzouk and Xiu, 2009; Goh et al., 2013). However, using a proxy forward solver in the inversion introduces model error, which has the potential to strongly bias posterior statistics (Laloy et al., 2013) and can lead to highly overconfident estimates of the wrong parameters (i.e., posterior collapse) if not accounted for.

To address the issue of model error arising from the use of proxy models in stochastic inversion, researchers have typically focused on two general approaches, both of which rely

upon pairs of detailed and proxy solver runs corresponding to different sets of model parameters. In the first approach, these pairs are used to construct a global error model, whose statistics are incorporated into the estimation procedure through, for example, the Bayesian likelihood function (e.g., Kaipio and Somersalo, 2007; Lehtikoinen et al., 2010; Schoups and Vrugt, 2010; Evin et al., 2014; Hansen et al., 2014; Smith et al., 2015; Piccolo and Cullen, 2016; Oliver and Alfonzo, 2018). Although this can be highly effective in some cases, we have found that the model errors for many inverse problems exhibit complex behavior that cannot be described in the same way over the entire parameter space. With the second approach, the aim is to construct a local error model, which is generally accomplished through some kind of interpolation between known model-error realizations (e.g., Kennedy and O'Hagan, 2001; Xu et al., 2014; Josset et al., 2015a; Cui et al., 2018). Although doing this effectively addresses the non-global nature of the model errors, it is implicitly assumed that the model-response surface is smooth enough for interpolation to be effective, and problems may arise in regions of the model parameter space that are not well-sampled by the model-error realizations.

Recently, Köpke et al. (2018) presented a new approach to account for model error arising from the use of proxy forward solvers in Bayesian-MCMC inversion, whereby information about the error is gathered during the inversion procedure through occasional runs of the proxy and detailed solvers together, the results of which are stored in a dictionary. In contrast to the existing methods mentioned above, the approach of Köpke et al. (2018) focuses on identifying by projection the model-error component of the residual through the construction of a local orthogonal model-error basis, rather than on the development of a global or local error model. In this paper, we adapt this methodology for use with ES parameter-estimation methods. In particular, we incorporate the related ideas into the ES-MDA algorithm, where for each ensemble member, the local basis is created using the K -nearest-neighbor (KNN) entries in the model-error dictionary. Doing this enables us to accurately solve the parameter-estimation problem using large ensembles, while at same time reduce computational costs through the use of a proxy solver.

The paper is organized as follows: In section 2 we begin with a short review of ensemble methods followed by the presentation of our approach to account for model error. In section 3, we then show the results of applying this methodology to the geophysical inverse problem of estimating spatially distributed radar-wave slowness from synthetic crosshole ground-penetrating radar (GPR) travel-time data. In this regard, results are compared with inversions based on the standard ES-MDA procedure for reference. Based on these findings, we discuss in section 4 how our results compare with standard MCMC sampling, the choice of parameters in our algorithm needed to provide an optimal balance between computational efficiency and accuracy, as well as how the inversion results progress as a function of ES-MDA iteration. Finally, in section 5, we conclude with some general comments on the methodology and provide directions for future research in this domain.

2. METHODOLOGY

2.1. Ensemble Methods

In a generic formulation of the ensemble Kalman filter (EnKF), the model state vector \mathbf{y}^n at data assimilation time step n is updated after the state forecast step. The update from forecast f to analysis a is carried out using the following equation (Emerick and Reynolds, 2012a):

$$\mathbf{y}_j^{n,a} = \mathbf{y}_j^{n,f} + \mathbf{K}^n(\mathbf{d}_{pert}^n - H(\mathbf{y}_j^{n,f})), \quad (1)$$

with Kalman matrix

$$\mathbf{K}^n = \mathbf{C}_{YD}^{n,f}(\mathbf{C}_{DD}^{n,f} + \mathbf{C}_D^n)^{-1}. \quad (2)$$

Here, $j = 1, 2, \dots, n_e$, where n_e is the number of ensemble members; $\mathbf{C}_{YD}^{n,f}$ is the calculated cross-covariance matrix between the forecast state vector $\mathbf{y}_j^{n,f}$ and the predicted data $\mathbf{d}_j^n = H(\mathbf{y}_j^{n,f})$ obtained through the observation operator $H(\cdot)$; $\mathbf{C}_{DD}^{n,f}$ is the calculated auto-covariance matrix of the predicted data; \mathbf{C}_D^n is the covariance matrix of the observed-data measurement errors; and \mathbf{d}_{pert}^n is the vector of perturbed observations. The latter is obtained using $\mathbf{d}_{pert}^n \sim \mathcal{N}(\mathbf{d}_{obs}^n, \mathbf{C}_D^n)$, where \mathbf{d}_{obs}^n denotes the observed data.

The ensemble smoother (ES) is a variation of the EnKF update formula presented in Equations (1) and (2) that is specifically formulated for parameter estimation problems. The general forward problem

$$\mathbf{d}_{obs} = F(\mathbf{m}_{true}) + \epsilon_d \quad (3)$$

links a set of observed data \mathbf{d}_{obs} to a set of “true” model parameters \mathbf{m}_{true} through the forward operator $F(\cdot)$ with measurement errors $\epsilon_d \sim \mathcal{N}(0, \mathbf{C}_D)$. The corresponding ES update equation is given by (Emerick and Reynolds, 2012a)

$$\mathbf{m}_j^a = \mathbf{m}_j^f + \mathbf{K}(\mathbf{d}_{pert} - F(\mathbf{m}_j^f)), \quad (4)$$

with

$$\mathbf{K} = \mathbf{C}_{MD}^f(\mathbf{C}_{DD}^f + \mathbf{C}_D)^{-1}. \quad (5)$$

Here, \mathbf{m}_j^f and \mathbf{m}_j^a denote the forecast and analyzed model-parameter vectors, respectively, which correspond to an update from prior to posterior; \mathbf{C}_{MD}^f is the cross-covariance matrix between \mathbf{m}_j^f and the predicted data $\mathbf{d}_j = F(\mathbf{m}_j^f)$; \mathbf{C}_{DD}^f is the auto-covariance matrix of the predicted data; and $\mathbf{d}_{pert} \sim \mathcal{N}(\mathbf{d}_{obs}, \mathbf{C}_D)$ is again a vector of perturbed observations. The idea with equations (4) and (5) is that, after defining an initial parameter ensemble by drawing from the Bayesian prior distribution, the ensemble members are updated to represent samples from the posterior distribution in a single analysis step that incorporates all of the available data.

2.2. The ES-MDA Algorithm

ES offers an efficient tool to solve parameter-estimation problems under the assumptions that the prior parameter distribution is Gaussian and the forward operator $F(\cdot)$ is linear. If these conditions are not satisfied, then ES can lead to unacceptable data matches and unphysical results (Aanonsen et al., 2009). To deal with this issue, we focus in this paper on a recent development by Emerick and Reynolds (2012a), namely the ensemble smoother with multiple data assimilation (ES-MDA). With this approach, one standard ES step, which is comparable to a single Gauss-Newton iteration when maximizing the posterior probability of the model parameters (Tarantola, 2005), is replaced by a number of smaller update steps (or assimilation iterations) based on a Kalman matrix and perturbed data vector that are recalculated at each iteration. In order to correctly sample from the posterior distribution, the measurement-error covariance matrix \mathbf{C}_D must be inflated in this procedure. Typically this is done by scaling \mathbf{C}_D by the number of assimilation iterations; however, more generalized inflation coefficients may be used (Emerick and Reynolds, 2012a). For linear forward solvers, the ES-MDA algorithm is theoretically equivalent to standard ES (Emerick and Reynolds, 2012b). For non-linear problems, it can be shown that the methodology has links to annealed importance sampling (Stordal and Elsheikh, 2015).

Algorithm 1 outlines the steps involved in the ES-MDA procedure where, for simplicity, the measurement-error covariance matrix is assumed diagonal with entries σ^2 and the corresponding inflation coefficient α is set to equal the number of assimilation iterations n_{iter} . To estimate the inverse of matrix $(\mathbf{C}_{DD} + \alpha \cdot \mathbf{C}_D)$ we use the truncated singular value decomposition (TSVD) and retain 99% of the total energy of the singular values (Emerick and Reynolds, 2012a).

