https://doi.org/10.35662/unine-thesis-2927

UNIVERSITY OF NEUCHÂTEL, SWITZERLAND

FACULTY of SCIENCE

CENTER FOR HYDROGEOLOGY AND GEOTHERMICS

CHYN

The Traveling Pilot Point Method for Solving Groundwater Inverse Problems in Aquifers with Categorical Parameter Distributions

---

A thesis presented for the degree of

DOCTOR of SCIENCES

by

Prashanth KHAMBHAMMETTU

EXAMINING COMMITTEE

Prof. Philippe RENARD

Prof. John DOHERTY

Prof. Sanjay SRINIVASAN

Dr. Mickaele LE RAVALEC

Dr. Julien STRAUBHAAR

Defended on *SEPTEMBER* 13^{th}, 2021

UNIVERSITÉ DE NEUCHÂTEL

Faculté des Sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel – Suisse
Tél : + 41 (0)32 718 21 00
E-mail : secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Monsieur Prashanth KHAMBHAMMETTU

Titre:

“The Travelling Pilot Point Method for solving groundwater inverse problems in aquifers with categorical distributions”

sur le rapport des membres du jury composé comme suit:

- Prof. Philippe Renard, directeur de thèse, Université de Neuchâtel, Suisse
- Dr Julien Straubhaar, Université de Neuchâtel, Suisse
- Prof. John Doherty, Flinders University, Watermark Numerical Computing, Australia
- Dr Mickaele Le Ravalec, IFP-EN, Rueil-Malmaison, France
- Prof. Sanjay Srinivasan, PennState University, USA

Neuchâtel, le 28 octobre 2021

Le Doyen, Prof. A. Bangerter

"If I have seen further than others, it is by standing upon the shoulders of giants."

Sir Isaac Newton

Abstract

Categorical parameter distributions are commonplace in hydrogeological systems consisting of geologic facies/categories with distinct properties, e.g., high-permeability channels embedded in a low-permeability matrix. Contaminated sites are often underlain by aquifers with distinct geological facies. Identifying low-permeability facies is very important at these sites because these facies store solute mass and act as secondary sources to higher-permeability facies, sustaining concentrations for decades while increasing risk and life-cycle costs. Parameter estimation is difficult in such systems because the discontinuities in the parameter space hinder the inverse problem. Previous research in this area has been focused on the use of stochastic methods. In this thesis, a novel approach based on Traveling Pilot points (TRIPS) to solve the categorical inverse problem is presented. In traditional implementations, aquifer properties (e.g., hydraulic conductivity) are estimated at fixed pilot point locations. In the TRIPS implementation, both the properties associated with the pilot points and their locations are estimated. Tikhonov regularization constraints are incorporated in the parameter estimation process to produce realistic parameter depictions. The TRIPS framework is alternatively combined with the Null Space Monte Carlo (NSMC) method and the pestpp-ies ensemble smoother to solve the categorical inverse problem for a hypothetical aquifer. While both NSMC and pestpp-ies produced posterior ensembles with similarities to an ensemble estimated using Rejection Sampling (RS), pestpp-ies was able to sample the posterior distribution with lesser number of forward run evaluations in a more comprehensive

manner. Model selection techniques, a combination of multi-dimensional scaling and K-Means clustering, were used to create a smaller but strategically diverse prior ensemble that when smoothed produced a posterior ensemble with properties similar to that of a larger posterior ensemble. Additionally, TRIPS and pestpp-ies together were used to develop categorical parameter ensembles that honor measured aquifer heads and concentrations simultaneously. The results indicated that even with an approximate geological prior model, a high degree of parameterization and history matching can lead to parameter ensembles that can be useful for making certain types of predictions (example: concentration predictions). However, for more demanding predictions (example: mass), an approximate geological prior is not adequate. The analysis was used to demonstrate how a framework with multiple puzzle pieces (geological parameterization, history matching, and remedial forecasts) could be efficiently assembled to guide decision makers at contaminated sites by quantifying the predictive uncertainty associated with parameter uncertainty. By pivoting from predictive models based on a single calibrated model to an ensemble-based approach, decision-makers can quantify uncertainty and take pragmatic decisions. Data worth analyses that can guide future data collection efforts could also be integrated into the framework to better future outcomes.

Keywords

categorical parameters; inverse problem; subspace methods; traveling pilot points; null space monte carlo; iterative ensemble smoothers; groundwater flow; solute transport; multi-dimensional scaling; k-means clustering; remediation; predictive uncertainty;

Résumé

Les distributions discrètes de paramètres sont courantes dans les systèmes hydrogéologiques constitués de faciès/catégories géologiques aux propriétés distinctes, par exemple, des chenaux de forte perméabilité noyés dans une matrice de faible perméabilité. Par exemple, les sites contaminés sont souvent constitués d'aquifères aux faciès géologiques distincts. L'identification des faciès de faible perméabilité est très importante sur ces sites car ces faciès stockent une partie de la masse de solutés et agissent comme des sources secondaires pour les faciès à perméabilité plus élevée, maintenant les concentrations pendant des décennies tout en augmentant le risque et les coûts de décontamination. L'estimation de ces paramètres est difficile car les discontinuités dans l'espace des paramètres entravent le problème inverse. Les recherches précédentes dans ce domaine se sont concentrées sur l'utilisation de méthodes stochastiques.

Dans cette thèse, une nouvelle approche basée sur des points pilotes voyageurs (TRIPS) pour résoudre le problème inverse catégorique est présentée. Dans les implémentations traditionnelles, les propriétés de l'aquifère (par exemple, la conductivité hydraulique) sont estimées aux emplacements fixes des points pilotes. Dans l'implémentation TRIPS, les propriétés associées aux points pilotes et leurs emplacements sont estimés. Des contraintes de régularisation de Tikhonov sont incorporées dans le processus d'estimation des paramètres pour produire des représentations réalistes des paramètres. Le cadre TRIPS est alternativement combiné avec la méthode Null Space Monte Carlo (NSMC) et le filtre d'ensemble pestpp-ies pour résoudre le problème inverse catégorique pour un aquifère hypothétique. Alors que la méthode NSMC et la méthode

pestpp-ies ont produit des ensembles a posteriori similaires à un ensemble estimé à l'aide de l'échantillonnage par rejet (RS), la méthode pestpp-ies a été capable d'échantillonner la distribution a posteriori avec un nombre inférieur d'évaluations, et ce de manière plus complète. Des techniques de sélection de modèles ont été utilisées pour créer un ensemble a priori plus petit mais stratégiquement diversifié qui, une fois lissé, a produit un ensemble a posteriori avec des propriétés similaires à celles d'un ensemble a posteriori plus grand. De plus, TRIPS et pestpp-ies ont été utilisés ensemble pour développer des ensembles de paramètres catégoriques qui honorent simultanément les charges hydrauliques et les concentrations mesurées dans les aquifères. Les résultats indiquent que même avec un modèle géologique préalable approximatif, un haut degré de paramétrisation et de correspondance de l'historique peut conduire à des ensembles de paramètres qui peuvent être utiles pour faire certains types de prédictions (exemple : prédictions de concentration). Cependant, pour des prédictions plus exigeantes (exemple : masse), un modèle géologique préalable approximatif n'est pas adéquat. L'analyse a été utilisée pour démontrer comment un cadre comportant plusieurs pièces de puzzle (paramétrage géologique, ajustemen des données historiques et prévisions de remédiation) pouvait être assemblé efficacement pour guider les décideurs sur les sites contaminés en quantifiant l'incertitude prédictive associée à l'incertitude des paramètres. En passant des modèles prédictifs basés sur un seul modèle calibré à une approche basée sur un ensemble, les décideurs peuvent quantifier l'incertitude et prendre des décisions pragmatiques. Les analyses de la valeur des données qui peuvent guider les futurs efforts de collecte de données pourraient également être intégrées au cadre afin d'améliorer les résultats futurs.

Mots clés

paramètres catégoriques; problème inverse; méthodes de sous-espace; points pilotes voyageurs; monte carlo à espace nul ; lisseurs d'ensemble itératifs ; écoulement des eaux souterraines ; transport de solutés ; mise à l'échelle multidimensionnelle ; regroupement par k-means ; assainissement ; incertitude prédictive;

Acknowledgements

This thesis has been one of the hardest things I have attempted in my life. Luckily, I wasn't alone in my journey.

I am deeply indebted to my advisors: Dr. Philippe Renard and Dr. John Doherty. Philippe and John did not shy away from working with an unconventional student, an Ocean or two away. Over the years, they have been extremely generous with their time, ideas, and mentorship. Their passion, constant encouragement, and willingness to entertain discussions at unorthodox times contributed a lot to the progression of this research. I would also like to thank my doctoral committee members, Dr. Mickaele Le Ravalec, Dr. Sanjay Srinivasan, and Dr. Julien Straubhaar for their valuable suggestions and insights.

Jeremy White has been a great friend, mentor, and coach. I learned a lot about ensemble smoothers from him. I would also like to thank Sorab Panday, Vivek Bedekar, Matt Tonkin, Chris Neville, Charles Andrews, Steve Larson, Marc Killingstad, Scott Potter, Mike Kladias, Matthew McCaughey, Jonathan Roller, Jack Wang, Fred Lont, Joseph Hughes, Christian Langevin, and others who have inspired me with their ideas/software, and aided me in this journey. I am also grateful to my former and current employers (S.S. Papadopulos Associates and Arcadis) for letting me use computing resources and devote some of my time to research.

Pursuing this research while balancing the responsibilities of a full-time job and family was very daunting. However, my wife, Subhadra Ponnada, has been my rock in this journey. This thesis would have never seen the light of the day without her unwavering love, unconditional support, encouragement, and positive attitude. My 7-year old daughter Ishaanvi, with her periodic "How to feel better and smile" cards, has been a great cheerleader.

I owe a lot to my parents (Gopalakrishna and Jayasree) for their love and emphasis on the importance of education and constant learning. My mother, who completed her Ph.D. at nearly the same age as me today, is a huge inspiration. My in-laws (Subrahmanyam and Padmavathi) have been a second set of parents with their wishes and encouragement. My brother Sumanth's help during my preparation for the defense was invaluable.

There are many others (family, friends, and colleagues) who have been with me throughout this journey. Thank you for being there and I am glad I wasn't alone. Finally, I owe a special thanks to Starbucks and Peets coffee, where I wrote a lot of this thesis while imbibing gallons of coffee.

Contents

List of Abbreviations

EnKF	**E**nsemble **K**alman **F**ilter
ES	**E**nsemble **S**moother
IES	**I**terative **E**nsemble **S**moother
ISR	**I**terative **S**patial **R**esampling
GDM	**G**radual **D**eformation **M**ethod
MODFLOW	**M**odular Groundwater **FLOW**
MPS	**M**ultiple-**P**oint **S**tatistics
NSMC	**N**ull **S**pace **M**onte **C**arlo
POPEX	**PO**sterior **P**opulation **EX**pansion
PC	**P**osterior **C**ovariance Estimation
PPM	**P**robability **P**erturbation **M**ethod
PEST	**P**arameter **EST**imation Software
RS	**R**ejection **S**ampling
TRIPS	**TR**avel**I**ng **P**ilot Point**S**

Dedicated to the memory of Dr. G.V. Loganathan - an exemplary teacher, mentor, and human being.

Chapter 1

Introduction

1.1 Context and Motivation

Groundwater flow and contaminant transport models are commonly used to answer questions pertaining, for example, to groundwater management and contaminant migration. According to Anderson, Woessner, and Hunt (2015), groundwater models provide a quantitative framework for synthesizing field information and conceptualizing hydrogeological processes so that they can answer questions about groundwater flow, well field capture, and contaminant migration. These models solve the forward problem to answer the question under investigation. The forward problem involves model parametrization followed by solving a partial differential equation to obtain a state vector $\mathbf{d}$ (representing, for example, the groundwater head or contaminant concentration) in response to specified boundary conditions. The inverse problem, on the other hand, involves identifying the model parameter vector $\mathbf{m}$ from the state vector $\mathbf{d}$.

Inverse problems in the groundwater modeling context have been studied extensively. Le Ravalec (2005) presents a comprehensive overview of solving inverse problems for modeling flow in porous media. Zhou, Gómez-Hernández,

and Li (2014) present a slightly more recent detailed discussion of the groundwater inversion problem and a review of historical and modern methods. The probabilistic formulation of the inverse problem (see, for example: Aster, Borchers, and Thurber 2013) can be expressed in terms of conditional probabilities as shown in equation (1.1).

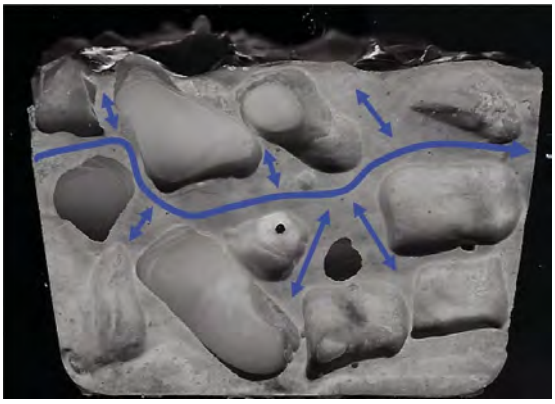
$$q(\mathbf{m} \mid \mathbf{d}) = \frac{f(\mathbf{d} \mid \mathbf{m}) p(\mathbf{m})}{c} \tag{1.1}$$

where the term $q(\mathbf{m} \mid \mathbf{d})$ is the posterior probability density function and represents the probability of occurrence of a parameter vector conditioned by the observed measured dataset. The term $p(\mathbf{m})$ is known as the prior and represents the probability of occurrence of any model based only on the initial information such as geological knowledge without considering the measurements of the state variables $\mathbf{d}$. The term $f(\mathbf{d} \mid \mathbf{m})$ known as the likelihood represents the probability of simulating the measured data $\mathbf{d}$ given a model vector $\mathbf{m}$.

The categorical inverse problem is a special case of the groundwater inverse problem, pertaining to aquifers that consist of discrete geological facies/categories. For example, consider a two-categories aquifer with fluvial high-permeability channels incised in a low-permeability matrix. At any location in this aquifer, we would find only one of the two facies - channel or matrix. The inverse problem, in this case, requires us to generate categorical aquifer distributions when presented with prior geologic information about borehole logs (static data) and measurements of aquifer state (e.g., groundwater heads).

Aquifers with categorical facies are abundant in nature and thus of great interest to researchers and practitioners. Classic examples include alluvial aquifers with sandy channel deposits in silt/clay and fractured limestone aquifers. For many of these cases, the variability between facies dominates the variability

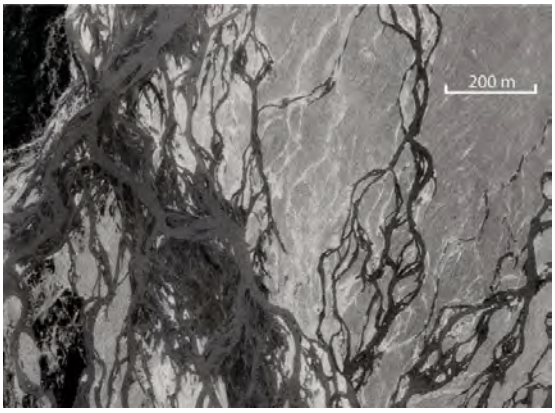
within the facies and could effectively be considered categorical. A few examples of categorical facies ranging from the pore scale to field scale are shown in Figure 1.1.



(A) Pore cross-section

(B) Sand and fine gravel

(C) Fractured limestone



(D) Braided channels

FIGURE 1.1: Examples of categorical geology in the real world. A) Physical pore cross-section showing flow channels and connecting dead-end pore spaces that provide solute mass storage. B) Poorly sorted sand and fine gravel (Quebec, Canada). C) Limestone epikarst in a dimension-stone quarry (Indiana, USA). D) Braided stream channels on the Delta River (Alaska, USA). Source: Payne, Quinnan, and Potter (2008)

Categorical inverse problems are often more challenging to solve than their continuous counterparts because the parameter space is discontinuous. Categorical inverse problems pertaining to contaminant migration are even more challenging because the solution would need to unravel both flow and transport

processes. Linde et al. (2017) presented an extensive review of existing methods for categorical inverse problems. We summarize a few of them here.

The gradual deformation method (GDM) generates a sequence of model realizations that converge to matching the measured data (Hu, Blanc, and Noetinger 2001). The key underlying concept in GDM is that the linear combinations of multiGaussian fields are also multiGaussian fields with similar statistics. It is, therefore, possible to explore a part of the model space by adjusting only one single parameter: a weight allowing to move between two pre-computed simulations. If the categorical field is obtained by truncation of one or several multiGaussian realizations, this process is straightforward and obtaining a discrete model that matches the measured observations can be treated as a usual continuous optimization problem.

Caers and Hoffman (2006) proposed the Probability Perturbation Method (PPM) when dealing with non-Gaussian priors and non-linear forward model responses. Rather than computing the posterior from the prior and likelihood, they instead decompose the posterior into a set of pre-posterior distributions containing facies and measurement data respectively. These pre-posterior probability distributions are perturbed until newer model realizations in the sequence increasingly converge to matching measurements. Ronayne, Gorelick, and Caers (2008) applied the PPM to a transient aquifer test model and generated a distribution of permeable discrete channels embedded within less permeable deposits.

Several authors developed an iterative Blocking Moving Window algorithm in conjunction with simulated annealing or a Markov chain Monte Carlo based method to guide a multiple-point statistics (MPS) model in reproducing state variables and honor facies data known from prior knowledge (Alcolea and Renard 2010; Hansen, Cordua, and Mosegaard 2012). Mariethoz, Renard, and

Caers (2010) proposed the Iterative Spatial Resampling (ISR) technique, a Markov chain-based method to sample from the posterior distribution. The transition from one element to the next in the Markov chain is based on sampling the values of the previous field at a set of random locations and using these points as conditioning data for the next iteration. While this procedure is straightforward and samples the posterior space in an unbiased manner, it is rather time-consuming. Jäggli, Straubhaar, and Renard (2017) proposed a faster approach named posterior population expansion (POPEX). The method is based on an adaptive importance sampling strategy expanding an initial ensemble of parameter models using MPS and local conditioning in such a manner that the new models are likely to belong to the posterior population. The POPEX approach was subsequently modified (Jäggli, Straubhaar, and Renard 2018) to overcome some biases in the posterior distribution by adapting the computation of the weights. More recently the approach was combined with Machine Learning technique to accelerate the algorithm (Juda and Renard 2021).

Several approaches to solve the categorical inverse problem based on pilot points have also been presented. Pilot points have been used to estimate heterogeneous hydrogeological parameter distributions for several decades (Certes and Marsily 1991; Lavenue 2001; Doherty 2003; Doherty and Hunt 2010). Pilot points are treated as surrogate parameters in the inverse modeling process for representing heterogeneity in a lower-dimensional space. In these applications, a location-specific hydrogeological attribute (e.g., porosity, hydraulic conductivity) is associated with the pilot point. A pre-determined number of pilot points are placed at strategic locations along the model domain to capture the heterogeneity in the system. An iteration of the forward problem involves estimation of properties associated with each pilot point followed by spatial interpolation to create a spatially continuous parameter distribution from the discrete pilot

point locations. The inverse problem involves the estimation of parameter values at the pilot point locations that honor the calibration constraints. Over the course of the parameter estimation, the locations of the pilot points remain static, but the parameters associated with the pilot points change. In the context of categorical fields, Li et al. (2013) used pilot points to guide an ensemble Kalman Filter approach to match dynamic (head) and geologic data simultaneously.

1.2 Scope of the Current Research

In this thesis, we develop and test a new approach where pilot points are used to parametrize the categorical inverse problem. In our approach, we use pilot points in a non-traditional manner, that we refer to as the "Traveling Pilot Points (TRIPS)" approach. We iteratively adapt their positions to define the geometries of the discrete categories. In other words, instead of adjusting the value of the hydraulic conductivity or any other physical parameter at the fixed location of the pilot points and then interpolating, we fix the category at the pilot points and adjust their locations before generating the categorical field. The new unknowns in the inverse problem formulation are the locations of the pilot point.

In the subsequent chapters, we describe this approach more precisely and explore if it could be used to solve the inverse problem for discrete/categorical parameter fields. In our opinion, there are two advantages to this approach. First, by using the positions of the pilot points, the categorical problem has been restated as a problem with continuous parameters which is easier to solve. Second, this approach allows us to infer the category geometries rather than to estimate them from spatial interpolation operations such as kriging indirectly.

The methodology described in this thesis has been tested so far on synthetic

problems with two facies. The tests considered both flow-only and flow-and-transport cases. The proposed method might require additional modifications for more complex problems with multiple facies.

1.3 Organization of the Thesis

This thesis is structured into three main chapters as follows:

- The TRIPS method for parametrizing the categorical inverse problem is described in chapter 2. In this chapter, TRIPS are used in conjunction with subspace methods (null space monte carlo (NSMC), posterior covariance estimation (PC)) to match measured heads and estimate posterior parameter ensembles for a categorical aquifer with two facies - permeable channels incised in a low-permeability matrix. The estimated posterior ensembles are compared against those estimated independently using rejection sampling (RS) - a surrogate for the "true posterior".

- In chapter 3, TRIPS are used in conjunction with an iterative ensemble smoother (IES) to estimate posterior parameter ensembles for the synthetic problem described in chapter 2. The resulting ensembles are compared against those estimated by other methods (NSMC, PC, RS). Prior parameter ensembles of different sizes are "smoothed" to understand how the estimated posterior ensemble depends on the size of the prior ensemble. Multi-dimensional scaling and K-Means clustering are used to demonstrate that it is possible to strategically select a smaller prior ensemble such that the resulting posterior ensemble would still have the statistical properties of a much larger posterior ensemble.

- In chapter 4, the TRIPS method in conjunction with IES are used to quantify predictive model uncertainty in solute concentration and mass forecasts at a hypothetical site with a categorical aquifer. The efficacy of the TRIPS and IES methods to assimilate hydraulic head and solute concentration measurements to estimate ensembles with categorical facies is demonstrated. The impacts of incorrectly assuming the aquifer heterogeneity to be continuous instead of categorical on the concentration and mass forecasts are investigated. Additionally, the size of the measurement dataset and its impact on the predictive forecasts is also discussed.

- Chapter 5 concludes the thesis by summarizing the key contributions and presents some perspectives for future research.

There are also three (3) appendices that provide supplemental information:

- Appendix A has the parameter mean values and the covariance matrix for the prior distribution used in chapters 2 and 3. This appendix also has a formula for estimating the covariance matrix of parameter differences.

- Appendix B pertains to chapter 3 and has a listing and explanation of the pestpp-ies control flags for the simulation with an initial ensemble of 10000.

- Appendix C has the parameter mean values and the covariance matrix for the prior distribution used in chapter 4.

Chapter 2

The Traveling Pilot Points Method*

2.1 TRIPS Methodology

In the following sections, we introduce the principle and present the details of the implementation of Traveling Pilot Points (TRIPS) to solve the categorical inverse problem. In contrast to traditional pilot points, TRIPS are not fixed in location but instead can travel to locations of interest in the model domain.

2.1.1 The Traveling Pilot Points Principle

Let us consider an aquifer containing permeable channels embedded within an impermeable matrix. It is possible to generate such discrete geological fields with different geostatistical techniques. One could use for example, transition probabilities (TProGS), plurigaussian simulations, object-based models, or multiple-point statistics to model these structures. For all these techniques, it is possible

*This chapter is based on Prashanth Khambhammettu, Philippe Renard, and John Doherty (2020). "The Traveling Pilot Point method. A novel approach to parameterize the inverse problem for categorical fields". In: *Advances in Water Resources* 138, p. 103556. ISSN: 0309-1708. DOI: https://doi.org/10.1016/j.advwatres.2020.103556. URL: https://www.sciencedirect.com/science/article/pii/S0309170819308401

to set a fixed number of locations where the type of geology is known (for example presence of a channel), but the locations themselves are unknown. Providing these locations as conditioning data to the geological simulation algorithm allows to change the parametrization of the geological simulation and to solve the inverse problem in this manner means to search for the optimal locations of these traveling pilot points. This approach modifies a discrete inverse problem into a continuous one and should, therefore, facilitate its resolution. This idea is very general and can have many applications.

2.1.2 An example of geological model

To test this idea in a simple situation, we consider a binary case with channels as illustrated in Figure 2.1. To constrain the geometry of the channels in a simple manner, we use a two-step approach based on object-based simulations constrained by a training image. On the one hand, the training image (Figure 2.1) provides in a graphical manner the size of the channels, their sinuosity, their spacing, and so on. This image can be drawn by hand based on a geological concept. It offers flexibility and simplicity. On the other hand, the object-based model ensures that all the channels are continuous and that the geological models are generated very rapidly.

For the object-based model, we consider that there is a fixed number of channels crossing the area from left to right (Figure 2.2). For each channel, we define a fixed number of traveling pilot points. For example, for a channel spanning an X distance of 100 m (100 meters), we can characterize it by 5 points spaced 20 m apart. If the aquifer domain is 100 m $\times$ 100 m with a typical distribution of 3 channels, 15 points are used to track all the channels. To simulate the entire domain, the channel central lines are interpolated with a spline function using

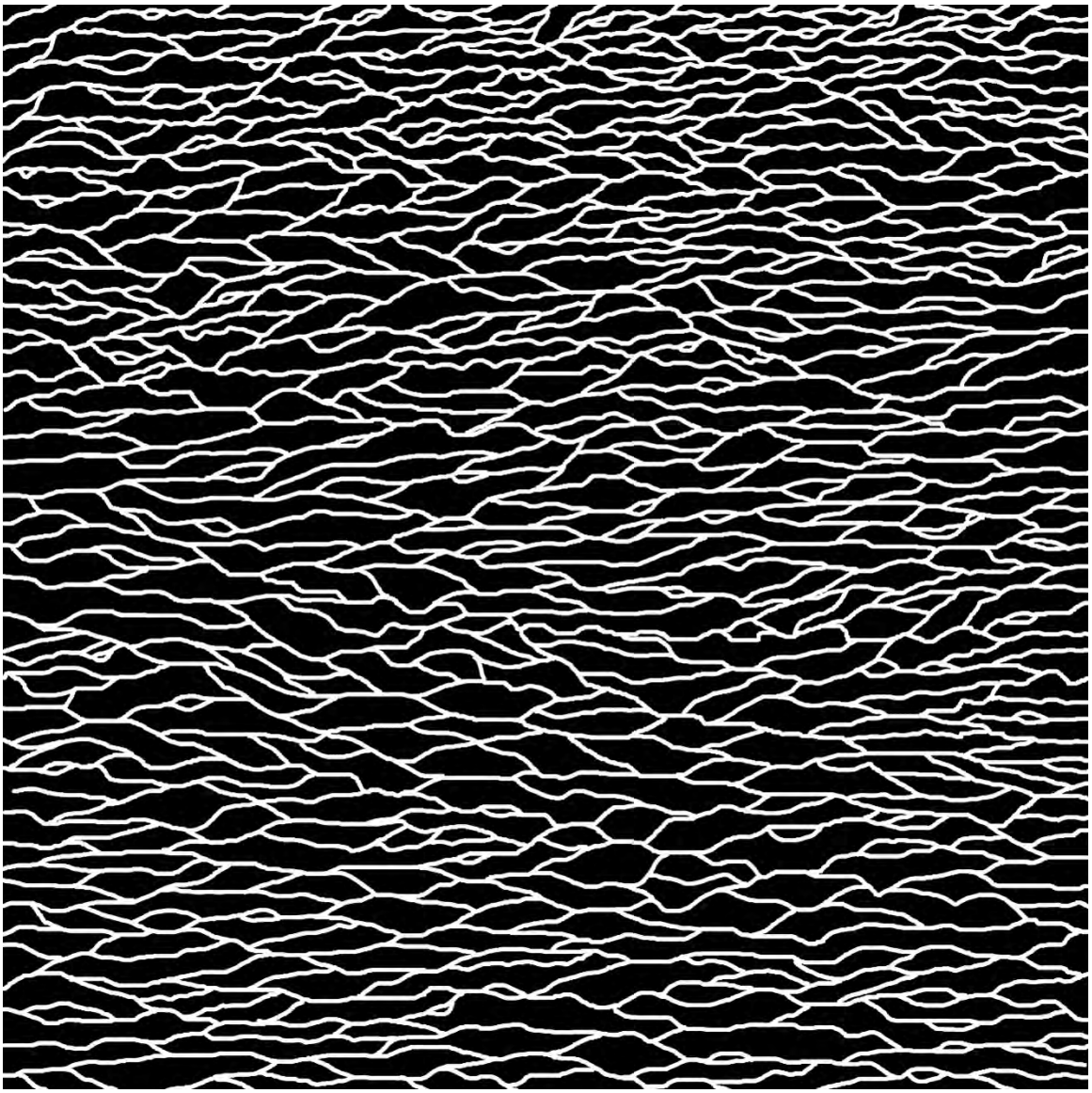


FIGURE 2.1: Training Image representative of the geology in the synthetic aquifer. The image has a size of 2500 by 2500 pixels. The black pixels represent the matrix, while the white represent the channels. This image is borrowed from Laloy et al. 2018

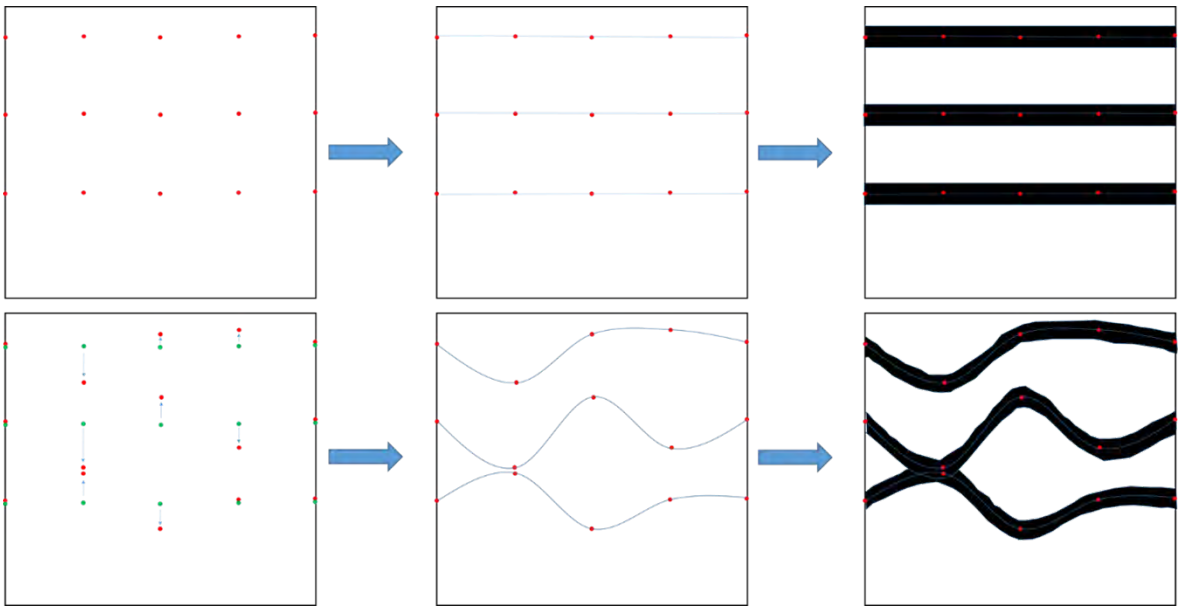


FIGURE 2.2: Development of object-based model from pilot points with (top) initial position of TRIPS and (bottom) updated positions of TRIPS. Splines connecting related TRIPS are shown in the central panel and the channel objects created by buffering the splines are shown in the right panel.

the position of the traveling pilot points as input. Then a constant thickness is applied along the central lines, and all pixels falling within this area are labeled as channel. We then have a simple function that relates the pilot point positions to channel geometry. To constrain the geometry of those channels in a simple manner and make the link with the training image, we assumed that we could reproduce reasonably well the variability of the channels, their shapes and their relative positions using a multiGaussian distribution of the position of the traveling pilot points. In this manner, the prior geological model is fully determined by a set of mean values and a prior covariance matrix.

To estimate the prior covariance matrix, the training image has been cut into many sub-images having the same lateral extension as the simulation domain. The vertical extension was taken larger in order to account for channels that would enter the domain from the top or the bottom of the domain but not being

entirely included in the simulation domain. For each sub-image, the positions of channels are tracked by recording the Y coordinate of the channel centerline at fixed intervals along the X axis. Then, the mean Y values $\overline{\mathbf{p}}$ for every traveling pilot point and an empirical covariance matrix, $\mathbf{C_p}$, representing the covariance of their position along the Y axis is calculated from these recorded Y coordinates. Once the covariance matrix is known, the generation of a geological model is obtained by first simulating a random vector $\mathbf{p}$ and then applying the procedure described above. The realizations of $\mathbf{p}$ are obtained using the discrete Karhunen-Loève expansion as shown in Equation 2.1 (Sarma, Durlofsky, and Aziz 2008).

$$\mathbf{p} = \overline{\mathbf{p}} + \mathbf{E}\mathbf{S}^{\mathbf{1/2}}\rho \tag{2.1}$$

In equation 2.1, $\mathbf{E}$ is the matrix of the eigenvectors of the covariance matrix $\mathbf{C_p}$, $\mathbf{S}$ is a diagonal matrix containing the eigenvalues of $\mathbf{C_p}$, and ρ is a vector of uncorrelated random normal variables (mean 0 and variance 1).

2.1.3 The Traveling Pilot Points approach

In the subsequent paragraphs, we present a more detailed description of the methodology. Let us consider a case where TRIPS are used to parametrize a property (e.g., hydraulic conductivity distribution) of a categorical aquifer containing f facies categories. Let n_i represent the number of TRIPS in facies i. The location of the j^{th} TRIP in the i^{th} facies in three-dimensional (3D) space is represented by (x_{ij}, y_{ij}, z_{ij}). The property value associated with the i^{th} facies category is represented by val_i. For example, if the x coordinates are known and the y and z coordinates are to be estimated, the vector $\mathbf{p}$, which contains all the unknowns (locations and category values) is represented by equation 2.2. This equation can be extended/modified for other problems with complex geometries.

$$\mathbf{p} = [y_{ij}, z_{ij}, val_{ij}, ...] \; where \; i \in [1, f] \; and \; j \in [1, n_i] \tag{2.2}$$

The model parameter vector $\mathbf{m}$ is then determined by a spatial interpolation operation, as shown in equation 2.3. For example, $\mathbf{m}$ could represent the hydraulic conductivity field containing typically on the order of several tens of thousands of values which can be categorical while $\mathbf{p}$ contains only a few tens of continuous unknowns.

$$\mathbf{m} = \mathbf{Z}(\mathbf{p}) \tag{2.3}$$

In equation 2.3, the operator $\mathbf{Z}$ could represent a spatial interpolation method such as kriging or inverse distance weighted interpolation, for example. In this thesis, this operator represents the mapping method illustrated in Figure 2.2 and described in Section 2.1.2. A groundwater flow/transport model uses the model parameter field $\mathbf{m}$ in conjunction with site-specific initial and boundary conditions to produce an output vector $\mathbf{d}$ of simulated heads/velocities/concentrations as represented in equation 2.4. The operator $\mathbf{g}$ in equation 2.4, an abstraction for the groundwater model, acts upon the parameter vector $\mathbf{m}$ to produce the output vector $\mathbf{d}$ of simulated heads/concentrations.

$$\mathbf{d} = \mathbf{g}(\mathbf{m}) = \mathbf{g}[\mathbf{Z}(\mathbf{p})] \tag{2.4}$$

If the vector $\mathbf{d_{obs}}$ represents the measured counterparts to $\mathbf{d}$, the measurement objective function, ϕ_m, which defines the misfit between the model and the measurements is calculated in equation 2.5 as

$$\phi_m = [\mathbf{d_{obs}} - \mathbf{g}(\mathbf{m})]^T \mathbf{C_D}^{-1} [\mathbf{d_{obs}} - \mathbf{g}(\mathbf{m})] \tag{2.5}$$

where the ^{T} superscript represents the matrix transpose operation and $\mathbf{C_D}^{-1}$ is a diagonal matrix with element q_{ii} (element in the i^{th} row and i^{th} column) containing the weight associated with the (i^{th}) measurement and equal (in this chapter) to the inverse of the measurement error variance (Doherty and Hunt 2010).