Algorithm 1: Standard ES-MDA

```

1 set  $\mathbf{C}_D = \sigma^2 \mathbf{I}$ ;  $\alpha = n_{iter}$ 
2 draw prior ensemble  $\mathbf{m}$  containing  $n_e$  sets of model
  parameters
3 for  $i = 1, \dots, n_{iter}$  do
4   | perturb observations  $\mathbf{d}_{pert} \sim \mathcal{N}(\mathbf{d}_{obs}, \alpha \cdot \mathbf{C}_D)$ 
5   | for  $j = 1, \dots, n_e$  do
6   |   | compute predicted data  $\mathbf{d}_j = F(\mathbf{m}_j)$ 
7   |   | end
8   | compute Kalman matrix  $\mathbf{K} = \mathbf{C}_{MD} \cdot (\mathbf{C}_{DD} + \alpha \cdot \mathbf{C}_D)^{-1}$ 
9   | for  $j = 1, \dots, n_e$  do
10  |   | compute residual  $\mathbf{r}_j = \mathbf{d}_{pert} - \mathbf{d}_j$ 
11  |   | update ensemble  $\mathbf{m}_j = \mathbf{m}_j + \mathbf{K} \mathbf{r}_j$ 
12  |   | end
13 end
```

2.3. Model Error

When working with a perfectly known forward solver $F(\cdot)$ in the ES-MDA procedure outlined above, the residual \mathbf{r}_j corresponding to the j th ensemble member \mathbf{m}_j , which quantifies

the misfit between the perturbed observations and the predicted (forward-calculated) data, is given by

$$\begin{aligned} \mathbf{r}_j &= \mathbf{d}_{pert} - F(\mathbf{m}_j) \\ &= \underbrace{F(\mathbf{m}_{true}) - F(\mathbf{m}_j)}_{\text{parameter-error component}} + \tilde{\epsilon}_d, \end{aligned} \quad (6)$$

where $\tilde{\epsilon}_d$ denotes the sum of the measurement errors and perturbation noise. In the case where $\mathbf{m}_j = \mathbf{m}_{true}$, we see from Equation (6) that the parameter-error term, which represents the component of the residual related to being at the wrong set of model-parameter values, will be zero and that the residual energy will tend to be minimized. In the case where a proxy forward solver $\hat{F}(\cdot)$ is used in the ES-MDA algorithm, however, the latter does not generally hold true because

$$\begin{aligned} \mathbf{r}_j &= \mathbf{d}_{pert} - \hat{F}(\mathbf{m}_j) \\ &= F(\mathbf{m}_{true}) - \hat{F}(\mathbf{m}_j) + \tilde{\epsilon}_d \\ &= \underbrace{F(\mathbf{m}_{true}) - F(\mathbf{m}_j)}_{\text{parameter-error component}} + \underbrace{F(\mathbf{m}_j) - \hat{F}(\mathbf{m}_j)}_{\text{model-error component}} + \tilde{\epsilon}_d. \end{aligned} \quad (7)$$

Indeed, the presence of an additional model-error component in Equation (7) compared to Equation (6) means that the residual energy may be minimized for model parameter vectors \mathbf{m}_j that are substantially different from \mathbf{m}_{true} , as such parameter sets well tend to compensate for the model errors. As mentioned previously, this can lead to strongly biased and overconfident posterior statistics.

In order to deal with model error in the ES-MDA procedure arising from use of a proxy solver, we build on the methodology presented in Köpke et al. (2018) for Bayesian-MCMC inversion, which focuses on identifying the model-error component of the residual using a projection-based method. We refer the reader to that paper for details beyond those given here. Algorithm 2 outlines the steps involved in our modified ES-MDA methodology, again assuming that $\alpha = n_{iter}$ and $\mathbf{C}_D = \sigma^2 \mathbf{I}$ for simplicity, where \mathbf{I} is the identity matrix. In addition we introduce n_d , defined as the number of detailed solver runs used to learn about the model error, and set it to a value to less than or equal to the number of ensemble members n_e .

In the modified ES-MDA algorithm, initial ensemble members \mathbf{m}_j are drawn from the prior parameter distribution and the corresponding predicted data $\hat{\mathbf{d}}_j = \hat{F}(\mathbf{m}_j)$ are computed using the proxy solver. In each assimilation iteration, a subset of the ensemble members having size n_d is randomly chosen, for which the detailed forward responses $\mathbf{d}_j = F(\mathbf{m}_j)$ are also calculated. The resulting n_d model-error vectors (i.e., $\mathbf{d}_j - \hat{\mathbf{d}}_j$) and corresponding parameter sets \mathbf{m}_j are stored in the dictionaries \mathbf{D}^E and \mathbf{D}^M , respectively. As \mathbf{D}^M and \mathbf{D}^E are further enriched with n_d entries in each ES-MDA iteration, more detailed information about the model error around the posterior solution is gathered.

For each ensemble member \mathbf{m}_j , the model-error component of the residual is identified and used to correct the proxy response in order to mimic the detailed forward solver. To this end,

the current model-parameter dictionary \mathbf{D}^M is searched for the K -nearest-neighbor (KNN) parameter sets to \mathbf{m}_j using a standard Euclidean distance measure (e.g., Hastie et al., 2009). An orthonormal basis \mathbf{B}_j for the model error at \mathbf{m}_j is then constructed from these parameter sets using the Gram-Schmidt technique (e.g., Strang et al., 1993). We assume in our work that the data-measurement-error and parameter-error components of the residual are orthogonal to the model-error component, and therefore cannot be well represented by the basis. An estimate of the model error $\tilde{\mathbf{e}}_j$ is thus obtained by projecting the residual onto \mathbf{B}_j

$$\tilde{\mathbf{e}}_j = \mathbf{B}_j \cdot \mathbf{B}_j^T \cdot \mathbf{r}_j. \quad (8)$$

This result, which represents the details missing in the proxy solution, is then added to the proxy response to obtain a corrected forward response

$$\tilde{\mathbf{d}}_j = \hat{\mathbf{d}}_j + \tilde{\mathbf{e}}_j \quad (9)$$

with corresponding corrected residual

$$\tilde{\mathbf{r}}_j = \mathbf{d}_{pert} - \tilde{\mathbf{d}}_j. \quad (10)$$

The corrected forward responses for all of the ensemble members are used to compute the corrected Kalman matrix

$$\tilde{\mathbf{K}} = \mathbf{C}_{M\tilde{D}} \cdot (\mathbf{C}_{\tilde{D}\tilde{D}} + \alpha \cdot \mathbf{C}_D)^{-1} \quad (11)$$

which is used with the corrected residuals to update the ensemble.

Under the stated assumptions and with appropriate choices of n_e , n_d , and K , Algorithm 2 allows us to effectively reduce the computational cost of ES-MDA when considering large ensembles through the use of a proxy solver. The dimensionality of the parameter-estimation problem and the difference in computational cost between the proxy and detailed forward solutions determine how much computational benefit is derived from this methodology. We refer the reader to Köpke et al. (2018) for a detailed discussion of the orthogonality assumption between the model-error and other components of the residual.

3. APPLICATION TO CROSSHOLE GPR TOMOGRAPHY

3.1. Experimental Design and Forward Models

As an example, we now apply our modified ES-MDA algorithm with model-error correction to the crosshole GPR travel-time tomography inverse problem. A transmitter and receiver antenna, located in two adjacent boreholes, are used to obtain the travel times of radar energy between the holes for different antenna positions. These times are linked to the spatial distribution of subsurface radar-wave velocity, the estimation of which is the goal of the inverse problem. Crosshole GPR travel-time tomography represents an excellent test problem for our purposes because (i) it has been extremely well-studied, most notably from a stochastic inverse standpoint (e.g., Giroux et al.,

Algorithm 2: ES-MDA with model-error correction

```

1 set  $C_D = \sigma^2 \mathbf{I}$ ;  $\alpha = n_{iter}$ ;  $n_d \leq n_e$ ;  $l = 0$ 
2 draw prior ensemble  $\mathbf{m}$  containing  $n_e$  sets of model
  parameters
3 for  $i = 1, \dots, n_{iter}$  do
4   perturb observations  $\mathbf{d}_{pert} \sim \mathcal{N}(\mathbf{d}_{obs}, \alpha \cdot C_D)$ 
5   select random subset  $\mathbf{m}_r$  of  $\mathbf{m}$  with  $r = 1, \dots, n_d$ 
6   for  $j = 1, \dots, n_e$  do
7     compute predicted data using proxy solver
8      $\hat{\mathbf{d}}_j = \hat{F}(\mathbf{m}_j)$ 
9     if  $\mathbf{m}_j \in \mathbf{m}_r$  then
10       $l = l + 1$ 
11      compute predicted data using detailed solver
12       $\mathbf{d}_l = F(\mathbf{m}_j)$ 
13      enrich model-error dictionary  $\mathbf{D}_l^E = \mathbf{d}_l - \hat{\mathbf{d}}_j$ 
14      enrich corresponding model-parameter dictionary
15       $\mathbf{D}_l^M = \mathbf{m}_j$ 
16    end
17  end
18 for  $j = 1, \dots, n_e$  do
19   search dictionary  $\mathbf{D}^M$  for  $K$ -nearest neighbors to  $\mathbf{m}_j$ 
20   take corresponding entries from  $\mathbf{D}^E$  and place in set
21    $\mathbf{D}^K(\mathbf{m}_j)$ 
22   build orthonormal basis  $\mathbf{B}$  having  $\text{span}\{\mathbf{D}^K(\mathbf{m}_j)\}$ 
23   compute residual  $\mathbf{r}_j = \mathbf{d}_{pert} - \hat{\mathbf{d}}_j$ 
24   project  $\mathbf{r}_j$  onto  $\mathbf{B}$  to estimate model error
25    $\tilde{\mathbf{e}}_j = \mathbf{B} \cdot \mathbf{B}^T \cdot \mathbf{r}_j$ 
26   correct proxy response  $\tilde{\mathbf{d}}_j = \hat{\mathbf{d}}_j + \tilde{\mathbf{e}}_j$ 
27 end
28 compute Kalman matrix  $\tilde{\mathbf{K}} = \mathbf{C}_{MD} \cdot (\mathbf{C}_{DD} + \alpha \cdot C_D)^{-1}$ 
29 for  $j = 1, \dots, n_e$  do
30   compute residual  $\tilde{\mathbf{r}}_j = \mathbf{d}_{pert} - \tilde{\mathbf{d}}_j$ 
31   update ensemble  $\mathbf{m}_j = \mathbf{m}_j + \tilde{\mathbf{K}}\tilde{\mathbf{r}}_j$ 
32 end

```

2007; Looms et al., 2008; Scholer et al., 2012; Hansen et al., 2013; Linde and Vrugt, 2013); (ii) it involves a high-dimensional and spatially distributed set of model parameters that must be estimated; and (iii) the forward problem can be solved in a variety of different ways using different physical approximations.