In equation 2.5, no consideration was given to the nature of the parameter vector. In cases where prior/preferred knowledge about the underlying parameter distribution exists, it is important to include that information to reduce the ill-posedness of the problem (Tonkin and Doherty 2005). We incorporate a plausibility/regularization term ϕ_r in equation 2.6 to represent the deviation of the parameter set from the prior knowledge about their preferred values.

$$\phi_r = [\mathbf{p} - \mathbf{p_i}]^T \mathbf{C_p}^{-1} [\mathbf{p} - \mathbf{p_i}] \tag{2.6}$$

In equation 2.6, the vector $\mathbf{p_i}$ represents our knowledge about preferred conditions. Here, we take for $\mathbf{p_i}$ the vector containing the simulated initial differences of the Y coordinates of the traveling pilot points obtained from the procedure defined in Section 2.1.2. Equation 2.6 ensures that the traveling pilot points can move around the initial position but in a manner that is compatible with the statistics derived from the analysis of the training image. Furthermore, to better constrain the relative positions of the traveling pilot points, we also considered the differences between the values in the regularization term. This is implemented by assuming that the differences between the updated and initial parameters should remain small. The covariance of the differences can be estimated from the covariance matrix of the parameter values as described in section A.1.

Note that for the sake of keeping the above explanations as simple as possible, we did not describe how the covariances and mean parameter values were included in the parameter for the hydraulic conductivities. This is done in a straightforward manner by assuming that the parameter values were uncorrelated to the positions. The final covariance matrix contains in this case two independent blocks: one for the position, one for the parameter values. The estimated mean and covariance values are shown in section A.2. Finally, the global objective function, ϕ_g, includes both measurement and parameter misfit:

$$\phi_g = \phi_m + \mu^2 \phi_r \tag{2.7}$$

Equation 2.7 represents a technique of regularization that was implemented in the parameter-estimation software, PEST (Doherty 2018). The factor μ^2 is a regularization weight multiplier, controlling the parameter misfit, and is explicitly estimated during the inversion process. The inverse problem in the current context is a minimization problem where the global objective function ϕ_g is minimized while keeping the channel geometry compatible with our prior knowledge expressed through regularization.

In summary, the overall flowchart for the TRIPS algorithm is presented in Figure 2.3. An initial vector $\mathbf{p_i}$ is generated using equation 2.1. This information is then transformed into a model parameter field (e.g., hydraulic conductivity) with the aid of the spatial interpolation operator. An initial forward model simulation is carried out. If the misfit is considered acceptable, the parameter estimation is stopped. Otherwise, an optimization method (in this chapter, gradient optimization in the PEST software) is used to minimize the global objective function ϕ_g and obtain better values of the TRIPS. An updated field of model parameters is created, forward model simulation is carried out, and the objective

function is re-evaluated. This process is repeated until the optimization objectives are met.

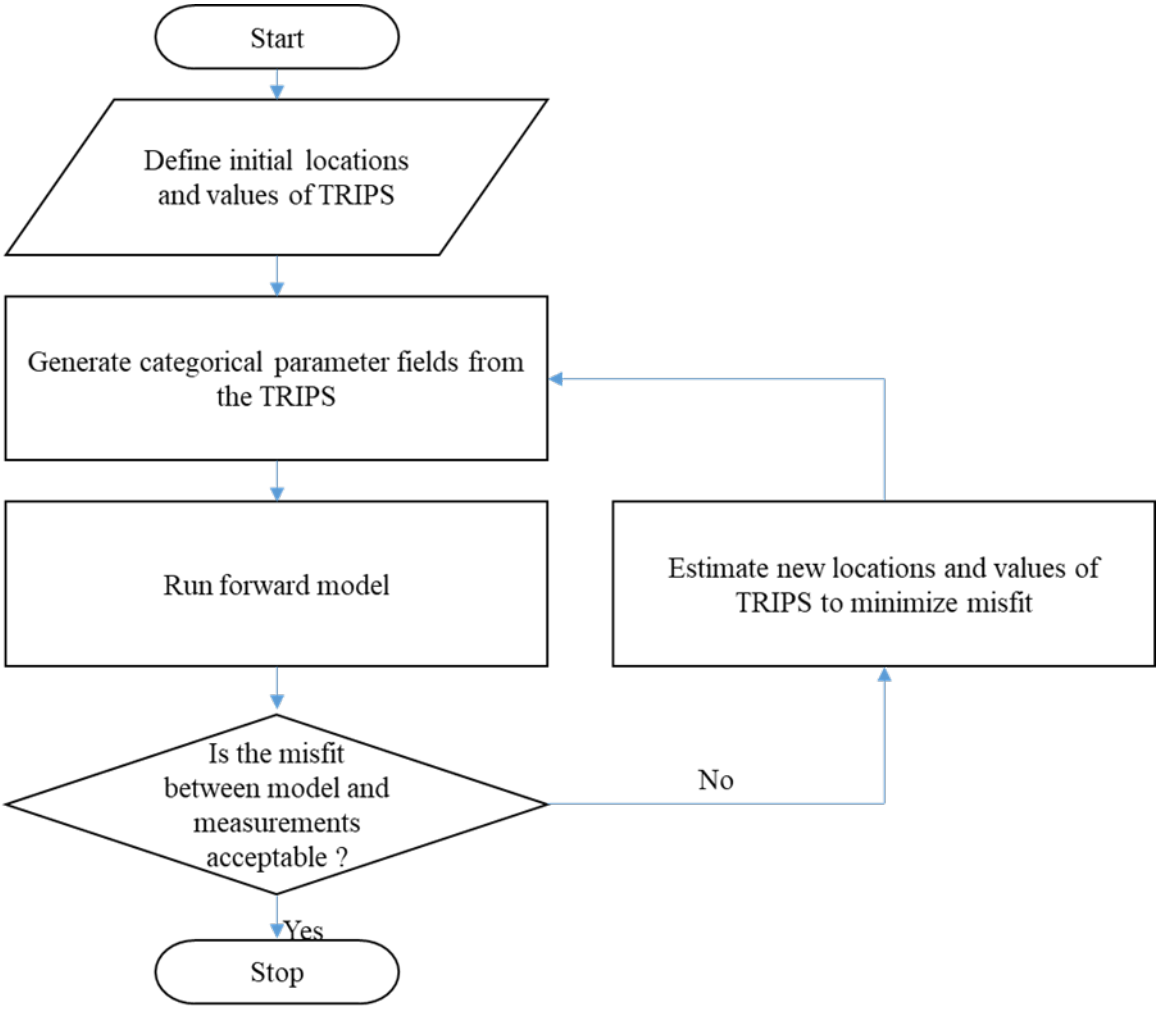


FIGURE 2.3: Flowchart depicting the TRIPS algorithm.

2.2 Generating ensembles of realizations

In this section, we describe three different approaches to generate ensembles of realizations. The first approach, rejection sampling (RS), is a simple but computationally expensive approach to sample from the posterior distribution. Since

this approach is capable of handling any kind of prior or posterior distributions, it serves as a benchmark for the other approaches which rely at least partly on a multiGaussian assumption. The second approach, Null Space Monte Carlo (NSMC), describes how the TRIPS methodology can be used in conjunction with subspace techniques which are computationally faster. The third approach, Posterior Covariance (PC), also a subspace technique, uses an alternate way to calculate the covariance matrix of the posterior and generates an en- semble rapidly. The second and third approaches are of interest in this chapter as they cannot be applied to categorical inverse problems without using an indirect parametrization such as the one proposed here with TRIPS .

In summary, the TRIPS approach provides a framework for generating a channelized categorical aquifer field from pilot points spaced along the channel centerlines. The NSMC and PC methods use subspace techniques to sample the positions of these pilot points and the hydraulic conductivities of the aquifer categories.

These three approaches are applied and compared in Section 2.4 on a synthetic problem.

2.2.1 Rejection sampling

Rejection sampling (RS) (Tarantola 2005; Mariethoz, Renard, and Caers 2010), is a simple but computationally expensive way of sampling the posterior distribution. In this method, many candidate parameters, $\mathbf{p}$, are generated by sampling from the prior distribution, as described in Section 2.1.2. These parameters are converted into model parameters $\mathbf{m}$. Forward simulations are carried out, and

misfit between the modeled and measured counterparts are tabulated. An acceptance probability, $P(\mathbf{m})$, defined in equation (2.8), is calculated for each candidate model based on the ratio of the likelihood function $\mathbf{L}(\mathbf{m}) = f(\mathbf{d} \mid \mathbf{m})$ to the maximum possible value of the likelihood function $\mathbf{L_{max}}$.

$$P(\mathbf{m}) = \frac{\mathbf{L}(\mathbf{m})}{\mathbf{L_{max}}} \tag{2.8}$$

In this chapter, $\mathbf{L_{max}}$ was determined as the maximum sampled value of the prior ensemble. The likelihood function is computed according to equation (2.9). It expresses the likelihood of a candidate model to reproduce the available data. It is inversely proportional to the measurement objective function and directly proportional to the standard deviation of the measurement error σ.

$$\mathbf{L}(\mathbf{m}) \propto exp[-\phi_m(\mathbf{m})] \tag{2.9}$$

For each candidate model, a random number from the Uniform distribution U(0,1) is concurrently generated along with the acceptance probability. If the acceptance probability is greater than this random number, the candidate model is accepted as a member of the posterior distribution. Otherwise, the candidate model is rejected. This method may reject many models, and therefore it is not computationally efficient, but the ensemble of accepted models represents the posterior distribution in an unbiased manner.

2.2.2 Null space Monte Carlo

The second method that we use in this chapter is the Null Space Monte Carlo (NSMC) methodology (Tonkin and Doherty 2009). It is a subspace-based pseudo-linear method capable of generating an ensemble of parameter realizations that

have a reasonable fit with the data by construction. The NSMC method is described below. The first step is to generate a single model with an acceptable level of misfit between the model and measurements. We do this using the gradient-based optimization method described in Section 2.1.3. The optimized parameter set from this model is denoted by the vector $\mathbf{p_c}$. The Jacobian matrix $\mathbf{X}$ is estimated. It contains the partial derivatives of the measured data with respect to the components of the vector $\mathbf{p_c}$. X_{ij} (representing the value in row i and column j) is calculated as the partial derivative of observation i with respect to parameter j. The weighted Jacobian matrix, $\mathbf{X^T C_D}^{-1}\mathbf{X}$ is computed. The matrix $\mathbf{C_D}^{-1}$ contains observation weights as defined in Section 2.1.3. This weighted Jacobian matrix is decomposed using singular value decomposition (Tonkin and Doherty 2009) as the product of three matrices in equation (2.10).

$$\mathbf{X^T C_D}^{-1}\mathbf{X} = \mathbf{USV^T} \tag{2.10}$$

In (2.10), $\mathbf{U}$ is an orthonormal matrix containing the basis vectors for the range space of the weighted Jacobian; $\mathbf{S}$ is a rectangular diagonal matrix containing eigenvalues of the weighted Jacobian matrix; $\mathbf{V}$ is an orthonormal matrix containing the basis vectors for the parameter solution space and parameter null space.

If there are n eigenvalues and the partition between the solution and null spaces is drawn after the first r eigenvalues, the matrix $\mathbf{V}$ can be thought of as $\mathbf{V} = [\mathbf{V_1}\ \mathbf{V_2}]$ where $\mathbf{V_2}$ has (n-r) columns which form the basis vectors for the null space. Moore and Doherty (Moore and Doherty 2005) present a discussion on the impact of this partition on predictive error variance. Next, we generate multiple parameter vectors by sampling from the prior distribution following the methodology described in Section 2.1.2. These parameter vectors constitute

the "uncalibrated parameters". The difference between each uncalibrated parameter set $\mathbf{p_u}$ and the calibrated parameter set is computed and projected into the parameter null space by multiplying with the null space projection matrix $\mathbf{V_2 V_2}^{\mathbf{T}}$. This projected parameter set will lie in the parameter null space if the model were linear and if the null space was delineated accurately. The projected differences are added to the calibrated parameter set to create a new parameter set $\mathbf{p_{u-new}}$. This process is described by (2.11) where

$$\mathbf{p_{u-new}} = \mathbf{p_c} + \mathbf{V_2 V_2}^{\mathbf{T}}(\mathbf{p_u} - \mathbf{p_c}) \tag{2.11}$$

Since the model is non-linear and there is uncertainty about the partition between the null and solution spaces, the parameter set from equation (2.11) does not often result in a calibrated model. Hence this parameter set is further updated using PEST (Doherty 2018) until the measurement mismatch is acceptable.

2.2.3 Posterior Covariance Calculation

In this method, the posterior covariance matrix $\mathbf{C}'$ is estimated from the prior covariance matrix under the assumption of linearity (Tarantola 2005).

$$\mathbf{C}' = \mathbf{C_p} - \mathbf{C_p X}^{\mathbf{T}}[\mathbf{X C_p X}^{\mathbf{T}} + \mathbf{C_D}]^{-1}\mathbf{X C_p} \tag{2.12}$$

In equation (2.12), $\mathbf{C_p}$ is the prior covariance matrix. The second term on the right-hand side represents the impact of calibrating the model. The term $\mathbf{C_D}$ represents the covariance of the measurement errors. The matrix $\mathbf{X}$ represents the Jacobian matrix of the calibrated model. After calculating $\mathbf{C}'$, several parameter sets are randomly generated using a random parameter generator as described in section 2.1.2. If the model were perfectly linear, each of these parameter sets

would reproduce the observed data. An inspection of the likelihood functions revealed that it is not the case. Hence, this parameter set is further updated using PEST (Doherty 2018) until the measurement mismatch is acceptable.

2.3 Synthetic problem

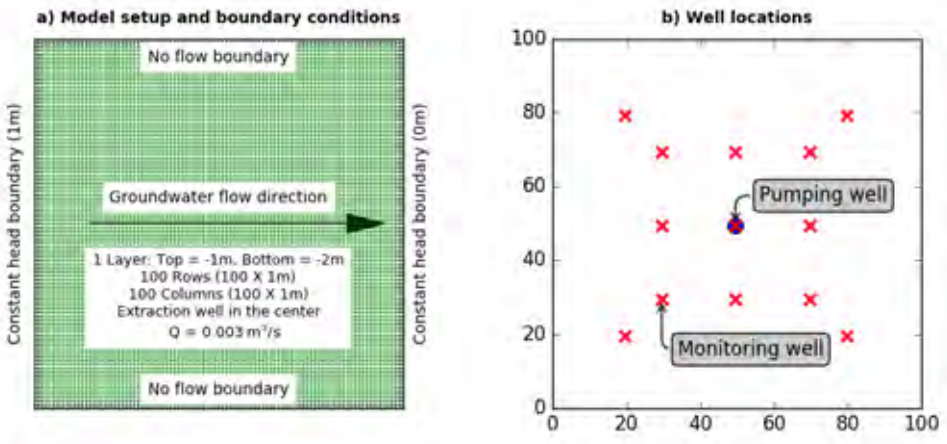


FIGURE 2.4: Model setup, boundary conditions, and well locations for the synthetic problem. The locations of the pumping well and monitoring wells are shown in the plot to the right.

A synthetic problem derived from Mariethoz and others (Mariethoz, Renard, and Straubhaar 2010) is analyzed in this chapter. A constant discharge pump test is conducted in a square-shaped (100 m $\times$ 100 m) confined aquifer. The pumping well extracts 0.003 m^3/s (cubic meters per second) from the center of the aquifer. The aquifer contains high-permeability fluvial channels embedded in a low-permeability matrix. Groundwater flow in the aquifer is two-dimensional flowing from left to right. A constant head boundary of 1 m is located on the left edge, and a constant head boundary of 0 m is located along the right edge. 12 monitoring wells are located around the pumping well. The aquifer schematic, boundary conditions, and well locations are shown in Figure 2.4. Aquifer heads are recorded at the pumping and monitoring wells once the system reaches

steady-state. The model representing this synthetic reality is referred to as the "reference model".

The facies distribution was developed following the approach described in Section 2.1.2. We used the training image (TI) introduced in Figure 2.1. It represents channels and matrix in a 2500m $\times$ 2500m area. A large number (30000) of sub-images were extracted from this TI. After visually inspecting a subset of these images, it was determined that there are typically three fluvial channels of width 13m in a 100m $\times$ 100m area. The covariance matrix $\mathbf{C_p}$ of the Y coordinates along the channel centerline was estimated according to the methodology described in Section 2.1.2. 15 points - 5 for each channel, were used to track the channels. The matrix scatter plot of the Y coordinates shown in Figure 2.5 depicts the correlation between the various coordinates. In this plot, the variables y11 to y15 represent the coordinates of the top channel in a left to right direction. The variables y21 to y25 represent the coordinates of the middle channel in a left to right direction and the variables y31 to y35 represent the coordinates of the bottom channel in a left to right direction. The off-diagonal plots show the scatter between two coordinates and the diagonal plots show the histogram of a single coordinate. The plot shows that each point is strongly correlated with its neighbors along the same channel and weakly correlated with points in the other channels.

The discrete Karhunen-Loève expansion, expressed in (2.1), was then used to generate 100'000 parameter realizations based on $\mathbf{C_p}$. For each realization, cubic B-splines were used to connect the channel coordinates. The splrep function in the scipy python library (Virtanen et al. 2020) was used to create the splines through the channel coordinates. A buffer of width 6.5 m around each of the splines was created to represent a channel of 13 m width. These channels were overlaid on the model grid and model cells fully covered by the channels were

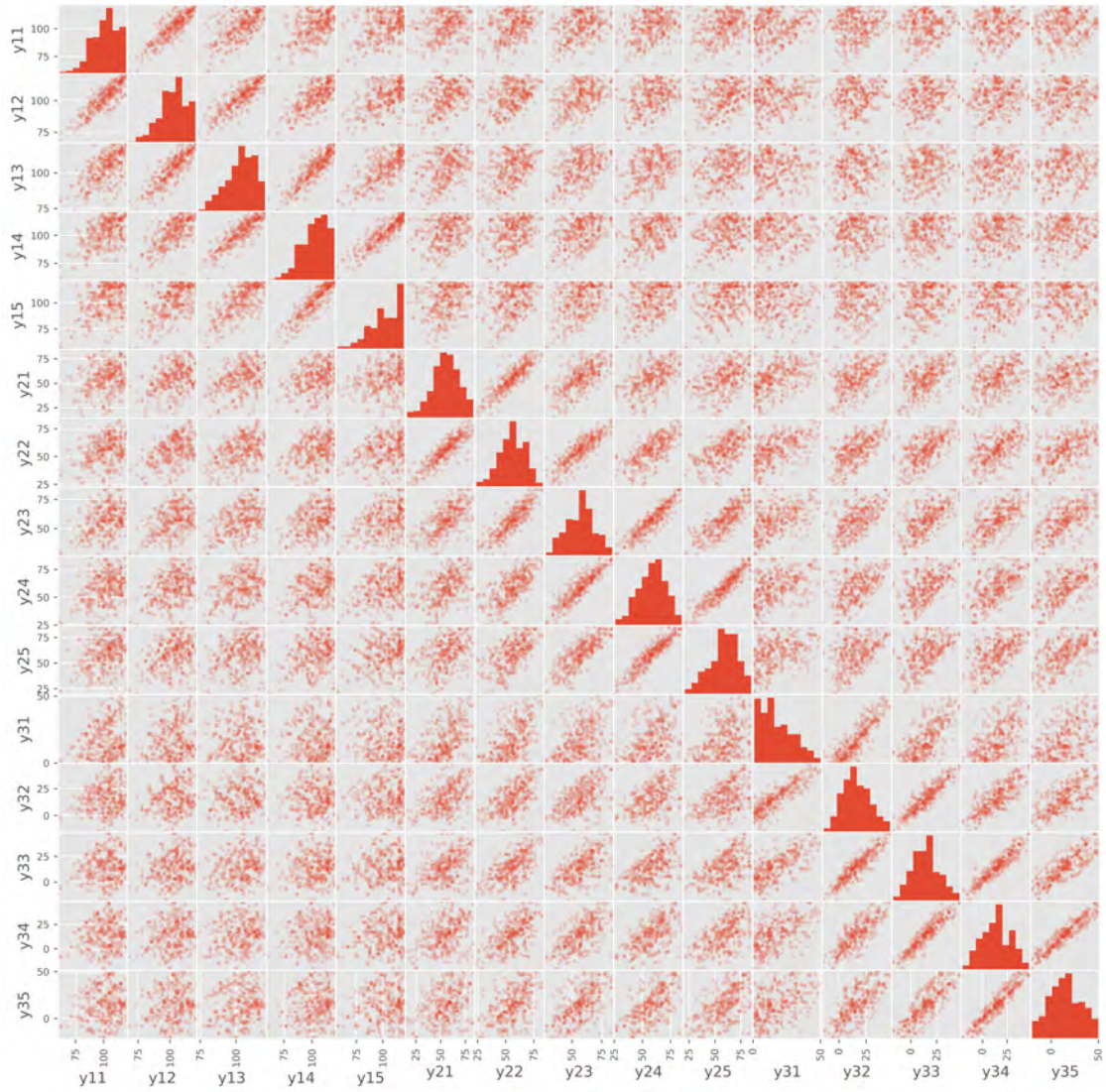


FIGURE 2.5: Matrix Scatter Plot of the sampled Y Coordinates

assigned a hydraulic conductivity value randomly generated within a lognormal distribution with mean = -2 log_{10} (m/s) and standard deviation = 0.1 log_{10} (m/s). The remaining cells were assumed to be a part of the matrix and were assigned a hydraulic conductivity randomly generated based on a lognormal distribution with a mean = -4 log_{10} (m/s), standard deviation = 0.1 log_{10} (m/s).

A realization was randomly selected to represent the synthetic reality. For this selected realization, the hydraulic conductivity values of the channel and matrix were 8.7 $\times$ 10^{-3} m/s and 1.1 $\times$ 10^{-4} m/s respectively.

Reference head observations were obtained in the following manner. Steady-state groundwater flow was simulated for the aquifer described above using the MODFLOW-NWT simulator (Niswonger, Panday, and Ibaraki 2011). The facies distribution and the head distribution of the reference model are shown in 2.6. The calculated head distribution was sampled at the 13 observation wells. Normally distributed random noise (mean = 0.0 m, standard deviation = 0.05 m) was added to the sampled heads to simulate measurement error. These 13 adjusted heads constituted the reference head measurements.

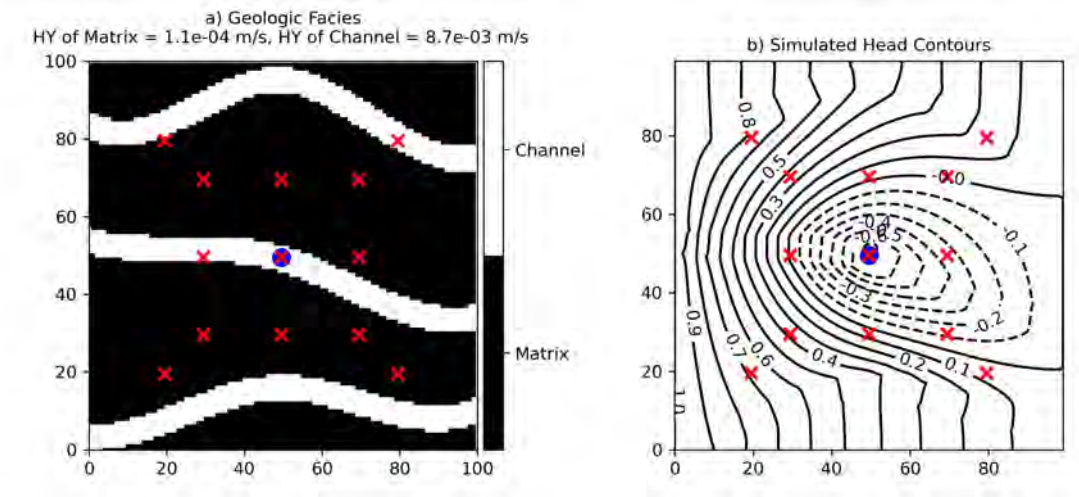


FIGURE 2.6: Facies Distribution (a) and Head Distribution (b) of the Reference Model

2.4 Results

In this section, we generate an ensemble of conditional parameter realizations for the synthetic problem using the various methods described in Section 2.2 (RS, NSMC, and PC). For each ensemble, the cell-by-cell probability of finding a channel in the model domain, mean ensemble head distribution, and standard deviation of the simulated head distribution were calculated.

2.4.1 Rejection Sampling

A large set of parameter fields were generated based on the prior covariance matrix described in Section 2.3. Forward simulations were undertaken for each of these simulations, and the results evaluated under the rejection sampling methodology described in Section 2.2.1. We could obtain 100 models in the posterior distribution by evaluating 100'050 models. Six of these models were randomly selected and the hydraulic conductivity distributions and the corresponding fits between observed and simulated heads are presented in Figure 2.7. These models demonstrate that several channel/hydraulic conductivity distributions can result in reasonable matches to the measurements. Five of the six realizations have three channels whereas one realization has only two channels.

2.4.2 NSMC Method

The first step to sample a posterior distribution using the NSMC methods is to obtain a model corresponding to a maximum value of the likelihood (2.2.2). Here, we describe how this model was obtained. A parameter set was first randomly generated from the prior covariance matrix $\mathbf{C_p}$. With the measurement and regularization constraints, parameter estimation was carried out by

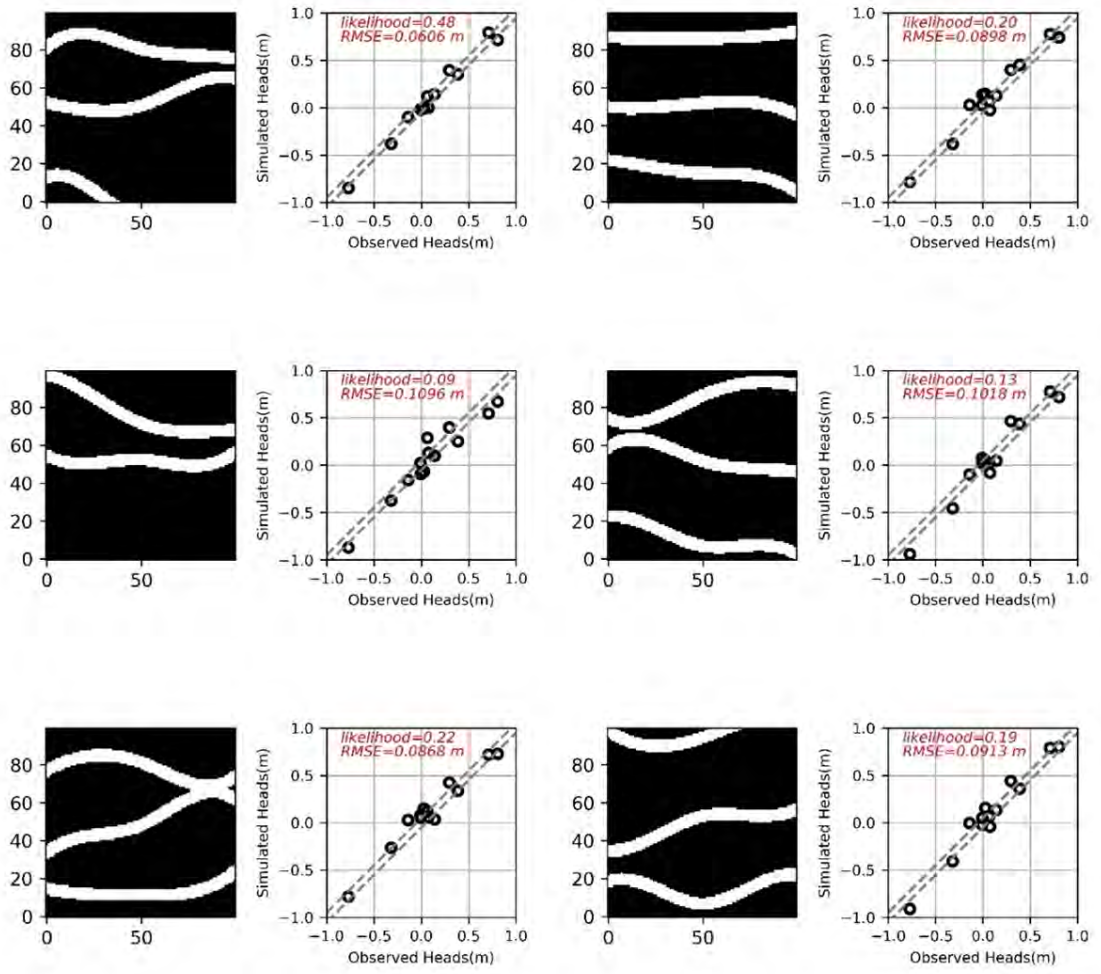


FIGURE 2.7: Six randomly selected realizations from the 100 realizations chosen using Rejection Sampling. For each realization, the hydraulic conductivity distribution is shown on the left and the fit between the observed and simulated heads is shown on the right.

maximizing the likelihood function. The optimization was stopped when the modeled heads were considered acceptable. Facies distribution, head contours, measured vs. simulated heads for the initial, two intermediate models, and the calibrated model are presented in Figure 2.8. This figure illustrates how the likelihood function increases as the central channel moves closer to the location of the pumping well (blue circle enclosing a red cross). A total of 484 groundwater flow model evaluations were required during this optimization to evaluate the misfit and the Jacobian. The resulting model is then used as the starting step for generating the ensemble.

The NSMC methodology described in Section 2.2.2 was used to generate 100 realizations. The computational cost for obtaining this ensemble was 12'754 forward model evaluations. Six of these models were randomly selected and the hydraulic conductivity distributions and the corresponding fits between observed and simulated heads are presented in Figure 2.9. All the six realizations have three channels with the central channel exhibiting more curvature than the top/bottom channels.

2.4.3 PC Posterior Ensembles

The PC methodology described in Section 2.2.3 was used to generate another ensemble consisting of 100 realizations. This method is much more computationally efficient, since we generated 100 models with only 505 forward model evaluations. Six of these models were randomly selected and the hydraulic conductivity distributions and the corresponding fits between observed and simulated heads are presented in 2.10. All the six realizations have three channels.

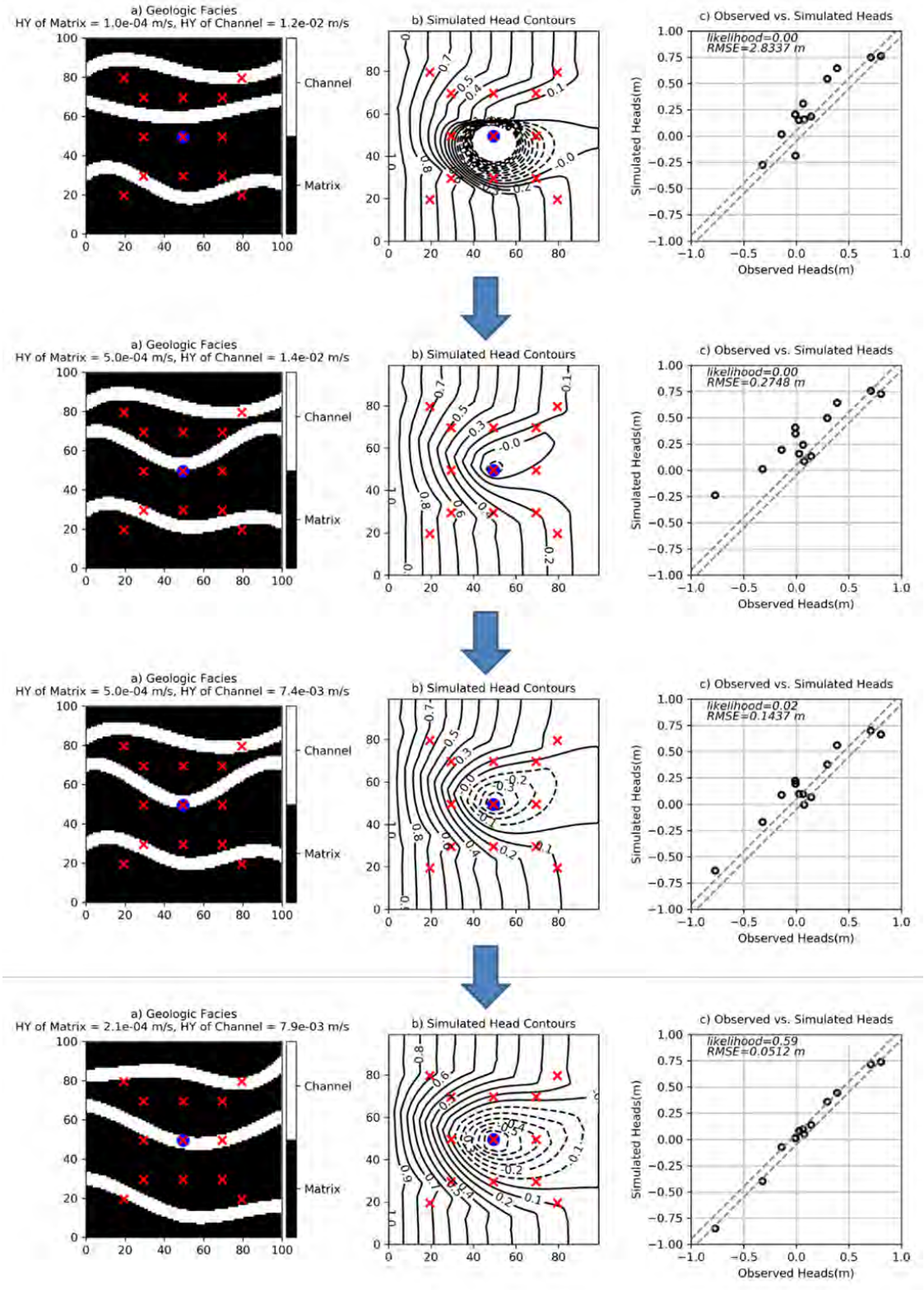


FIGURE 2.8: Facies Distribution (a), Head Distribution (b), and Observed vs. simulated heads of the uncalibrated model (top panel), intermediate models (middle panels) and the calibrated model (bottom panel).

2.4.4 Comparison of Posterior Ensembles

The characteristics of the parameter ensembles obtained using the various methods are presented in Figure 2.11. Information is presented in a grid with four rows and four columns. Each row represents the characteristics of a parameter ensemble. In the first column of each row, cell-by-cell probabilities of finding a channel for that ensemble are depicted. Reddish colors imply a higher probability of finding a channel and bluish colors imply a lower probability. The probabilities for the prior distribution (first row) are low everywhere. When the head data are not accounted for, the proposed geological model can place the channels anywhere in the domain in a uniform manner.

All the methods that are conditioned by the head data show that they can locate the presence of a channel at the location of the pumping well with a high probability. They also indicate a high probability to find matrix above and below the pumping well as well as two (2) other channels in the top and bottom of the field. However, there are some differences in the values of the probabilities when moving away from the location of the pumping well.

The ensemble obtained with RS is unbiased and considered as the reference in this experiment. The RS and the NSMC ensembles exhibit more similarities than RS and PC. The PC method shows higher values for the probabilities than the RS and NSMC methods. It means that the subspace methods did not capture the complete variability of the posterior ensemble. They are much more efficient numerically but this comes at the cost of an underestimation of the uncertainty. Some channel configurations that can reproduce the data and belong to the prior geological model are not identified in that case.

Histograms for the log-transformed hydraulic conductivities presented in

the second column exhibit a bimodal distribution. The channel hydraulic conductivities are shown in red and the matrix hydraulic conductivities are in blue. The hydraulic conductivities of the reference model are shown as black dots on the histogram. The matrix hydraulic conductivities vary more than their channel counterparts. As with the ensemble probability, the PC method has narrower histograms implying a lack of variability. We also see on these graphs that all methods identify properly the hydraulic conductivity of the channels, while the matrix conductivity may be overestimated as compared to the reference by the NSMC and PC methods. A possible reason for this overestimation could be to compensate for deviations from the geometry in the reference field for maintaining the observed gradients.

Mean and standard deviation of the ensemble head distributions are in the third and fourth columns. The prior mean and variance are showing symmetry around the pumping well. The uncertainty on the head value is high as shown by the high values of the standard deviation. The effect of conditioning to the head values reduces significantly the uncertainty for the three ensembles. The ensemble mean head from the NSMC method resembles the ensemble mean head from the RS method and the head from the reference model (Figure 2.6). The mean head estimated with the PC ensemble does not show as clearly as the two other methods, the shift of the cone of depression toward the bottom of the image. Furthermore, in terms of standard deviation and uncertainty estimation the PC method has a very low ensemble standard deviation lending further credence to the lack of variability in realizations. As compared to the reference method (RS), the NSMC is closer to it but still underestimates the variability especially in the upper part of the domain.

2.5 Summary and Discussion

In this chapter, we propose a new approach involving Traveling Pilot Points (TRIPS) and linear subspace methods to solve the categorical inverse problem in a probabilistic framework. We summarize some of the main findings below.

The first key proposition that we make in this chapter is in letting pilot points travel and estimating both the locations of the channels and associated properties like hydraulic conductivity. We then propose to estimate the prior covariance matrix of the position of the pilot points from a training image. The advantage of that approach is in its simplicity. The user can provide an image of the type of channels that they want to model, and the covariance will be inferred directly. If the training image is too small, it is possible to use a MPS simulation algorithm and generate an ensemble of simulations and derive the covariance matrix from the analysis of this simulation in the same manner as we analyze the sub-images in this work. In addition, we use first-order (difference) regularization constraints to preserve the curvature of the channels in the inversion process.

The proposed TRIPS parametrization was integrated in an optimization framework based on linear subspace methods allowing to obtain solutions of the categorical inverse problem for a synthetic aquifer. A posterior ensemble obtained with the rejection sampling method was considered to represent the reference solution and compared against the NSMC and PC ensembles. The comparisons indicate that these parameter ensembles exhibit similarities with the reference distribution. The PC method was much more efficient in estimating members of the posterior ensemble. However, the variability was underestimated (Figure 2.11, bottom row). The NSMC method was comparably slower because

more model evaluations were required. However, the ensemble probability estimated by this method is closer to the ensemble from RS. The NSMC method provides a balance between computational efficiency and representation of the posterior ensemble. The number of model evaluations required by the NSMC method were comparable to stochastic approaches like ISR (Mariethoz, Renard, and Caers 2010) and POPEX (Jäggli, Straubhaar, and Renard 2017), while the PC method is much faster.

Overall, we believe the TRIPS methodology to be a promising entrant in the field of categorical inversion. While the example problem presented in this chapter considers only two-dimensional channels traversing the domain, the methodology can be extended to real-world three-dimensional datasets with a larger number of facies and more complex geometries. For example, the TRIPS method could be used to estimate the complex channel framework at a real site such as the one discussed by Ronayne and others (Ronayne, Gorelick, and Caers 2008). For this case, the pilot points would represent the positions of channels in three dimensions and three-dimensional splines passing through the pilot points could be used to delineate the channels.

More generally, the extension of the proposed methodology is straightforward for all object-oriented geological modeling techniques (Pyrcz and Deutsch 2014) since the positions of the objects are controlled in these models by seed points which can be considered as Traveling Pilot Points. The prior statistics on the number of objects and relative locations of these points can be derived from a set of initial simulations. The proposed algorithm described could then be used to update these locations and solve the inverse problem. The shape parameters concerning the three-dimensional size and orientation of the objects can be handled as well easily since these are continuous parameters that an inversion code like PEST can optimize. This step would be analogous to the identification of the

hydraulic conductivity values within the channels as illustrated in the example treated in this chapter. For objects having a flexible shape such as channels with varying width, traditional techniques such as the standard pilot points can be coupled with TRIPS: one can attach a width parameter to every traveling pilot point and interpolate the width along the channel length and update these parameters during the inversion.

The TRIPS method could also be used with pixel based geostatistical methods such as plurigaussian or MPS simulations (Pyrcz and Deutsch 2014). Starting from one or a set of initial realizations, we could extract a set of conditioning locations and the corresponding categories from the realization. TRIPS would then proceed by moving the locations of these points, keeping the value of the categories and simulating again the complete field using these new conditioning data as input in the geostatistical algorithm. In this last case, developing the appropriate parameter covariance matrix remains a challenge.

Considering the subspace methods tested in this chapter, one aspect which works to their favor is the large null space. As the size of the null space increases, TRIPS/NSMC methods could prove to be computationally parsimonious in comparison with other methods. We have shown in this chapter, on a simple synthetic problem, that the gain in numerical efficiency comes at the cost of an underestimation of the overall uncertainty.

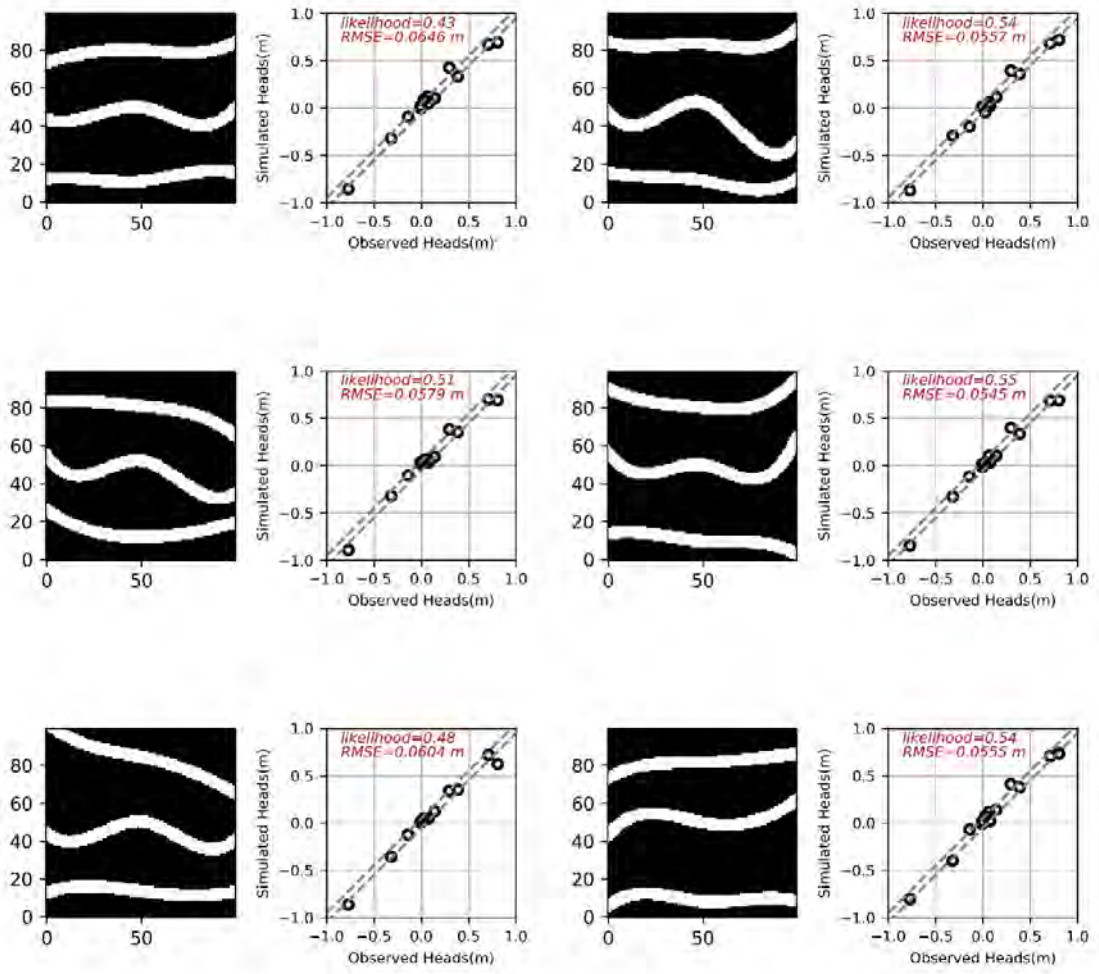


FIGURE 2.9: Six randomly selected realizations from the 100 realizations generated using NSMC. For each realization, the hydraulic conductivity distribution is shown on the left and the fit between the observed and simulated heads is shown on the right.

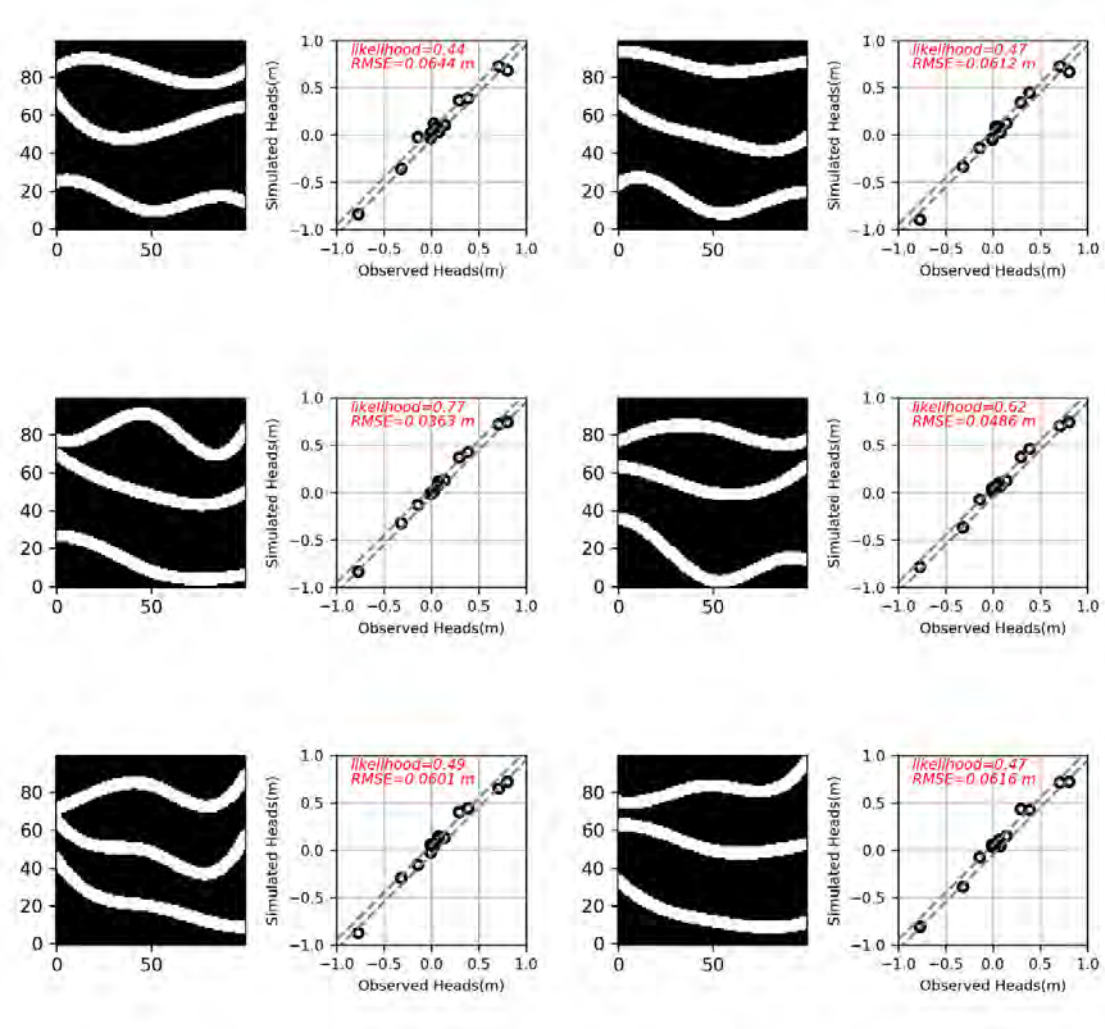


FIGURE 2.10: Six randomly selected realizations from the 100 realizations chosen using the Posterior Covariance method. For each realization, the hydraulic conductivity distribution is shown on the left and the fit between the observed and simulated heads is shown on the right.

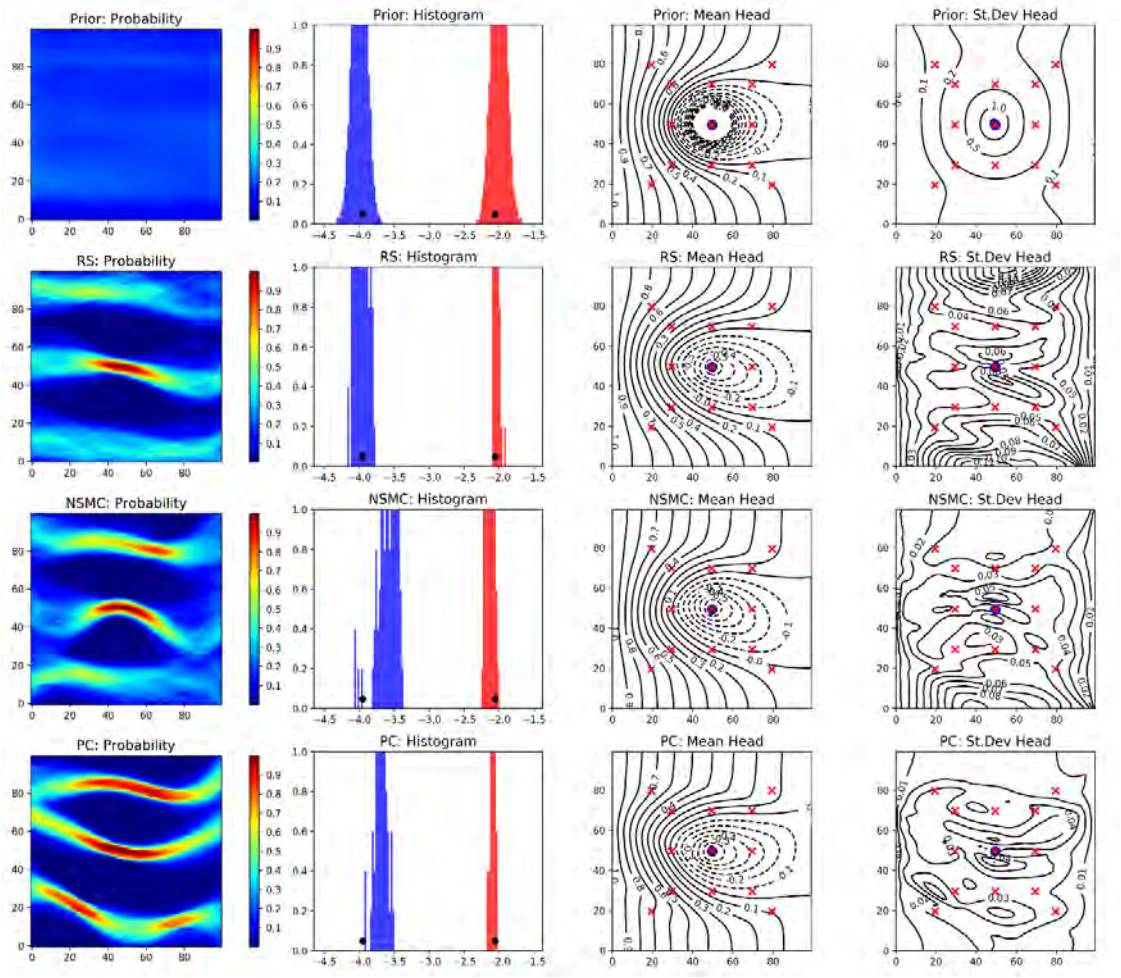


FIGURE 2.11: Parameter ensemble characteristics from various methods, Prior (top row), Rejection sampling (second row), NSMC (third row) and PC (bottom row). In each row, ensemble parameter probability, histogram of ensemble conductivities, mean and standard deviation of simulated heads are shown.

Chapter 3

Coupling TRIPS with an Iterative Ensemble Smoother*

3.1 Introduction

In chapter 2, the TRIPS method was presented and used in conjunction with linear subspace methods to parametrize the categorical inverse problem, and estimate an ensemble of likely categorical parameter distributions for a hypothetical problem representing groundwater flow through an aquifer with two distinct facies. It was shown in particular that this simple way of framing the inverse problem had the advantages to reduce significantly the number of unknowns, to allow the use of continuous optimization techniques as well as linear subspace methods to obtain an ensemble of models fitting the data and representing some uncertainty on the solution of the inverse problem. However, we also showed by comparing the results of this approach with an unbiased sampler of the posterior space (the rejection sampling (RS)) that the ensemble obtained was not covering well the true posterior distribution (Figure 2.11). There was a tendency to be over-confident in the results: the zones of very high probability of presence of channels were generally larger in the results obtained with

*This chapter corresponds to a manuscript in preparation

NSMC or PC than with the unbiased rejection sampling. In addition, the histograms of the posterior hydraulic conductivity values for the matrix and channel were also slightly biased from the reference and tending to underestimate the hydraulic conductivity contrast . These results showed therefore that even if the TRIPS approach was promising, there was room for improvements.

Linear subspace methods require the estimation of a sensitivity (Jacobian) matrix **G**, containing the partial derivatives of the measured data with respect to the parameters in the model. The estimation of the Jacobian matrix is a computationally intensive undertaking, especially, for highly parametrized models with several hundreds to thousands of parameters. Ensemble-based methods offer a computationally efficient method for estimating an approximate form of the Jacobian matrix, **G**.

Since their introduction about three decades ago, ensemble-based methods are increasingly being adopted in a wide variety of fields such as petroleum reservoir modeling, hydrology, hydrogeology, meteorology, oceanography. In this chapter, the feasibility of using the TRIPS method with ensemble-based methods to improve the sampling of the posterior distribution for the same problem as in chapter 2 is explored. Two classes of ensemble methods, the Ensemble Kalman Filter (EnKF) and the Ensemble Smoother (ES), are briefly summarized here.

Kalman filtering (Kalman 1960) was a technique developed to estimate the current state of a linear dynamical system from uncertain observations. In the environmental modeling context, the Kalman filter is used in conjunction with an environmental model that calculates a series of updated system states over a series of successive time steps; this process is referred to as "recursive data assimilation". The estimated state at any time step depends on the previous state of the system, and on the properties of the system, as embodied in model

parameters. A limitation of the Kalman filtering is the requirement to explicitly compute and store the error covariance matrix of size $n \times n$ with n being the size of the state vector. For higher-dimensional problems encountered in the real world, this limitation is often hard to overcome.

The EnKF (Evensen 1994; Evensen 2009) was proposed to overcome the limitations of the Kalman filter. The EnKF uses an ensemble of realizations to estimate an approximate form of the covariance matrix based on the size of the ensemble. Since this computation is tractable, it has gained wide acceptance. In the groundwater and reservoir modeling context, several researchers (Heidari et al. 2011; Tavakoli, Srinivasan, and Wheeler 2013; Li et al. 2015; Kumar 2019, e.g) have applied the EnKF to solve inverse problems.

The Ensemble Smoother (ES) is an alternative data assimilation method proposed by Leeuwen and Evensen (1996). An ES "smooths" all members of an ensemble by assimilating data in a single step and computing a global member update. Skjervheim and Evensen (2011) demonstrated the applicability of ensemble smoothers in characterizing the posterior distribution for reservoir history-matching problems. They also compare and contrast the ES with the EnKF. In their words, both methods use an ensemble of realizations to represent uncertainty and an updating scheme based on minimizing variance. However, while the the EnKF computes recursive updates in the time-domain, the ES computes a global update in the space-time domain avoiding multiple restarts.

The equations which are used by the ES method are based on the Bayes theorem which implies a linear relationship between model parameters and estimated states. For non-linear models encountered in the real world, this assumption of linearity often leads to an incorrect characterization of the posterior distribution. To remedy this problem, Emerick and Reynolds (2013) proposed an iterative ensemble smoother (IES) by assimilating the same data several times

to iteratively improve the match obtained by the ES method. An IES repeatedly smooths ensembles of prior parameter distributions until they represent a posterior parameter ensemble. Chen and Oliver (2013) proposed an improved IES for generating a posterior parameter ensemble by starting with a prior parameter ensemble and repeatedly iteratively assimilating data and smoothing. This method, referred to as Levenberg Marquardt Ensemble Random Maximum Likelihood (LM-EnRML), uses the Levenberg-Marquardt method (Hanke 1997) to regularize the parameter update vector for an ensemble. White (2018) implemented the IES algorithm in an open-source software called pestpp-ies and used the software to efficiently generate several likely parameter distributions for a real-world groundwater flow and transport model. Cao, Li, and Chen (2018) used pilot points to bridge IES and multiple-point-statistics to condition parameter fields for a synthetic 2D flow and transport model. Recently, Lam et al. (2020) used IES coupled with a multi-resolution MPS approach to tackle the categorical inverse problem and showed an application for a 2D groundwater flow model. White, Knowling, and Moore (2020) used pestpp-ies to quantify parameter and model uncertainty for a model of the Edwards Aquifer in Texas, USA.

Motivated by the documented success of the IES method and the ease-of-use of the pestpp-ies implementation, this chapter explores whether TRIPS and pestpp-ies together can be used to solve the categorical inverse problem previously described in chapter 2. In addition to using pestpp-ies to sample the posterior, numerical experiments were also carried out to explore the dependence of the estimated posterior on the size of the prior ensemble. Model selection techniques are used to investigate if it is possible to strategically select a smaller prior ensemble, that in turn would lead to a smaller posterior ensemble, while still maintaining the diversity of a larger posterior ensemble.

The subsequent sections of this chapter are organized as follows. In section 3.2, the various methods used in this chapter are described. The approach to develop a posterior parameter ensemble for the synthetic problem is described in section 3.2.1. The results are presented in section 3.3. The feasibility of model selection techniques to develop posterior ensembles at a lesser computational cost is also explored in this section. A summary of key findings is presented in section 3.4.

3.2 Methods

The IES formulation used in this paper is described in section 3.2.1. The flow formulation for the problem in this paper is summarized in section 3.2.2. Section 3.2.3 has the methodology for constructing a smaller prior ensemble from a larger prior ensemble by strategic selection.

3.2.1 Iterative Ensemble Smoother

The LM-EnRML algorithm from Chen and Oliver (Chen and Oliver 2013) is briefly summarized here. For a comprehensive description, readers are advised to consult the original publication.

A groundwater model acting upon a parameter vector $\mathbf{m}$ to produce simulated states can be represented mathematically as $\mathbf{g}(\mathbf{m})$ where the operator $\mathbf{g}(.)$ is an abstraction for the groundwater model. Assuming an ensemble with N_e realizations is used to assimilate N_{obs} measurements into a model with N_{par} parameters, the objective function to be minimized (ϕ_m) for every member of the ensemble at the l^{th} IES iteration is shown in equation 3.1.

$$\phi_m = [\mathbf{g}(\mathbf{m_l}) - \mathbf{d_j}^{\mathbf{0}}]^T \mathbf{C_D}^{-1} [\mathbf{g}(\mathbf{m_l}) - \mathbf{d_j}^{\mathbf{0}}] \tag{3.1}$$

$\mathbf{m_l}$ represents the parameters in the l^{th} iteration, $\mathbf{C_D}$ is the covariance matrix of the data, $\mathbf{d_j}^{\mathbf{0}}$ is a perturbed observation from a multiGaussian probability density function with mean $\mathbf{d_{obs}}$ and covariance $\mathbf{C_D}$. In the TRIPS context, $\mathbf{m_l}$ contains the positions of the traveling pilot points and the hydraulic conductivity values of the channel and matrix facies.

To minimize ϕ_m, the maximum a posteriori Gauss-Levenberg-Marquardt parameter update in the l^{th} IES iteration ($\delta \mathbf{m_l}$) is calculated as:

$$\begin{aligned} \delta \mathbf{m_l} = -[(1+\lambda_l)\mathbf{C_M}^{-1} + \mathbf{G_l}^T \mathbf{C_D}^{-1} \mathbf{G_l}]^{-1} \times \\ [\mathbf{C_M}^{-1}(\mathbf{m_l} - \mathbf{m_j}^{\mathbf{pr}}) + \mathbf{G_l}^T \mathbf{C_D}^{-1}(\mathbf{g}(\mathbf{m_l}) - \mathbf{d_j}^{\mathbf{0}})] \end{aligned} \tag{3.2}$$

In equation 3.2, λ_l is the Marquardt damping parameter or Marquardt lambda at iteration l. $\mathbf{C_M}$ is the covariance matrix of model parameters. $\mathbf{m_j}^{\mathbf{pr}}$ is a matrix containing an ensemble of model parameter vectors sampled from the prior distribution. The dimensions of this and the other key matrices are provided in table 3.1.

Vector/Matrix	Dimensions
$\delta \mathbf{m_l}$, $\mathbf{m_l}$, $\mathbf{m_j}^{\mathbf{pr}}$	$N_e \times N_{par}$
$\mathbf{g}(\mathbf{m_l})$, $\mathbf{d_j}^{\mathbf{0}}$	$N_e \times N_{obs}$
$\mathbf{G_l}$	$N_{obs} \times N_{par}$
$\mathbf{C_M}$	$N_{par} \times N_{par}$
$\mathbf{C_D}$	$N_{obs} \times N_{obs}$

TABLE 3.1: Dimensions of matrices and vectors in the LM-EnRML parameter update

The Jacobian matrix $\mathbf{G_l}$ represents the sensitivities of the data to the parameters at iteration l. Sensitivity of the i^{th} data measurement to the j^{th} parameter is contained in the row i and column j of $\mathbf{G_l}$. Parameter estimators like PEST (Doherty 2018), or pestpp-glm (White et al. 2020) estimate $\mathbf{G}$ using a first-order finite-difference approximation to the partial derivative of the incremental change in the simulated equivalent of measurement i to the incremental change in parameter j. This process is computationally expensive, especially for models with large number of parameters. The LM-EnRML algorithm on the other hand infers an empirical form of the $\mathbf{G}$ using information from the ensemble as shown in equation (3.3).

$$\mathbf{G_l} = \mathbf{C_D}^{(1/2)} \Delta \mathbf{d_l}^{\mathbf{e}} (\Delta \mathbf{m_l}^{\mathbf{e}})^{-1} \mathbf{C_{SC}}^{-1/2} \tag{3.3}$$

where,

$$\Delta \mathbf{d_l}^{\mathbf{e}} = \mathbf{C_D}^{-1/2} (\mathbf{d_l}^{\mathbf{e}} - \overline{\mathbf{d_l}^{\mathbf{e}}}) / \sqrt{N_e - 1} \tag{3.4}$$

$$\Delta \mathbf{m_l}^{\mathbf{e}} = \mathbf{C_{SC}}^{-1/2} (\mathbf{m_l}^{\mathbf{e}} - \overline{\mathbf{m_l}^{\mathbf{e}}}) / \sqrt{N_e - 1} \tag{3.5}$$

In equation (3.5), $\mathbf{C_{SC}}$ is a scaling matrix for model variables typically chosen to be a diagonal matrix with diagonal elements equal to the variance of the prior distribution for each type of the model variables. $\mathbf{m_l}^{\mathbf{e}}$ and $\mathbf{d_l}^{\mathbf{e}}$ are the ensemble of the model variables at the l^{th} iteration and simulated data at the l^{th} iteration; the overbar denotes the ensemble mean. When there is a linear relationship between the ensemble parameters and the simulated values, the empirical form of $\mathbf{G}$ is a good approximation to the $\mathbf{G}$ matrix.

Similar to equation (3.5), the dimensionless prior model deviations from the ensemble mean are written as

$$\Delta \mathbf{m_{pr}}^{\mathbf{e}} = \mathbf{C_{SC}}^{-1/2}(\mathbf{m_{pr}}^{\mathbf{e}} - \overline{\mathbf{m_{pr}}^{\mathbf{e}}})/\sqrt{N_e - 1} \tag{3.6}$$

$\mathbf{m_{pr}}^{\mathbf{e}}$ is the ensemble of the prior model variables and the overbar denotes the ensemble prior mean. The truncated singular value decomposition (TSVD) forms of the deviations in data and prior model variables are written as

$$\Delta \mathbf{d_l}^{\mathbf{e}} = \mathbf{U_d}^{\mathbf{P_d}} \mathbf{W_d}^{\mathbf{P_d}} \mathbf{V_d}^{\mathbf{P_d}\,T} \tag{3.7}$$

$$\Delta \mathbf{m_{pr}}^{\mathbf{e}} = \mathbf{U_{m_o}}^{\mathbf{P_{m_o}}} \mathbf{W_{m_o}}^{\mathbf{P_{m_o}}} \mathbf{V_{m_o}}^{\mathbf{P_{m_o}}\,T} \tag{3.8}$$

P_d and P_{m_o} are the number of singular values retained in the TSVD in equations (3.7) and (3.8).

The final form of equation (3.2) after dropping the iteration index l on all terms except λ_l and using the relationships in the subsequent equations is shown in equation (3.9). Equation (3.9) is the parameter update from the LM-EnRML algorithm.

$$\begin{aligned}
\delta \mathbf{m} = {} & \mathbf{C_{SC}}^{1/2} \Delta \mathbf{m^e} \mathbf{V_d}^{\mathbf{P_d}} [(1+\lambda_l)\mathbf{I_{P_d}} + \mathbf{W_d}^{P_d{}^2}]^{-1} \\
& \times \mathbf{V_d}^{P_d{}^T} \Delta \mathbf{m_e}^T \mathbf{U_{m_o}}^{P_{m_o}} (\mathbf{W_{m_o}}^{P_{m_o}})^{-2} \\
& \times \mathbf{U_{m_o}}^{P_{m_o}{}^T} \mathbf{C_{SC}}^{-1/2} (\mathbf{m} - \mathbf{m_j}^{pr}) \\
& - \mathbf{C_{SC}}^{1/2} \Delta \mathbf{m^e} \mathbf{V_d}^{\mathbf{P_d}} \mathbf{W_d}^{\mathbf{P_d}} [(1+\lambda_l)\mathbf{I_{P_d}} + \mathbf{W_d}^{P_d{}^2}]^{-1} \\
& \times \mathbf{U_d}^{P_d{}^T} \mathbf{C_D}^{-1/2} (\mathbf{g(m)} - \mathbf{d_j}^0)
\end{aligned} \tag{3.9}$$

where, $\mathbf{I_{P_d}}$ is an identity matrix of dimension P_d. An approximate form of equation (3.9) neglecting the second term (model mismatch) is known as the approximate LM-EnRML and shown in equation (3.10).

$$\begin{aligned}\delta\mathbf{m} = {\mathbf{C_{SC}}}^{1/2}\Delta\mathbf{m^e}\mathbf{V_d}^{\mathbf{P_d}}[(1+\lambda_l)\mathbf{I_{P_d}} + {\mathbf{W_d}}^{P_d{}^2}]^{-1} \\ \times\ {\mathbf{V_d}}^{P_d{}^T}\Delta{\mathbf{m_e}}^T{\mathbf{U_{m_o}}}^{P_{m_o}}({\mathbf{W_{m_o}}}^{P_{m_o}})^{-2} \\ \times\ {\mathbf{U_{m_o}}}^{P_{m_o}{}^T}{\mathbf{C_{SC}}}^{-1/2}(\mathbf{m} - \mathbf{m_j}^{pr})\end{aligned} \tag{3.10}$$

White (2018) incorporated the LM-EnRML method into the PEST++ framework (White et al. 2020) and developed the pestpp-ies software. The pestpp-ies implementation of LM-EnRML (equation 3.9) is used in this chapter.

The Marquardt lambda, λ_l, has an important role in LM-EnRML and pestpp-ies. A high value of λ_l results in calculation of a damped (i.e. shortened) parameter upgrade vector aligned with the gradient of the objective function (ϕ_m). This allows rapid decrease in the objective function where the latter is far above its minimum. A low value of λ_l allows better navigation of an objective function surface in which the minimum lies at the bottom of a narrow valley, the existence of which is an outcome of a high degree of post-calibration parameter correlation. Ideally, λ should fall from a high value to a low value as the iterative parameter adjustment process progresses.

pestpp-ies employs a trial and error procedure to find the best value of λ_l to employ at any iteration of the ensemble-based inversion process. Because of computational constraints, the best λ_l cannot be determined by testing on each member of the ensemble. Instead, λ_l values are tested using only a user-specified subset of the ensemble. Once the best λ_l value (that which produces the largest decrease in the ensemble mean value of ϕ_m) is identified through this process, the remaining members of the ensemble are adjusted using this λ_l. Therefore, in pestpp-ies, both the ensemble size and the size of the subset used for testing the λ_l evaluations play a role in determining the parameter updates.

pestpp-ies also allows the users to discard ensemble members with large simulation times (compared to the mean simulation time) and those that produce ϕ_m values far from the ensemble mean ϕ_m. In this chapter, realizations that took more than four (4) times the ensemble mean simulation time and realizations that had ϕ_m values more than two standard deviations (computed from the ensemble) from the ensemble mean ϕ_m were discarded for practical considerations.

To make the software useful for highly parametrized models, a built-in TCP/IP run manager allows parallel ensemble evaluation allowing the user to distribute their ensemble evaluation across multiple processors in a local cluster or public/private cloud.

3.2.2 Simulation of Groundwater Flow

Groundwater flow in this chapter is simulated using the MODFLOW 6 hydrologic model (Langevin et al. 2021). The groundwater flow equation in MODFLOW 6 is discretized using a control-volume finite-difference (CVFD) method. It solves the partial differential equation (3.11) that describes the distribution of hydraulic head (h) based on Darcy's law for flow through a porous medium.

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right)+Q^{'}{}_{s}=S_{S}\frac{\partial h}{\partial t} \tag{3.11}$$

where K_{xx} (LT^{-1}), K_{yy} (LT^{-1}), and K_{zz} (LT^{-1}) are hydraulic conductivities along the x, y, and z coordinate axes assumed to be oriented along the major axes of hydraulic conductivity. $Q^{'}{}_{s}$ (T^{-1}) is a volumetric flux per unit volume representing sources or sinks, S_S (L^{-1}) is the specific storage of the porous medium, and t (T) represents time. In this chapter, only steady-state flow is considered. Hence the flow equation is reduced to the form shown in equation (3.12).

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right)+Q'_s=0 \tag{3.12}$$

3.2.3 Realization Selection for Ensemble Smoothing

All ensemble-based methods rely on the use of a set of models generated from a given distribution. Usually, when this set is too small, there is a risk that the covariances between parameters and data are not estimated accurately. On the other hand, the computational cost with ensemble-based methods is directly proportional to the number of models in the set. Therefore to minimize the computational cost, one obvious approach is to use a small set. And if we use such a small set, there is a high risk that the models do not cover and represent properly the prior distribution.

For this reason, this section describes a methodology to select a subset of realizations (models) from a larger set with a goal that the subset would produce a response with the same statistical properties as the larger set of models. The method is not new, it was used in previous studies to select models efficiently for uncertainty quantification and to visualize the distribution of models in low dimensions (Scheidt and Caers 2009; Scheidt, Li, and Caers 2018). Here, to our knowledge, this is the first time that the method is used in conjunction with Ensemble Smoothing to ensure that the initial model ensemble covers properly the prior distribution even if a small number of models are used.

Note that the method will also be used to better visualize and understand the evolution of the ensemble during the smoothing process. These are two different applications of the same method but with different objectives.

A large set of realizations is first generated from the prior distribution of

model parameters. Then, the procedure requires the computation of a dissimilarity distance between realizations of the ensemble. Scheidt and Caers (2009) present a comprehensive overview of this process and also a number of metrics to compute the dissimilarity distance (distance). Assuming that a flow simulation was carried out for all the models, the distance between realizations ii and jj, $d(ii, jj)$, is defined as shown below.

$$d(ii, jj) = \|h^{ii} - h^{jj}\|_2 = \left(\sum_{n=1}^{N} |{h_n}^{ii} - {h_n}^{jj}|^2 \right)^{1/2} \tag{3.13}$$

where the model has N cells and ${h_n}^{ii}$ represents the simulated hydraulic head of model ii at cell n.