GPR travel times are linked to the spatial distribution of electrical properties between the two boreholes, predominantly the dielectric permittivity, through Maxwell's equations. Numerical solution of these equations represents the most accurate means of calculating the travel times, but at the same time it is highly computationally expensive. To reduce the computational cost, the physics of the electromagnetic wave propagation can be approximated using ray theory, whereby the effects of frequency are ignored and we solve the eikonal equation (e.g., Nowack, 1992). To decrease the computational cost even further, the straight-ray approximation may also

be considered, which means that the ray paths that connect transmitter and receiver locations are assumed to be straight lines (e.g., Cordua et al., 2008). The latter approximation is typically applied in cases where contrasts in velocity do not exceed 10%; however it is only truly valid when the subsurface is homogeneous. Here, we consider the eikonal equation to be our detailed forward solver $F(\cdot)$ and the straight-ray approach to be our proxy solver $\hat{F}(\cdot)$. This choice was made for demonstration purposes, as it allows us to compare the results of ES-MDA inversions obtained using our approach to those obtained using standard ES-MDA, as well as MCMC, based on the detailed solver alone. That is, the eikonal solution is fast enough to allow it to be used in the standard ES-MDA algorithm with a large number of ensemble members, as well as for MCMC posterior sampling. Note that, instead of estimating directly velocity from GPR travel times in our analysis, we focus on the estimation of subsurface slowness (the reciprocal of velocity) which makes the straight-ray forward problem linear.

The survey configuration for our synthetic experiments consists of two boreholes that are 8-m deep and 4-m apart (Figure 1). Transmitter and receiver antenna positions are distributed equally in depth every 0.2 m down the left and right boreholes, respectively. Sending a radar pulse from all transmitter positions to all receiver positions yields 1,600 travel-time data. We consider a pixel-based parameterization of the subsurface whereby the region between the boreholes by discretized into 20×40 square cells of constant-slowness and side length 0.2 m. The synthetic "true" subsurface and initial prior ensemble members are generated by sequential Gaussian simulation using the GSLIB software package (Deutsch and Journel, 1992). The mean slowness is set to 10 ns/m and an exponential autocovariance kernel having a standard deviation of 1.7 ns/m is assumed, with horizontal and vertical correlation lengths of 6 m and 1.5 m, respectively. The corresponding synthetic observed data are generated by solving the eikonal equation and adding measurement errors, the latter of which are simulated as Gaussian random noise having covariance matrix $C_D = \sigma^2 \mathbf{I}$ with standard deviation $\sigma = 0.2$ ns.

3.2. ES-MDA Results

Our goals in this analysis are to (i) study the effects of model error on ES-MDA inversions; (ii) investigate the influence of the ensemble size on the accuracy of the results obtained; and (iii) explore how the parameters of our modified ES-MDA procedure with model-error correction can be chosen to provide an optimal balance between computational efficiency and accuracy. To this end, we compare parameter-estimation results for different numbers of ensemble members when (i) there is no model error, meaning that the detailed (eikonal-equation) forward solver is used within the standard ES-MDA procedure (Algorithm 1); (ii) model error is present but not accounted for, meaning that the proxy (straight-ray) forward solver is used within the standard ES-MDA procedure; and (iii) model error is present and accounted for through the use of Algorithm 2. In each case, we examine the combined results from 10 ES-MDA inversions obtained using different initial ensembles and $n_{iter} =$

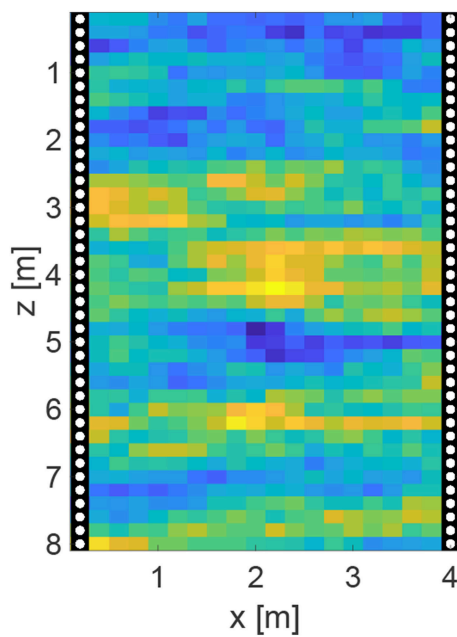


FIGURE 1 | Considered crosshole GPR survey configuration with boreholes shown as black lines. Transmitter and receiver positions are shown as white dots and distributed every 0.2 m down the boreholes.

8 assimilation iterations. Ensemble sizes of $n_e = 20, 40, 80, 160, 320$, and 640 are considered in our analysis.

To assess the quality of the inversion results, we consider two metrics. The average root-mean-square (RMS) travel-time misfit, which quantifies globally the ability of the posterior ensemble to represent the observed data, is defined for ES-MDA run i ($i = 1, 2, \dots, 10$) as follows:

$$M_i^T = \frac{1}{n_e} \sum_{j=1}^{n_e} \frac{1}{\sqrt{n_T}} \|\mathbf{d}_{obs} - \mathbf{d}_{ij}\|_2. \quad (12)$$

where n_T is the number of travel-time data. For the case where model error is absent and data errors are zero-mean and Gaussian distributed with covariance matrix $\mathbf{C}_D = \sigma^2 \mathbf{I}$, the expected value of M_i^T will be σ . Note that, in the case where model error is present but not accounted for, the detailed forward solver \mathbf{d}_{ij} in equation (12) is replaced with the proxy solver $\hat{\mathbf{d}}_{ij}$.

We also consider in our analysis the average RMS slowness misfit, defined by

$$M_i^S = \frac{1}{n_e} \sum_{j=1}^{n_e} \frac{1}{\sqrt{n_S}} \|\mathbf{m}_{true} - \mathbf{m}_{ij}\|_2, \quad (13)$$

where n_S is the number of slowness cells. This metric quantifies how well the posterior ensemble captures the true underlying model parameters, and can only be employed in the case of synthetic data where the true subsurface slowness distribution is known. In addition to the two metrics in Equations (12) and (13), we plot the mean slowness fields over all ensemble members and

all ES-MDA runs in order to visually compare them with the true slowness distribution.

Figure 2 summarizes the parameter-estimation results obtained for the case where there is no model error. In **Figure 2A** we observe that the average RMS travel-time misfit decreases consistently with larger numbers of ensemble members toward the expected value of $\sigma = 0.2$ ns which reflects the prescribed data errors. After around 320 ensemble members, adding more members is seen to only slightly further improve the results. **Figure 2B** shows that the slowness misfit also decreases consistently as a function of ensemble size. This is supported by **Figure 2C**, which shows that the mean slowness fields become increasingly detailed and similar to the true subsurface distribution as the number of ensemble members increases. The increasing overall accuracy of the ES-MDA results with larger ensemble size is based on the reduction of sampling errors following the central limit theorem (Evensen, 2007). We can conclude that larger ensembles combined with the detailed forward solver enable us to obtain more reliable posterior parameter estimates, but at the cost of significantly greater computational effort when the detailed solver is computationally expensive.

Figure 3 summarizes the parameter-estimation results obtained for the case where model error is present but not accounted for. In **Figure 3A** we observe that, in accordance with **Figure 2A**, the travel-time misfit consistently decreases with larger ensemble size. However, it approaches a stable value that is well above the target value of $\sigma = 0.2$ ns, because the presence of model error does not allow data fitting to a level that is in accordance with the prescribed data errors. In addition, unlike in **Figure 2B**, the slowness misfit now decreases only until $n_e = 40$, after which it increases again (**Figure 3B**). For $n_e \leq 40$, we do not have enough ensemble members to resolve the details of the posterior distribution and therefore only the posterior mean can be represented in the parameter estimation results (Chen and Zhang, 2006). Conversely, when $n_e > 40$, the solution moves toward a biased posterior distribution. That is, with more ensemble members the parameters have the ability to compensate for the model error and the data become over-fitted, meaning that a better data match is achieved but the parameters do not represent the true subsurface model. The mean slowness fields in **Figure 3C** allow us to see how the model error introduces bias in the parameter-estimation results as n_e increases; strong artifacts are clearly observed in these fields for ensemble sizes larger than 160.

Finally, we examine the inversion results obtained for the case where model error is present and accounted for through our modified ES-MDA approach. We first consider inversions where $n_d = 20$ detailed solver runs per iteration are used to build the model-error dictionary and $K = 20$ KNN from this dictionary are used to construct the model-error basis for each ensemble member. The corresponding results are shown in **Figure 4**. Note that, in the case with $n_e = 20$, the detailed solver is executed for each ensemble member, which corresponds to the standard ES-MDA procedure in Algorithm 1. In **Figures 4A,B** we see that both the travel-time and slowness misfit decrease

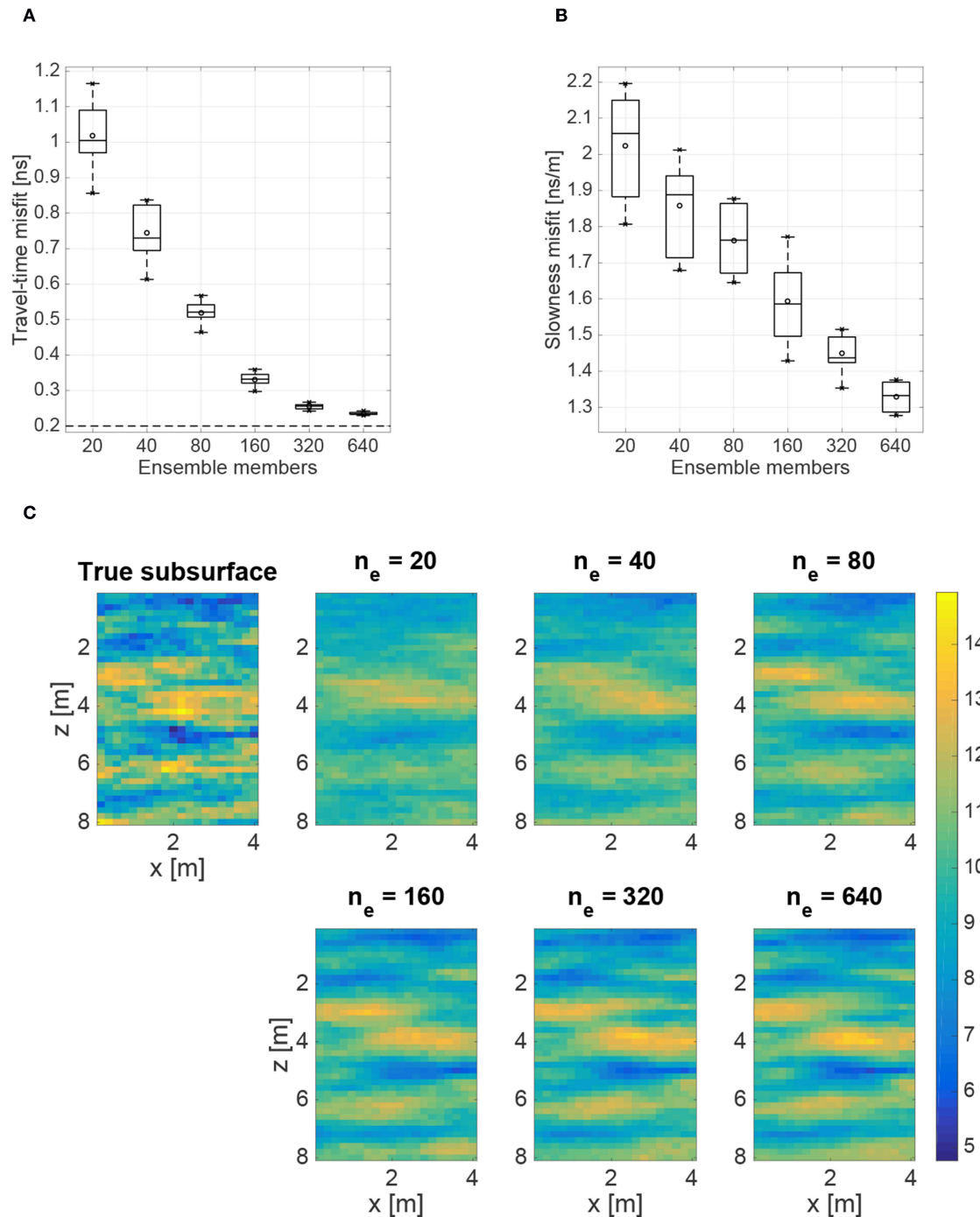


FIGURE 2 | Results of 10 standard ES-MDA runs for the case of no model error. Shown as a function of the number of ensemble members, n_e , are **(A)** box plots of the average RMS travel-time misfit [ns]; **(B)** box plots of the average RMS slowness misfit [ns/m]; and **(C)** the mean posterior slowness fields [ns/m]. Added to the box plots are the mean (circles), and minimum and maximum values (crosses). The dashed horizontal line in **(A)** represents the expected value of the travel-time misfit assuming that the residuals follow the prescribed Gaussian distribution for the data errors.

from $n_e = 20$ until they reach a minimum at around 80–160 ensemble members. This demonstrates that the consideration of larger ensembles through use of a proxy solver combined with our model-error correction can lead to more accurate

results compared to standard ES-MDA based on small ensembles and the detailed forward solver. These results are confirmed in **Figure 4C**, where we observe that the bias is largely removed from the mean slowness fields for $n_e \leq 160$ in comparison to

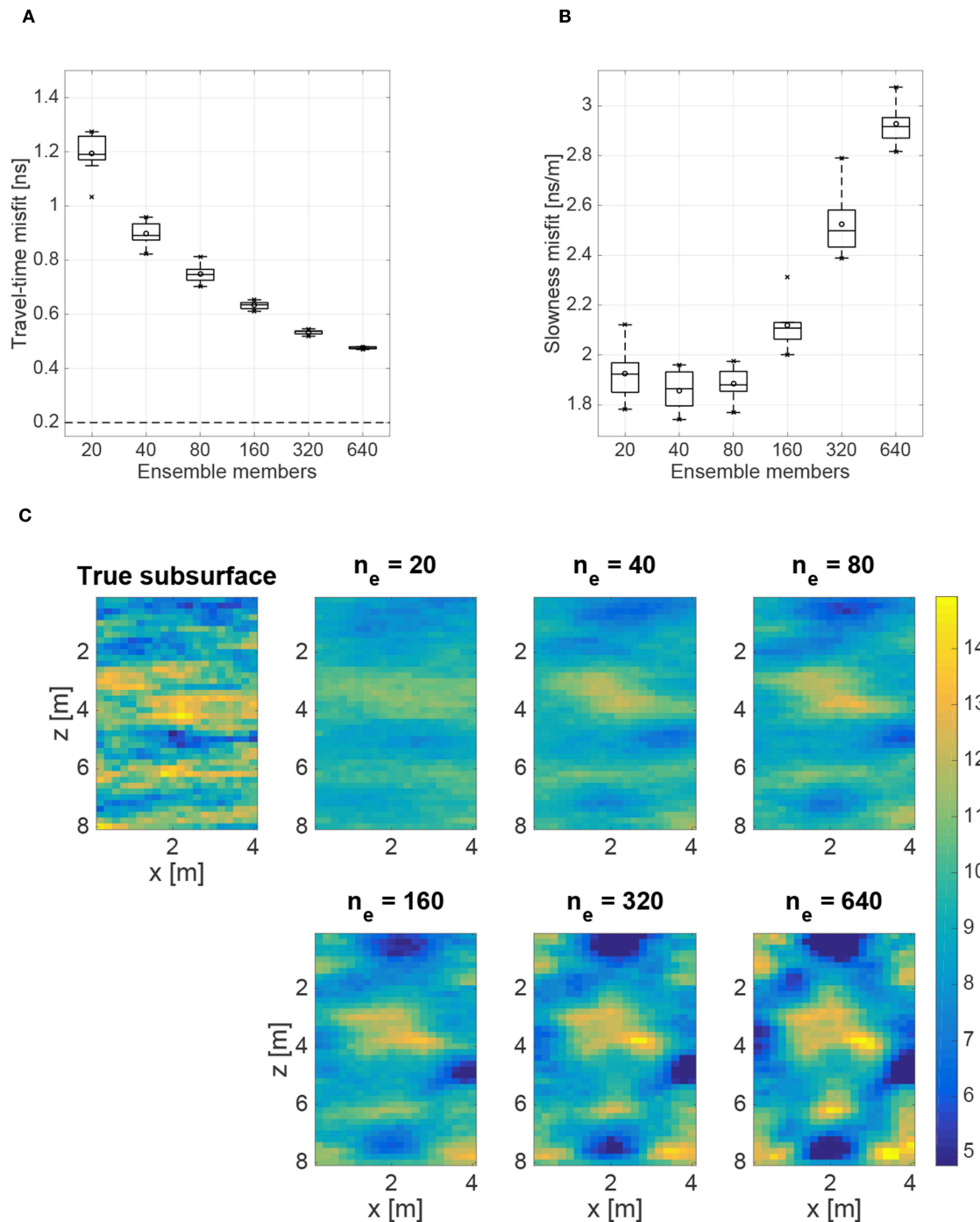


FIGURE 3 | Results of 10 standard ES-MDA runs for the case where model error is present but not accounted for. Shown as a function of the number of ensemble members, n_e , are **(A)** box plots of the average RMS travel-time misfit [ns]; **(B)** box plots of the average RMS slowness misfit [ns/m]; and **(C)** the mean posterior slowness fields [ns/m]. Added to the box plots are the mean (circles), and minimum and maximum values (crosses). The dashed horizontal line in **(A)** represents the expected value of the travel-time misfit assuming that the residuals follow the prescribed Gaussian distribution for the data errors.