Smaller distances between realizations imply that their simulated heads are similar. For a set of N realizations, a distance matrix of size $(N \times N)$ is computed. Next, the technique of multi-dimensional scaling (MDS) (Borg and Groenen 1997; Cox and Cox 1994) is applied to project the dissimilarity matrix into lower-dimensional Euclidean space. In this projected space, the Euclidean distance between realizations corresponds to their dissimilarity distance. Thus, "similar" realizations are projected close to each other and "dissimilar" realizations are spread out. In this chapter, the dissimilarity matrix is projected into two-dimensional Euclidean space using the MDS function in the scikit-learn python library (Pedregosa et al. 2011) . This projected matrix is classified into multiple clusters using the K-Means function in scikit-learn (Pedregosa et al. 2011). The centroid of each cluster is identified and the realization closest to the centroid is selected as the representative member of the cluster.

This process is illustrated with an example. For example, let the prior ensemble size be 10'000. A flow simulation would be carried out for each member of this ensemble. A dissimilarity distance matrix of dimension 10'000 $\times$

10'000 would be computed. This distance matrix would be projected into two-dimensional Euclidean space using MDS. If we are interested in selecting 100 realizations to capture the diversity of the ensemble, the MDS-scaled realizations are clustered into 100 groups using the K-Means algorithm. 100 realizations, each corresponding to the centroid of a K-Means cluster will be selected to form a smaller ensemble.

In this process, all the selected realizations represent a cluster of original realizations. But they may not correspond to the same probability density. Some of the selected realizations may represent a zone of the prior distribution corresponding to a high density of probability and may therefore represent a large number of original realizations. Other selected realizations may represent more extreme and more rare simulations. It would be possible, therefore, to weight those different simulations to account for the number of underlying simulations that they represent, when computing covariances and mean values. However, since those realizations will be updated during the smoothing process, it is not obvious how the updating of their initial weights should be computed. Therefore, in the following analysis, we did not account for this effect, but we need to remember that the initial set does not represent an ensemble of equiprobable realizations in a large set.

3.3 Results

In this section, TRIPS and pestpp-ies together are used to generate posterior parameter ensembles for the inverse problem previously described in section 2.3. As previously stated, this inverse problem required the estimation of 17 parameters (locations of 15 traveling pilot points and the hydraulic conductivities of the channel and matrix facies) based on 13 head observations. The target objective

function was set at 22.362 which corresponds to a chi-square distribution with 13 degrees of freedom and a one-sided error (α) of 0.05.

3.3.1 Posterior Parameter Ensemble Estimation with pestpp-ies

To understand the role of the initial ensemble size, ensemble smoothing was carried out with progressively larger ensemble sizes of 100, 1'000, and 10'000 for a maximum of 20 IES iterations. These prior parameter ensembles were generated using the RANDPAR program (Doherty 2018) based on the prior mean and covariance matrix $\mathbf{C_M}$ for this problem (Appendix A.2).

As previously described in section 3.2.1, the Marquardt lambda λ_l has a key role in determining the parameter upgrades between smoothing iterations. In the pestpp-ies implementation, the subset size used for testing various values of λ_l also influence the parameter upgrades. Subset sizes of 10, 20, and 20 were chosen to evaluate the optimal λ_l values for the ensembles with sizes of 100, 1'000, and 10'000. While the user specifies the subset size for λ_l evaluation, pestpp-ies chooses the members of that subset either randomly or based on the highest values of ϕ_m. For this analysis, the random option was chosen.

Realizations that took more than four times the average run time or whose objective function was two standard deviations outside the ensemble mean were discarded for practical considerations. Additionally, some realizations exceeded the target objective function of 22.362. These realizations were excluded from the posterior ensemble. The updating of the traveling pilot point locations with corresponding changes in the geologic facies and simulated heads is illustrated for two realizations (ensemble of 100) in figures 3.1 and 3.2. These two figures depict the traveling pilot points within each channel using different symbols and show how the traveling pilot points change positions to improve the fits

between measured and simulated heads. It should be noted that pestpp-ies performs the parameter updates for these realizations simultaneously. All the pestpp-ies control flags for the largest ensemble are presented in Appendix B.

Prior ensemble size	100	1'000	10'000
Optimal number of smoothing iterations	8	18	16
Realizations discarded	50	480	4091
Realizations with $\phi_m > 22.362$	2	0	4287
Total number of forward model runs	2598	21638	128885
Posterior ensemble size	48	520	1622

TABLE 3.2: Pestpp-ies performance statistics for three different ensembles. For each ensemble, the optimal smoothing iteration, realizations discarded, number of forward model evaluations, sizes of prior and posterior ensembles are shown.

The mean objective function and number of posterior models which match the target criterion for the three ensembles are shown in Figure 3.3. For the three ensembles, the mean objective function decreases gradually in the initial stages. This period of gradual decrease is followed by a period of sharp decrease and then eventual stabilization. For the smallest ensemble of 100, 48 posterior realizations are obtained after eight smoothing iterations and 2'598 forward model evaluations. After the eighth iteration, there was a decrease in the number of the realizations which met the target objective criterion. For the medium sized ensemble of 1'000, 520 posterior realizations are obtained after 18 smoothing iterations and 21'638 forward model evaluations. For the largest ensemble of 10'000, 1'622 posterior models are obtained after 16 smoothing iterations and 128'885 forward model evaluations. The performance statistics for the three ensembles are presented in Table 3.2.

A metric called "posterior percentage" is defined as the ratio of the number of models meeting the target objective criterion to the prior ensemble size. The posterior percentages for the small, medium, and large ensembles are 48%, 52%,

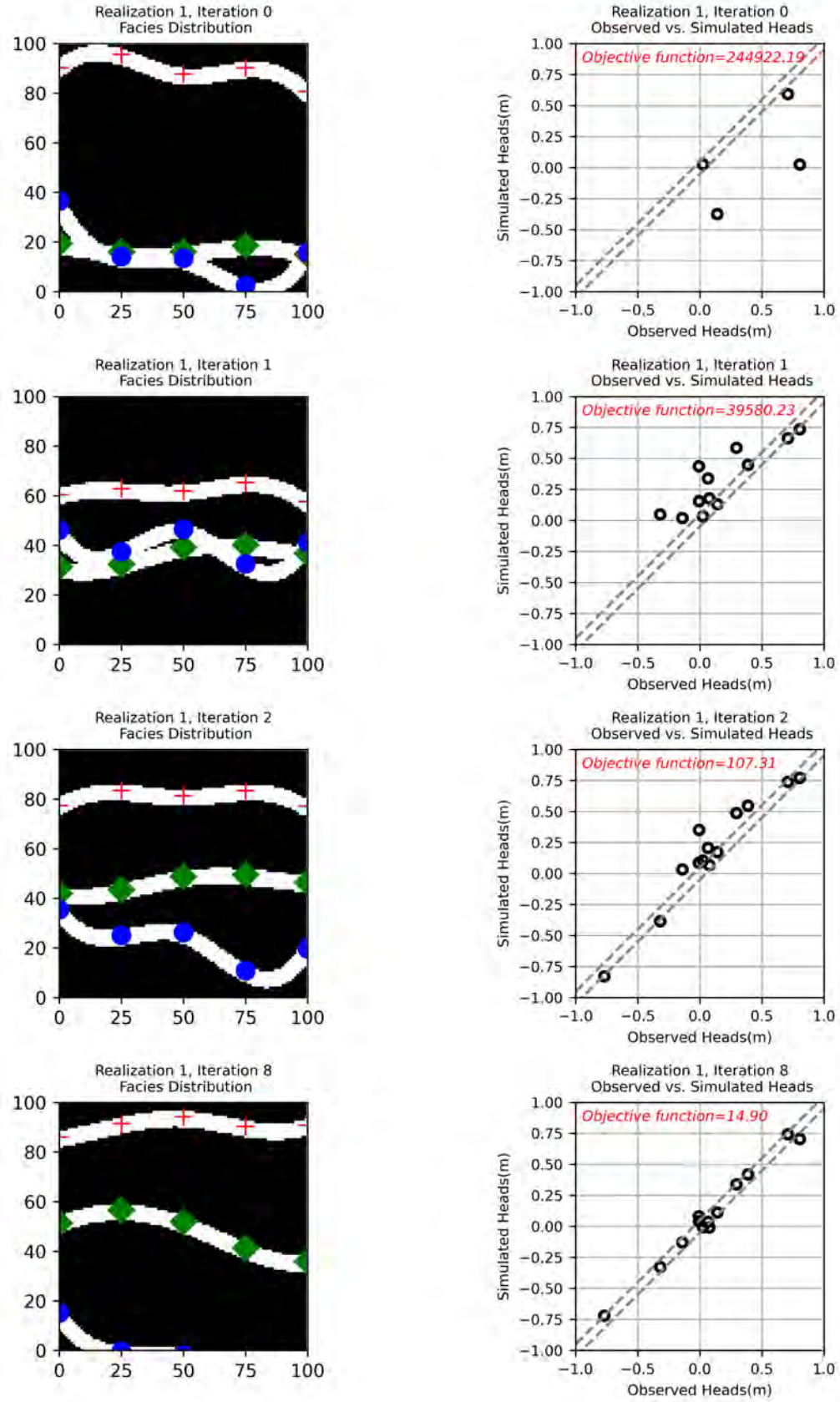


FIGURE 3.1: Facies distribution for realization 1 in the 100 ensemble simulation. The positions of the traveling pilot points comprising the three channels are shown using different symbols.

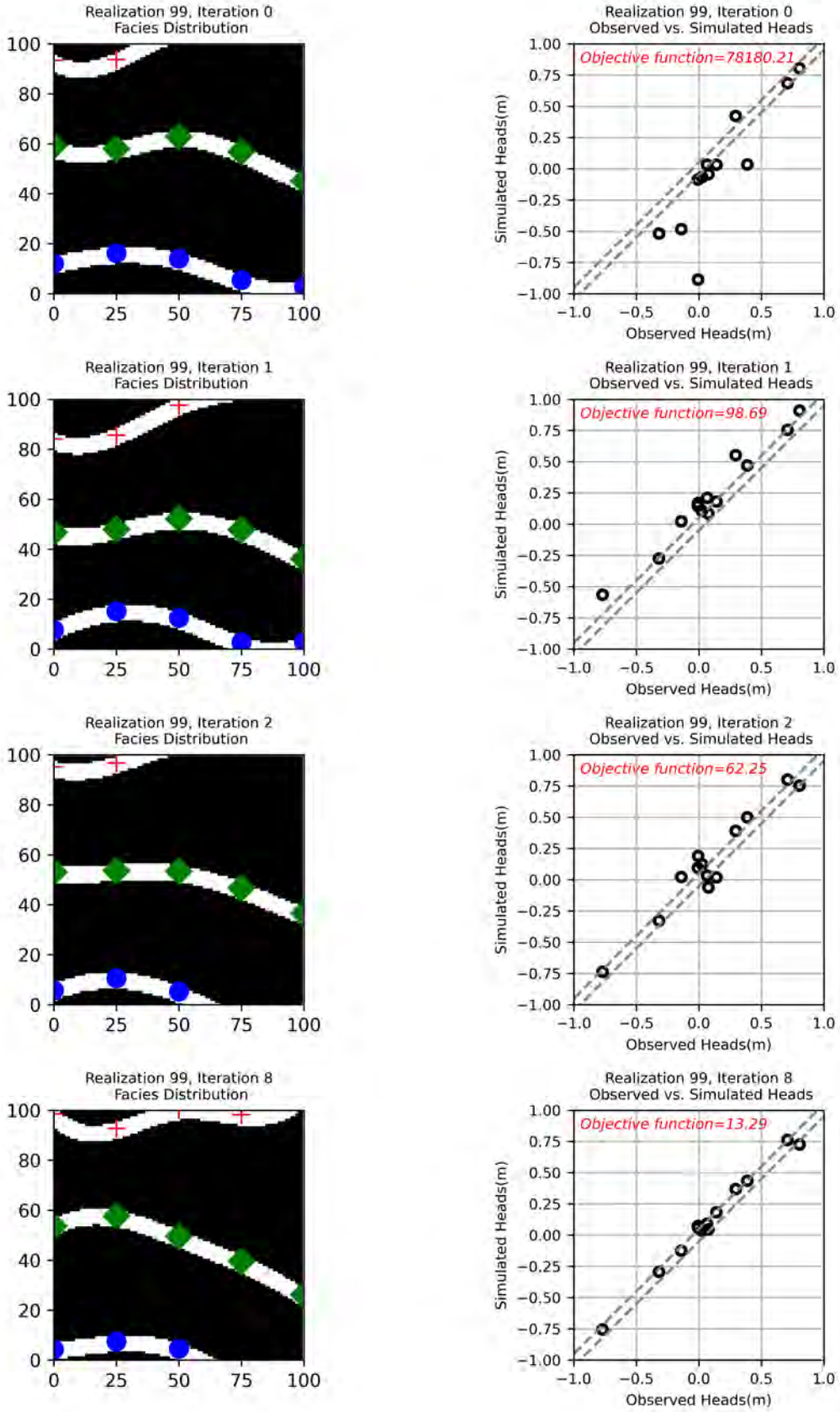


FIGURE 3.2: Facies distribution for realization 99 in the 100 ensemble simulation. The positions of the traveling pilot points comprising the three channels are shown using different symbols.

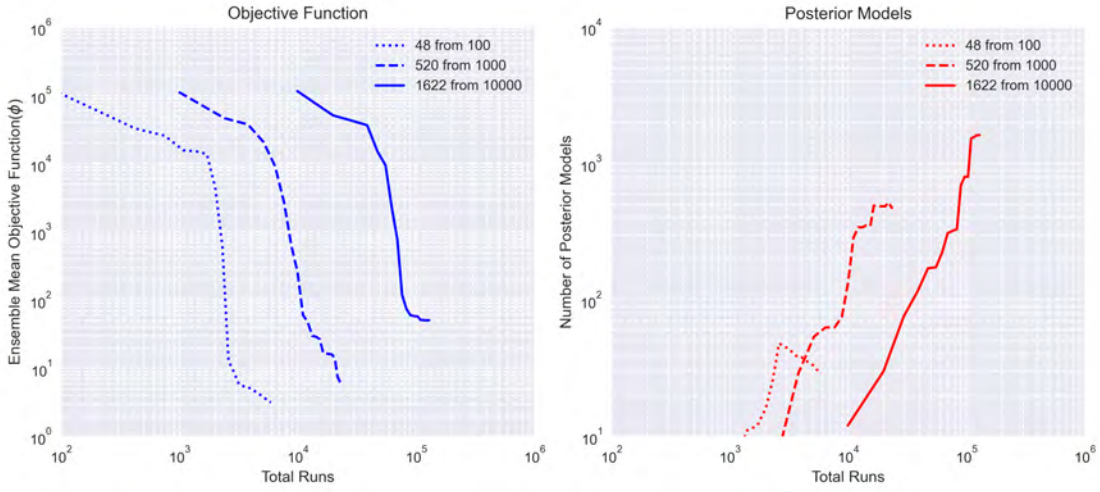


FIGURE 3.3: Ensemble mean objective Function (left) and the number of posterior models (right) for the three ensemble sizes (100, 1'000, and 10'000) plotted against the number of forward model evaluations.

and 16% respectively. The abrupt decrease in this metric for the largest ensemble is noteworthy and worthy of further exploration. At this stage, there are two possible explanations. First, it is possible that larger ensembles have more realizations with extreme parameters that are not amenable to smoothing by pestpp-ies. These realizations either get discarded (objective functions outside two standard deviations from the ensemble mean or simulation times greater than four times the ensemble mean simulation time) or don't get smoothed enough to become part of the posterior ensemble. Another way of saying this is that with a large ensemble, pestpp-ies finds it harder to converge to the ensemble mean. Another possible reason is that the size of the subset (20) that is used for determining the optimal λ_l is too small. Using a larger subset for λ_l search may have identified different λ_l values that may have yielded a higher posterior percentage. The subset size may be proportionally related to the size of the ensemble.

The ensemble characteristics for the three ensembles are shown in Figure 3.4 in a grid with three rows and four columns. The ensemble characteristics

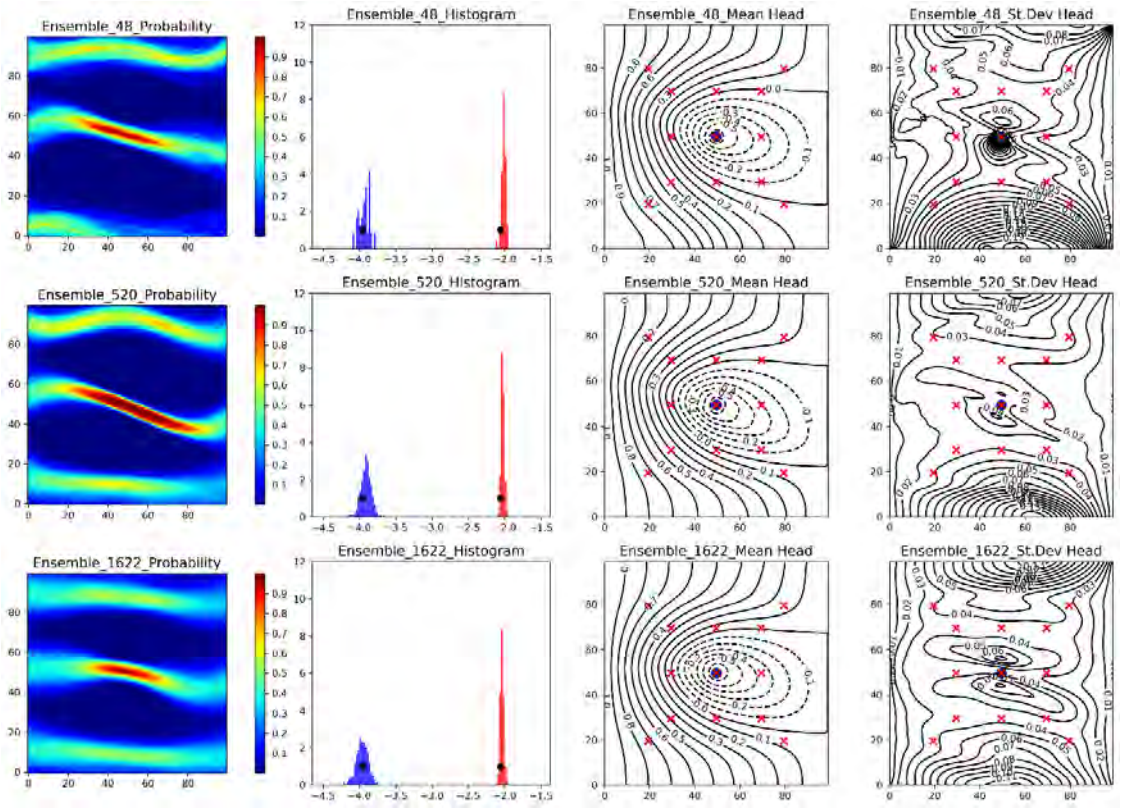


FIGURE 3.4: Parameter ensemble characteristics from ensembles of different sizes, 48 models from 100 (top row), 520 models from 1'000 (second row), and 1'622 models from 10'000 (bottom row). In each row, ensemble parameter probability, histogram of ensemble conductivities (log), mean and standard deviation of simulated heads are shown.

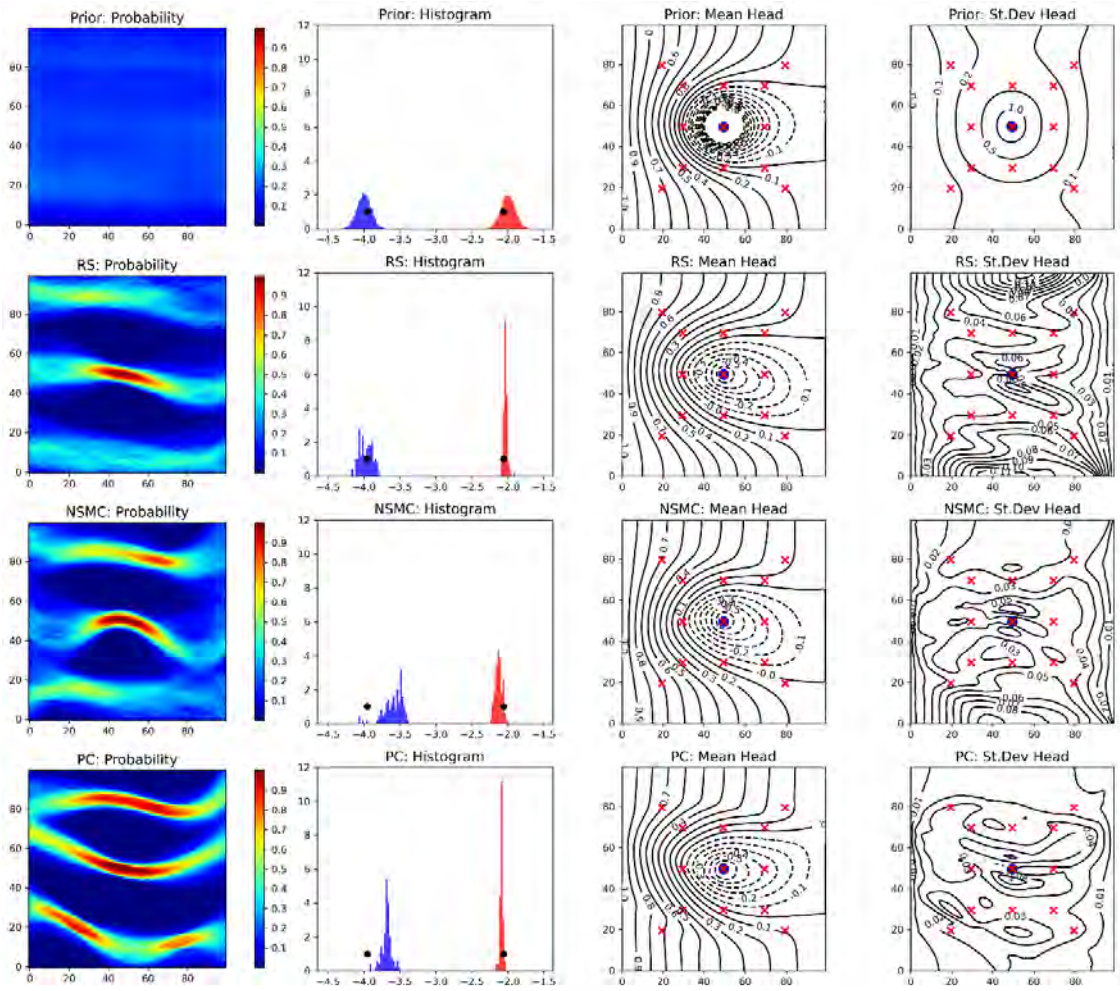


FIGURE 3.5: Parameter ensemble characteristics from various methods, Prior (top row), RS sampling (second row), NSMC (third row) and PCE (bottom row). In each row, ensemble parameter probability, histogram of ensemble conductivities (log), mean and standard deviation of simulated heads are shown. Same as Figure 2.11

for the smallest ensemble (prior size of 100) are shown in the top row and the bottom row contains the ensemble characteristics for the largest ensemble (prior size of 10'000). Since this problem was previously solved in chapter 2 using RS, NSMC, and PC methods, those results are presented again in Figure 3.5 for easy comparison.

In the first column of each row, the probability of finding a channel in a model cell is depicted using a color scale. Lower probabilities are shown in blue while a red color indicates higher probabilities. Ensemble histograms of the log-transformed hydraulic conductivities are shown in the second column. The channel hydraulic conductivities are shown in red and those of the matrix are shown in blue. The hydraulic conductivities of the reference model shown as black dots on the histograms are enclosed by the estimated posterior distribution. Mean and standard deviations of ensemble heads are calculated and shown in the third and fourth columns respectively.

The results indicate that as the number of posterior models in the ensemble increased, the match between the ensemble probability (Figure 3.4) and that estimated by rejection sampling (Figure 3.5) improved. For the largest ensemble, 1'622 posterior members were estimated from a prior ensemble of size 10'000. In this case, the orientation and location of the channels in the ensemble mean hydraulic conductivity distribution for this ensemble closely resemble those in the ensemble mean estimated by RS. The bimodal distributions of hydraulic conductivity histograms also approach the reference posterior distributions obtained by rejection sampling and are not biased anymore. These two observations suggest that this posterior ensemble sampled the true posterior ensemble quite well and better than what was obtained in chapter 2. The use of pestpp-ies seems to improve significantly the results and obtain a more reliable uncertainty quantification. Interestingly, even the smallest posterior ensemble with 48

members got some of the key details correct about the locations and orientation of the channels as well as the values of the hydraulic conductivities. When compared against the ensemble mean hydraulic conductivity distribution from NSMC and PC, all the three pestpp-ies ensembles better resemble the ensemble mean from RS and do not show a systematic bias and underestimation of the hydraulic conductivity contrast between the channels and the matrix.

3.3.2 Ensemble Dissimilarity and Strategic Selection

The similarity/dissimilarity of the three ensembles (prior sizes of 100, 1'000, and 10'000) is explored in this section. For the three ensembles, dissimilarity distance matrices were computed using equation 3.13 for both the prior and posterior realizations. These distances were projected into two-dimensional space using MDS as described in section 3.2.3. The projected distances are shown in Figures 3.6, 3.7, and 3.8.

There are three panels in each of these figures. The left-most panel depicts the projected distance between the prior model realizations. Each point in this panel represents a prior model realization. As the ensemble size increases, the prior distribution begins to resemble a hockey stick with the reference model (shown with a black asterisk) at the tip of the stick. The shape of the prior distribution may be an artifact of the assumed prior covariance matrix and may be a topic for future exploration.

The central panel in these plots is the same as the left panel with one small difference. Each realization in this panel is colored to indicate whether pestpp-ies was able to perturb its parameters to eventually make it a member of the posterior ensemble. Green dots indicate a successful transition from the prior to posterior distribution in this panel while red dots indicate failure. The spatial

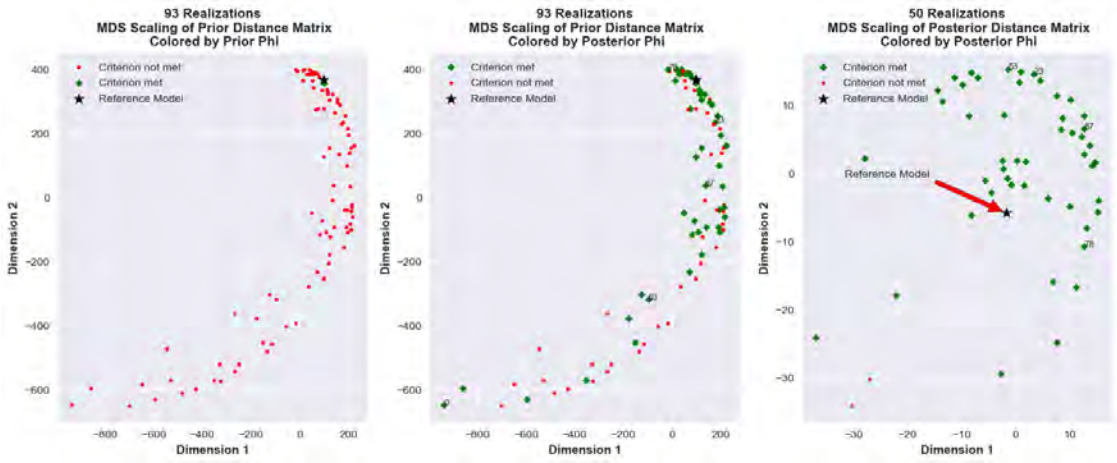


FIGURE 3.6: MDS projected distances for ensemble with an initial size of 100. The rows contain a three-panel plot for each ensemble. The left panel depicts the projected distance between the prior model realizations. The central panel depicts the prior distance colored by whether a prior realization was perturbed enough to become a member of the posterior. The right panel depicts the MDS projected distance between the posterior realizations. Select realizations are labeled to depict transition from the prior to the posterior distributions.

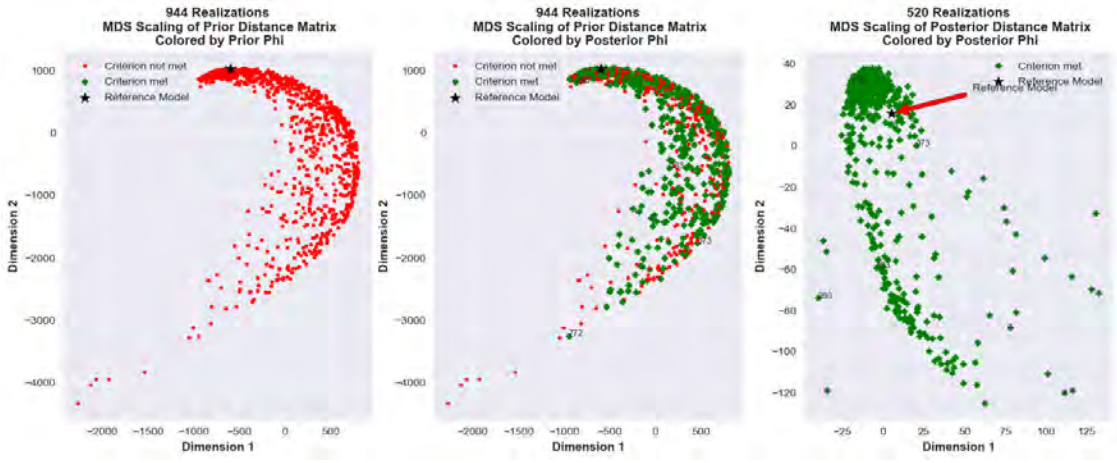


FIGURE 3.7: MDS projected distances for ensemble with an initial size of 1′000. The rows contain a three-panel plot for each ensemble. The left panel depicts the projected distance between the prior model realizations. The central panel depicts the prior distance colored by whether a prior realization was perturbed enough to become a member of the posterior. The right panel depicts the MDS projected distance between the posterior realizations. Select realizations are labeled to depict transition from the prior to the posterior distributions.

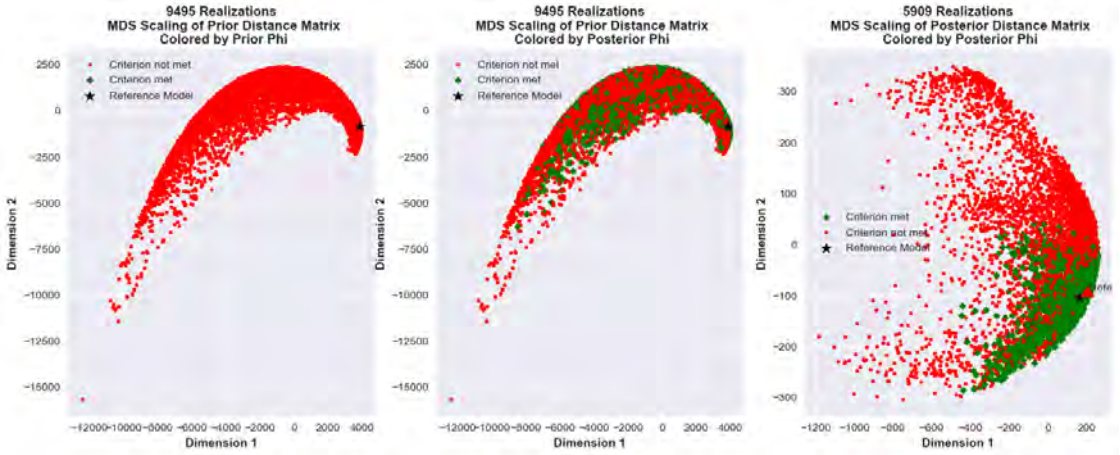


FIGURE 3.8: MDS projected distances for ensemble with an initial size of 10'000. The rows contain a three-panel plot for each ensemble. The left panel depicts the projected distance between the prior model realizations. The central panel depicts the prior distance colored by whether a prior realization was perturbed enough to become a member of the posterior. The right panel depicts the MDS projected distance between the posterior realizations. Select realizations are labeled to depict transition from the prior to the posterior distributions.

extent of the green dots throughout the plot indicates that the pestpp-ies algorithm is not spatially biased to prior realizations that are close to the reference (shown with a black star).

The right-most panel in these plots depicts the distance between realizations of the posterior ensemble. This panel shows that the posterior ensemble encloses the reference model but also has members that are farther from the reference model indicating that the ensemble is not biased towards the reference model. One can also notice prior realizations that were originally very far from the reference coming close to the reference after ensemble smoothing. For example, realization 772 in Figure 3.7 (middle panel) is quite far from the reference model in the prior distribution. However, pestpp-ies brings this realization into the posterior as shown in the right most panel of Figure 3.8. Realization 772 can be seen to the upper-left of the reference model (black star).

In section 3.3.1, the results indicated that an ensemble consisting of 1'622 realizations best represents the posterior ensemble properties estimated using rejection sampling. 128'885 forward simulations were required to generate this ensemble from an initial ensemble size of 10'000. This section explores whether it is possible to replicate this match (visual) with a smaller ensemble. Specifically, could a strategically-selected smaller ensemble, that when subject to pestpp-ies, visually replicate the properties of the posterior ensemble? Numerical experiments performed to explore this possibility are described here.

Three prior ensembles of sizes 100, 1'000, and 5'000 were selected from the largest ensemble of 10'000 realizations by following the selection methodology (MDS and K-Means clustering) described in section 3.2.3. The MDS projections organized into clusters for these three ensembles are shown in Figures 3.9, 3.10, and 3.11. Ensemble smoothing was performed with pestpp-ies for the three selected ensembles.

The results were insightful. For the case where 100 realizations were selected from 10'000, ensemble smoothing could not produce even a single posterior realization. Experience with pestpp-ies suggests that it works best when there are several realizations clustered around the ensemble mean. However, with a small and diverse ensemble, the possibility of finding several realizations clustered around the mean would be low. As a result, pestpp-ies could have struggled to smooth this ensemble.

For the remaining ensembles, the ensemble smoothing yielded 598 (out of 1'000), and 1'274 (out of 5'000) posterior realizations. For these ensembles, the probability of finding a channel, hydraulic conductivity histogram, mean and standard deviation of heads are shown in Figure 3.12. Remarkably, the properties of the ensemble shown in the bottom row (1'274 out of 5'000) are very similar

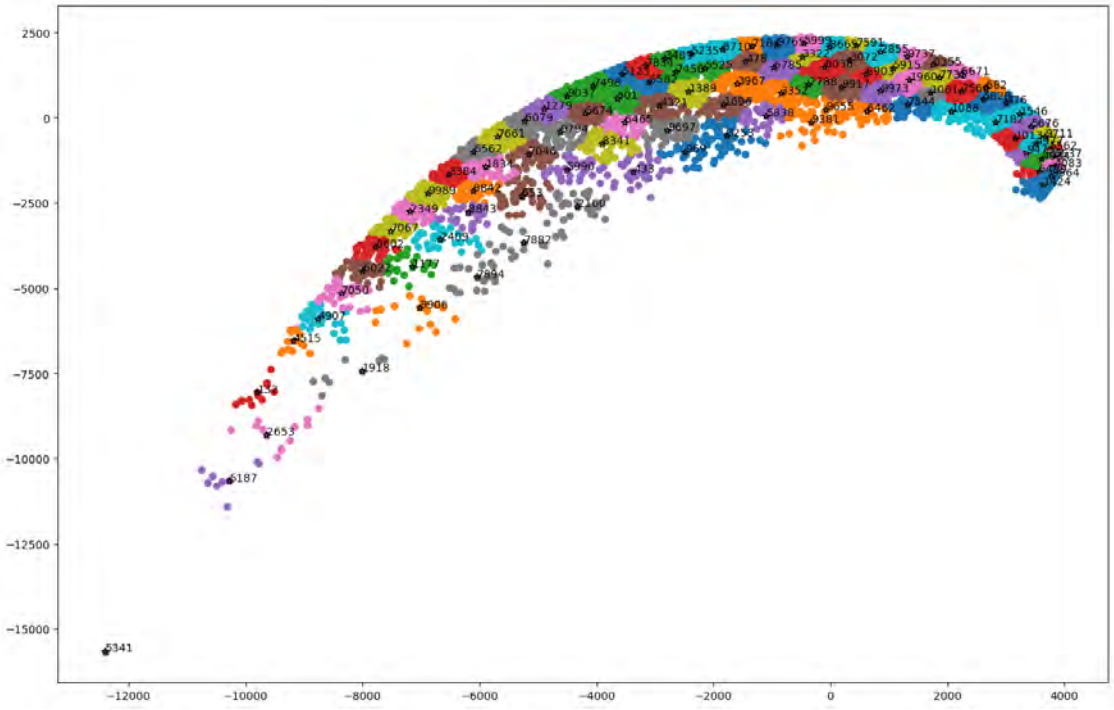


FIGURE 3.9: Model selection using K-Means clustering of an MDS projected Distance Matrix. The MDS-projected distance matrix is partitioned into 100 clusters. Model realizations corresponding to the centroid of each cluster are selected.