Figure 4C. However, for ensemble sizes larger than around 160, the travel-time and slowness misfit are seen to again increase, meaning that the data become under-fitted. That is, the ensemble size becomes too large compared to the number of detailed solver

calculations for the model error to be well represented in the dictionary, meaning that projection onto the model-error basis will not properly identify the model-error component of the residual. This has the effect of introducing bias into the inversion

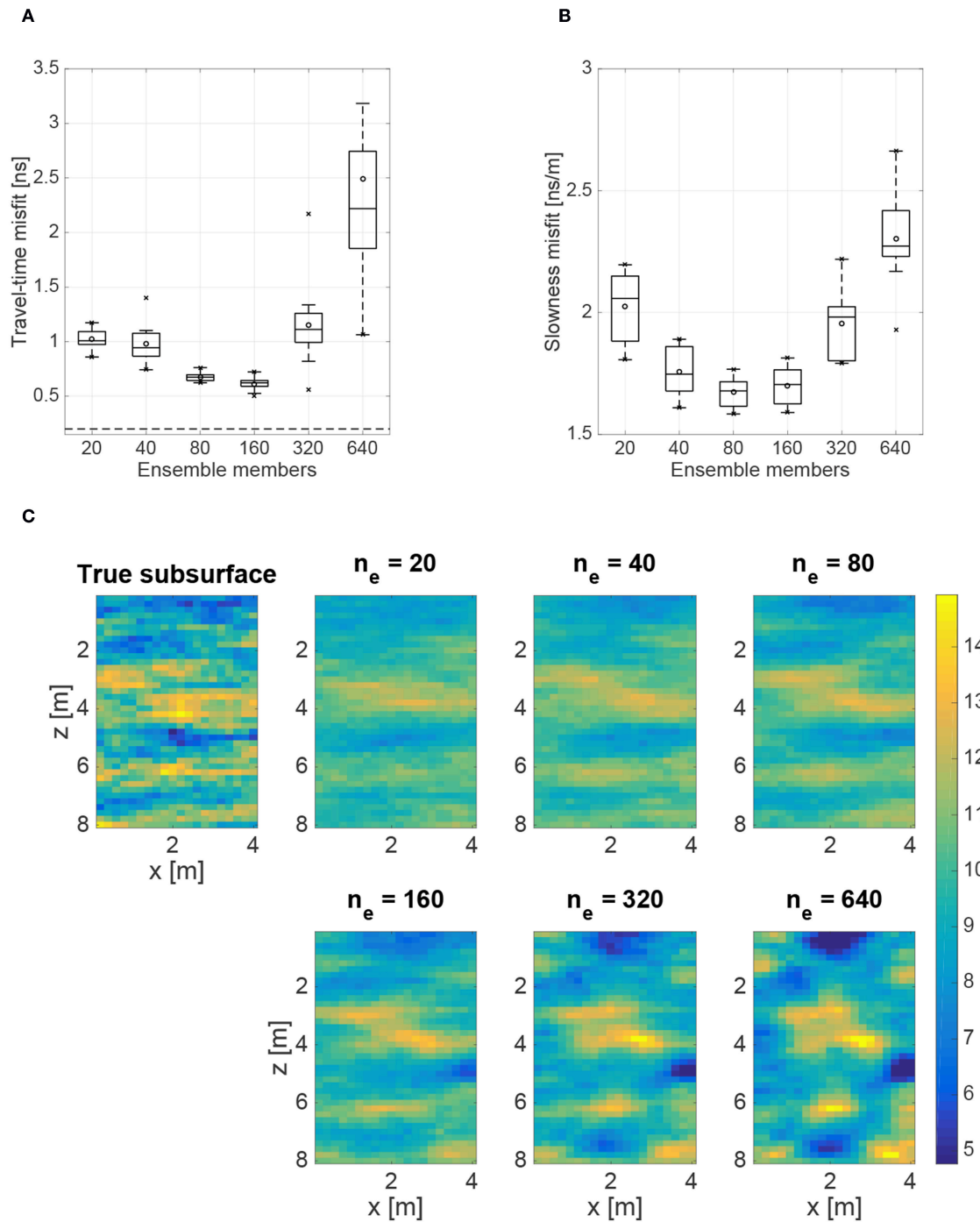


FIGURE 4 | Results of 10 runs of our modified ES-MDA algorithm to account for model error, with 20 detailed solver calculations and 20 KNN. Shown as a function of the number of ensemble members, n_e , are **(A)** box plots of the average RMS travel-time misfit [ns]; **(B)** box plots of the average RMS slowness misfit [ns/m]; and **(C)** the mean posterior slowness fields [ns/m]. Added to the box plots are the mean (circles), and minimum and maximum values (crosses). The dashed horizontal line in **(A)** represents the expected value of the travel-time misfit assuming that the residuals follow the prescribed Gaussian distribution for the data errors.

results, which is clearly seen in the mean slowness fields in **Figure 4C** when $n_e > 160$.

To explore the latter findings, we consider again Algorithm 2, but this time using $n_d = 40$ detailed solver runs per iteration and

$K = 40$ KNN to build the model-error dictionary and construct the model-error basis, respectively. Here, when $n_e = 40$, the detailed solver is executed for each ensemble member, which again corresponds to the standard ES-MDA procedure. Similar

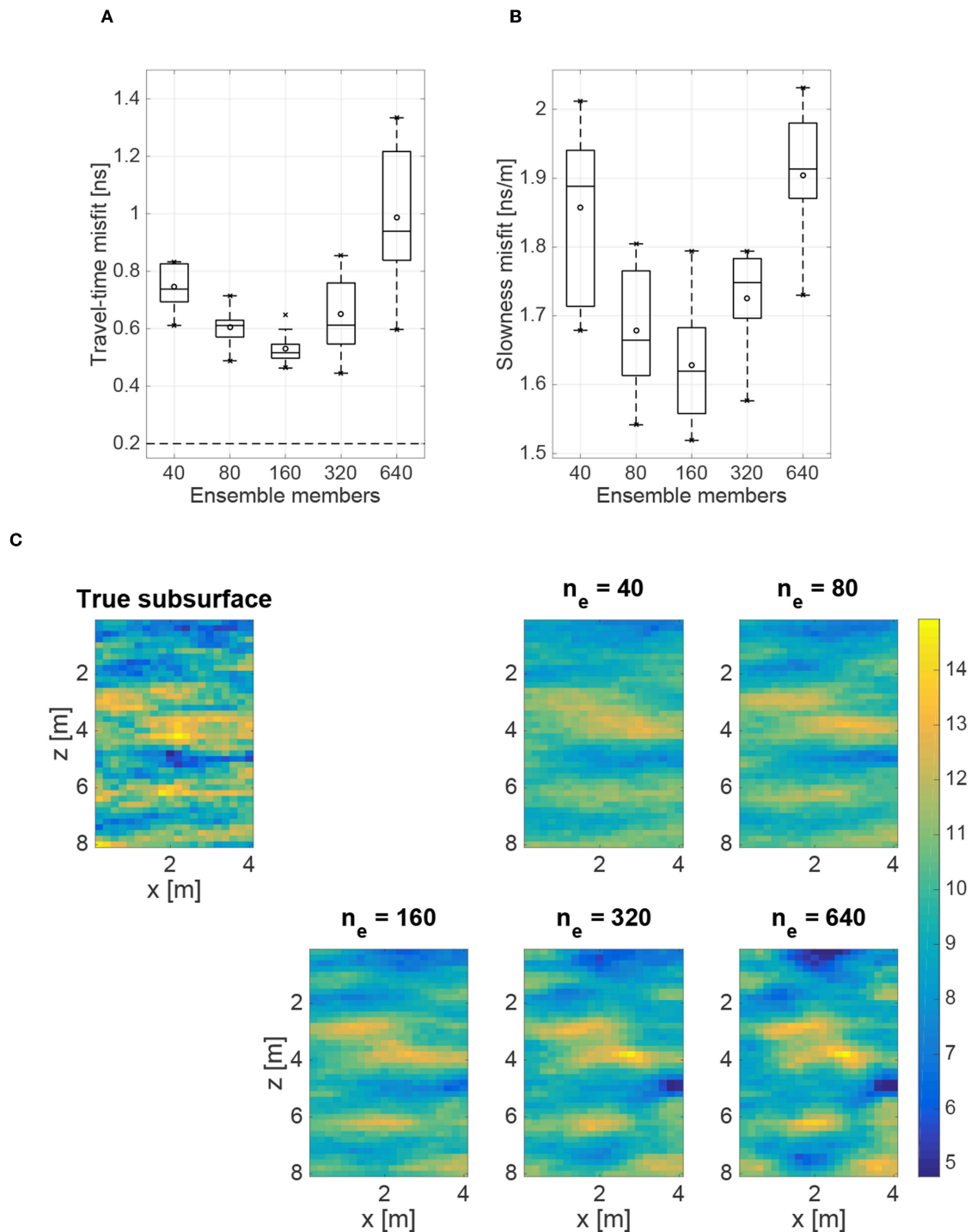


FIGURE 5 | Results of 10 runs of our modified ES-MDA algorithm to account for model error, with 40 detailed solver calculations and 40 KNN. Shown as a function of the number of ensemble members, n_e , are **(A)** box plots of the average RMS travel-time misfit [ns]; **(B)** box plots of the average RMS slowness misfit [ns/m]; and **(C)** the mean posterior slowness fields [ns/m]. Added to the box plots are the mean (circles), and minimum and maximum values (crosses). The dashed horizontal line in **(A)** represents the expected value of the travel-time misfit assuming that the residuals follow the prescribed Gaussian distribution for the data errors.

to before, we observe in **Figures 5A,B** that the travel-time and slowness misfit decrease from $n_e = 40$ until a minimum is reached. Although it is difficult to determine the exact position

of this minimum due to the limited discretization, we see that it falls somewhere around 160 ensemble members. After the minimum value, the travel-time and slowness misfit are again

seen to increase and the data become under-fitted. This behavior is well reflected in the mean slowness fields in **Figure 5C**, which show good agreement with the true field for $n_e \leq 320$, but clearly contain model-error-related artifacts when $n_e = 640$.

4. DISCUSSION

We saw above that use of the modified ES-MDA approach described in Algorithm 2 can allow for a significant reduction in posterior bias when employing a proxy forward solver compared to the standard ES-MDA procedure. This offers the possibility of considering large ensemble sizes within ES-MDA, which can be computationally prohibitive in the context of an expensive detailed forward solver. One issue requiring further discussion, however, is the balance between (i) the number of ensemble members considered n_e , which in the case of no model error controls the accuracy of the results obtained; and (ii) the number of detailed solver runs n_d , which determines the success of the model-error correction. We saw in **Figure 4** that, when $n_d = 20$ detailed solver runs per iteration were considered in the modified ES-MDA procedure, use of ensemble sizes between 40 and 160 allowed for an improvement in parameter estimates compared to standard ES-MDA based on the detailed solver with $n_e = 20$. When $n_d = 40$ detailed solver runs per iteration were considered, on the other hand, a corresponding improvement was found for ensemble sizes between 80 and 320. These results suggest that, at least for the application presented in this paper, Algorithm 2 can be successfully applied only for ensembles having size less than 8 times the number of detailed solver runs per iteration. Past this number, there will not be enough entries in the model-error dictionary to allow for an accurate correction of the model error for all ensemble members, and the benefits of using an approximate solver with model-error correction will be compromised. Further exploration of these findings in the context of other inverse problems is required.

Another issue in need of some discussion is how the results of using the standard and modified ES-MDA algorithms presented in **Figures 2–5** compare with samples from the “true” posterior distribution, the latter of which we assume to be available through MCMC sampling based on the detailed forward solver. To this end, we show in **Figure 6** five randomly chosen posterior slowness realizations obtained via MCMC sampling based on the eikonal equation (**Figure 6A**); standard ES-MDA based on both the eikonal equation and straight-ray approximation (**Figures 6B,C**); and our modified ES-MDA procedure with $n_d = 20$ and $n_d = 40$ (**Figures 6D,E**). The point-wise posterior mean and standard deviation, computed over all available samples, are also shown for reference. The results in **Figure 6A** were obtained using the sequential geostatistical simulation technique (e.g., Ruggeri et al., 2015), where after burn-in, the results of 140,000 MCMC iterations were thinned to provide 140 posterior samples. For **Figures 6B–E**, the number of ensemble members considered was chosen to be the maximum investigated value ($n_e = 640$) for standard ES-MDA, whereas for our modified ES-MDA procedure it was set equal to 8 times the number of detailed solver runs, as discussed above.

In comparing the posterior realizations in **Figures 6A,B**, we see that they are highly similar, which suggests that ES-MDA based on the detailed forward solver and using a large number of ensemble members allows for adequate sampling of the Bayesian posterior distribution. The corresponding standard deviation images generally show a pattern that reflects the degree of ray coverage; regions of higher slowness contain a smaller ray density. However, the ES-MDA solution is seen to contain more variability, which may arise because the 140,000 MCMC iterations utilized were not enough to adequately explore the posterior space. In examining the stochastic realizations in **Figure 6C**, the proxy-related bias in this solution is clearly apparent. Here, the symmetric pattern of variability reflects variations in ray density that are controlled solely by the antenna locations in the straight-ray case. Finally, in comparing the results in **Figures 6D,E** with those in **Figure 6A**, we see that our modified ES-MDA algorithm largely removes the proxy-related bias and allows for the generation of posterior samples that are close in appearance to the MCMC solution, which again validates our approach. These samples do, however, show a slightly higher degree of variability with less correlation compared to **Figures 6A,B**, with the $n_d = 40$ solution providing a better match than the $n_d = 20$ solution. As discussed above, with a fixed number of detailed solver runs per iteration, there is an upper limit to the number of ensemble members that can be effectively considered in our procedure, which in turn may not be enough to characterize exactly the posterior distribution (see **Figure 2**). More accurate results would thus likely require greater numbers of detailed solver runs to allow for an increase in the ensemble size. The higher degree of variability in these results may also reflect imperfect removal of the model error, or the lesser number of samples used to compute the point-by-point mean and standard deviation.

Lastly, we wish to elaborate on the number of internal ES-MDA assimilation iterations considered in our approach, which was held constant at a value of $n_{iter} = 8$ for all of the results presented section 3.2. To this end, we study in **Figure 7** the travel-time and slowness misfit as a function of iteration for an ensemble size of $n_e = 160$ when (i) no model error is present; (ii) model error is present but not accounted for; and (iii) model error is present and accounted for using $n_d = 40$ detailed solver runs per iteration and $K = 40$ KNN. We observe overall that the first assimilation iteration has the largest influence on reducing the travel-time and slowness misfit from prior to posterior. This arises because of the linearity of the proxy (straight-ray) solver and the weak non-linearity of the detailed (eikonal) solver in our travel-time tomography application. Indeed, Emerick and Reynolds (2012b) proved that, for linear problems, one ES update using the measurement noise covariance matrix is equivalent to multiple ES-MDA updates using the inflated covariance matrix. When using the detailed solver in the inversion where there is no model error, for example, **Figures 7A,D** show a large decrease in travel-time and slowness misfit after one iteration and a slow decrease from iterations 2–8. In this case 4 iterations would be enough to obtain similar parameter-estimation results to those obtained using 8 iterations. When model error is present but not accounted