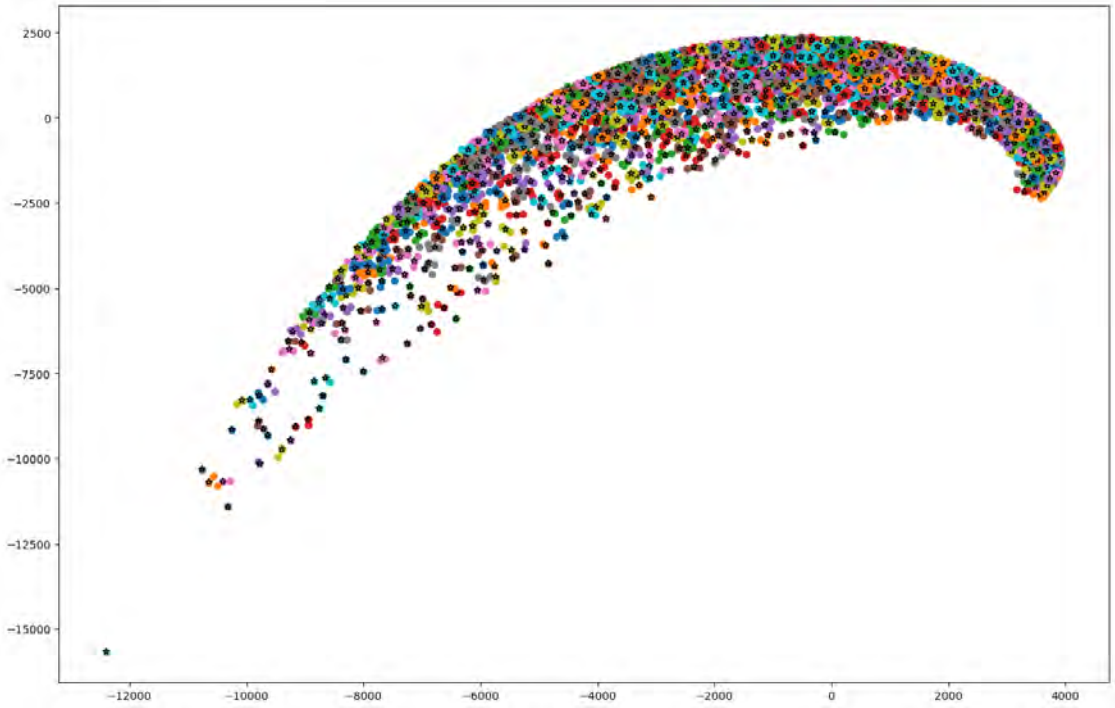


FIGURE 3.10: Model selection using K-Means clustering of an MDS projected Distance Matrix. The MDS-projected distance matrix is partitioned into 1'000 clusters. Model realizations corresponding to the centroid of each cluster are selected.

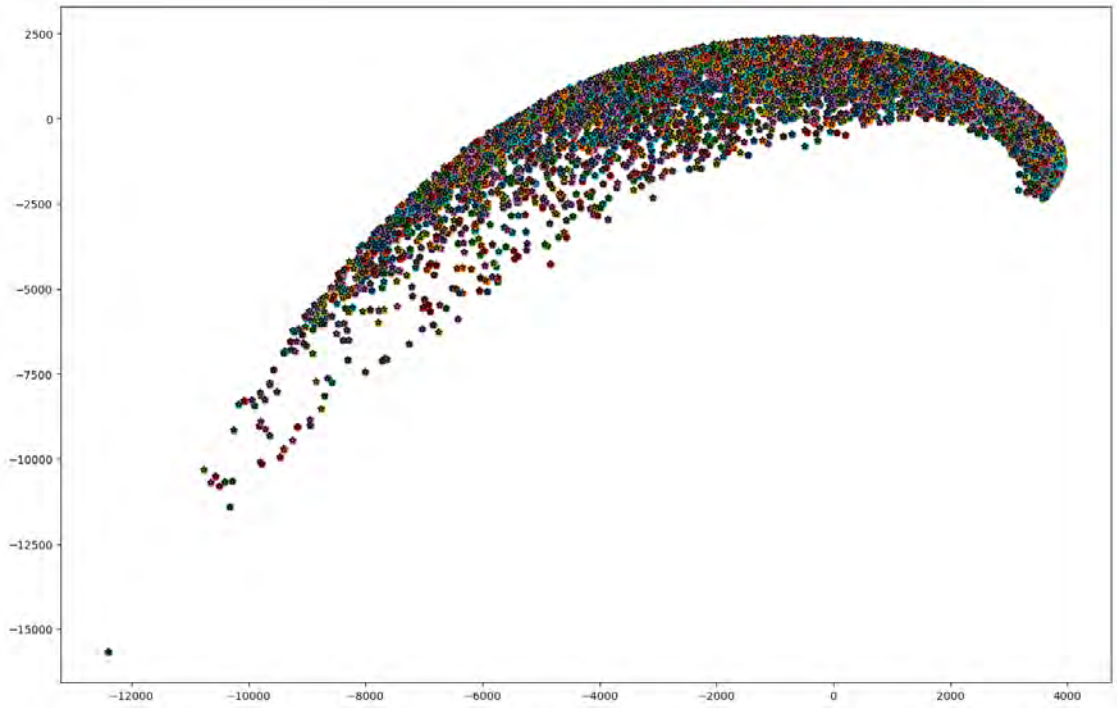


FIGURE 3.11: Model selection using K-Means clustering of an MDS projected Distance Matrix. The MDS-projected distance matrix is partitioned into 5'000 clusters (top row). Model realizations corresponding to the centroid of each cluster are selected.

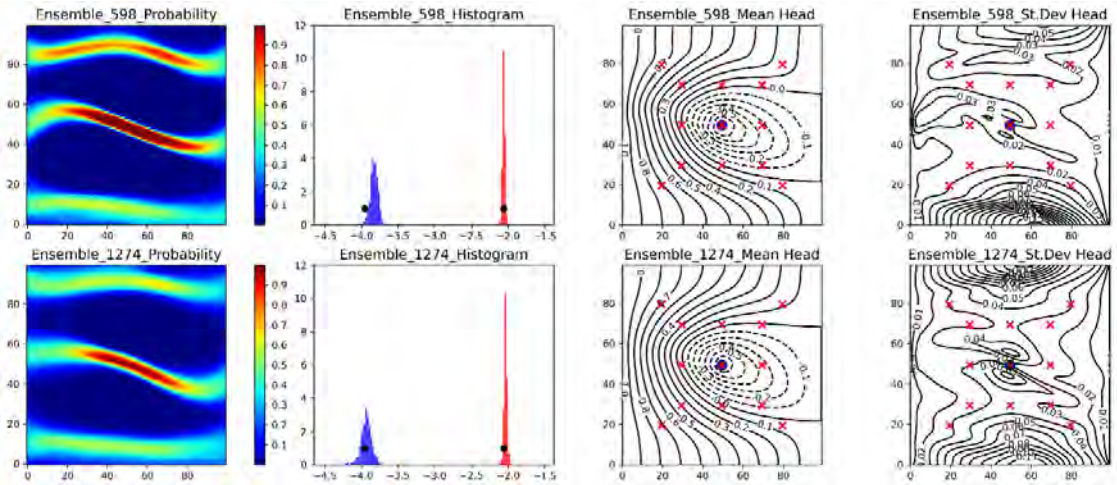


FIGURE 3.12: Parameter ensemble characteristics from the selected ensemble consisting of 598 and 1274 models. Ensemble parameter probability, histogram of ensemble conductivities (log), mean and standard deviation of simulated heads are shown.

to those of the RS ensemble (Figure 3.5, second row) and the larger posterior ensemble (1'622 out of 10'000) shown in Figure 3.4, bottom row). Moreover, 55'785 forward evaluations were needed to generate the posterior ensemble. In other words, model selection helped reduce the required number of forward evaluations by 56%, a huge improvement. It remains to be explored whether other dissimilarity/clustering techniques could lead to a larger improvement.

This numerical exercise suggests that the current approach of selecting a diverse ensemble is feasible. There seems to be a tradeoff between ensemble sizes and the number of posterior realizations. For small ensembles that have realizations far from each other, pestpp-ies seems to struggle in finding even a single posterior realization. The performance improves for larger ensembles. This suggests that ensemble smoothing may require a minimum number of realizations in each cluster to resolve the parameter to observation correlations in a meaningful manner. There could be ways to circumvent the problem with smaller ensemble sizes by modifying the model parameter covariance matrix $\mathbf{C_M}$ by accounting for the clustering of the realizations.

3.4 Summary and Discussion

In this chapter, the pestpp-ies ensemble smoother is used in conjunction with the TRIPS method to sample the posterior distribution for a synthetic problem. Progressively larger prior ensemble sizes are considered until the ensemble mean of the estimated posterior ensemble resembles an ensemble mean independently obtained from RS. Model selection techniques are used to demonstrate that it is possible to select a smaller but diverse prior ensemble whose mean is visually similar to that estimated with a larger prior ensemble.

The synthetic problem in this chapter required the estimation of 17 parameters from the assimilation of 13 observations. While ensemble smoothers were developed to optimally smooth ensembles with large number of parameters, the success of pestpp-ies here is not a foregone conclusion as the synthetic problem is highly non-linear. The first contribution of this chapter is the demonstration that TRIPS and pestpp-ies can be combined to successfully produce a large posterior parameter ensemble for a categorical inverse problem. For this specific case, the posterior distribution is closer to the reference distribution obtained with a rejection sampling method than the ones obtained with the methods presented in the previous chapter. The results are closer to the reference values and the uncertainty is better quantified, while the number of required forward evaluations with pestpp-ies were found to be comparable to other subspace methods like NSMC.

The second main contribution of this analysis is that we show that it is possible to use MDS projection and K-Means clustering to strategically select a smaller number of prior realizations such that the resulting posterior ensemble would still have the properties obtained from a much larger posterior ensemble. This is an important finding to reduce computational costs and could have positive implications for complex real-world problems with large simulation times.

Finally, the use of MDS distance visualization shows how pestpp-ies is able to assimilate data and smooth ensemble realizations throughout the prior distribution.

Tests with varying ensemble sizes indicated that a large posterior ensemble (1'622 realizations) was required to match the properties of an ensemble generated using RS. The nonlinearity of the problem may have dictated the larger size of this ensemble. But it is worth noting that results better than those obtained with NSMC or PC were already obtained with a small ensemble size.

In general, two aspects of the TRIPS and pestpp-ies combination are attractive for solving nonlinear inverse problems for highly parametrized real-world problems. Firstly, the number of forward evaluations to compute the sensitivity matrix **G** is not a function of the model parametrization but is instead influenced by the size of the prior ensemble. Secondly, the parallel implementation of pestpp-ies allows the user to distribute ensemble estimation across several processors or computers to generate a large ensemble in a short period of time.

There are several directions for exploring the implications of the findings in this chapter.

- For the synthetic problem, the largest ensemble yielded only a posterior percentage of 16%. A small subset (size 20) was used to determine λ_l values. Ensemble smoothing tests with larger subset sizes could reveal whether a larger subset for evaluating λ_l values could yield better outcomes. It will also be interesting to explore if there is a relationship between the optimal subset size and the size of the prior ensemble.

- There could be other ways to improve the selection of realizations. It is possible that a better dissimilarity distance matrix could be more efficient. Additionally, there could be better options to cluster the projected distances than the K-Means algorithm.

- The relationship between the prior covariance matrix and the projected distance distribution could be investigated.

- Finally there could be ways to improve the selection of realizations by weighting the selected realizations based on the number of underlying simulations that they represent, when computing covariances and mean values.

The combination of these three ideas (TRIPS, pestpp-ies, and realization/model selection) holds a lot of promise for efficiently solving complex real-world problems.

Chapter 4

TRIPS method in a framework to quantify remedial uncertainty*

4.1 Introduction

In chapters 2 and 3, the TRIPS method was applied on models that only simulated groundwater flow in aquifers with categorical facies. In chapter 3, the combination of the TRIPS method and the pestpp-ies ensemble smoother was found to be effective at inferring the facies positions and sampling the posterior distribution. In this chapter, we explore whether TRIPS and pestpp-ies could be used to assimilate head and concentration observations at contaminated sites to infer facies locations. Why is this important? We offer two reasons.

- Groundwater contamination remains above cleanup goals at hundreds of thousands of hazardous waste sites across the United States (National Research Council 2013) and the rest of the world. According to the U.S. Environmental Protection Agency, expenditures for soil and groundwater cleanup at over 300'000 sites in the United States may exceed $200 billion (not adjusted for inflation) by the year 2033 (United States Environmental

*This chapter corresponds to a manuscript in preparation for Groundwater Monitoring & Remediation.

Protection Agency 2004) – a significant amount. Techniques that can better characterize the contamination extents would lead to targeted cleanup efforts resulting in significant cost savings and benefit human health and the environment.

- Contaminated sites are often underlain by heterogeneous formations with distinct geological facies of different hydraulic conductivities (Figure 4.1). The low-permeability facies in these formations can serve as secondary contaminant sources to higher permeability zones over time (e.g. Zheng and Gorelick 2003; Chapman and Parker 2005; Chapman et al. 2012; Sale et al. 2013; Horst et al. 2017; Farhat et al. 2020). Therefore, characterizing low-permeability facies is important for accurately forecasting concentrations at contaminated sites. Horst et al. (2017) proposed a three-compartment model (Figure 4.2) to conceptualize solute transport in an alluvial aquifer. Their model included: high-permeability facies contributing to rapid advective transport; moderate-permeability facies with slow advective transport/storage; and low-permeability facies where mass is stored.

Managers and stakeholders at these sites rely on various tools, including numerical models, to assess the impact of remedial alternatives, predict future conditions, and manage remediation costs. These numerical models rely on parameter vectors $\mathbf{m}$ and allow to compute a state variable vector $\mathbf{d}$ (containing for example groundwater heads, groundwater flow velocities and/or solute concentrations) in response to initial and boundary conditions by solving the forward problem using partial differential equations representing groundwater flow and solute transport.

Traditionally, the numerical model is initially parametrized based on a conceptual site model that describes the geological heterogeneity, initial/boundary

FIGURE 4.1: Subsurface settings with distinct geological facies
Source: Sale et al. 2013

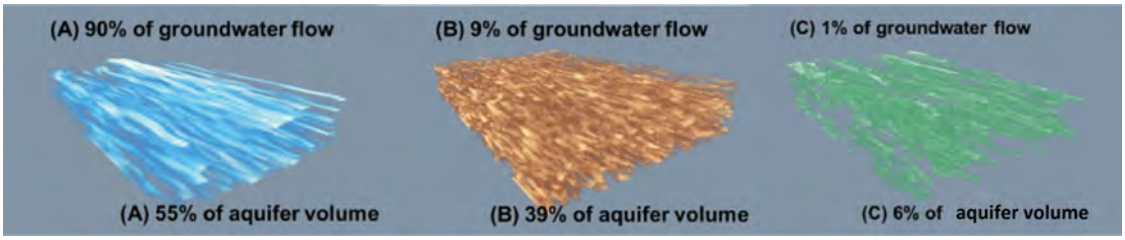


FIGURE 4.2: Separation of an alluvial aquifer based on order-of-magnitude contrasts in groundwater flux: (A) advective/ transport zones, (B) slow advective/storage zones, and (C) storage zones.
Source: Horst et al. 2017

conditions. Subsequently, the values of the parameter vector **m** are calibrated (adjusted) by solving a deterministic inverse problem, such that the model can match measurements of aquifer states (**d**) at the site. Once the match between simulated and measured aquifer states is deemed acceptable, this unique calibrated numerical model is used in a predictive context. A single model, using parameters from the calibrated model, is then often used for making predictions; however, the consequences of an inaccurate prediction can have human important health, ecological, or monetary ramifications.

Because the inverse problem does not have a unique solution (Tarantola 2005), a single model that is calibrated to meet certain goodness-of-fit metrics has very little chances to result in good predictions (Freyberg 1988; Moore and Doherty 2005). Since it is nearly impossible to fully characterize heterogeneous subsurface environments, Rajaram (2016, and many others) suggest that predicting solute behavior in such environments for deterministic predictions at relevant scales necessitates a stochastic approach; which is in line with a typical decision-making agency's expectation that predictions should be accompanied by uncertainty measures that allow risk assessment. Recognizing the non-uniqueness of the inverse problem, practitioners are increasingly using ensembles of calibrated models to characterize predictive uncertainty (see for example Moreno and Paster 2018; Hayley et al. 2019; Ahmmed, James, and Yelderman 2020; Moeck, Molson, and Schirmer 2020).

In chapter 3, the pestpp-ies implementation (White 2018) of the IES was used to assimilate head measurements and condition parameter ensembles for a model representing groundwater flow in a categorical aquifer with two facies. Encouraged by the results presented in section 3.3.1, and also the documented use of iterative ensemble smoothers for conditioning groundwater parameter fields to aquifer state measurements (for example Li et al. 2015; Cao, Li, and

Chen 2018; Knowling, White, and Moore 2019; Lam et al. 2020; White, Knowling, and Moore 2020), the applicability of combining the TRIPS framework with pestpp-ies for assimilating both head and concentration measurements (typically available at contaminated sites), is explored in this chapter.

To this end, the fate and transport of a conservative solute (no sorption/decay) in a synthetic aquifer with two distinct geological facies (permeable channels incised in a low-permeability matrix), is analyzed in this chapter. Since it is well known that groundwater heads alone are not adequate for accurately characterizing aquifer heterogeneity (Rajaram 2016; Fiori et al. 2016; Schilling, Cook, and Brunner 2019), a measurement vector $\mathbf{d}$ comprising of both groundwater solute concentrations and groundwater heads is considered. TRIPS and pestpp-ies are used to assimilate this measurement vector and generate posterior parameter ensembles under various assumptions of heterogeneity and size of the measurement vector $\mathbf{d}$.

Since the categorical parameter ensembles explicitly represent both permeable channels and low-permeability matrix, they are well suited to represent both the rapid advective transport through the permeable channels and the delayed diffusive transport from the solute stored in the low-permeability matrix. The applicability of these posterior parameter ensembles in predicting solute concentrations and mass at a synthetic contaminated site is evaluated. The evaluation also includes an assessment of how the number of measurements and the choice of the geological prior determine the characteristics of the posterior ensemble and the resulting predictions.

The subsequent sections of this chapter are organized as follows. In section 4.2, the various methods used in this chapter are presented. The synthetic problem used in this paper is described in section 4.3. Posterior parameter ensembles for the synthetic problem are presented in section 4.4. The implications of the

number of observations and the choice of the geological conceptual model on the accuracy of predictions is discussed. A summary of findings is presented in section 4.5.

4.2 Methods

The various methods used in this chapter are described in this section. Since the TRIPS method (section 2.1) and the pestpp-ies implementation (section 3.2.1) were previously described, readers are requested to refer to those sections. The formulation of the groundwater flow and transport equations and the simulation method is described in section 4.2.1. The equation for evaluating the spread (variability) of an ensemble is presented in section 4.2.2.

4.2.1 Simulation of Groundwater Flow and Transport

Groundwater flow and solute transport in this chapter is simulated using the MODFLOW 6 hydrologic model (Langevin et al. 2021). The groundwater flow equation in MODFLOW 6 is discretized using a control-volume finite-difference (CVFD) method. It solves the partial differential equation (4.1) that describes the distribution of hydraulic head (h) based on Darcy's law for flow through a porous medium.

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right)+Q^{'}{}_{s}=S_{S}\frac{\partial h}{\partial t} \tag{4.1}$$

where K_{xx} (LT^{-1}), K_{yy} (LT^{-1}), and K_{zz} (LT^{-1}) are hydraulic conductivities along the x, y, and z coordinate axes assumed to be oriented along the major axes of hydraulic conductivity. $Q^{'}{}_{s}$ (T^{-1}) is a volumetric flux per unit volume representing sources or sinks and $S_S(L^{-1})$ is the specific storage of the porous

medium and t (T) represents time. In this chapter, only steady-state flow is considered. Hence the flow equation is reduced to the form shown in equation (4.2).

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right)+Q^{'}{}_{s}=0 \tag{4.2}$$

The dual-porosity/domain formulation (Deans 1963; Coats and Smith 1964; Van Genuchten, Davidson, and Wierenga 1974) for solute transport in groundwater is described by a partial differential equation based on mass conservation as shown in equation (4.3). In this formulation, a heterogeneous aquifer is assumed to comprise of a mobile domain where advective transport is dominant and overlapping immobile domain(s) where molecular diffusion is dominant. A mass-transfer coefficient is used to control mass transfer between the mobile and immobile domains.

$$\begin{aligned}\frac{\partial(S_w\theta C)}{\partial t} &= -\nabla\,.\,(\mathbf{q}C)+\nabla\,.\,(S_w\theta\mathbf{D}\nabla C)+q^{'}{}_{s}C_s+M_s-\lambda_1\theta S_w C-\gamma_1\theta S_w \\ &\quad -f_m\rho_b\frac{\partial(S_w\overline{C})}{\partial t}-\lambda_2 f_m\rho_b S_w\overline{C}-\gamma_2 f_m\rho_b S_w-\sum_{im=1}^{nim}\zeta_{im}S_w(C-C_{im})\end{aligned} \tag{4.3}$$

where S_w is the water saturation (dimensionless) defined as the volume of water per volume of voids, θ is the effective porosity of the mobile domain (dimensionless), C is the volumetric concentration of the solute in the mobile domain expressed as mass of dissolved solute per unit volume of fluid (ML^{-3}), t is time (T), $\mathbf{q}$ is the vector of specific discharge (LT^{-1}), $\mathbf{D}$ is the second-order tensor of hydrodynamic dispersion coefficients (L^2T^{-1}), $q^{'}{}_{s}$ is the volumetric flow rate per unit volume of aquifer for mass sources and sinks (T^{-1}), C_s is the volumetric

solute concentration of the source or sink fluid (ML^{-3}), M_s is the rate of solute mass loading per unit volume of aquifer ($ML^{-3}T^{-1}$), λ_1 is the first-order decay rate coefficient for the liquid phase (T^{-1}), γ_1 is the zero-order decay rate coefficient for the liquid phase ($ML^{-3}T^{-1}$), f_m is the fraction of aquifer solid material available for sorptive exchange with the mobile phase under fully saturated conditions, ρ_b is the bulk density of the aquifer material (ML^{-3}), $\overline{C}$ is the sorbed concentration of solute mass in the mobile domain (M/M), λ_2 is the first-order decay rate coefficient (T^{-1}) for the sorbed phase of the mobile domain, γ_2 is the zero-order decay rate coefficient for the sorbed phase of the mobile domain ($MM^{-1}T^{-1}$), nim is the number of immobile domains, ζ_{im} is the rate coefficient for the transfer of mass between the mobile domain and immobile domain (T^{-1}), and C_{im} is the solute concentration for immobile domain (ML^{-3}).

In this chapter, solute transport for a conservative solute (no sorption, decay) is evaluated assuming a single immobile domain. Dispersion is also ignored for the sake of simplicity. The simplified transport equation for this case is given below in equation (4.4).

$$\frac{\partial(S_w\theta C)}{\partial t} = -\nabla \cdot (\mathbf{q}C) + q'_s C_s + M_s - \zeta_{im} S_w (C - C_{im}) \tag{4.4}$$

4.2.2 Ensemble Spread

One important issue with ensemble-based methods is that the ensemble may collapse and fail to represent correctly the uncertainty about the posterior distribution. To monitor how the spread varies, the ensemble spread (S_{ens}) metric was defined (Houtekamer and Mitchell 1998). It represents the average variance of a log-transformed parameter in an ensemble. For a model with a total of N_m grid

cells and ${\sigma^2}_{ens,i}$ representing the variance estimated from the ensemble of log-transformed parameter realizations at a single grid cell i, the ensemble spread is calculated as shown in equation (4.5).

$$S_{ens} = \sqrt{\frac{1}{N_m} \sum_{i=1}^{N_m} {\sigma^2}_{ens,i}} \tag{4.5}$$

4.3 Synthetic Problem

A synthetic problem was designed to illustrate the decision-making process for evaluating the uncertainty in forecasting/predicting solute concentrations and mass at contaminated sites. This problem, while synthetic, has several elements in common with complex real-world sites: a mix of low and high permeability facies, decade-long solute transit times, head and concentration measurements, sensitivity to the size of the measurement dataset and the assumed geological model of hetrerogeneity. The problem is described below and contains two data sets used for the inversion, plus one data set to assess the quality of predictive uncertainty assessment. Posterior parameter ensembles are developed for this problem using TRIPS in conjunction with pestpp-ies. These parameter ensembles are applied in a predictive context as described in section 4.4.

4.3.1 Hydraulic conductivity field

In this problem, two-dimensional groundwater flow and transport is simulated in a synthetic square shaped confined aquifer (100 m $\times$ 100 m $\times$ 1 m) containing two facies; high-permeability channels incised in a low-permeability matrix.

As in heterogeneous real-world contaminated sites, the low-permeability matrix facies in this problem can serve as a secondary contaminant source to the higher-permeability channel facies over time.

A reference hydraulic conductivity distribution was developed for this aquifer using the TRIPS method. The reference aquifer was assumed to have a maximum of three channels with five TRIPS per channel. The previously-developed prior covariance matrix (A.2) was modified to represent the prior covariance for the transport parameters. The mean and covariance used to generate the prior parameter ensemble are shown in Appendix C. A member of this ensemble with hydraulic conductivities of 8.7×10^{-3} m/s for the channel facies and 1.1×10^{-4} m/s for the matrix facies was selected as the reference hydraulic conductivity distribution (Figure 4.3).

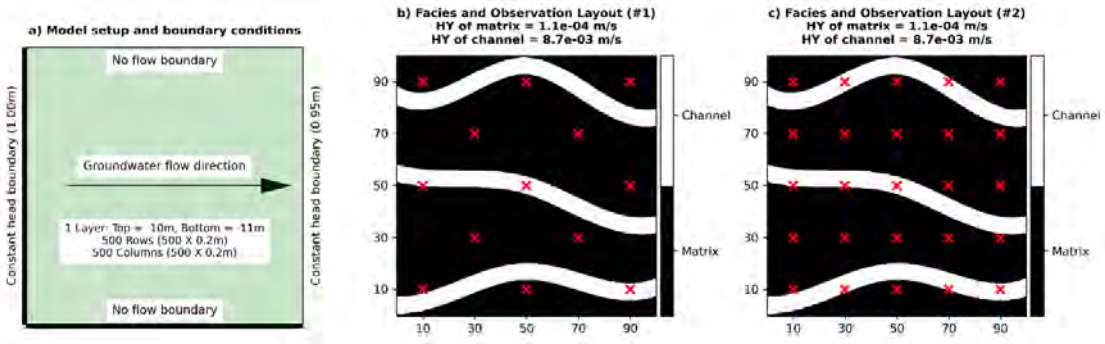


FIGURE 4.3: Model setup, boundary conditions, and monitoring well locations for the groundwater flow problem. The model setup and boundary conditions are shown in panel a. The facies distribution of the reference model and the sparse (13 locations) and dense (25 locations) monitoring networks are shown in panels b and c respectively.

4.3.2 Groundwater Flow and Transport Problem

Constant head boundaries of 1 m and 0.95 m were enforced on the left and right edges of the model respectively to simulate steady-state two-dimensional

groundwater flow from left to right. The low gradient across the model grid (5×10^{-4} m/m) increases solute travel times through the model mimicking those of a larger field scale problem. Since the simulation of solute transport requires a finely spaced grid (Konikow 2011), a numerical grid with 1 layer (thickness of 1 m), 500 rows, 500 columns (row and column spacings of 0.2 m) was used in this model.

Two different monitoring scenarios are considered in this evaluation. In the first monitoring scenario, groundwater heads and concentrations are monitored at 13 monitoring locations in the Site. This scenario is assumed to represent a relatively sparse monitoring network. The second monitoring scenario considers a denser monitoring network with 25 monitoring locations – almost double the number of locations in the sparse scenario. The model grid, boundary conditions, and monitoring locations are shown in Figure 4.3.

A dissolved solute source (dissolved concentration of 10^4 μg/l (micrograms per liter)) is assumed to be present in the second column of the model along all the rows for a period of two (2) years. In the authors' experience, constant concentration boundaries are one of the types of boundary conditions used to represent continuous solute sources at real-world sites. After two years, the source is assumed to be removed instantaneously, and the remaining solute is allowed to transit through the aquifer and exit through the downgradient specified head boundary.

The mobile-immobile domain formulation of the transport equation, as implemented in MODFLOW 6 (section 4.2.1, equation 4.4) is used to simulate solute transport. For this problem, the solute is assumed to be conservative, and sorption and decay processes are not represented. Additionally, hydrodynamic dispersion is ignored in this evaluation for the sake of simplicity. Incidentally,

for field-scale problems, inherent numerical dispersion from the grid discretization may be large (Peaceman 1977). An attempt was also made to explicitly incorporate diffusive transport, but very little difference was observed in the simulated concentrations with and without diffusion. This observation is consistent with the findings of Chapman et al. (2012), who demonstrate that simulation of diffusive mass transport required centimeter scale grid cells with very small time steps. Ideally, a more refined grid with very small timesteps would be better suited for representing the diffusion processes. However, that spatial and temporal refinement would have resulted in very long simulation times complicating the task of uncertainty analysis, which requires several model evaluations. As a compromise, the dual-domain formulation discussed in section 4.2.1 with the upstream finite difference scheme and adaptive time stepping is used in this evaluation.

In the reference model, the total porosity of the channel facies (θ_{ch}) is assumed to be 35% and the total porosity of the matrix facies (θ_{mx}) is assumed to be 43%. According to Payne, Quinnan, and Potter (2008), these porosities are representative for high and low permeability materials. The mobile porosities for the channel ($\theta_{m,ch}$) and matrix facies ($\theta_{m,mx}$) were assumed to be 10% and 1% respectively. The mass-transfer coefficient (ζ_{im}) between the mobile and immobile domains is assumed to be 10^{-3}/day.

In the reference model, the hydraulic conductivity contrast ratio between the channel and matrix facies is 79 $\left(\frac{8.7 \times 10^{-3}}{1.1 \times 10^{-4}}\right)$. This ratio compares well with the observations of Zheng and Gorelick (2003) who noted that when the hydraulic conductivity in permeable channels approaches 100 times that of the matrix's hydraulic conductivity, calculated concentration distributions start exhibiting non-gaussian patterns with pronounced tails which are incidentally observed in real-world contaminated sites.

Solute transport was simulated for a period of thirty years assuming steady-state groundwater flow. It was assumed that measurement data exists for the first four years and that the remaining 26 years represent future (predictive) conditions. Because of the steady-state flow assumption, a single head measurement was taken at the monitoring locations (13 or 25 depending on the scenario). At each monitoring location, solute concentrations were measured quarterly, resulting in a total of 16 measurements (4×4) at each location over a 4-year period. Given the head and concentration measurements over the first four years, the inverse problem requires the estimation of parameter ensembles that honor the measured data. The objective function, (ϕ_m), minimized for the inverse problem is represented mathematically as shown below in equation 4.6.

$$\phi_m = [\mathbf{d_{h,mes}} - \mathbf{d_{h,sim}}]^T[\mathbf{d_{h,mes}} - \mathbf{d_{h,sim}}] + [\mathbf{d_{c,mes}} - \mathbf{d_{c,sim}}]^T[\mathbf{d_{c,mes}} - \mathbf{d_{c,sim}}] \tag{4.6}$$

$\mathbf{d_{h,mes}}$ and $\mathbf{d_{h,sim}}$ represent the measured and simulated heads respectively. These vectors have sizes of 13 and 25 for the sparse and dense monitoring scenarios respectively. $\mathbf{d_{c,mes}}$ and $\mathbf{d_{c,sim}}$ represent the log_{10} transformed measured and simulated concentrations. These vectors have sizes of 208 and 400 for the sparse and dense monitoring scenarios respectively. Measurement noise was ignored in this evaluation for simplicity and hence the covariance matrix of the measurement noise error is not shown in the equation.

Since aquifer heterogeneity is a key driver of solute transport (Zheng and Gorelick 2003; Konikow 2011), two different approaches to model the geological heterogeneity for the inverse problem (two different priors) are evaluated: the discrete and the continuous cases.

4.3.3 Discrete case

In the discrete case, a categorical hydraulic conductivity distribution is created using the TRIPS method assuming three channels with five TRIPS per channel resulting in a total of 15 TRIPS.

Besides the locations of the TRIPS, the channel and matrix hydraulic conductivities are also estimated using pestpp-ies. Two additional parameters are also estimated. These were the dual domain mass-transfer coefficient (ζ_{im}) and the mobile porosity for the channel facies ($\theta_{m,ch}$). The mobile porosity for the matrix facies, $\theta_{m,mx}$, was constrained to be $1/10^{th}$ of the $\theta_{m,ch}$.

The discrete approach has a clear advantage as it assumes a geological prior for the aquifer heterogeneity that is the same as the reference model.

4.3.4 Continuous case

One of the aims of the numerical experiments presented in this chapter is to investigate how the inverse problem and uncertainty quantification will perform if instead of using the proper conceptual model for the geological heterogeneity, a simpler multiGaussian model is used. This mismatch of the geological priors is designed to illustrate real-world conditions where one often works with a limited understanding of the subsurface heterogeneity and has to rely on this type of multiGaussian models.

In this case, the hydraulic conductivity is evaluated at 250'000 (500×500) grid block locations and is assumed to vary between 5×10^{-5} m/s and 5×10^{-2} m/s. Since the number of model evaluations for calculating the parameter sensitivity matrix (**G**, equation 3.3) in the pestpp-ies implementation is not governed by the number of parameters, it is feasible to estimate the sensitivity matrix for a large number of parameters efficiently.

The parameters of the multiGaussian model were obtained by analyzing the reference hydraulic conductivity field in order to match as well as possible the statistics of the reference even if the conceptual heterogeneity is different. The hydraulic conductivity distribution for the reference model was log-transformed and a spherical semivariogram with a sill of 0.583, a range of 140 along the X direction, and an anisotropy ratio of 4 (Figure 4.4) was fitted to the empirical variogram computed from the reference hydraulic conductivity field.

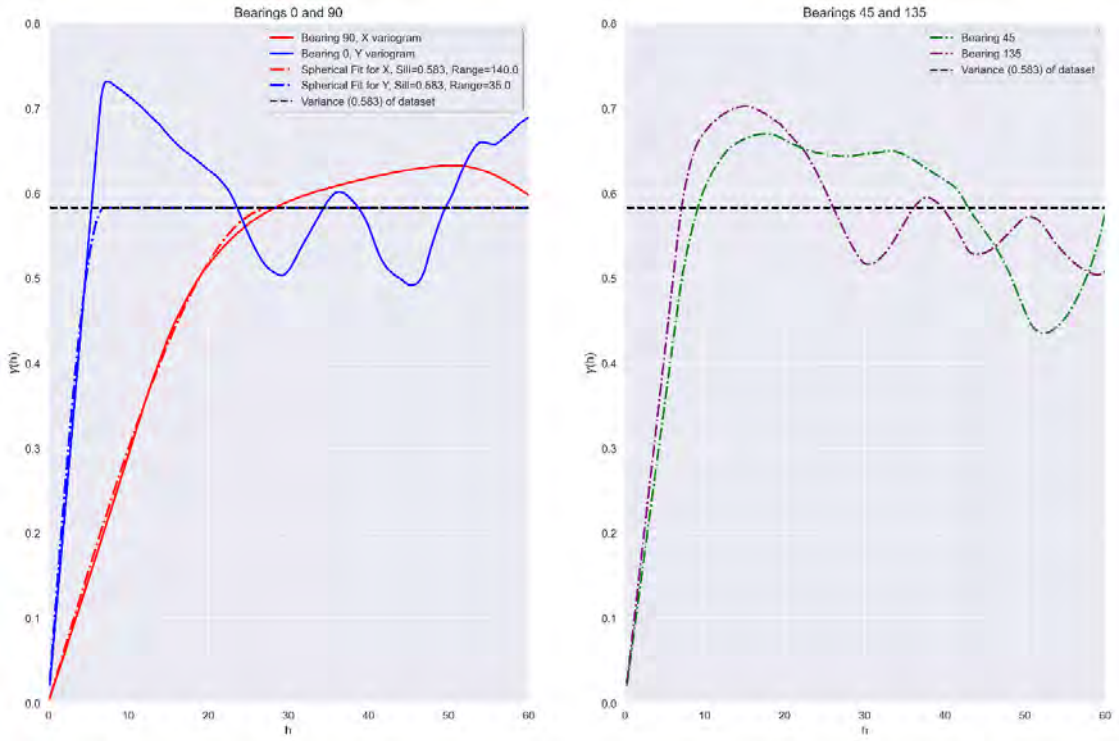


FIGURE 4.4: Variogram of the log-transformed hydraulic conductivity distribution in the reference model. The left panel shows the fitted variogram (dashed lines) and the calculated variogam (solid lines) for bearings of 0 and 90. The right panel shows the calculated variogram for bearings of 45 and 135. An anisotropy of 4 is inferred between the X (bearing of 0) and Y (bearing of 90) directions.

This variogram was then used to generate realizations of the logarithm of the hydraulic conductivity using the FIELDGEN utility from the PEST groundwater utility suite (Doherty 2018). FIELDGEN generates these parameter distributions

based on the Sequential Gaussian Simulation technique (Deutsch and Journel 1998).

Three additional parameters are also estimated for the continuous case. These were the dual domain mass-transfer coefficient (ζ_{im}), the mobile porosity for the channel facies ($\theta_{m,ch}$), and a parameter identifying the transition between channel and matrix facies (K_{tran}). Model cells with estimated hydraulic conductivities greater than K_{tran} were assumed to be channel cells and assigned a total porosity of 35% (θ_{ch}) and the estimated mobile porosity for the channel facies ($\theta_{m,ch}$). Model cells with estimated hydraulic conductivities less than K_{tran} were assumed to be facies cells and assigned a total porosity of 43% (θ_{mx}) and the estimated mobile porosity for the matrix facies $\theta_{m,mx}$. To simplify the problem, $\theta_{m,mx}$ was constrained to be $1/10^{th}$ of the $\theta_{m,ch}$.

4.3.5 Summary

Heterogeneity	Heterogeneity Parameters	Other Parameters
Continuous	250000	ζ_{im}, $\theta_{m,ch}$, K_{tran}
Discrete	17	ζ_{im}, $\theta_{m,ch}$

Monitoring Network	Head Measurements	Concentration Measurements
Sparse	13	208 (13×16)
Dense	25	400 (25×16)

TABLE 4.1: Parameters and Observations for Inverse Problem

In summary, four different cases are considered, two with a sparse data set and with continuous or discrete hydraulic conductivity fields, and again two with a dense data set with continous or discrete hydraulic conductivity fields.

Table 4.1 summarizes the number of parameters and observations used in these four cases.

In subsequent sections of this chapter, parameter ensembles that honor the measured heads and concentrations in these different cases are presented and analyzed.

4.4 Results

The results are presented in three sub-sections. In sub-section 4.4.1, posterior parameter ensembles are presented for the continuous case. In sub-section 4.4.2, posterior parameter ensembles are presented for the categorical case. And finally, sub-section 4.4.3, presents the results of the predictive uncertainty analysis.

4.4.1 Results for the Continuous Case

The pestpp-ies software was used to estimate the posterior parameter ensemble. As shown in Table 4.1, this inverse problem required the estimation of 250'003 parameters based on 13 (or 25) head observations and 208 (or 400) concentration observations. An initial ensemble of 100 realizations was progressively smoothed for a maximum of 15 iterations. Models with realizations that took more than four times the average run time or whose objective function was two standard deviations outside the ensemble mean objective function (defining the misfit between the simulated and measured values) were discarded to keep the estimation stable.

The ensemble spread (section 4.2.2) and mean objective function are shown in Figure 4.5 for both monitoring scenarios (sparse and dense). The ensemble

mean objective function (left panel of Figure 4.5) decreases rapidly for the first three iterations and then decreases gradually. The ensemble spread (right panel of Figure 4.5) keeps decreasing indicating that the parameter ensemble becomes less diverse with increasing iterations. Based on this graph and an evaluation of the simulated heads and concentrations, parameters corresponding to iterations 13 and 10 were selected to be the posterior ensembles for the sparse and dense monitoring scenarios respectively. The process of selecting the iteration was a compromise between ensemble diversity and matching the measured data. The posterior ensemble for the sparse monitoring scenario had 45 members while the posterior ensemble for the dense monitoring scenario had 65 members. Some of the ensemble statistics for the selected iterations are summarized in Table 4.3.

Parameter	Ensemble	Mean	Standard Deviation
ζ_{im} $(log_{10}\ day^{-1})$	Reference	-3.00	-
	Prior	-2.94	0.6851
	Posterior (Sparse)	-3.00	0.0207
	Posterior (Dense)	-3.00	0.0379
$\theta_{m,ch}$	Reference	0.10	-
	Prior	0.10	0.0280
	Posterior (Sparse)	0.14	0.0090
	Posterior (Dense)	0.12	0.0157

TABLE 4.2: Parameter Ensemble Characteristics for the Continuous Case

Monitoring	Iteration	Size	Mean ϕ_m	Std. dev. ϕ_m
Sparse	13	45	59275.5	4984.97
Dense	10	65	570669.0	73343.9

TABLE 4.3: Ensemble statistics for the continuous case

The match between the measured and simulated heads for the prior and posterior ensembles are shown in Figure 4.6. Assimilation of the measured data led to a better match between the measured and simulated heads for the posterior ensemble. However, there seems to be room for improvement. The match

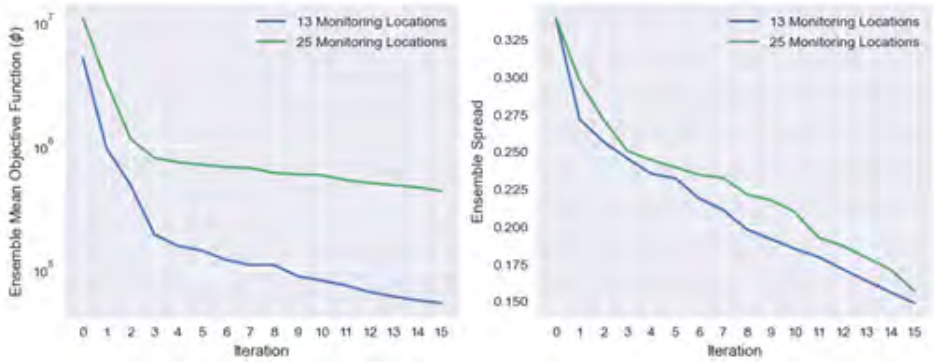


FIGURE 4.5: Ensemble Spread vs. mean objective function for the continuous parametrization case. The left panel shows the ensemble mean objective function and the right panel shows the ensemble spread.

for the sparse monitoring scenario looks better than that for the dense monitoring scenario, mostly because there were fewer observations to match. The match between the measured and simulated concentrations for the prior ensemble are shown in Figure 4.7. The simulated prior concentrations (green lines) exhibit a variability spanning several orders of magnitude while encompassing the measured concentrations (red line). The matches for the sparse and dense monitoring scenarios are shown in Figures 4.8 and 4.9 respectively. In both the monitoring cases, data assimilation led to a reduction in the variability of simulated concentrations. For both monitoring scenarios, good fits are obtained at several monitoring locations (row 50 and column 250 for example). However, the ensemble smoothing process struggles to match the concentrations at locations where the solute concentration drops from the maximum value of 10^4 μg/l rapidly first and then gradually afterwards (row 450 and column 450 for example). This inability to simultaneously match all the concentration observations is most likely a consequence of incorrectly assuming a continuous parameter distribution model.

The characteristics of the estimated ensembles are shown in Figure 4.10. The

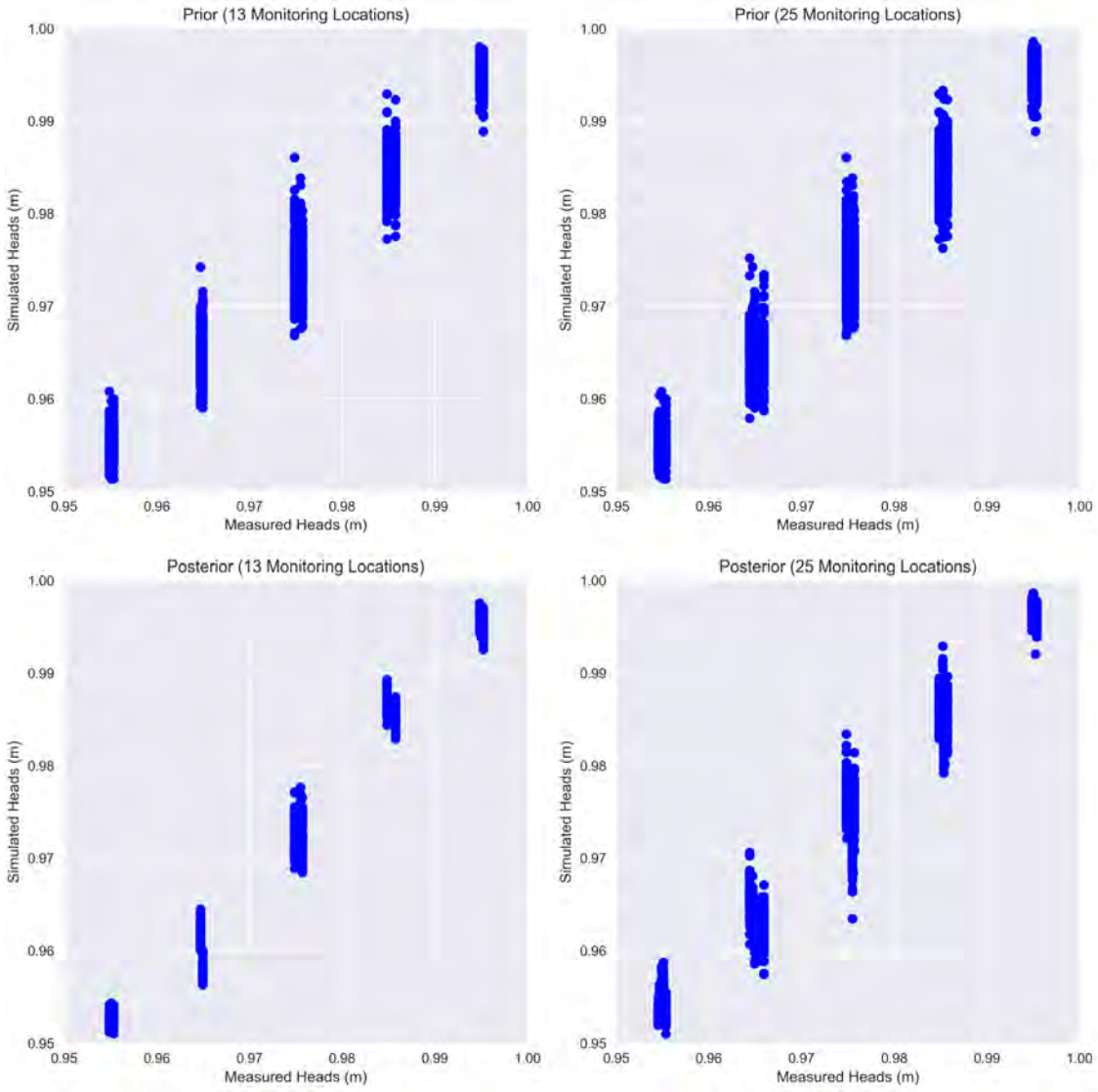


FIGURE 4.6: Measured vs. simulated Heads for the continuous parametrization case. The top panel shows the relationship between the measured and simulated heads in the prior ensembles for the sparse (13) and dense (25) monitoring scenarios. The bottom panel shows the relationship between the measured and simulated heads in the posterior ensembles for the sparse and dense ensembles.

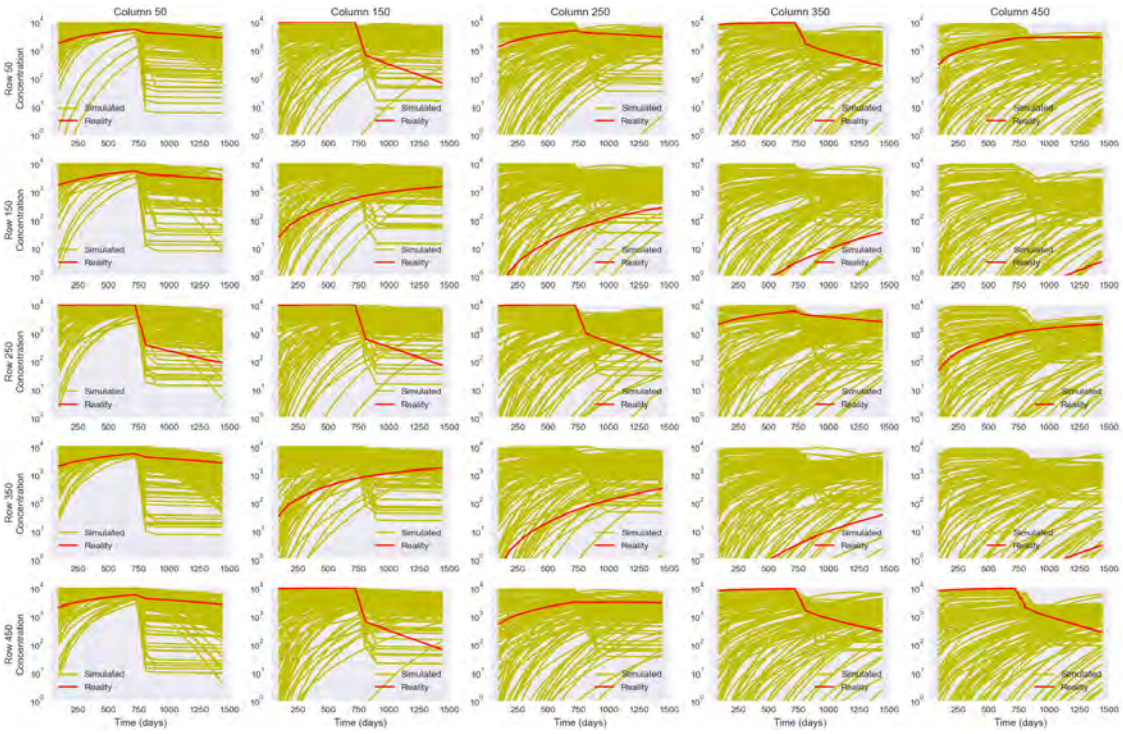


FIGURE 4.7: Measured vs. Simulated concentrations (prior ensemble) for the continuous parametrization case. The measured concentrations are shown using solid red lines while the simulated concentrations are shown using green lines. The model row and column for each observation are shown in the plot.

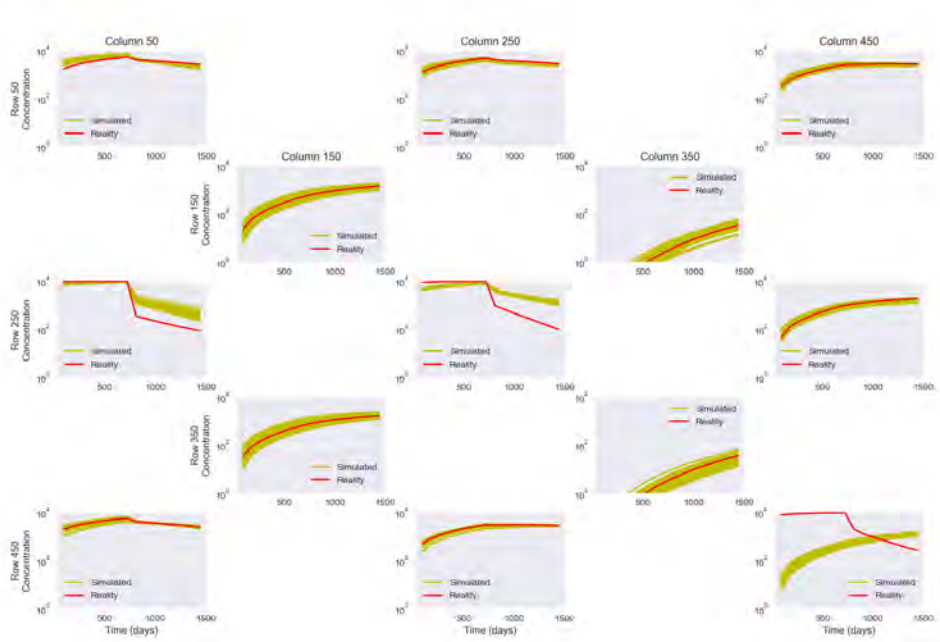


FIGURE 4.8: Measured vs. simulated concentrations (posterior ensemble for the sparse monitoring scenario) for the continuous parametrization case. The measured concentrations are shown using solid red lines while the simulated concentrations are shown using green lines. The model row and column for each observation are shown in the plot.

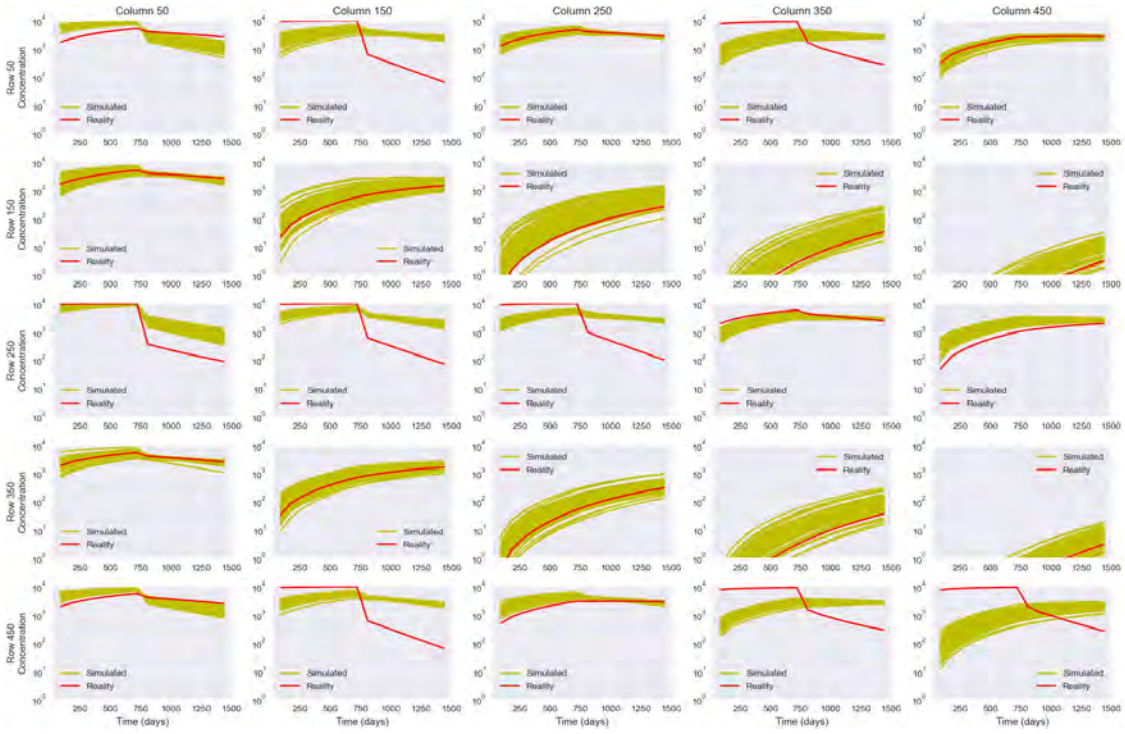


FIGURE 4.9: Measured vs. simulated concentrations (posterior ensemble for the dense monitoring scenario) for the continuous parametrization case. The measured concentrations are shown using solid red lines while the simulated concentrations are shown using green lines. The model row and column for each observation are shown in the plot.

hydraulic conductivity and simulated heads for the reference model are shown in the top row. The characteristics of the prior ensemble (mean and standard deviation of the hydraulic conductivity and simulated heads) are shown in the second row. The characteristics of the posterior ensemble for the sparse and dense monitoring scenarios are shown in the third and fourth rows respectively. While neither monitoring scenario could faithfully produce the characteristics of the original ensemble, they were still able to produce certain characteristics of the reference model. The ensemble mean hydraulic conductivity for both monitoring scenarios has highs in the center and along the north and south edges and lows in the center – a pattern observed in the reference model. However, the mean hydraulic conductivity from the posterior ensemble estimated using the denser monitoring network seems to represent the orientation of the channel facies better. The ensemble standard deviation of hydraulic conductivity for the posterior ensemble was lower than that of the prior ensemble. The ensemble mean head exhibited some differences when compared against the reference head distribution particularly in the southern regions of the aquifer. The ensemble standard deviation of simulated head for the posterior ensembles is lowest near the left and right edges because of the specified boundary heads.

The simulated concentrations for the reference model are compared against the ensemble mean concentrations of the posterior ensembles obtained from the sparse and dense monitoring networks in Figure 4.11. Concentrations are compared at times of 0, 2, 5, 10, 20, and 30 years. The reference concentrations are shown in the first column. The high-permeability regions allow the solute to quickly exit the domain whereas the low-permeability regions retain solute mass for decades. The posterior ensembles generated from both the sparse and dense monitoring networks exhibit similarities in terms of where the high and low concentrations occur in the model. However, the concentration results from the

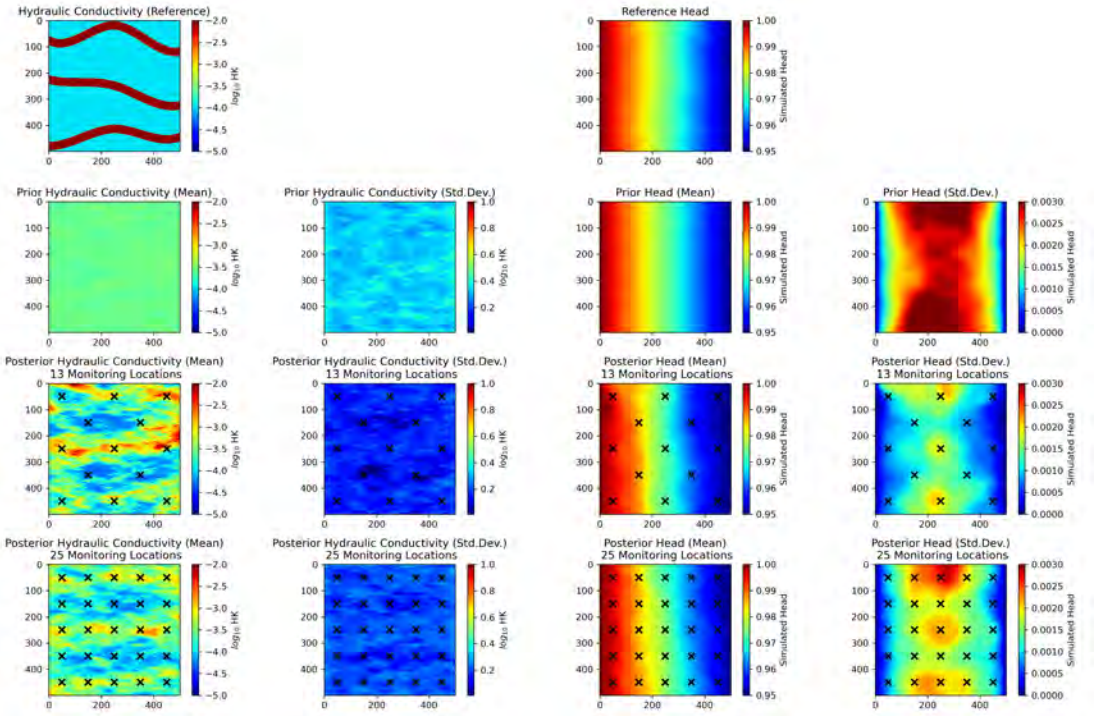


FIGURE 4.10: Ensemble Hydraulic Conductivity and Heads for the continuous parametrization case. The hydraulic conductivity and simulated heads for the reference model are shown in the top row. The characteristics of the prior ensemble (mean and standard deviation of the hydraulic conductivity and simulated heads) are shown in the second row. The characteristics of the posterior ensemble for the sparse and dense monitoring scenarios are shown in the third and fourth rows respectively.

denser monitoring network compare better because the orientation of the channel facies in this ensemble is closer to the reference distribution. The ensemble characteristics for the other parameters are shown in Table 4.2. In Table 4.2, the ensemble mean and standard deviation for mass transfer coefficient (ζ_{im}) are estimated on log-transformed parameters.

4.4.2 Results for the Categorical Case

The pestpp-ies software was used to estimate the posterior parameter ensemble. As shown in Table 4.1, this inverse problem required the estimation of 19 parameters based on 13 (or 25) head observations and 208 (or 400) concentration observations. An initial ensemble of 100 realizations was progressively smoothed for 10 iterations. Models with realizations that took more than four times the average run time or whose objective function was two standard deviations outside the ensemble mean objective function were discarded to keep the estimation stable.

The ensemble spread (section 4.2.2) and mean objective function are shown in Figure 4.12 for both monitoring scenarios (sparse and dense). The objective function (left panel of Figure 4.12) decreases with successive smoothing iterations; this behavior is in in contrast to the objective function behavior of the continuous ensembles where the decrease stalled after a few iterations. The ensemble spread (right panel of Figure 4.12) also keeps decreasing indicating that the parameter ensemble becomes less diverse with increasing iterations. Based on this graph and an evaluation of the simulated heads and concentrations, parameters corresponding to iterations 6 and 7 were selected to be the posterior ensembles for the sparse and dense monitoring scenarios, respectively. As in the continuous case, this selection was performed to maintain diversity in the

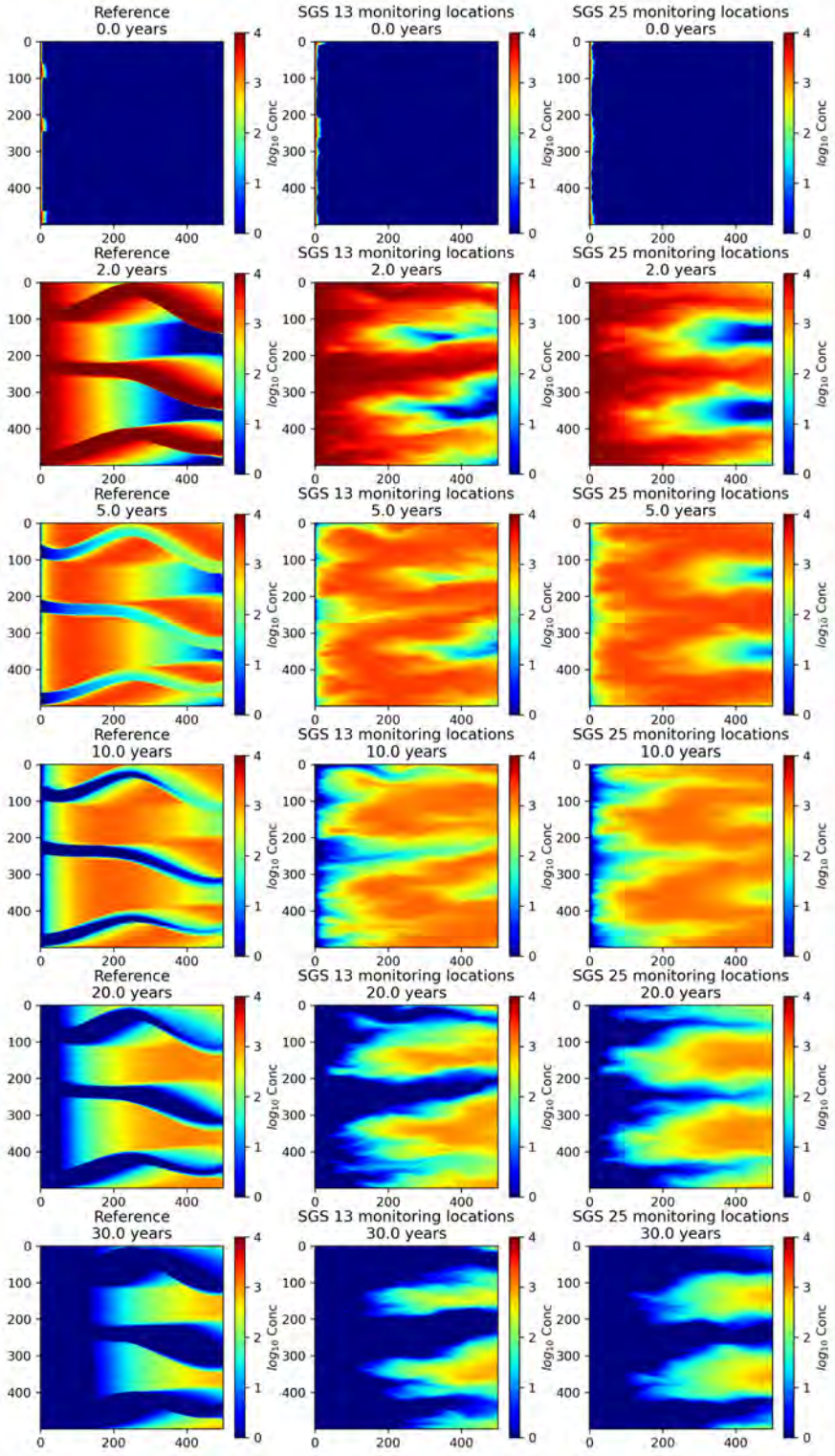


FIGURE 4.11: Simulated concentrations for the continuous parametrization case. Reference concentrations, ensemble average for the sparse and dense monitoring scenarios are shown in the first, second, third columns respectively. Each row represents a time between 0 and 30 years where the source was introduced at 0 years and removed after 2 years.

ensemble while also matching the measured data. The posterior ensemble for the sparse monitoring scenario had 41 members while the posterior ensemble for the dense monitoring scenario had 45 members. Some of the ensemble statistics for the selected iterations are summarized in Table 4.5.

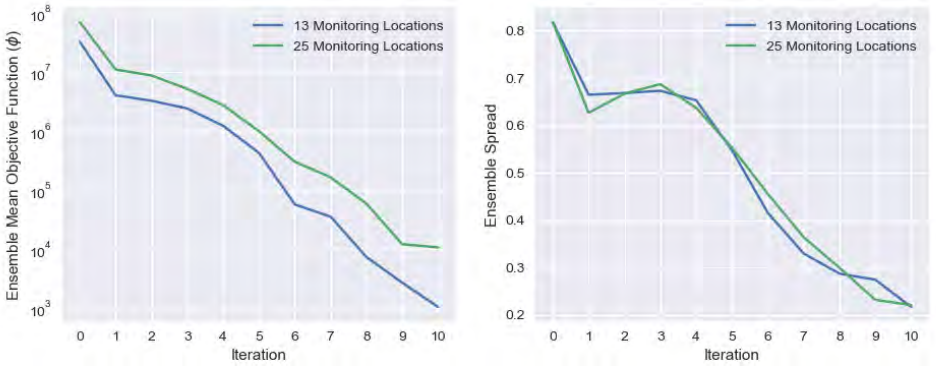


FIGURE 4.12: Ensemble Spread vs. mean objective function for the discrete parametrization case. The left panel shows the ensemble mean objective function and the right panel shows the ensemble spread.

The match between the measured and simulated heads for the prior and posterior ensembles are shown in Figure 4.13. Ensemble smoothing led to a better match between the observed and simulated as seen in the bottom panel. Moreover, the match for the discrete case is a lot better than the corresponding match for the continuous case shown in Figure 4.6. The match between the measured and simulated concentrations for the prior ensemble are shown in Figure 4.14. The simulated prior concentrations (green lines) exhibit a variability spanning several orders of magnitude while encompassing the measured concentrations (red line). The matches for the sparse and dense monitoring scenarios are shown in Figures 4.15 and 4.16 respectively. In both the monitoring cases, ensemble smoothing led to a reduction in the variability of simulated concentrations and led to good matches at all the monitoring locations.

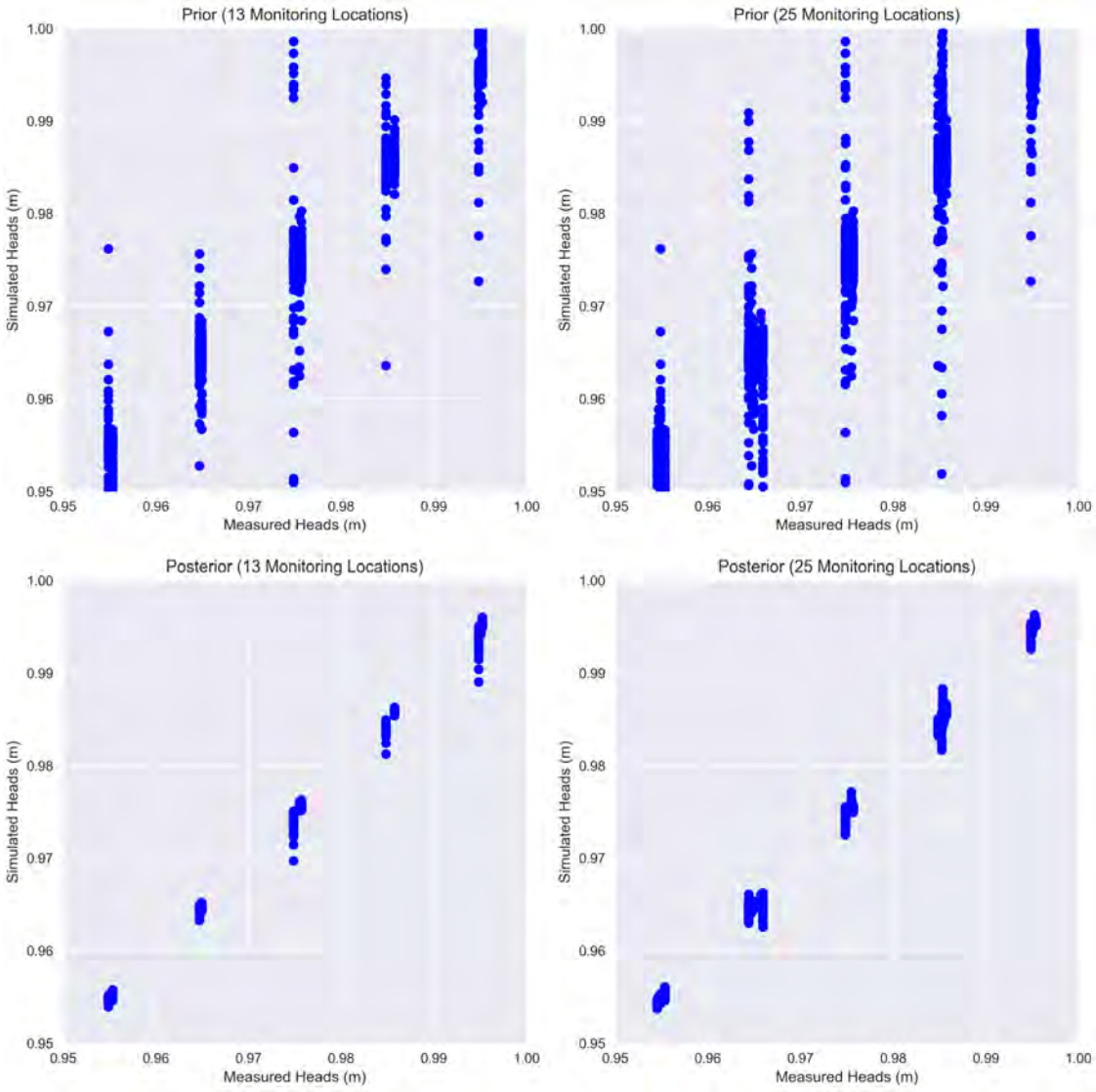


FIGURE 4.13: Measured vs. simulated heads for the discrete parametrization case. The top panel shows the relationship between the measured and simulated heads in the prior ensembles for the sparse (13) and dense (25) monitoring scenarios. The bottom panel shows the relationship between the measured and simulated heads in the posterior ensembles for the sparse and dense ensembles.

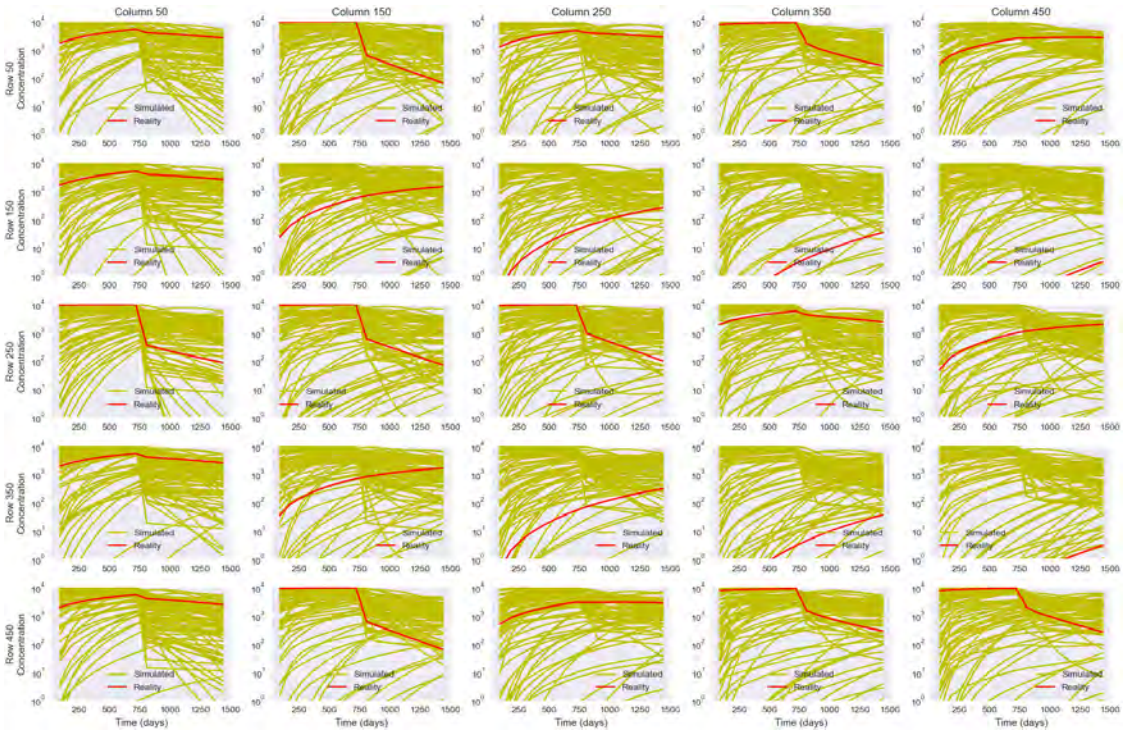


FIGURE 4.14: Measured vs. simulated concentrations (prior ensemble) for the discrete parametrization case. The measured concentrations are shown using solid red lines while the simulated concentrations are shown using green lines. The model row and column for each observation are shown in the plot.

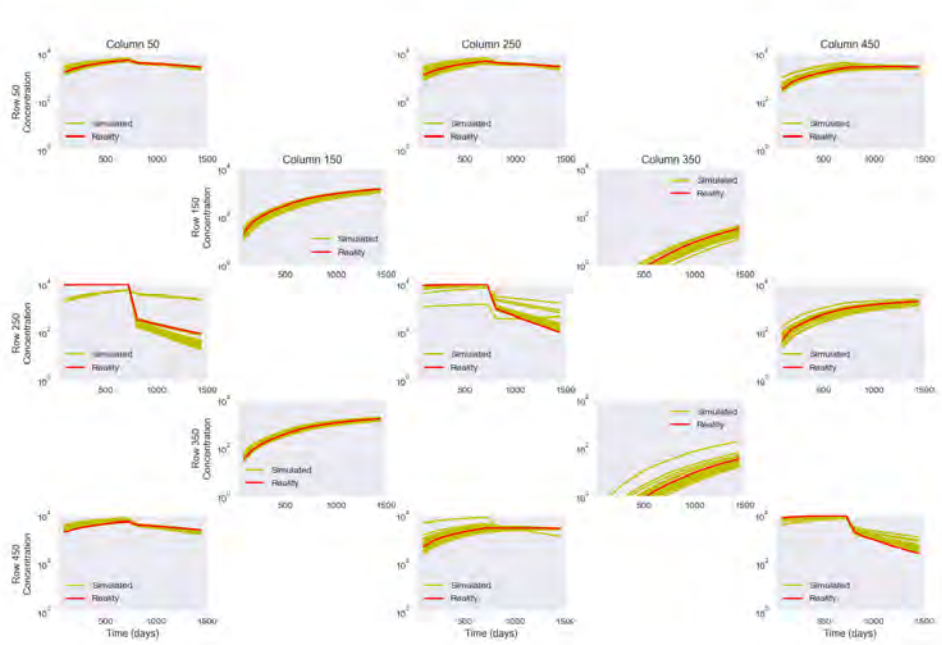


FIGURE 4.15: Measured vs. simulated concentrations (posterior ensemble for the sparse monitoring scenario) for the discrete parametrization case. The measured concentrations are shown using solid red lines while the simulated concentrations are shown using green lines. The model row and column for each observation are shown in the plot.

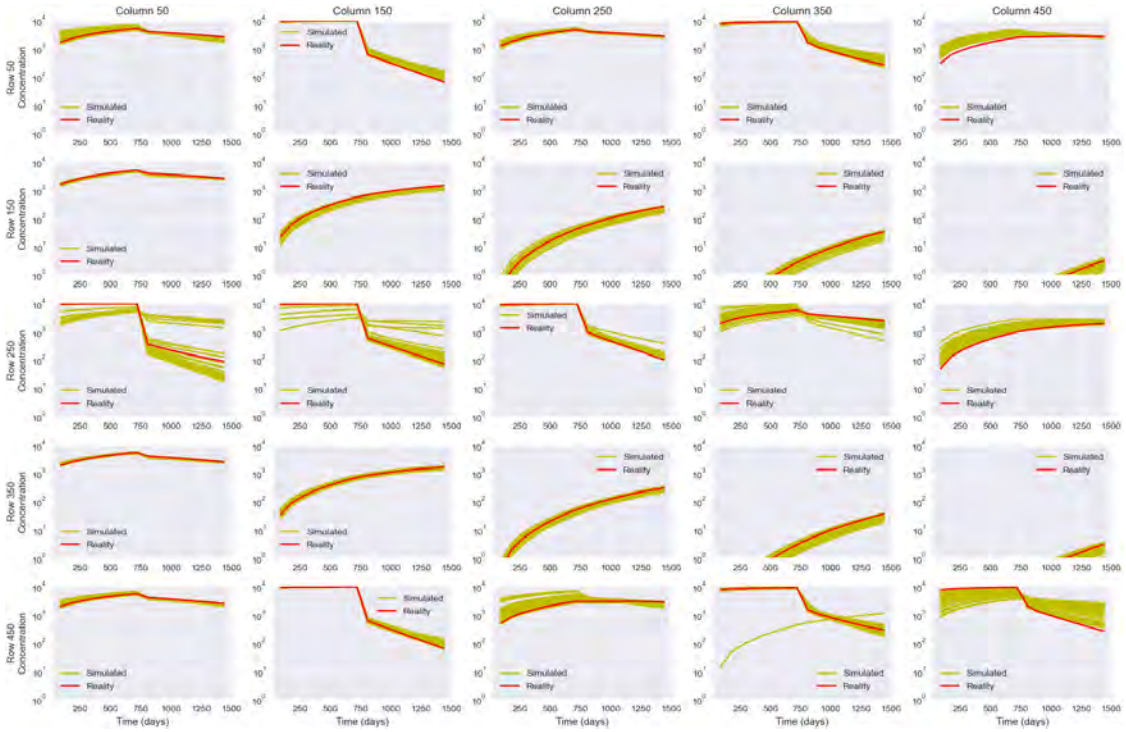


FIGURE 4.16: Measured vs. simulated concentrations (posterior ensemble for the dense monitoring scenario) for the discrete parametrization case. The measured concentrations are shown using solid red lines while the simulated concentrations are shown using green lines. The model row and column for each observation are shown in the plot.

The characteristics of the estimated ensembles are shown in Figure 4.17. The hydraulic conductivity and simulated heads for the reference model are shown in the top row. The features of the prior ensemble (mean and standard deviation of the hydraulic conductivity and simulated heads) are shown in the second row. The characteristics of the posterior ensemble for the sparse and dense monitoring scenarios are shown in the third and fourth rows, respectively. The ensemble mean hydraulic conductivities and simulated heads closely resemble the reference distribution for both the sparse and dense monitoring networks. The standard deviation for the posterior ensembles is highest at the boundaries of the channel facies suggesting that the ensemble members differ in the thickness of the channels. The ensemble standard deviation of simulated head is very small indicating similarity in the simulated heads.

The simulated concentrations for the reference model are compared against the ensemble mean concentrations of the posterior ensembles obtained from the sparse and dense monitoring networks in Figure 4.18. Concentrations are compared at times of 0, 2, 5, 10, 20, and 30 years. The reference concentrations are shown in the first column. The posterior ensembles generated from both the sparse and dense monitoring are very much like the concentrations from the reference model. The ensemble characteristics for the other parameters are shown in Table 4.4. In Table 4.4, the ensemble mean and standard deviation for the mass transfer coefficient and hydraulic conductivities of the channel and matrix are estimated on log-transformed parameters.

4.4.3 Predictive Evaluation

In this section, the posterior ensembles developed in sections 4.4.1 and 4.4.2 are used to support the remedial decision-making process. The prior and posterior

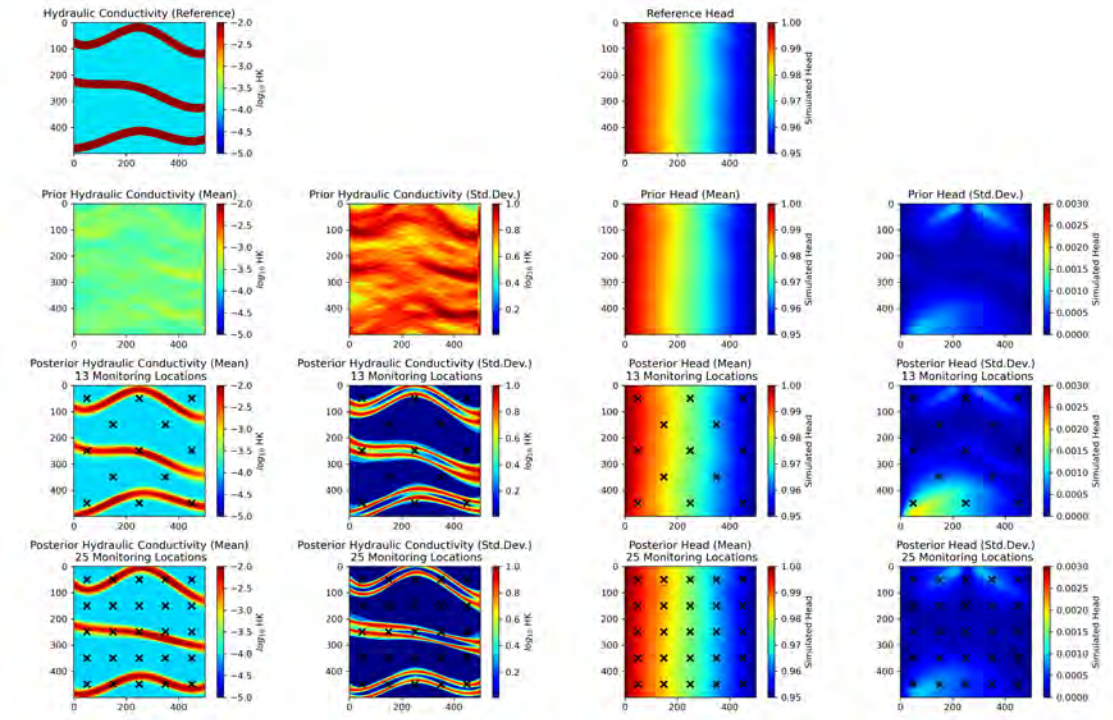


FIGURE 4.17: Ensemble Hydraulic Conductivity and Heads for the discrete parametrization case. The hydraulic conductivity and simulated heads for the reference model are shown in the top row. The characteristics of the prior ensemble (mean and standard deviation of the hydraulic conductivity and simulated heads) are shown in the second row. The characteristics of the posterior ensemble for the sparse and dense monitoring scenarios are shown in the third and fourth rows respectively.

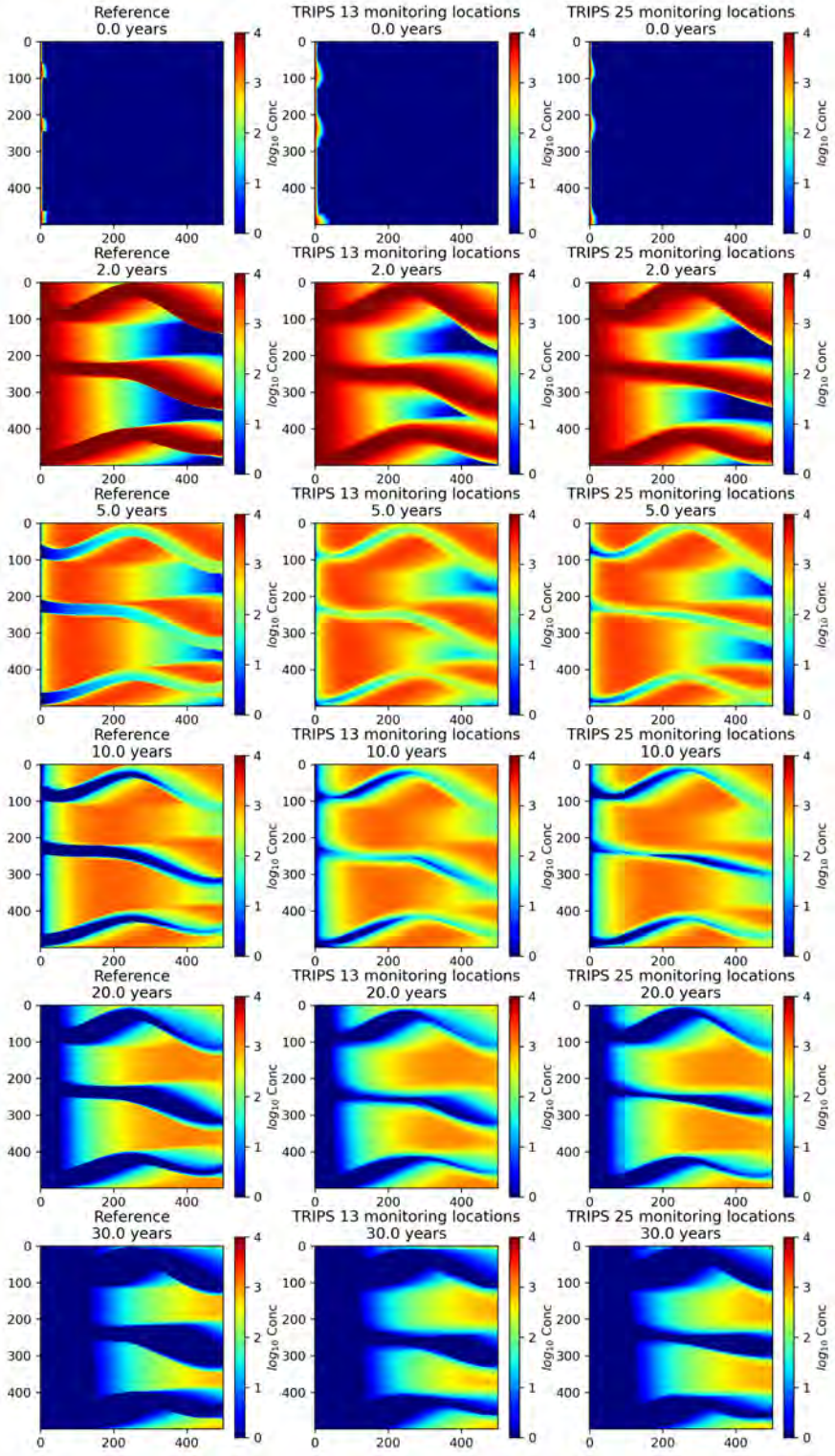


FIGURE 4.18: Simulated concentrations for the discrete parametrization case. Reference concentrations, ensemble average for the sparse and dense monitoring scenarios are shown in the first, second, third columns respectively. Each row represents a time between 0 and 30 years where the source was introduced at 0 years and removed after 2 years.

Parameter	Ensemble	Mean	Standard Deviation
ζ_{im} ($log_{10}\ day^{-1}$)	Reference	-3.00	-
	Prior	-3.00	0.5281
	Posterior (Sparse)	-3.01	0.0315
	Posterior (Dense)	-3.04	0.0254
Mobile Porosity (Channel Facies)	Reference	0.10	-
	Prior	0.10	0.0450
	Posterior (Sparse)	0.07	0.0072
	Posterior (Dense)	0.10	0.0092
Channel Facies Hydraulic Conductivity (log_{10} m/s)	Reference	-2.06	-
	Prior	-1.95	0.03421
	Posterior (Sparse)	-2.11	0.0543
	Posterior (Dense)	-2.13	0.0465
Matrix Facies Hydraulic Conductivity (log_{10} m/s)	Reference	-3.96	-
	Prior	-3.89	0.03627
	Posterior (Sparse)	-3.98	0.0222
	Posterior (Dense)	-3.98	0.0175

TABLE 4.4: Parameter Ensemble Characteristics for the Discrete Case

Monitoring	Iteration	Size	Mean ϕ_m	Std. dev. ϕ_m
Sparse	6	41	56292.3	59865.6
Dense	7	45	162363.0	90416.4

TABLE 4.5: Ensemble statistics for the discrete case

ensembles discussed in the previous sections (4.4.1 and 4.4.2) were used to make two kinds of predictions. First, the ensembles were used to predict maximum concentrations at the down-gradient boundary (column 499 of the model). At a real site, this type of prediction is typically performed for areas with sensitive receptors where concentrations are evaluated against a regulatory threshold limit. Second, the ensembles were used to predict the total mass exiting the aquifer. At a real site, this type of prediction is typically used to design treatment systems and to quantify the mass remaining in the system. Between the two predictions, the mass prediction is more challenging because it is a spatial aggregation of both the concentration distribution and the volumetric flux.

The predicted maximum concentrations for the various ensembles are shown in 4.19. The maximum concentrations from the reference model are shown with a dashed line in this figure. The predictions from the prior ensembles for both the continuous and discrete cases are shown in the first column. The prior predictions vary several orders of magnitude while encompassing the reference prediction. The predictions with the posterior ensembles for the continuous case are shown in the top row and the predictions with the posterior ensembles for the discrete case are shown in the bottom row. The second column represents the ensembles based on the sparse monitoring network while the third column represents the ensembles based on the denser monitoring network.

For the continuous case, the ensemble for the denser monitoring network does a better job at bounding the reference prediction than the ensemble for the sparser monitoring network. This is interesting given the discrepancies between the posterior ensembles for the continuous case and the reference distribution. It is possible that for some predictions, the act of assimilating additional head/concentration data could lead to a reduction in predictive uncertainty even when working with an imperfect geological prior.

For the discrete case, both sets of posterior ensembles (sparse and dense) performed well in bounding the reference prediction and reproducing the features of the reference prediction like the kink observed between 10 and 15 years. The close correspondence between the ensembles for the two different monitoring networks indicates that when the geological prior is representative, additional head/concentration data do not significantly add value to the process of quantifying the predictive uncertainty.

The predicted mass estimates (cumulative over time) in kilograms (Kg) for the various ensembles are shown in Figure 4.20. The mass exiting the reference model is shown with a dashed line in this figure. The predictions from the prior

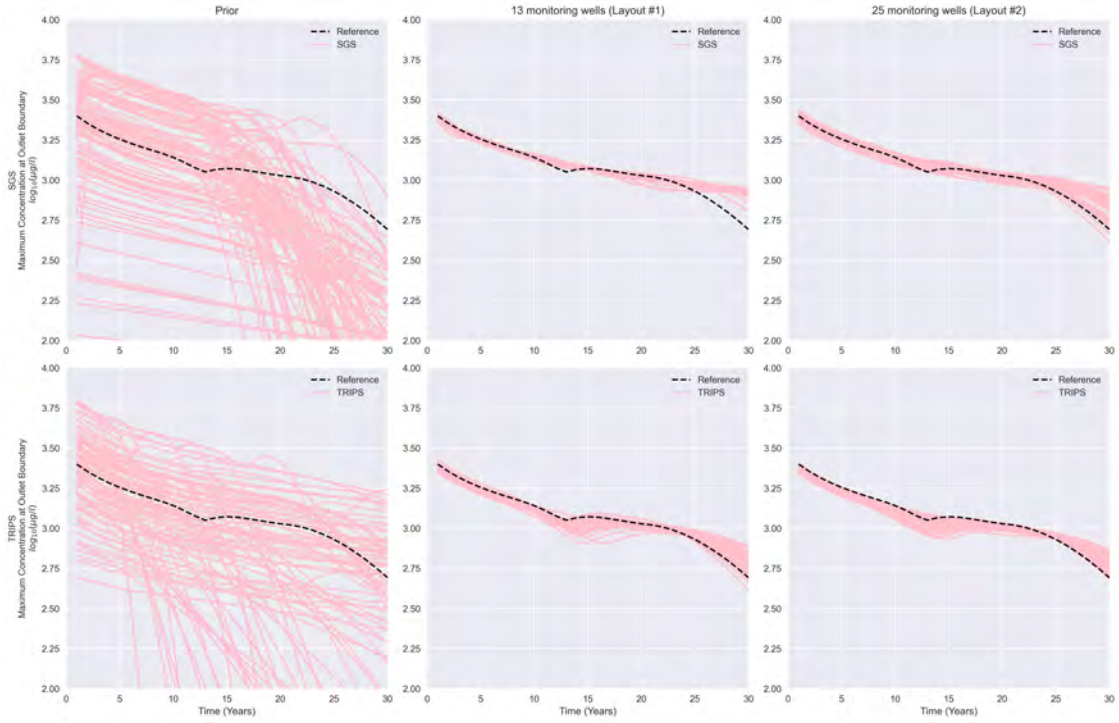


FIGURE 4.19: Predicted maximum concentrations at the downgradient boundary (column 499). Predictions from the ensembles for the continuous case are shown in the top row and predictions for the ensembles for the discrete case are shown in the bottom row. The first column depicts the concentrations from the prior ensembles. The second column depicts the concentrations from the scarce monitoring scenario. The last column depicts the concentrations from the dense monitoring scenario.

ensembles for both the continuous and discrete cases are shown in the first column. The predictions with the posterior ensembles for the continuous case are shown in the top row and the predictions with the posterior ensembles for the discrete case are shown in the bottom row. The prior predictions consistently underpredict the reference estimate. The second column represents the ensembles based on the sparse monitoring network while the third column represents the ensembles based on the denser monitoring network.

The reference model estimates that nearly 54 Kg of solute would exit the aquifer at the end of 30 years. The reference model also predicts that most of the mass would have exited the aquifer within the first five (5) years. Interestingly, none of the continuous prior ensembles (first row, first column) can match this behavior. The highest mass estimate from the continuous prior ensemble is about 10 Kg - an underprediction by a factor of five (5). The primary reason for the discrepancy is that the high-permeability areas in the continuous case are not well connected. As a result, there is an accumulation of mass within the aquifer. This in turn implies that lesser amount of mass inflow (as compared to the discrete case) is required into the system to maintain similar concentrations. This discrepancy in the mass estimates is also seen in the continuous posterior ensembles for the two monitoring scenarios (second and third columns, first row). On the other hand, the estimates from the discrete TRIPS prior ensemble encapsulate the reference estimate. Assimilation of head and concentration measurements lead to the posterior ensembles encapsulating the reference estimate (second and third columns, second row).

This analysis thus demonstrates that the ensemble smoother can assimilate head and concentration data to develop parameter ensembles that could be used for evaluating prediction uncertainty. More importantly, these results demonstrate that all predictions are not the same. For the concentration predictions, the

continuous ensembles could "reasonably" simulate the spatial patterns of heads and concentrations and also bound the reference prediction. However, the continuous ensembles were not suited to estimate the mass exiting the system. The discrete TRIPS ensembles on the other hand do a good job of simulating both heads and concentrations and are also well suited as a predictive tool for this problem.

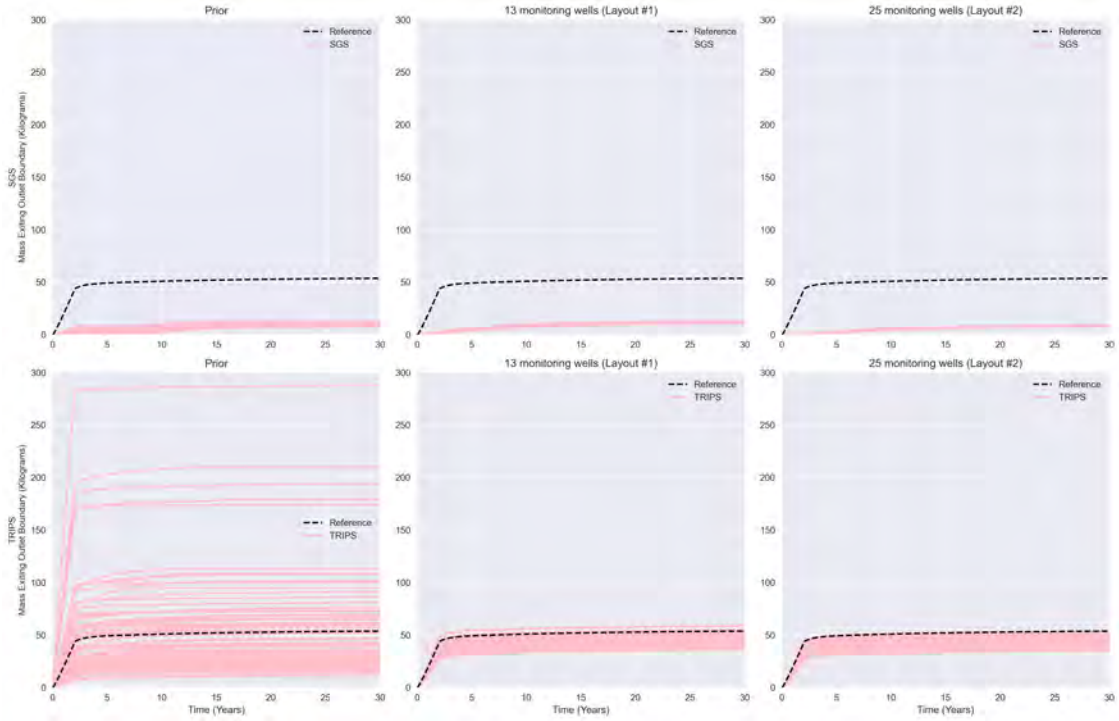


FIGURE 4.20: Predicted mass (kilograms) exiting the downgradient boundary (cumulative over time). Predictions from the ensembles for the continuous case are shown in the top row and predictions for the ensembles for the discrete case are shown in the bottom row. The first column depicts the mass from the prior ensembles. The second column depicts the concentrations from the scarce monitoring scenario. The last column depicts the concentrations from the dense monitoring scenario.

4.5 Summary and Discussion

Decision makers at contaminated sites are often required to make complex decisions regarding the fate and transport of the contaminants at their sites. They often rely on a single calibrated groundwater model to make forecasts several years/decades into the future to guide the decision-making process. However, the non-unique nature of the inverse problem can lead to predictions that may not hold true. This chapter presented an approach for developing an ensemble of models that honor measured aquifer states and can be used to develop multiple remedial forecasts to quantify the predictive uncertainty arising from parameter uncertainty.

The primary contribution of this chapter is to demonstrate that TRIPS and pestpp-ies together can be used to develop categorical parameter ensembles that honor measured aquifer heads and concentrations simultaneously. Additionally, the analysis demonstrates how multiple puzzle pieces (geological parametrization, history matching, and remedial forecasts) could be efficiently assembled to guide decision makers at contaminated sites by quantifying the predictive uncertainty associated with parameter uncertainty. The results indicate that even with an approximate geological prior model, a high degree of parametrization and history matching can lead to parameter ensembles that can be useful for making certain predictions (example: concentration predictions). However, for more demanding predictions (example: mass), an approximate geological prior is not adequate.

Uncertainty in predicting solute transport can arise from an imperfect understanding of: model parametrization (heterogeneity for example), initial aquifer conditions, boundary stresses, solute sources and release history. This chapter only focuses on the predictive uncertainty arising from the possible variations

in parametrization. However, the uncertainty in initial/boundary conditions could also be incorporated into the uncertainty framework based on "soft" information like the site history and climatic variations.

Previous attempts at solving the categorical inverse problem with only head measurements in chapter 3 produced parameter ensembles that had a lot more variability than the ensembles described in this chapter. By incorporating concentration data which are directly affected by permeability connections and contrasts in the subsurface, the posterior parameter uncertainty was decreased. Despite using a large number of parameters for the continuous case, the resulting hydraulic conductivity distribution(s) did not have the connectedness of the reference model's hydraulic conductivity distribution. This discrepancy arises from representing the non-Gaussian variance of hydraulic conductivity with a multiGaussian assumption (variogram). For example, Gómez-Hernández and Wen (1998) showed for a hypothetical site that using a multiGaussian model resulted in over-estimating source-to-receptor travel times because of incomplete permeability connections. Other researchers (Zinn and Harvey 2003; Renard and Allard 2013) also convey similar warnings.

The synthetic problem analyzed in this paper has several similarities to real-world sites namely facies with high and low permeability, decade long solute transit times, and a partial understanding of the subsurface. Therefore, the approach presented here could be adopted at large contaminated sites where the cost of an incorrect prediction far outweighs the costs associated with sampling the posterior and generating an ensemble of likely models.

Site managers, regulators, stakeholders, decision makers can thus make informed decisions by evaluating the spectrum of likely predictions rather than relying on a single prediction. The framework should include periodic data collection efforts which could then be used to update the predictions. Data worth

analyses (Dausman et al. 2010; Wöhling, Geiges, and Nowak 2016, for example) which have shown to reduce predictive uncertainty, could be used to decide the locations of future monitoring wells. A pragmatic framework that integrates data acquisition efforts with an ensemble-based approach could lead to better future outcomes.

Chapter 5

Conclusions

5.1 Introduction

Several researchers have developed methods to solve the categorical inverse problem, a commonly encountered problem in the field of geosciences. In this thesis, a new method known as the Traveling Pilot Point Method (TRIPS) involving pilot points that move (travel), is developed to parametrize the categorical inverse problem. A summary of the key contributions and perspectives for future research are presented in this chapter.

5.2 Summary of Contributions

The most important proposition of this thesis is in letting pilot points travel to estimate both the locations of the channels and associated properties like hydraulic conductivity. The positions of the pilot points and the associated values are governed by a prior covariance matrix inferred from a training image. The advantage of the TRIPS approach is in its simplicity. The user can provide an image of the type of channels that they want to model, and the covariance can be inferred directly. For small training images, it is possible to use a multiple-point statistics simulation algorithm and generate an ensemble of simulations

and derive the covariance matrix from the analysis of this simulation in the same manner as the sub-images from the training image were analyzed.

The TRIPS method was applied in conjunction with NSMC and PC methods to solve the categorical inverse problem for a synthetic aquifer. A posterior ensemble obtained with the rejection sampling method was considered to represent the reference solution. First-order (difference) regularization constraints were found to be useful in preserving the curvature of the channels in the inversion process when TRIPS were coupled with NSMC and PC. The analysis showed that the NSMC method provides a balance between computational efficiency and representation of the posterior ensemble and the number of forward evaluations were comparable to stochastic approaches like ISR (Mariethoz, Renard, and Straubhaar 2010) and POPEX (Jäggli, Straubhaar, and Renard 2017), while the PC method is much faster.

The TRIPS method when coupled with the pestpp-ies smoother was successfully able to produce a large posterior parameter ensemble for the synthetic aquifer. The combination of the two methods resulted in a posterior ensemble comparable to one estimated using RS. Computationally, this combination required lesser number of forward evaluations than the combination of TRIPS and NSMC. Another contribution of this analysis is in the exploration of the dissimilarity in the prior and posterior ensembles. Using MDS projection and K-Means clustering, a method to strategically select a smaller number of prior realizations such that the resulting posterior ensemble would still have the properties obtained from a much larger posterior ensemble was demonstrated. This is an important finding and could have positive implications for complex real-world problems with large simulation times.

Decision makers at contaminated sites are often required to make complex decisions regarding the fate and transport of the contaminants at their sites.

They often rely on a single calibrated groundwater model to make forecasts several years/decades into the future to guide the decision-making process. However, the non-unique nature of the inverse problem can lead to predictions that may not hold true. TRIPS and pestpp-ies together were used to develop categorical parameter ensembles that honor measured aquifer heads and concentrations simultaneously. The analysis also demonstrated how multiple puzzle pieces (geological parameterization, history matching, and remedial forecasts) could be efficiently assembled to guide decision makers at contaminated sites by quantifying the predictive uncertainty associated with parameter uncertainty. The results indicate that even with an approximate geological prior model, a high degree of parameterization and history matching can lead to parameter ensembles that can be useful for making certain predictions (example: concentration predictions). However, for more demanding predictions (example: mass), an approximate geological prior is not adequate.

5.3 Perspectives for Future Research

In this section, a few ideas for future research, that build on the current analysis and numerical experiments are offered.

5.3.1 Extension to real-world problems

The work presented in this thesis has proven that TRIPS can be used to solve the inverse problem for a very simplified geological setting: a small number of two-dimensional channels of fixed width and completely traversing the domain.

One of the main question is whether the TRIPS method can be applicable on more complex real-world situations. It is quite clear that it should not be difficult

to add some additional TRIPS in a 3D space and control the positions of channels in three dimensions. Three-dimensional splines passing through the pilot points could be used for example to delineate the channels. Such an approach could be used to estimate the channel framework at a real site such as the one discussed by Ronayne, Gorelick, and Caers (2008).

More generally, the extension of the proposed methodology seems to be rather straightforward for object-oriented geological modeling techniques (Pyrcz and Deutsch 2014) since the positions of the objects are controlled in these models by seed points which can be considered as Traveling Pilot Points. The prior statistics on the number of objects and relative locations of these points can be derived from a set of initial simulations. The proposed algorithm could be used to update these locations and solve the inverse problem. This approach is similar to what was proposed by Hu and Jenni (2005) or Jenni et al. (2007). The shape parameters concerning the three-dimensional size and orientation of the objects can be handled as well easily since these are continuous parameters that an inversion code like PEST can optimize. This step would be analogous to the identification of the hydraulic conductivity values within the channels as illustrated in the example treated in this paper. For objects having a flexible shape such as channels with varying width, traditional techniques such as the standard pilot points can be coupled with TRIPS: one can attach a width parameter to every traveling pilot point and interpolate the width along the channel length and update these parameters during the inversion. A difficulty however with this approach could be that the number of parameters to update could become very large. However, the pestpp-ies ensemble smoother (or other ensemble-based methods) which is efficient at solving inverse problems with large numbers of parameters could be quite useful here.

Since most categorical simulation techniques (object based, sequential indicator or truncated plurigaussian, multiple-point statistics, etc.) can be conditioned (see for example Lantuéjoul 2013), it could also be possible (at least in principle) to select a fixed and limited number of locations on a simulation grid, run one simulation and fix the values of the category from that simulation at these locations. In a second step, the positions of these TRIPS could be updated while keeping the categories unchanged. This approach clearly requires further analysis since the covariance matrix which is used in the current version of the TRIPS method would need to be defined and the way to infer those covariances is still unclear.

5.3.2 Integration with Iterative Ensemble Smoothers

For the synthetic problem evaluated in chapter 3, a large ensemble (1'622 members) produced an ensemble mean comparable to the ensemble mean generated by RS - a surrogate for the "true posterior". It is suspected that the non-linearity of the categorical inverse problem necessitated a large ensemble to sample the posterior. This relationship could be investigated further.

In the pestpp-ies implementation, the largest prior ensemble yielded only a small percentage of posterior realizations. One of the possible reasons for this low yield is thought to be the small subset (size 20) used to test λ_l values. Further research could reveal whether a larger subset for evaluating λ_l values could yield a higher number of posterior realizations.

Model selection techniques indicate that a strategically selected prior ensemble could sample the posterior distribution as well as a larger less-diverse ensemble. It remains to be see whether different dissimilarity distance metrics and clustering techniques could yield better results. Additionally, there could be

ways to improve the selection of realizations by weighting the selected realizations based on the number of underlying simulations that they represent, when computing covariances and mean values.

5.3.3 Improving remedial outcomes

The synthetic problem analyzed in this paper has several similarities to real-world sites namely facies with high and low permeability, decade long solute transit times, and a partial understanding of the subsurface. Therefore, the approach presented here could be adopted at large contaminated sites where the cost of an incorrect prediction far outweighs the costs associated with sampling the posterior and generating an ensemble of likely models. Site managers, regulators, stakeholders, decision makers can thus make informed decisions by evaluating the spectrum of likely predictions rather than relying on a single prediction. The framework should include periodic data collection efforts which could then be used to update the predictions. Data worth analyses (Dausman et al. 2010; Wöhling, Geiges, and Nowak 2016, for example) which have shown to reduce predictive uncertainty could be used to decide the locations of future monitoring wells. A pragmatic framework that integrates data acquisition efforts with an ensemble-based approach could lead to better future outcomes.

Model selection techniques could be incorporated into the remedial uncertainty framework to quantify uncertainty using a smaller number of realizations. This could be an attractive prospect because models that simulate both flow and transport have large run times. With a strategically selected smaller ensemble, uncertainty quantification could be performed in a more efficient manner.

Appendix A

Covariance Matrix

A.1 Calculation of a covariance matrix for parameter differences

Given the covariance matrix **C** for a parameter vector **p**, calculation of the covariance matrix $\mathbf{C}(\mathbf{p}-)$ for the parameter difference vector $\mathbf{p}-$ is described in this section with an example. If the parameter vector **p** had three parameters *y1, y2, and y3*, the parameter difference vector $\mathbf{p}-$ would have the differences *y1-y2, y1-y3, and y2-y3*. In matrix form, this relationship can be expressed by the equation

$$\begin{pmatrix} y1-y2 \\ y1-y3 \\ y2-y3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} y1 \\ y2 \\ y3 \end{pmatrix} \tag{A.1}$$

Equation A.1 could be generalized for an arbitrary number of parameters/parameter differences by the equation

$$\mathbf{p}- = \mathbf{A}\mathbf{p} \tag{A.2}$$

If we have $\mathbf{y} = \mathbf{A}\mathbf{x}$, the covariance matrix of $\mathbf{y}$, $\mathbf{C}(\mathbf{y})$ can be calculated (Aster, Borchers, and Thurber 2013) by the equation

$$\mathbf{C}(\mathbf{y}) = \mathbf{A}\mathbf{C}(\mathbf{x})\mathbf{A}^T \tag{A.3}$$

combining equations A.2 and A.3, the covariance matrix for the parameter difference vector can be calculated by the equation

$$\mathbf{C}(\mathbf{p}-) = \mathbf{A}\mathbf{C}(\mathbf{p})\mathbf{A}^T \tag{A.4}$$

A.2 Mean and Covariance of Prior Distribution (Chapters 2 and 3)

The mean values of parameters in the prior distribution for the problem in Chapters 2 and 3 are shown in Table A.1 below.The covariance matrix for the parameters in the prior distribution is shown in Table A.2.

Parameter	Description	Mean
y11	y coordinate	92.44
y12	y coordinate	92.24
y13	y coordinate	91.45
y14	y coordinate	91.22
y15	y coordinate	90.48
y21	y coordinate	54.73
y22	y coordinate	53.81
y23	y coordinate	53.29
y24	y coordinate	52.65
y25	y coordinate	51.85
y31	y coordinate	21.57
y32	y coordinate	20.44
y33	y coordinate	19.62
y34	y coordinate	19.43
y35	y coordinate	19.45
$log_{10}(K_c)$	Hydraulic Conductivity of channel	10^{-2} m/s
$log_{10}(K_m)$	Hydraulic Conductivity of matrix	10^{-4} m/s

TABLE A.1: Descriptions and mean values of parameters in the prior distribution. yij represents the j^{th} TRIP in the i^{th} channel.The origin for the y coordinates is at the lower left corner of the model grid.

	y11	y12	y13	y14	y15	y21	y22	y23	y24	y25	y31	y32	y33	y34	y35	$log_{10}(K_c)$	$log_{10}(K_m)$
y11	253.8	231.1	203.5	186.3	173.8	107.7	102.4	92.3	90.9	85.0	52.2	34.5	28.6	36.7	31.1	0.0	0.0
y12	231.1	244.5	219.6	206.4	193.9	104.8	98.3	89.3	90.4	85.5	45.6	32.2	24.5	31.8	25.0	0.0	0.0
y13	203.5	219.6	247.7	237.2	222.4	98.1	97.4	96.1	96.7	91.8	44.0	34.7	32.4	38.6	34.2	0.0	0.0
y14	186.3	206.4	237.2	273.1	262.2	89.3	92.0	95.0	100.5	94.8	37.1	32.0	35.2	39.9	38.4	0.0	0.0
y15	173.8	193.9	222.4	262.2	295.2	82.6	85.7	96.8	99.6	95.1	25.9	23.4	26.1	26.3	27.4	0.0	0.0
y21	107.7	104.8	98.1	89.3	82.6	231.6	217.7	204.9	187.1	176.5	103.2	94.5	85.2	80.7	68.7	0.0	0.0
y22	102.4	98.3	97.4	92.0	85.7	217.7	240.3	225.6	207.3	198.4	102.0	95.3	84.7	82.0	69.6	0.0	0.0
y23	92.3	89.3	96.1	95.0	96.8	204.9	225.6	262.0	243.8	227.9	95.9	91.6	90.7	97.0	87.4	0.0	0.0
y24	90.9	90.4	96.7	100.5	99.6	187.1	207.3	243.8	270.0	252.5	87.5	84.6	87.8	103.6	95.6	0.0	0.0
y25	85.0	85.5	91.8	94.8	95.1	176.5	198.4	227.9	252.5	273.4	86.5	83.2	81.2	94.9	89.7	0.0	0.0
y31	52.2	45.6	44.0	37.1	25.9	103.2	102.0	95.9	87.5	86.5	168.2	142.0	129.0	115.5	110.6	0.0	0.0
y32	34.5	32.2	34.7	32.0	23.4	94.5	95.3	91.6	84.6	83.2	142.0	167.6	155.6	146.1	137.8	0.0	0.0
y33	28.6	24.5	32.4	35.2	26.1	85.2	84.7	90.7	87.8	81.2	129.0	155.6	186.8	185.1	172.2	0.0	0.0
y34	36.7	31.8	38.6	39.9	26.3	80.7	82.0	97.0	103.6	94.9	115.5	146.1	185.1	235.3	222.5	0.0	0.0
y35	31.1	25.0	34.2	38.4	27.4	68.7	69.6	87.4	95.6	89.7	110.6	137.8	172.2	222.5	253.4	0.0	0.0
$log_{10}(K_c)$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01	0
$log_{10}(K_m)$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	0.01

TABLE A.2: Covariance Matrix for prior parameters

Appendix B

pestpp-ies Control Variables

B.1 Control Variables

In this appendix, we document the pestpp-ies control variables for the simulation with an initial ensemble of 10000 (section 3.3.1) along with the reason for their inclusion. The control variables are shown in Table B.1.

TABLE B.1: pestpp-ies Control Variables for Ensemble of 10,000

pestpp-ies Control Variable	**Explanation**
++ ies_num_reals(10000)	Generate a prior parameter ensemble with 10000 realizations
++ parcov(priorcovmat.unc)	The initial parameter ensemble is generated based on the prior parameter covariance matrix contained in the file called priorcovmat.unc.
++ ies_parameter_ensemble(randpar.csv)	Initial parameter realizations are in a comma separated value file called randpar.csv.
++ ies_add_base(false)	A base parameter realization is not provided to pestpp-ies.
++ ies_use_prior_scaling(false)	Parameters are used as-is without any scaling.
++ ies_subset_size(20)	A subset of 20 realizations will be tested with various Marquardt lambdas.
++ ies_subset_how(random)	The subset of 20 realizations will be randomly selected.
++ ies_lambda_mults(0.01,0.1,1.0,5.0,10.0)	The best lambda from the previous iteration will be multiplied by these factors and tested during the current iteration.
++ lambda_scale_fac(0.1,0.5,1.0,1.5,2.0)	These are the line search factors along parameter upgrade directions computed with the various Marquardt lambdas.
++ ies_no_noise(true)	Measurement noise is not added to the observations.
++ overdue_giveup_fac(4.0)	If an overdue simulation takes more than 4 times the average run time, the run manager will consider that run to have failed and exclude that realization from further parameter estimation.
++ ies_bad_phi_sigma(2.0)	If the objective function of a realization is more than 2 standard deviations away from the mean of the ensemble, the run manager will consider that run to have failed.

Appendix C

Mean and Covariance Matrix for Chapter 4

The mean values of parameters in the prior distribution for the problem in Chapter 4 are shown in Table C.1 below.The covariance matrix for the parameters in the prior distribution are shown in Tables C.2 and C.3.

Parameter	Description	Mean
y11	y coordinate	92.44
y12	y coordinate	92.24
y13	y coordinate	91.45
y14	y coordinate	91.22
y15	y coordinate	90.48
y21	y coordinate	54.73
y22	y coordinate	53.81
y23	y coordinate	53.29
y24	y coordinate	52.65
y25	y coordinate	51.85
y31	y coordinate	21.57
y32	y coordinate	20.44
y33	y coordinate	19.62
y34	y coordinate	19.43
y35	y coordinate	19.45
$log_{10}(K_c)$	Hydraulic Conductivity of channel	10^{-2} m/s
$log_{10}(K_m)$	Hydraulic Conductivity of matrix	10^{-4} m/s
$log_{10}(\zeta_{im})$	Mass transfer coefficient	10^{-3} /day
$log_{10}(\theta_{m,ch})$	Mobile porosity of channel facies	10%

TABLE C.1: Descriptions and mean values of parameters in the prior distribution. yij represents the j^{th} TRIP in the i^{th} channel.The origin for the y coordinates is at the lower left corner of the model grid.

	y11	y12	y13	y14	y15	y21	y22	y23	y24	y25	y31	y32	y33	y34	y35	$log_{10}(K_c)$	$log_{10}(K_m)$
y11	253.8	231.1	203.5	186.3	173.8	107.7	102.4	92.3	90.9	85.0	52.2	34.5	28.6	36.7	31.1	0.0	0.0
y12	231.1	244.5	219.6	206.4	193.9	104.8	98.3	89.3	90.4	85.5	45.6	32.2	24.5	31.8	25.0	0.0	0.0
y13	203.5	219.6	247.7	237.2	222.4	98.1	97.4	96.1	96.7	91.8	44.0	34.7	32.4	38.6	34.2	0.0	0.0
y14	186.3	206.4	237.2	273.1	262.2	89.3	92.0	95.0	100.5	94.8	37.1	32.0	35.2	39.9	38.4	0.0	0.0
y15	173.8	193.9	222.4	262.2	295.2	82.6	85.7	96.8	99.6	95.1	25.9	23.4	26.1	26.3	27.4	0.0	0.0
y21	107.7	104.8	98.1	89.3	82.6	231.6	217.7	204.9	187.1	176.5	103.2	94.5	85.2	80.7	68.7	0.0	0.0
y22	102.4	98.3	97.4	92.0	85.7	217.7	240.3	225.6	207.3	198.4	102.0	95.3	84.7	82.0	69.6	0.0	0.0
y23	92.3	89.3	96.1	95.0	96.8	204.9	225.6	262.0	243.8	227.9	95.9	91.6	90.7	97.0	87.4	0.0	0.0
y24	90.9	90.4	96.7	100.5	99.6	187.1	207.3	243.8	270.0	252.5	87.5	84.6	87.8	103.6	95.6	0.0	0.0
y25	85.0	85.5	91.8	94.8	95.1	176.5	198.4	227.9	252.5	273.4	86.5	83.2	81.2	94.9	89.7	0.0	0.0
y31	52.2	45.6	44.0	37.1	25.9	103.2	102.0	95.9	87.5	86.5	168.2	142.0	129.0	115.5	110.6	0.0	0.0
y32	34.5	32.2	34.7	32.0	23.4	94.5	95.3	91.6	84.6	83.2	142.0	167.6	155.6	146.1	137.8	0.0	0.0
y33	28.6	24.5	32.4	35.2	26.1	85.2	84.7	90.7	87.8	81.2	129.0	155.6	186.8	185.1	172.2	0.0	0.0
y34	36.7	31.8	38.6	39.9	26.3	80.7	82.0	97.0	103.6	94.9	115.5	146.1	185.1	235.3	222.5	0.0	0.0
y35	31.1	25.0	34.2	38.4	27.4	68.7	69.6	87.4	95.6	89.7	110.6	137.8	172.2	222.5	253.4	0.0	0.0
$log_{10}(K_c)$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
$log_{10}(K_m)$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	0.3

TABLE C.2: Covariance Matrix for prior flow and transport parameters

	$log_{10}(\zeta_{im})$	$\theta_{m,ch}$
$log_{10}(\zeta_{im})$	0.3	0.0
$\theta_{m,ch}$	0.0	0.3

TABLE C.3: Covariance Matrix for prior transport-only parameters

Bibliography

Ahmmed, B., S. C. James, and J. Yelderman (2020). "Post-Calibration Uncertainty Analysis for Travel Times at a Naval Weapons Industrial Reserve Plant". In: *Water* 12.12, p. 3428. DOI: 10.3390/w12123428.

Alcolea, A. and P. Renard (2010). "Blocking Moving Window algorithm: Conditioning multiple-point simulations to hydrogeological data". In: *Water Resources Research* 46.8, pp. 1–18. DOI: 10.1029/2009WR007943.

Anderson, Mary P., William W. Woessner, and Randall J. Hunt (2015). *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*. 2nd edition. Elsevier.

Aster, Richard C., Brian Borchers, and Clifford H. Thurber (2013). *Parameter estimation and inverse problems*. Second Edi. ISBN: 9780128100929.

Borg, Ingwer and Patrick F. Groenen (1997). *Modern Multidimensional Scaling - Theory and Applications*. New York: Springer.

Caers, Jef and T. Hoffman (2006). "The probability perturbation method: A new look at Bayesian inverse modeling". In: *Mathematical Geology* 38.1, pp. 81–100. ISSN: 08828121. DOI: 10.1007/s11004-005-9005-9.

Cao, Zhendan, Liangping Li, and Kang Chen (2018). "Bridging iterative Ensemble Smoother and multiple-point geostatistics for better flow and transport modeling". In: *Journal of Hydrology* 565, pp. 411–421. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2018.08.023. URL: https://doi.org/10.1016/j.jhydrol.2018.08.023.

Certes, Catherine and Ghislain de Marsily (1991). "Application of the pilot point method to the identification of aquifer transmissivities". In: *Advances in Water Resources* 14.5, pp. 284–300. DOI: https://doi.org/10.1016/0309-1708(91)90040-U.

Chapman, Steven W. and Beth L. Parker (2005). "Plume persistence due to aquitard back diffusion following dense nonaqueous phase liquid source removal or isolation". In: *Water Resources Research* 41.12, pp. 1–16. ISSN: 00431397. DOI: 10.1029/2005WR004224.

Chapman, Steven W., Beth L. Parker, Tom C. Sale, and Lee Ann Doner (2012). "Testing high resolution numerical models for analysis of contaminant storage and release from low permeability zones". In: *Journal of Contaminant Hydrology* 136-137, pp. 106–116. ISSN: 01697722. DOI: 10.1016/j.jconhyd.2012.04.006. URL: http://dx.doi.org/10.1016/j.jconhyd.2012.04.006.

Chen, Yan and Dean S. Oliver (2013). "Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification". In: *Computational Geosciences* 17.4, pp. 689–703. ISSN: 14200597. DOI: 10.1007/s10596-013-9351-5.

Coats, K.H. and B.D. Smith (1964). "Dead-End Pore Volume and Dispersion in Porous Media". In: *Society of Petroleum Engineers Journal* March, pp. 73–84. DOI: 10.2118/647-PA.

Cox, Michael A.A. and Trevor F. Cox (1994). *Multidimensional Scaling*. London: Chapman and Hill.

Dausman, Alyssa M., John Doherty, Christian D. Langevin, and Michael C. Sukop (2010). "Quantifying data worth toward reducing predictive uncertainty". In: *Ground Water* 48.5, pp. 729–740. ISSN: 0017467X. DOI: 10.1111/j.1745-6584.2010.00679.x.

Deans, H.A. (1963). "A Mathematical Model for Dispersion in the Direction Of Flow in Porous Media". In: *Society of Petroleum Engineers Journal* 3.01, pp. 49–52. ISSN: 0197-7520. DOI: 10.2118/493-pa.

Deutsch, Clayton V. and A.G. Journel (1998). *GSLIB: Geostatistical software library and users's guide. New York: Oxford University Press*. 6th editio. Oxford University Press.

Doherty, John E. (2003). "Ground water model calibration using pilot points and regularization". In: *Ground Water* 41.2, pp. 170–177. ISSN: 0017467X. DOI: 10.1111/j.1745-6584.2003.tb02580.x.

— (2018). *PEST (Model-Independent Parameter Estimation), 7th Edition*. Brisbane, Australia.

Doherty, John E. and Randall J. Hunt (2010). "Approaches to highly parameterized inversion: a guide to using PEST for groundwater-model calibration". In: *U. S. Geological Survey Scientific Investigations Report 2010-5169*, p. 70. URL: http://pubs.usgs.gov/sir/2010/5169/.

Emerick, Alexandre A and Albert C Reynolds (2013). "Ensemble Smoother with Multiple Data Assimilation". In: 55, pp. 3–15. DOI: 10.1016/j.cageo.2012.03.011.

Evensen, G. (1994). "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics". In: *Journal of Geophysical Research* 99.C5, pp. 10143–10162. ISSN: 0148-0227. DOI: 10.1029/94JC00572. URL: http://doi.wiley.com/10.1029/94JC00572.

— (2009). *Data assimilation: The ensemble Kalman filter*, pp. 1–307. ISBN: 9783642037108. DOI: 10.1007/978-3-642-03711-5.

Farhat, Shahla K., David T. Adamson, Arun R. Gavaskar, Sophia A. Lee, Ronald W. Falta, and Charles J. Newell (2020). "Vertical Discretization Impact in Numerical Modeling of Matrix Diffusion in Contaminated Groundwater". In:

Groundwater Monitoring and Remediation 40.2, pp. 52–64. ISSN: 17456592. DOI: 10.1111/gwmr.12373.

Fiori, A., V. Cvetkovic, G. Dagan, S. Attinger, A. Bellin, P. Dietrich, A. Zech, and G. Teutsch (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?" In: *Water Resources Research* 52.12, pp. 9228–9234. ISSN: 19447973. DOI: 10.1002/2015WR017525.

Freyberg, David L. (1988). *an Exercise in Ground-Water Model Calibration and Prediction*. DOI: 10.1111/j.1745-6584.1988.tb00399.x.

Gómez-Hernández, J. Jaime and Xian Huan Wen (1998). "To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology". In: *Advances in Water Resources* 21.1, pp. 47–61. ISSN: 03091708. DOI: 10.1016/S0309-1708(96)00031-0.

Hanke, Martin (1997). "A regularizing Levenberg - Marquardt scheme , with applications to inverse groundwater filtration problems". In: *Inverse Problems* 13, pp. 79–95. DOI: https://doi.org/10.1088/0266-5611/13/1/007.

Hansen, T. M., K. S. Cordua, and K. Mosegaard (2012). "Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling". In: *Computational Geosciences* 16.3, pp. 593–611. ISSN: 14200597. DOI: 10.1007/s10596-011-9271-1.

Hayley, Kevin, Alexis Valenza, Emma White, Bruce Hutchison, and Jens Schumacher (2019). "Application of the Iterative Ensemble Smoother method and cloud computing: A groundwater modeling case study". In: *Water (Switzerland)* 11.8. ISSN: 20734441. DOI: 10.3390/w11081649.

Heidari, L., V Gervais, M. Le Ravalec, and H. Wackernagel (2011). "History matching of reservoir models by ensemble Kalman filtering: The state of the

art and a sensitivity study". In: *AAPG Memoir* 96, pp. 249–264. DOI: 10.1306/13301418M963486.

Horst, John, Scott Potter, Matthew Schnobrich, Nicklaus Welty, Ankit Gupta, and Joseph Quinnan (2017). "Advancing Contaminant Mass Flux Analysis to Focus Remediation: The Three-Compartment Model". In: *Groundwater Monitoring and Remediation* 37.4, pp. 15–22. ISSN: 17456592. DOI: 10.1111/gwmr.12250.

Houtekamer, P.L. and Herschel L. Mitchell (1998). "Data assimilation using an ensemble Kalman filter technique." In: *Monthly Weather Review* 126.3, pp. 796–811. ISSN: 00270644. DOI: 10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Hu, Lin Y., Georges Blanc, and Benoît Noetinger (2001). "Gradual deformation and iterative calibration of sequential stochastic simulations". In: *Mathematical Geology* 33.4, pp. 475–489. ISSN: 08828121. DOI: 10.1023/a:1011088913233.

Hu, Lin Y. and Sandra Jenni (2005). "History matching of object-based stochastic reservoir models". In: *SPE Journal* 10.03, pp. 312–323. DOI: 10.2118/81503-PA.

Jenni, S., L. Ying Hu, R. Basquet, G. de Marsily, and B. Bourbiaux (2007). "History matching of a stochastic model of field-scale fractures: Methodology and case study". In: *Oil and Gas Science and Technology* 62.2 SPECIAL ISSUE, pp. 265–276. ISSN: 12944475. DOI: 10.2516/ogst:2007022.

Juda, P. and P. Renard (2021). "An Attempt to Boost Posterior Population Expansion Using Fast Machine Learning Algorithms". In: *Frontiers in Artificial Intelligence* 4, p. 25. DOI: 10.3389/frai.2021.624629.

Jäggli, C., J. Straubhaar, and P. Renard (2017). "Posterior population expansion for solving inverse problems". In: *Water Resources Research* 53.4, pp. 2902–2916. ISSN: 19447973. DOI: 10.1002/2016WR019550.

Jäggli, C., J. Straubhaar, and P. Renard (2018). "Parallelized Adaptive Importance Sampling for Solving Inverse Problems". In: *Frontiers in Earth Science* 6.November, pp. 1–15. ISSN: 2296-6463. DOI: 10.3389/feart.2018.00203.

Kalman, Rudolph Emil (1960). "A New Approach to Linear Filtering and Prediction Problems". In: *Transactions of the ASME–Journal of Basic Engineering* 82.Series D, pp. 35–45.

Khambhammettu, Prashanth, Philippe Renard, and John Doherty (2020). "The Traveling Pilot Point method. A novel approach to parameterize the inverse problem for categorical fields". In: *Advances in Water Resources* 138, p. 103556. ISSN: 0309-1708. DOI: https://doi.org/10.1016/j.advwatres.2020.103556. URL: https://www.sciencedirect.com/science/article/pii/S0309170819308401.

Knowling, Matthew J., Jeremy T. White, and Catherine R. Moore (2019). "Role of model parameterization in risk-based decision support: An empirical exploration". In: *Advances in Water Resources* 128.February, pp. 59–73. ISSN: 03091708. DOI: 10.1016/j.advwatres.2019.04.010. URL: https://doi.org/10.1016/j.advwatres.2019.04.010.

Konikow, Leonard F. (2011). "The Secret to Successful Solute-Transport Modeling". In: *Ground Water* 49.2, pp. 144–159. ISSN: 0017467X. DOI: 10.1111/j.1745-6584.2010.00764.x.

Kumar, Devesh (2019). "Ensemble-Based Assimilation of Non-Linearly Related Dynamic Data in Reservoir Models Exhibiting Non-Gaussian Characteristics". PhD thesis. Penn State University, p. 140. ISBN: 9781626239777.

Laloy, Eric, Romain Hérault, Diederik Jacques, and Niklas Linde (2018). "Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network". In: *Water Resources Research* 54.1, pp. 381–406. ISSN: 19447973. DOI: 10.1002/2017WR022148. arXiv: 1708.04975.

Lam, D.-T., P. Renard, J. Straubhaar, and J. Kerrou (2020). "Multiresolution Approach to Condition Categorical Multiple-Point Realizations to Dynamic Data With Iterative Ensemble Smoothing". In: *Water Resources Research* 56.2, p. 29. DOI: 10.1029/2019WR025875.

Langevin, C.D., J.D. Hughes, E.R. Banta, A.M. Provost, R.G. Niswonger, and S. Panday (2021). *MODFLOW 6 Modular Hydrologic Model version 6.2.1: U.S. Geological Survey Software Release, 18 February 2021*. DOI: https://doi.org/10.5066/F76Q1VQV.

Lantuéjoul, Christian (2013). *Geostatistical simulation: models and algorithms*. Springer Science & Business Media.

Lavenue, M. (2001). "Three-dimensional interference test interpretation in a fractured-unfractured aquifer using the pilot point inverse method". In: *Water Resources Research* 37.11, pp. 2659–2676. DOI: 10.1029/2000WR000289.

Le Ravalec, Mickaele (2005). *Inverse Stochastic Modeling of Flow in Porous Media: Applications to Reservoir Characterization*. Technip, France: IFP Publications, p. 194. ISBN: 2-7108-0864-1.

Leeuwen, P.J. van and Geir Evensen (1996). "Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation". In: *Monthly Weather Review* 124, pp. 2898–2913.

Li, L., S. Srinivasan, H. Zhou, and J. J. Gomez-Hernandez (2015). "Two-point or multiple-point statistics? A comparison between the ensemble Kalman filtering and the ensemble pattern matching inverse methods". In: *Advances in Water Resources* 86, pp. 297–310. ISSN: 03091708. DOI: 10.1016/j.advwatres.2015.05.014.

Li, Liangping, Sanjay Srinivasan, Haiyan Zhou, and J. J. Gómez-Hernández (2013). "A pilot point guided pattern matching approach to integrate dynamic data into geological modeling". In: *Advances in Water Resources* 62.PA, pp. 152–

128. ISSN: 03091708. DOI: 10.1016/j.advwatres.2013.10.008. URL: http://dx.doi.org/10.1016/j.advwatres.2013.10.008.

Linde, Niklas, Philippe Renard, Tapan Mukerji, and Jef Caers (2017). "Geological realism in hydrogeological and geophysical inverse modeling: A review". In: *arXiv* 86, pp. 86–101. ISSN: 23318422.

Mariethoz, G., P. Renard, and J. Caers (2010). "Bayesian inverse problem and optimization with iterative spatial resampling". In: *Water Resources Research* 46.11, pp. 1–17. DOI: 10.1029/2010WR009274.

Mariethoz, G., P. Renard, and J. Straubhaar (2010). "The direct sampling method to perform multiple-point geostatistical simulations". In: *Water Resources Research* 46.11, pp. 1–14. DOI: 10.1029/2008WR007621.

Moeck, Christian, John Molson, and Mario Schirmer (2020). "Pathline Density Distributions in a Null-Space Monte Carlo Approach to Assess Groundwater Pathways". In: *Groundwater* 58.2, pp. 189–207. ISSN: 17456584. DOI: 10.1111/gwat.12900.

Moore, C. and J. Doherty (2005). "Role of the calibration process in reducing model predictive error". In: *Water Resources Research* 41.5, pp. 1–14. ISSN: 00431397. DOI: 10.1029/2004WR003501.

Moreno, Ziv and Amir Paster (2018). "Prediction of pollutant remediation in a heterogeneous aquifer in Israel: Reducing uncertainty by incorporating lithological, head and concentration data". In: *Journal of Hydrology* 564.June, pp. 651–666. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2018.07.012.

National Research Council (2013). *Alternatives for Managing the Nation's Complex Contaminated Groundwater Sites*. Washington, D.C.: The National Academies Press, pp. 1–407. ISBN: 9780309278102. DOI: 10.17226/14668.

Niswonger, R.G., Sorab Panday, and Motomu Ibaraki (2011). *MODFLOW-NWT, A Newton formulation for MODFLOW-2005: U.S. Geological Survey Techniques and Methods 6–A37*. Reston, Virginia.

Payne, Fred C., Joseph A. Quinnan, and Scott T. Potter (2008). *Remediation Hydraulics*. First. Boca Raton, FL: CRC Press, p. 439.

Peaceman, D.W. (1977). *Fundamentals of Numerical Reservoir Simulation*. Ed. by Elsevier, p. 174.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). "Scikit-learn: Machine Learning in {P}ython". In: *Journal of Machine Learning Research* 12, pp. 2825–2830. URL: https://scikit-learn.org/stable/about.html{\#}citing-scikit-learn.

Pyrcz, M. and C. Deutsch (2014). *Geostatistical Reservoir Modeling*. Second Edition. Oxford University Press, p. 433. ISBN: 978-0199731442.

Rajaram, Harihar (2016). "Debates—Stochastic subsurface hydrology from theory to practice: Introduction". In: *Water Resources Research* 52.12, pp. 9215–9217. DOI: https://doi.org/10.1002/2016WR020066.

Renard, P. and D. Allard (2013). "Connectivity metrics for subsurface flow and transport". In: *Advances in Water Resources* 51, pp. 168–196. DOI: 10.1016/j.advwatres.2011.12.001. URL: http://dx.doi.org/10.1016/j.advwatres.2011.12.001.

Ronayne, Michael J., Steven M. Gorelick, and Jef Caers (2008). "Identifying discrete geologic structures that produce anomalous hydraulic response: An inverse modeling approach". In: *Water Resources Research* 44.8. DOI: https://doi.org/10.1029/2007WR006635. eprint: https://agupubs.onlinelibrary.

wiley.com/doi/pdf/10.1029/2007WR006635. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007WR006635.

Sale, T. C., B.L. Parker, C.J. Newell, and J.F. Devlin (2013). *STATE-OF-THE-SCIENCE REVIEW Management of Contaminants Stored in Low Permeability Zones (SERDP Project ER-1740)*. October. SERDP/ESTCP, p. 348. ISBN: 1120091020. URL: https://www.serdp-estcp.org/Program-Areas/Environmental-Restoration/Contaminated-Groundwater/Persistent-Contamination/ER-1740/ER-1740-TR.

Sarma, Pallav, Louis J. Durlofsky, and Khalid Aziz (2008). "Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics". In: *Mathematical Geosciences* 40.1, pp. 3–32. ISSN: 18748961. DOI: 10.1007/s11004-007-9131-7.

Scheidt, C. and J. Caers (May 2009). "Representing Spatial Uncertainty Using Distances and Kernels". In: *Mathematical Geosciences* 41.4, pp. 397–419. ISSN: 1874-8961. DOI: 10.1007/s11004-008-9186-0. URL: http://link.springer.com/10.1007/s11004-008-9186-0.

Scheidt, C., L. Li, and J. Caers (2018). *Quantifying uncertainty in subsurface systems*. Vol. 236. John Wiley & Sons.

Schilling, O. S., P. G. Cook, and P. Brunner (2019). "Beyond Classical Observations in Hydrogeology: The Advantages of Including Exchange Flux, Temperature, Tracer Concentration, Residence Time, and Soil Moisture Observations in Groundwater Model Calibration". In: *Reviews of Geophysics* 57.1, pp. 146–182. DOI: 10.1029/2018RG000619.

Skjervheim, J. A. and G. Evensen (2011). "An ensemble smoother for assisted history matching". In: *Society of Petroleum Engineers - SPE Reservoir Simulation Symposium 2011* 2.2003, pp. 1049–1063. DOI: 10.2118/141929-ms.

Tarantola, Albert (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia: Society for Industrial and Applied Mathematics, p. 358. ISBN: 0-89871-572-5.

Tavakoli, Reza, Sanjay Srinivasan, and Mary F. Wheeler (2013). "Rapid updating of stochastic models using an ensemble filter approach". In: *Society of Petroleum Engineers - SPE Reservoir Simulation Symposium 2013* 2.February, pp. 1342–1353. DOI: 10.2118/163673-ms.

Tonkin, Matthew and John Doherty (2009). "Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques". In: *Water Resources Research* 45.1, pp. 1–17. ISSN: 00431397. DOI: 10.1029/2007WR006678.

Tonkin, Matthew James and John Doherty (2005). "A hybrid regularized inversion methodology for highly parameterized environmental models". In: *Water Resources Research* 41.10, pp. 1–16. ISSN: 00431397. DOI: 10.1029/2005WR003995.

United States Environmental Protection Agency (2004). *Cleaning Up the Nation's Waste Sites: Markets and Technology Trends*. Tech. rep. United States Environmental Protection Agency, p. 338. URL: https://nepis.epa.gov/Exe/ZyPDF.cgi/30006II3.PDF?Dockey=30006II3.PDF.

Van Genuchten, M. Th., J. M. Davidson, and P. J. Wierenga (Jan. 1974). "An Evaluation of Kinetic and Equilibrium Equations for the Prediction of Pesticide Movement Through Porous Media". In: *Soil Science Society of America Journal* 38.1, pp. 29–35. ISSN: 0361-5995. DOI: 10.2136/sssaj1974.03615995003800010016x. URL: https://onlinelibrary.wiley.com/doi/10.2136/sssaj1974.03615995003800010016x.

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert

Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

White, Jeremy T. (2018). "A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions". In: *Environmental Modelling and Software* 109.March, pp. 191–201. ISSN: 13648152. DOI: 10.1016/j.envsoft.2018.06.009. URL: https://doi.org/10.1016/j.envsoft.2018.06.009.

White, Jeremy T., Randall J. Hunt, Michael N. Fienen, and John E. Doherty (2020). *PEST++, a Software Suite for Parameter Estimation, Uncertainty Analysis, Management Optimization and Sensitivity Analysis*. Tech. rep. Reston, Virginia: U.S. Geological Survey, p. 52. DOI: https://doi.org/10.3133/tm7C26. URL: https://www.usgs.gov/software/pest-parameter-estimation-code-optimized-large-environmental-models.

White, Jeremy T., Matthew J. Knowling, and Catherine R. Moore (2020). "Consequences of Groundwater-Model Vertical Discretization in Risk-Based Decision-Making". In: *Groundwater* 58.5, pp. 695–709. ISSN: 17456584. DOI: 10.1111/gwat.12957.

Wöhling, Thomas, Andreas Geiges, and Wolfgang Nowak (2016). "Optimal Design of Multitype Groundwater Monitoring Networks Using Easily Accessible Tools". In: *Groundwater* 54.6, pp. 861–870. ISSN: 17456584. DOI: 10.1111/gwat.12430.

Zheng, Chunmiao and Steven M. Gorelick (2003). "Analysis of solute transport in flow fields influenced by preferential flowpaths at the decimeter scale".

In: *Ground Water* 41.2, pp. 142–155. ISSN: 0017467X. DOI: 10.1111/j.1745-6584.2003.tb02578.x.

Zhou, Haiyan, Jaime Gómez-Hernández, and Liangping Li (2014). "Inverse methods in hydrogeology: Evolution and recent trends". In: *Advances in Water Resources* 63, pp. 22–37. DOI: 10.1016/j.advwatres.2013.10.014. URL: http://dx.doi.org/10.1016/j.advwatres.2013.10.014.

Zinn, Brendan and Charles F. Harvey (2003). "When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields". In: *Water Resources Research* 39.3, pp. 1–19. ISSN: 00431397. DOI: 10.1029/2001WR001146.