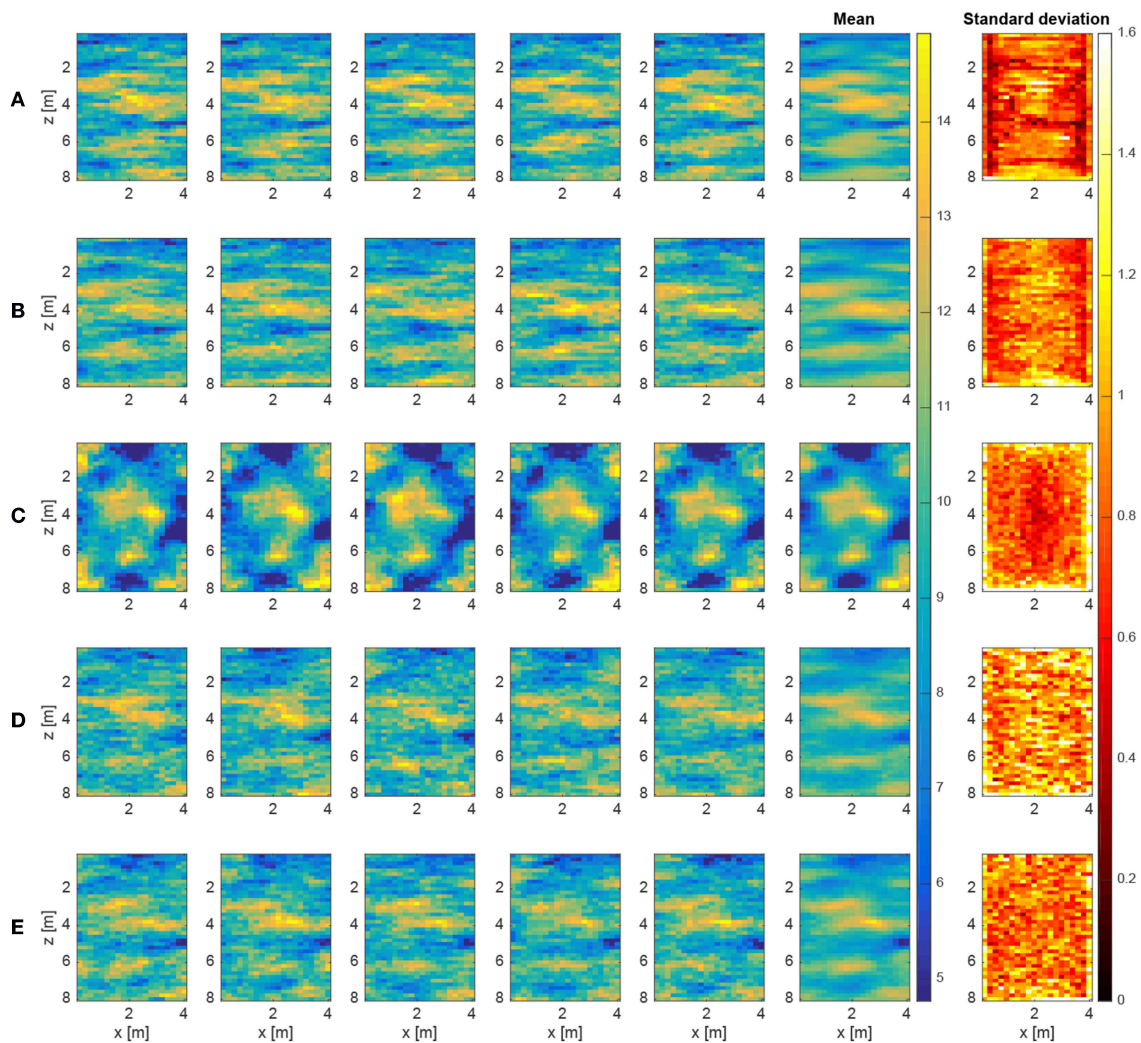


FIGURE 6 | Five randomly chosen posterior slowness realizations along with the point-wise mean and standard deviation [ns/m] obtained via **(A)** MCMC sampling based on the detailed eikonal solver (140 samples total); **(B)** standard ES-MDA based on the detailed eikonal solver with $n_e = 640$ (**Figure 2**; 6,400 samples total); **(C)** standard ES-MDA based on the proxy straight-ray solver with $n_e = 640$ (**Figure 3**; 6,400 samples total); **(D)** our modified ES-MDA algorithm with 20 KNN, $n_d = 20$, and $n_e = 160$ (**Figure 4**; 1,600 samples total); and **(E)** our modified ES-MDA procedure with 40 KNN, $n_d = 40$, and $n_e = 320$ (**Figure 5**; 3,200 samples total).

for, we see in **Figure 7B** that a good travel-time data match is achieved and no further improvement is observed after one iteration. However, **Figure 7E** shows that the corresponding slowness misfit is still large after one iteration compared to **Figure 7D**, which arises because of overfitting; that is, the inversion attempts to fit the model error. Applying our proposed method, we see in **Figure 7C** that the travel-time misfit is again primarily reduced in the first iteration and behaves similarly to the case where no model error is present (**Figure 7A**). More importantly, over-fitting is significantly reduced (**Figure 7F**) and the slowness misfit after only 4 iterations is similar to that seen in **Figure 7B**. This again confirms that employing our proposed approach can effectively remove proxy-related bias and allow the ES-MDA procedure to yield results that are comparable to inversions when no model error is present.

Although it may be possible to arrive at these results in less iterations than the 8 considered in this paper, it is difficult to know in advance how the combined approach of proxy solver and model-error correction behaves in terms of the internal ES-MDA iterations.

5. CONCLUSIONS

We have presented in this paper an approach that builds on the work of Köpke et al. (2018) in order to remove the bias associated with the use of proxy forward solvers in ES-MDA inversions. This allows for the consideration of larger ensemble sizes, which help to improve the accuracy of the parameter estimates obtained. Instead of constructing a local or global error

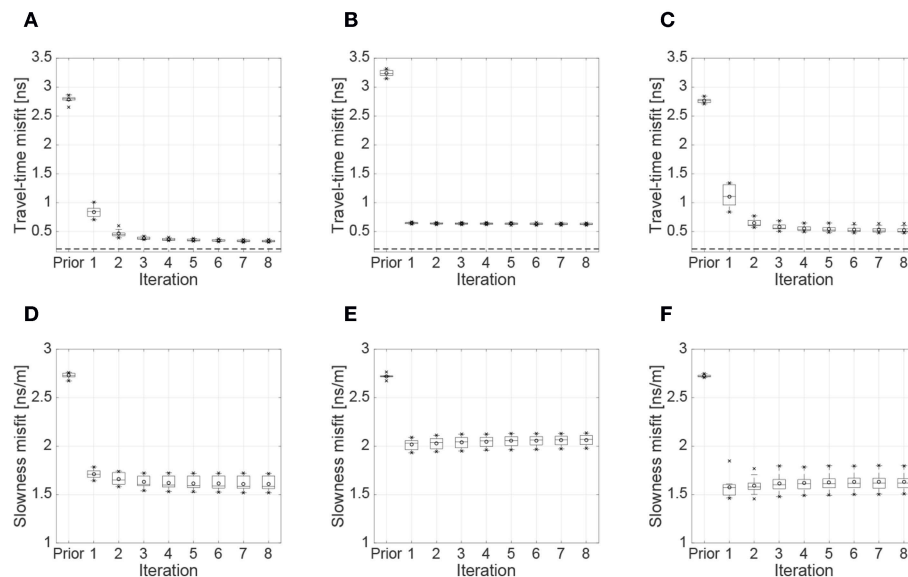


FIGURE 7 | Box plots showing the results of 10 ES-MDA runs using 160 ensemble members, as a function of the number of assimilation iterations, for the cases of **(A,D)** no model error; **(B,E)** model error present but not accounted for; and **(C,F)** model error present and accounted for using Algorithm 2 with 40 detailed solver calculations and 40 KNN. Added to the box plots are the mean (circles), and minimum and maximum values (crosses). The dashed horizontal line in **(A–C)** represents the expected value of the travel-time misfit assuming that the residuals follow the prescribed Gaussian distribution for the data errors.

model, our approach importantly aims to identify the model-error component of the residual during the ES-MDA procedure, which is used to correct the proxy forward response. This is accomplished through construction of an orthonormal model-error basis for each ensemble member and at each iteration based on a prescribed number of KNN entries selected from a model-error dictionary. The latter is created as the inversion proceeds, and thus no prior information about the model error is required before running the procedure.

With regard to the considered example problem of estimating the spatial distribution of subsurface slowness from crosshole GPR travel times, we saw that our modified ES-MDA approach allows us to obtain accurate posterior estimates characteristic of large ensembles with a computational cost comparable to a small number of runs of the detailed forward solver. The results did show, however, that the success of the approach depends on the ratio between the number of ensemble members and the number of detailed solver runs per iteration used to learn about the model error. In particular, for the crosshole GPR tomographic example considered, this ratio should not exceed a value of approximately 8.

Despite the successful application of our model-error approach, there remain a number of topics that should be investigated further. For example, in the work presented here, we set the number of KNN equal to the number of detailed solver runs per ES-MDA iteration used to learn about the model

error. In this case, the same model-error basis is constructed in the first iteration for each ensemble member. In subsequent iterations and with a growing dictionary, for each ensemble member the KNN are used to extract local information about the model error from the model-error dictionary and a local model-error basis is calculated, respectively. However, this choice could be validated and optimized to improve the identification of the model-error component of the residual. Further, we assume with our method that the latter is approximately orthogonal to the data-measurement and parameter-error components, which allows for its identification using a projection approach. Although we have found this assumption to yield acceptable results for a range of test problems examined so far, it requires further investigation.

AUTHOR CONTRIBUTIONS

CK: developing ideas, setting up the codes and writing the manuscript; AE: discussing ideas and revising the manuscript; JI: discussing ideas, revising the manuscript, and obtaining funding for the research.

FUNDING

This work was supported by a research grant to JI from the Swiss National Science Foundation (number 200021_140864).

REFERENCES

- Aanonsen, S. I., Nævdal, G., Oliver, D. S., Reynolds, A. C. and Vallès, B. (2009). The ensemble Kalman filter in reservoir engineering—a review. *Spe J.* 14, 393–412. doi: 10.2118/117274-PA
- Arridge, S., Kaipio, J., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T., et al. (2006). Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Prob.* 22:175. doi: 10.1088/0266-5611/22/1/010
- Beskos, A., Girolami, M., Lan, S., Farrell, P. E., and Stuart, A. M. (2017). Geometric MCMC for infinite-dimensional inverse problems. *J. Comput. Phys.* 335, 327–351. doi: 10.1016/j.jcp.2016.12.041
- Buizza, R., and Palmer, T. N. (1998). Impact of ensemble size on ensemble prediction. *Month. Weather Rev.* 126, 2503–2518. doi: 10.1175/1520-0493(1998)126<2503:IOESOE>2.0.CO;2
- Calvetti, D., Ernst, O., and Somersalo, E. (2014). Dynamic updating of numerical model discrepancy using sequential sampling. *Inverse Prob.* 30:114019. doi: 10.1088/0266-5611/30/11/114019
- Chen, Y., Keyes, D., Law, K. J., and Ltaief, H. (2016). Accelerated dimension-independent adaptive Metropolis. *SIAM J. Sci. Comput.* 38, S539–S565. doi: 10.1137/15M1026432
- Chen, Y., and Zhang, D. (2006). Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Adv. Water Resour.* 29, 1107–1122. doi: 10.1016/j.advwatres.2005.09.007
- Cordua, K. S., Looms, M. C., and Nielsen, L. (2008). Accounting for correlated data errors during inversion of cross-borehole ground penetrating radar data. *Vadose Zone J.* 7, 263–271. doi: 10.2136/vzj2007.0008
- Cotter, S. L., Roberts, G. O., Stuart, A. M., and White, D. (2013). MCMC methods for functions: modifying old algorithms to make them faster. *Stat. Sci.* 28, 424–446. doi: 10.1214/13-STS421
- Cui, T., Fox, C., and O'Sullivan, M. J. (2018). Adaptive approximation error models for efficient uncertainty quantification with application to multiphase subsurface fluid flow. *arXiv:1809.03176*.
- Deutsch, C. V., and Journel, A. G. (1992). *GSLIB: Geostatistical Software Library and User's Guide*. New York, NY: Oxford University Press.
- Elsheikh, A. H., Wheeler, M. F., and Hoteit, I. (2013). An iterative stochastic ensemble method for parameter estimation of subsurface flow models. *J. Comput. Phys.* 242, 696–714. doi: 10.1016/j.jcp.2013.01.047
- Emerick, A. A., and Reynolds, A. C. (2012a). Ensemble smoother with multiple data assimilation. *Comput. Geosci.* 55, 3–15. doi: 10.1016/j.cageo.2012.03.011
- Emerick, A. A., and Reynolds, A. C. (2012b). History matching time-lapse seismic data using the ensemble kalman filter with multiple data assimilations. *Comput. Geosci.* 16, 639–659. doi: 10.1007/s10596-012-9275-5
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* 99, 10143–10162. doi: 10.1029/94JC00572
- Evensen, G. (2007). *Data Assimilation: The Ensemble Kalman Filter*. Heidelberg: Springer Science & Business Media.
- Evin, G., Thyer, M., Kavetski, D., McInerney, D., and Kuczera, G. (2014). Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resour. Res.* 50, 2350–2375. doi: 10.1002/2013WR014185
- Giroux, B., Gloaguen, E., and Chouteau, M. (2007). bh_tomoa Matlab borehole georadar 2D tomography package. *Comput. Geosci.* 33, 126–137. doi: 10.1016/j.cageo.2006.05.014
- Goh, J., Bingham, D., Holloway, J. P., Grosskopf, M. J., Kuranz, C. C., and Rutter, E. (2013). Prediction and computer model calibration using outputs from multifidelity simulators. *Technometrics* 55, 501–512. doi: 10.1080/00401706.2013.838910
- Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* 7, 223–242. doi: 10.2307/3318737
- Hansen, T. M., Cordua, K. S., Jacobsen, B. H., and Mosegaard, K. (2014). Accounting for imperfect forward modeling in geophysical inverse problems - Exemplified for crosshole tomography. *Geophysics* 79, H1–H21. doi: 10.1190/geo2013-0215.1
- Hansen, T. M., Cordua, K. S., Looms, M. C., and Mosegaard, K. (2013). Sippi: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: part 2—application to crosshole GPR tomography. *Comput. Geosci.* 52, 481–492. doi: 10.1016/j.cageo.2012.10.001
- Hansen, T. M., Cordua, K. S., and Mosegaard, K. (2012). Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling. *Comput. Geosci.* 16, 593–611. doi: 10.1007/s10596-011-9271-1
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning, Data Mining, Inference and Prediction, 2nd Edn.* New York, NY: Springer Series in statistics.
- Josset, L., Demyanov, V., Elsheikh, A. H., and Lunati, I. (2015a). Accelerating Monte Carlo Markov chains with proxy and error models. *Comput. Geosci.* 85, 38–48. doi: 10.1016/j.cageo.2015.07.003
- Josset, L., Ginsbourger, D., and Lunati, I. (2015b). Functional error modeling for uncertainty quantification in hydrogeology. *Water Resour. Res.* 51, 1050–1068. doi: 10.1002/2014WR016028
- Kaipio, J., and Somersalo, E. (2007). Statistical inverse problems: discretization, model reduction and inverse crimes. *J. Comput. Appl. Math.* 198, 493–504. doi: 10.1016/j.cam.2005.09.027
- Kennedy, M. C., and O'Hagan, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 63, 425–464. doi: 10.1111/1467-9868.00294
- Khu, S.-T., and Werner, M. G. (2003). Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling. *Hydrol. Earth Syst. Sci.* 7, 680–692. doi: 10.5194/hess-7-680-2003
- Köpke, C., Irving, J., and Elsheikh, A. H. (2018). Accounting for model error in Bayesian solutions to hydrogeophysical inverse problems using a local basis approach. *Adv. Water Resour.* 116, 195–207. doi: 10.1016/j.advwatres.2017.11.013
- Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., and Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resour. Res.* 49, 2664–2682. doi: 10.1002/wrcr.20226
- Lehikoinen, A., Huttunen, J. M. J., Finsterle, S., Kowalsky, M. B., and Kaipio, J. P. (2010). Dynamic inversion for hydrological process monitoring with electrical resistance tomography under model uncertainties. *Water Resour. Res.* 46:W04513. doi: 10.1029/2009WR008470
- Linde, N., and Vrugt, J. A. (2013). Distributed soil moisture from crosshole ground-penetrating radar travel times using stochastic inversion. *Vadose Zone J.* 12:vzj2012.0101. doi: 10.2136/vzj2012.0101
- Looms, M. C., Binley, A., Jensen, K. H., and Nielsen, L. (2008). Identifying unsaturated hydraulic parameters using an integrated data fusion approach on cross-borehole geophysical data. *Vadose Zone J.* 7, 238–248. doi: 10.2136/vzj2007.0087
- Marzouk, Y., and Xiu, D. (2009). A stochastic collocation approach to Bayesian inference in inverse problems. *Commun. Comput. Phys.* 6, 826–847. doi: 10.4208/cicp.2009.v6.p826
- Nowack, R. L. (1992). Wavefronts and solutions of the eikonal equation. *Geophys. J. Int.* 110, 55–62. doi: 10.1111/j.1365-246X.1992.tb00712.x
- Oliver, D. S., and Alfonzo, M. (2018). Calibration of imperfect models to biased observations. *Comput. Geosci.* 22, 145–161. doi: 10.1007/s10596-017-9678-4
- Piccolo, C., and Cullen, M. (2016). Ensemble data assimilation using a unified representation of model error. *Month. Weather Rev.* 144, 213–224. doi: 10.1175/MWR-D-15-0270.1
- Rasmussen, C. E., and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Cambridge: MIT Press.
- Reynolds, A. C., Zafari, M., and Li, G. (2006). “Iterative forms of the ensemble kalman filter,” in *ECMOR X-10th European Conference on the Mathematics of Oil Recovery*, A030. Amsterdam.
- Robert, C. P., and Casella, G. (2004). *Monte Carlo Statistical Methods*. New York, NY: Springer.
- Ruggeri, P., Irving, J., and Holliger, K. (2015). Systematic evaluation of sequential geostatistical resampling within MCMC for posterior sampling of near-surface geophysical inverse problems. *Geophys. J. Int.* 202, 961–975. doi: 10.1093/gji/ggv196
- Scholer, M., Irving, J., Looms, M. C., Nielsen, L., and Holliger, K. (2012). Bayesian Markov-chain-monte-carlo inversion of time-lapse crosshole GPR data to characterize the vadose zone at the arrenaes site, Denmark. *Vadose Zone J.* 11. doi: 10.2136/vzj2011.0153

- Schoups, G., and Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resour. Res.* 46:W10531. doi: 10.1029/2009WR008933
- Smith, T., Marshall, L., and Sharma, A. (2015). Modeling residual hydrologic errors with Bayesian inference. *J. Hydrol.* 528, 29–37. doi: 10.1016/j.jhydrol.2015.05.051
- Stordal, A. S., and Elsheikh, A. H. (2015). Iterative ensemble smoothers in the annealed importance sampling framework. *Adv. Water Resour.* 86, 231–239. doi: 10.1016/j.advwatres.2015.09.030
- Strang, G., Strang, G., Strang, G., and Strang, G. (1993). *Introduction to Linear Algebra*, Vol. 3. Wellesley, MA: Wellesley-Cambridge Press.
- Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: SIAM.
- van Leeuwen, P. J. (2001). An ensemble smoother with error estimates. *Month. Weather Rev.* 129, 709–728. doi: 10.1175/1520-0493(2001)129<0709:AESWEE>2.0.CO;2
- Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: theory, concepts, and MATLAB implementation. *Environ. Modell. Softw.* 75, 273–316. doi: 10.1016/j.envsoft.2015.08.013
- Xu, T., Valocchi, A. J., Choi, J., and Amir, E. (2014). Use of machine learning methods to reduce predictive error of groundwater models. *Groundwater* 52, 448–460. doi: 10.1111/gwat.12061

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Köpke, Elsheikh and Irving. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership