ENHANCING INVERSE MODELING IN HYDROGEOLOGY WITH MODERN MACHINE LEARNING ALGORITHMS

A Dissertation
Presented to
The Academic Faculty

by

Ming Liu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in
Computational Science and Engineering

Georgia Institute of Technology
August 2021

ENHANCING INVERSE MODELING IN HYDEOGEOLOGY WITH MODERN MACHINE LEARNING ALGORITHMS

Approved by:

Dr. Jian Luo, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Jingfeng Wang
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Mahdi Roozbahani
School of Computational Science and Engineering
Georgia Institute of Technology

Dr. Emanuele Di Lorenzo
School of Earth and Atmospheric Sciences
Georgia Institute of Technology

Dr. Jimeng Sun
The Grainger College of Engineering, Computer Science
University of Illinois, Urbana-Champaign

Date Approved: 07/29/2021

To my parents and sister

ACKNOWLEDGEMENTS

Many years pursuing the Ph.D. degree at Georgia Tech will be always one of my best experiences in my life. There are many people who deserve recognition and thanks for their time and efforts in leading to the completion of this thesis. Professor Jian Luo has always been providing me with continuous support, guidance, and equally importantly, an environment of high degree of autonomy to explore novel ideas and topics, and for conducting quality research. It is truly a privilege to spend these years learning from such a great mentor and researcher.

Special thanks also go to my committees, Dr. Jimeng Sun, Dr. Jingfeng Wang, Dr. Emanuele Di Lorenzo and Dr. Mahdi Roozbahani for taking the time to engage with my work. This dissertation would not have been completed without their patient guidance and suggestions. Their constructive feedbacks and critical assessment will continue to be invaluable for my future work.

I would also like to acknowledge my colleagues in the School of Computational Science and Engineering and the School of Civil and Environmental Engineering at Georgia Tech: Yue Zhao, Yuzhi Guo, Fangzhou Liu, Tongtong Xu, Saubhagya Singh Rathore, Yuening Tang, Quan Guo, Yu He, Xiaoyang Meng, Ardavan (Ari) Afshar, Yanbo Xu, Sungtae An. It has been always entertaining to grow with them, share research ideas with them, and get through hard times with them. There are so many other friends who have been always helpful in many ways in Atlanta, I would like to express my wholehearted gratitude although their names are not listed here.

Finally, and most importantly, my deepest gratitude and love go to my parents and sister who have cared, loved and believed in me for all these years. Without their unconditional love, support, and sacrifices, my life would never have been pleasant and fulfilling. Their vision with persistence, courage, tenacity, and enthusiasm continue to inspire me in life. This dissertation is dedicated to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

$\mathbf{y}$	AVAILABLE MEASUREMENT.
$\mathbf{f}$	FORWARD MODEL.
$\boldsymbol{\epsilon}$	ERROR INVOLVED IN THE FORWARD PROCESS.
$\mathbf{s}$	UNKNOWN RANDOM VARIABLE IN THE INVERSE PROBLEM.
$\mathcal{R}$	REGULARIZATION TERM.
$\mathbf{T}$	TIKHONOV REGULARIZATION MATRIX.
$\boldsymbol{\nu}$	PARAMETER FOR REGULARIZATION TERM.
$\mathbf{R}$	COVARIANCE MATRIX OF THE ERROR TERM $\boldsymbol{\epsilon}$.
Σ^2	VARIANCE OF THE ERROR TERM ϵ WHEN IT IS ASSUMED AS I.I.D..
$\boldsymbol{\ell}$	LOSS FUNCTION.
$\mathbf{C}$	COVARIANCE MATRIX OF $\mathbf{s}$.
$\mathbf{V}$	ALL THE PRINCIPAL COMPONENT.
$\boldsymbol{\beta}$	DRIFT COEFFICIENT OF THE DRIFT FUNCTION.
$\mathbf{X}$	DRIFT OF MEAN.
$\mathbf{v}$	ONE PRINCIPAL COMPONENT.
$\mathbf{D}$	DIAGONAL MATRIX STORING ALL EIGENVALUES AFTER EIGEN DECOMPOSITION.

$\mathbf{V}_{\mathrm{K}}$	THE FIRST K PRINCIPAL COMPONENTS.
$\boldsymbol{\mu}_{\mathbf{s}}$	MEAN VECTOR OF THE INTERESTED RANDOM VARIABLE.
$\mathbf{a}$	REDUCED RANDOM VARIABLE, PRINCIPAL COMPONENT COEFFICIENT.
$\boldsymbol{\Sigma}_{\mathrm{A}}$	COVARIANCE MATRIX OF $\mathbf{a}$.
$\mathbf{f}^{-1}$	INVERSE MODEL
$\boldsymbol{\epsilon}'$	ERROR INVOLVED IN THE INVERSE PROCESS
$\boldsymbol{\kappa}$	HYDRAULIC CONDUCTIVITY.
$\boldsymbol{s}$	STOREAGE COEFFICENT.

SUMMARY

Inverse estimation of spatially distributed parameter fields plays an important role in many scientific disciplines including hydrogeology, geophysics, earth science, environmental engineering, etc. Classic stochastic sampling approaches such as Markov Chain Monte Carlo (MCMC) and optimization approaches such as geostatistical approach (GA) can solve inverse problems with a modest number of unknowns. However, we may face challenges when it comes to large-scale, highly heterogeneous fields or fields with special characteristics, such as connected preferential paths.

In this thesis, firstly, we develop a new data augmentation approach, i.e., fast conditional image quilting to synthesize realizations based on limited measurements; and this approach is later used to generate channelized training images to support the inverse modeling research study.

Secondly, unlike MCMC and optimization approaches that require many forward model evaluations in each iteration, we develop two neural network inverse models on full dimensions (NNI) and principal components (NNPCI) to directly explore the inverse relationships between indirect measurements such as hydraulic heads and the underlying parameter fields such as hydraulic conductivity. We successfully apply our neural network models to large-scale hydraulic tomography experiments to estimate spatially distributed hydraulic conductivity. In particular, with the help of principal component analysis (PCA), the number of neurons in the last layer of NNPCI is the same as that of retained principal components, thus further accelerating the algorithm and making the system scalable regardless of large-scale unknown field parameters. NNI also demonstrates satisfactory

inverse results on full dimensions for both Gaussian and non-Gaussian fields with channelized patterns. The major computational advantage for NNI and NNPCI is that the training data can be generated by independent forward model simulations that can be done efficiently using parallel computing.

Finally, to account for errors from different sources, including input errors, model structure errors, model parameters errors, etc., we incorporate Bayesian theorem to the neural network models for uncertainty analysis. The system behaves more stably and consistently on varying spatial and temporal scales. The developed approaches are successfully validated with synthetic and field cases.

CHAPTER 1. INTRODUCTION

In groundwater hydrology, aquifer simulation is necessary to better understand groundwater flow and transport behavior, and can be used to guide effective *in-situ* remediation and groundwater resources management. Aquifer simulation requires the characterization of hydrogeologic properties, such as hydraulic conductivity and specific storage coefficient, which are inherently heterogeneous in space. In hydrogeology, local-scale estimates of hydraulic conductivities can be obtained by laboratory measurements of core samples (A. Klute & R. C. Dinauer, 1986). However, high costs of drilling wells on highly discretized grids hinder the broader application of this approach in field practice. Therefore, characterization of such spatially distributed parameters is usually achieved by inference from indirect measurements such as hydraulic heads from slug tests and pumping tests, and concentrations from tracer tests, etc. (Cardiff et al., 2009; Cirpka et al., 2007; Fienen et al., 2006). The inference approach is also called as the inverse modeling approach because it derives input parameters from indirect measurable outputs. This thesis mainly investigates inverse modeling of large-dimensional hydrogeologic parameter fields such as hydraulic conductivity, which is usually considered as the major uncertainty source in aquifer simulations. The proposed research approaches and computational algorithms can be readily extended to invert random fields of other physical properties in a variety of scientific disciplines, including hydrology, geology, geophysics, earth science and environmental engineering, etc.

In order to handle large-scale and highly heterogeneous hydrogeological inverse problems, we usually apply stochastic sampling methods such as Markov Chain Monte

Carlo (MCMC) and gradient-based optimization methods such as geostatistical approach (GA). Both these methods are capable of inverting unknown parameter fields replying on repeated forward model evaluations. Thus, the computational cost often becomes unaffordable for large-dimensional inverse problems when a large number of conditional realizations need to be generated or large-dimensional Jacobian matrices need to determined iteratively. Moreover, spatial covariance based on two-point geostatistics is typically used to regularize the random fields that are the potential solutions. Such geostatistical regularizations may not be applicable for random fields with special features such as connected paths.

The main goals of this thesis include:

- **Goal 1**: Development of efficient data augmentation algorithms that could characterize specific properties such as better connectivity of the field;
- **Goal 2**: Development of efficient and accurate algorithms by combining state-of-the-art neural network machine learning with inverse modeling for inverse estimation of large-dimensional hydrogeological parameter fields.

To achieve Goal 1, the method of fast conditional image quilting (FCIQ) is developed to help substantiate the research of inverse modeling. For Goal 2, the neural network machine learning methods consider cases with and without geostatistical regularizations. In the absence of spatial covariance, neural network is designed to learn the entire image, i.e., full dimensional machine learning. For given spatial covariance, a neural network principal component inverse framework is developed based on reduced dimensions, named as NNPCI framework. Moreover, to address the errors from various

sources, variational Bayesian neural principal component inverse (VBNPCI) is developed to conduct uncertainty analysis by adding Bayesian distribution for the neurons of each hidden layer in the NN architecture on top of NNPCI. The proposed algorithms are validated by various numerical experiments on both synthetic data and field data.

1.1 Fast Conditional Image Quilting (FCIQ)

FCIQ were based on multiple point statistics (MPS) and texture synthesis (TS) methods to generate spatially distribution parameter fields. Due to their limitations, neither of these two methods can be generalized to all random fields. Specifically, MPS was proposed to address the incapability of two-point geostatistical characterization using training images (TIs), from which more explicit representation of higher order statistics of measured field data can be extracted. TS was further developed from expert understanding and geologist sketching. However, both methods are computational expensive, particularly for large-scale fields with special distribution features. For example, TS uses pixel as the unit of synthesis and exhaustive search across the entire image is required for conditional distribution computing. In this thesis research, to further increase the computational efficiency, two approaches are incorporated in the FCIQ algorithm:

(1) Use of the extended patch approach and multiple path variations to help FCIQ in searching a good patch for its neighboring block and increasing the availabilities of pattern selections.
(2) Development of a multiple resolution approach with Fast Fourier Transform (FFT) to accelerate the search process.

1.2 Neural Network Inverse (NNI) on Full Dimensions and Neural Network Principal Components Inversion (NNPCI)

The classic geostatistical approach is a gradient-based optimization approach for Bayesian inverse problems. Instead of characterizing the entire posterior distribution, the geostatistical approach only finds the maximum a posteriori (MAP) estimate, usually denoted as "the best estimates, and provides the associated uncertainty by its posterior covariance matrix (Kitanidis, 1996). The geostatistical approach may require a large number of forward model runs in each iteration to determine Jacobian matrix for nonlinear inverse problems, which is a large obstacle to solve large-dimensional inverse problems such as hydraulic tomography and generate conditional realizations or posterior realizations. In contrast, the use of neural networks to simulate the inverse modeling (NNI) of hydraulic tomography can be much more efficient by avoiding the determination of the Jacobian matrix and generating conditional realization upon well-trained framework. However, for an inverse problem with highly parameterized conductivity fields, the computational overhead of full-dimensional NNI is still a practical issue. NNPCI mainly makes three contributions to the geostatistical approach:

(1) Using neural networks to bridge the inverse connection between measurements and the underlying parameter field.
(2) Providing a more intuitive method, without computing Jacobian matrices, and quickly generating conditional realizations.
(3) Incorporating PCA to make the framework scalable, because the dimensionality of the last neural layer is independent of the number of unknown parameters.

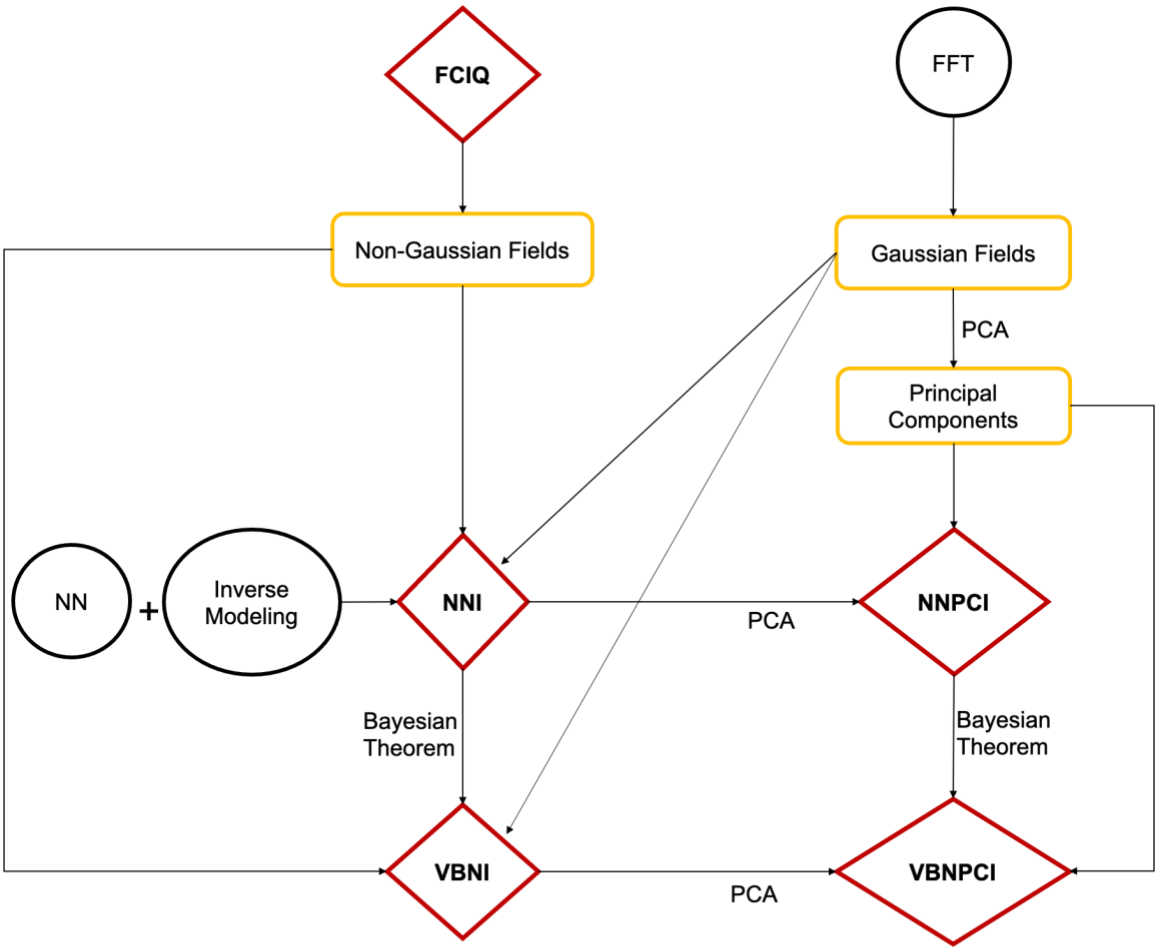
This approach is more advantageous to the inversion of large-scale and highly heterogeneous fields.

1.3 Variational Bayesian Neural Inverse (VBNI) on Full Dimensions and Variational Bayesian Neural Principal Components Inversion (VBNPCI)

VBNI and VBNPCI are built on top of NNI and NNPCI respectively by adding uncertainty to all neurons in hidden layers of the framework. Uncertainty is created on top of the weights instead of roughly adding variations to the last layer of the neural network. This method can create reasonable conditional realizations while preserving the dominant field patterns.

1.4 Organization of Thesis

This thesis is composed of seven chapters. In CHAPTER 2, a literature review regarding Bayesian inverse modeling, geostatistical approach and machine learning methods in Hydrology is provided. In CHAPTER 3, we propose a new data augmentation approach for random field characterization and validate it by several experiments. In CHAPTER 4, we develop the method of full-dimensional neural network and neural network principal component inverse approach. In CHAPTER 5, uncertainty analysis is introduced to the neural network methods and validated by three numerical experiments with different synthetic data. In CHAPTER 6, the performance of NNPCI is applied to inverting real world data. In CHAPTER 7, we conclude the discussion by summarizing current work and providing potential future research directions. Figure 1 visualizes the flowchart for the thesis structure.



- FCIQ: Fast Conditional Image Quilting
- FFT: Fast Fourier Transform
- PCA: Principal Component Analysis
- NN: Neural Network
- NNI: Neural Network Inverse
- VBNI: Variational Bayesian Neural Inverse
- VBNPCI: Variational Bayesian Neural Principal Components Inverse
- Non-Gaussian Fields: Channelized domain, connected
- Gaussian Fields: Gaussian fields with Gaussian or exponential covariance function, non-connected

Figure 1. Thesis Structure Flowchart.

CHAPTER 2. LITERATURE REVIEW

2.1 Bayesian Inverse Problems

A measurement equation is fundamental to understand the concept of inverse problem, which can be generally described as:

$$\mathbf{y} = \mathbf{f}(\mathbf{s}) + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^{n\times 1}$ represents the available data; $\mathbf{f}$ represents the forward model, which is mainly a physical process described by a group of governing equations and solved by numerical solvers; $\mathbf{s} \in \mathbb{R}^{m\times 1}$ represents the interested variables to be estimated by the inverse approach and $\boldsymbol{\epsilon}$ represents errors involved in the forward process, which can include the errors from oversimplification of physical models, random measurement errors, and numerical solver errors, etc. In hydrogeological applications, variable $\mathbf{s}$, with hydraulic conductivity as the most representative member, is often recognized as the main source of uncertainty (Carrera & Neuman, 1986; Kitanidis, 1996, 1997), whose characterization is required for a trustworthy model simulation. In practice, one way to estimate hydraulic conductivity in a local scale is by core samples (A. Klute & R. C. J. P. Dinauer, 1986). However, the high expenses of well drilling prevent this method from broader applications. On the other hand, the acquisition of $\mathbf{y}$ data, which could be head data, tracer concentration data, or temperature data, is more convenient compared with that of $\mathbf{s}$. Therefore, in a real-world application, the problem is usually processed in an 'inverse' way: the estimate and the uncertainty of $\mathbf{s}$ are usually inferred by the available measurements $\mathbf{y}$. The inverse problem can be viewed as an optimization problem as

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathrm{f}(\mathbf{s})\|^2 \tag{2}$$

where $\hat{\mathbf{s}} \in \mathbb{R}^{m \times 1}$ represents the best estimate of $\mathbf{s}$ given the available data $\mathbf{y}$. However, in real world applications, the high-dimensional unknown $\mathbf{s}$ and the sparse observational measurements $\mathbf{y}$ incurs extraordinary high parameter-to-data sensitivity and ill-posed properties to this optimization problem. To mitigate the inherent instability and non-uniqueness of inverse problems, regularizations are required.

The form of the regularization term $\mathcal{R}(\mathbf{s})$ is problem dependent. Tikhonov regularization is one of the most popular ones in inverse problems. A Tikhonov regularization term can be formulated as:

$$\mathcal{R}(\mathbf{s}) = \|\mathbf{T}\mathbf{s}\|^2 \tag{3}$$

where $\mathbf{T}$ is called Tikhonov matrix. The Tikhonov regularization term imposes smoothness onto the solution of the optimization problem in Equation (4) and thus compensate its non-uniqueness. As a smoothing operator, matrix $\mathbf{T}$ has various forms in different applications. After adding regularization, the inverse problem can be formulated as

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathrm{f}(\mathbf{s})\|^2 + \upsilon \|\mathbf{T}\mathbf{s}\|^2 \tag{4}$$

whose loss function is

$$\ell(\mathbf{s}) = \frac{1}{2} \|\mathbf{y} - \mathrm{f}(\mathbf{s})\|^2 + \upsilon \|\mathbf{T}\mathbf{s}\|^2 \tag{5}$$

where υ is usually a tuning parameter that controls the relative weights between the fitting to the observational data and the smoothness of the solution **s**.

The formulation of the inverse problem can also be interpreted from a Bayesian view. According to Bayes' rule, the posterior distribution $\mathrm{p}(\mathbf{s}|\mathbf{y})$ of the unknown parameter **s** is proportional to the product of its prior distribution $\mathrm{p}(\mathbf{s})$ and the likelihood of the observational data $\mathrm{p}(\mathbf{y}|\mathbf{s})$

$$\mathrm{p}(\mathbf{s}|\mathbf{y}) \propto \mathrm{p}(\mathbf{y}|\mathbf{s})\mathrm{p}(\mathbf{s}) \tag{6}$$

The likelihood encodes information from observational data and the prior provides the former beliefs about **s**, both of which constrain the posterior distribution. For a zero-mean Gaussian error term, the likelihood function is

$$\mathrm{p}(\mathbf{y}|\mathbf{s}) \propto \exp\left(-\frac{1}{2}\big(\mathbf{y}-\mathrm{f}(\mathbf{s})\big)^{\mathrm{T}}\mathbf{R}^{-1}\big(\mathbf{y}-\mathrm{f}(\mathbf{s})\big)\right) \tag{7}$$

where **R** represents the covariance matrix of the error **ϵ**. A zero-mean Gaussian prior can be described as

$$\mathrm{p}(\mathbf{s}) \propto \exp\left(-\frac{1}{2}\mathbf{s}^{\mathrm{T}}(\mathbf{T}^{\mathrm{T}}\mathbf{T})\mathbf{s}\right) \tag{8}$$

where $(\mathbf{T}^{\mathrm{T}}\mathbf{T})$ acts as the inverse of the covariance matrix. Following this equation, the posterior distribution is

$$p(\mathbf{s}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{y}-f(\mathbf{s})\right)^{T}\mathbf{R}^{-1}\left(\mathbf{y}-f(\mathbf{s})\right)-\frac{1}{2}\mathbf{s}^{T}(\mathbf{T}^{T}\mathbf{T})\mathbf{s}\right) \quad (9)$$

For $\mathbf{R}=\sigma^2\mathbf{I}$, it can be simplified as

$$p(\mathbf{s}|\mathbf{y}) \propto \exp\left(-\left(\frac{1}{2\sigma^2}\|\mathbf{y}-f(\mathbf{s})\|^2+\|\mathbf{T}\mathbf{s}\|^2\right)\right) \quad (10)$$

If $\sigma^2=\upsilon$, the posterior distribution is proportional to the exponential of a negative loss function

$$p(\mathbf{s}|\mathbf{y}) \propto \exp\left(-\ell(\mathbf{s})\right) \quad (11)$$

Therefore, to find the parameter **s** that minimizes the loss function in Equation (5) equals to find the **s** that maximizes the posterior in Equation (10), and the inverse problem given an explicit loss function is formulated as a Bayesian inverse problem.

The regularization terms in inverse problems, or the prior distribution in Bayesian inverse problems have various forms. Chen et al. (2008) solved an inverse problem by an optimization method, where the regularization term is selected as the second norm of the interested variable. In this case, the Tikhonov matrix is a scaled identity matrix $\mathbf{T}=\lambda\mathbf{I}$, and the regularization term is $\mathcal{R}(\mathbf{s})=\lambda^2\|\mathbf{s}\|^2$. From a Bayesian perspective, it equals to applying a zero-mean multi-Gaussian prior onto the variable **s** with covariance $\mathbf{C_s}=\frac{1}{\lambda}\mathbf{I}$. Liu and Kitanidis (2011) leverage the connection between optimization and Bayesian inverse problems to employ a sparse Tikhonov matrix to enforce the smoothness of the inversion result.

Another widely accepted form of the covariance matrix, especially in ground water research area, the one defined by geostatistical covariance function (Matheron, 1963). This form has been applied to plenty research in hydrogeological applications (Kitanidis, 1995, 1996; Jonghyun Lee et al., 2018; S. Liu et al., 2002; Liu et al., 2007; Nowak & Cirpka, 2006).

By adding regularization terms or priors, a well-posed Bayesian inverse problem is formulated and has a unique solution. The extra advantage of using prior in Bayesian approach is to provide uncertainty quantifications for the best estimate. Various computational approaches have been developed to seek for solutions of Bayesian inverse problems, which can be divided into two categories: the ensemble-based approach and the non-ensemble-based approach.

2.2 Connectivity Measures in Hydrogeology

In the past decades, many geologists have demonstrated that flow and transport behaviors are significantly different in disconnected heterogenous fields and connected heterogenous fields, i.e. fields with preferential flow paths of connected high hydraulic conductivities (Fogg, 1986; Gong, 2013; Luo & Cirpka, 2011; Silliman & Wright, 1988). A good example is that connected high-conductivity paths might lead to early breakthrough of contaminant plumes and enhance tailing behavior due to kinetic mass transfer between fast and slow flow zones, which cannot be quantified by classical advection-dispersion models and macro dispersion theory (Zinn & Harvey, 2003). Geologists use the degree to which they are interconnected instead of the sizes or directions to represent a major feature of connected paths.

Quantification of connectivity properties in hydrogeology utilizes two-point cluster functions and two-cut indicators, as shown in Equation (12), which has been successfully applied to many implications such as soil moisture patterns and hydraulic conductivity in aquifer formations (Gong, 2013; Neuweiler & Cirpka, 2005; Torquato et al., 1988; Western et al., 2001).

$$C(h) = \frac{< I(x)P(x,x') >_{|x-x'|=h}}{< I(x) >} \tag{12}$$

$$P(x,x') = \begin{cases} 1, & Cl(x) = Cl(x') \\ 0, & otherwise \end{cases} \tag{13}$$

where $C(h)$ represents the probability that two points x and x' with distance $h = |x - x'|$ belongs to the same cluster. The two-cut indicator field $I(x)$ is equal to 1 if the original value at location x is between two thresholds defined by users and taking $I(x)$ as 0 otherwise. $Cl(x)$ denotes the indexed cluster at location x.

Figure 2 shows an example of heterogeneous fields with different connectivity. The first row represents a regular random field with the spatial correlation described by a Gaussian model. The cluster analysis indicates that the values near the mean are connected. The second and third rows represent fields with high-value and low-value connected paths, respectively. The cluster analysis indicates that the probabilities of elements with connected high values and low values are higher than other types of patterns, respectively. We also need to notice that all fields shown in Figure 2 have the same two-point spatial correlations. The right column represents the two-point cluster function for each field: the red line, the blue line and the green line indicate low permeability, intermediate

permeability, and high permeability respectively. Taking the normal random field as an example, the red line and the green line tend to be zero when the distance between two points increases, which means two points with very low or high conductivity values are not likely to be connected. In contrast, the blue line is stabilized at 0.5 when the distance is very large, this means most intermediate-valued points are connected, no matter how far away they are. Clearly, the connected fields cannot be sufficiently characterized by such two-point spatial correlations.

Traditional inverse modeling in Hydrogeology typically applies two-point spatial correlations to regulate the smoothness of underlying parameter fields. Two-point based approaches work well for random fields that can be sufficiently characterized by spatial covariance models as a function of distance or oriented spacing (anisotropic fields). However, such approaches may fail for spatial fields with special characteristics, such as highly connected preferential paths, because two-point spatial correlations cannot sufficiently describe those properties.

In recent decades, many state-of-the-art training-image based multiple points statistics (MPS) methods emerge to formalize spatial patterns of specific phenomena in geological area, for example, SNESIM (Strebelle, 2002), FILTERSIM (Wu et al., 2008) and Direct Sampling (Rezaee et al., 2013). Among these approaches, they were developed aiming to generate non-Gaussian random fields or reflect higher-order statistics that cannot be extracted by two-point statistics. However, they suffer from huge computational and storage cost when it comes to deal with large-scale inverse problem in real-world applications. It is therefore worth putting more efforts into advancing MPS research study and developing more efficient inverse algorithms to handle spatially-correlated fields and

complex random fields well and better understand flow and transport behavior in heterogeneous fields.

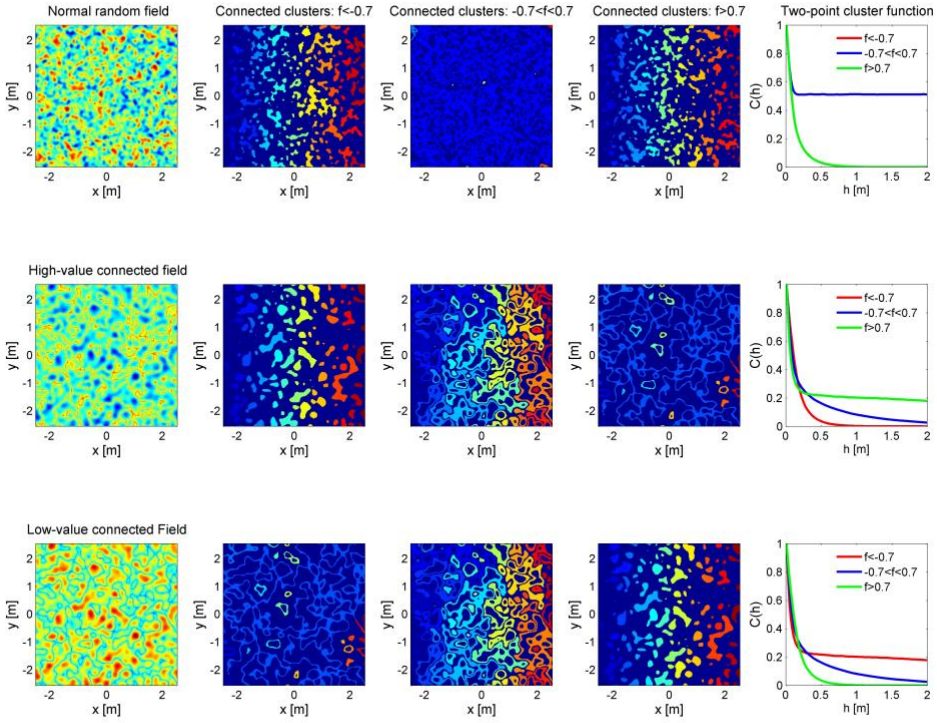


Figure 2. Spatially-Correlated Random Fields and Cluster Analysis.

2.3 Geostatistical Inverse Modeling

Geostatistical inverse as an optimization method directly searches for the MAP estimate of the posterior distribution without simulating any conditional realizations. The associated uncertainty of the MAP estimate is mostly computed by its local covariance matrix. It typically requires a large number of forward model runs in each iteration to correct the values of unknown parameters until convergence.

Researchers have made continued dedications to improve the efficiency of these optimizations algorithms. Hu et al. (2011) proposes an approximated Hessian preconditioned conjugate gradient approach to solve seismic wave inversion problems. Bui-Thanh et al. (2012) proposes a conjugate gradient approach that combined with a randomized approximating algorithm for Hessian. This approach is also validated in a seismic wave Bayesian inverse problem.

In hydrogeology, by embracing a geostatistical description of the prior covariance matrix (Matheron, 1963), the geostatistical approach was proposed to solve Bayesian inverse problems. Two representative geostatistical approaches are the quasi-linear geostatistical approach and the Successive Linear Estimator (SLE). SLE is a cokriging-like geostatistical approach that produces a linear estimator with successive assimilation of nonlinear relationship in the forward model (Yeh et al., 1995; Yeh et al., 1996). Its effectiveness and efficiency have been validated in various applications, especially in hydraulic tomography problems (S. Liu et al., 2002; Yeh & Liu, 2000; Zhu & Yeh, 2005; Zhu & Yeh, 2006). Recently, to expand the application realm of SLE, Karhunen-Loeve

expansion is incorporated into it to construct a reduced-order SLE for large-scale hydraulic tomography problems (Zha et al., 2018).

Geostatistical Approach (GA) proposed by Hoeksema and Kitanidis (1984); Kitanidis and Vomvoris (1983); Kitanidis (1995) is an iterative quasi-linear approach. The quasi-linear geostatistical approach also has a close relation to Gauss-Newton optimization and cokriging (Nowak & Cirpka, 2004), and thus has high efficiencies in solving Bayesian inverse problems. However, for highly-parameterized Bayesian inverse problems that admit high-resolution hydraulic conductivity fields, the geostatistical approach still calls for special implementations to abate the computation overhead (Lee & Kitanidis, 2014; Liu & Kitanidis, 2011). The reasons lie in two aspects: firstly, numerical simulation of forward models needs to be performed on a high-resolution grid for multiple times; and secondly, the large number of unknowns increases the cost of matrix computation during the inversion.

In the past decades, many researchers have devoted to reducing the computational cost of the quasi-linear approach. Nowak et al. (2003) employed an FFT algorithm for efficient computations of the cross-covariance matrix, which was applied to a sandbox inverse problem (Nowak & Cirpka, 2006). Liu and Kitanidis (2011) proposed a sparse prior covariance matrix that can be recognized as a general smoothing operator, which facilitates the storage and the computation of the covariance matrix. This sparsely formulated approach was validated in a sandbox hydraulic tomography problem. For large-scale linear geostatistical inverse problems, Ambikasaran et al. (2013) and Saibaba et al. (2012) proposed hierarchical matrices to accelerate the solving process. For Bayesian inverse problems with large volume of observational data, Klein et al. (2017) proposed a

preconditioned conjugate gradient algorithm, and Lin et al. (2017) proposed a randomized data reduction algorithm to condition the quasi-linear approach. For inversion of time series data, such as pumping and tracer tests, the concept of temporal moments was applied. Instead of running the transient forward model hundreds or even thousands of times, the temporal moment approach transforms them to be steady-state for faster forward model evaluations (Li et al., 2005; Zhu & Yeh, 2006; Nowak & Cirpka, 2006; Pollock & Cirpka, 2008; Yin & Illman, 2009).

Recently, Kitanidis and Lee (2014) introduced the principal component analysis (PCA) in the quasi-linear geostatistical approach to ease the computation of Bayesian inverse problems, named as principal component geostatistical approach (PCGA), and demonstrated its effectiveness and efficiency in several hydrogeological applications (Lee & Kitanidis, 2014; Lee et al., 2016; Kang et al., 2017). PCGA substantially reduces the computational cost associated with the covariance matrix and the computation of the Jacobian matrices. Besides, unlike other optimization approaches that rely on adjoint-solvers, PCGA only requires forward model itself to compute the Jacobian matrix. Later Zhao and Luo (2020) proposed the reformulated geostatistical approach (RGA) which possessed a more intuitive way of computing Jacobian matrices, and yielded a scalable normal equation system that particularly benefits inverse problems with a large volume of measurements. On top of RGA, upscaling principal component geostatistical approach (UPCGA) further takes advantage of RGA and recognized the principal component coefficients as latent variable to connect the coarse grid and the fine grid solution.

2.4 Application of Machine Learning to Hydrogeology

Recent years have witnessed the fast development of machine learning, deep learning and artificial intelligence across many different scientific disciplines including computer science, mathematics, biology, physics, etc. In hydrogeology area, many physical models and theories have been developed rapidly for years. There could be significantly more information in large-scale hydrological data sets than hydrologists have been able to translate into theory or physical model. Thus, more and more geologists have dived into the data-driven algorithms, i.e., machine learning, to facilitate maximal extraction of information from spatial data, temporal data, or any other type of source. For example, the use of long short-term memory (LSTM) by Fang et al. (2017) has ignited a surge in machine learning applications across all domains of hydrology. A variety of machine learning techniques have been adopted to address challenges in hydrogeology area (Shen et al., 2021).

Precipitation and snow are important components on earth and also major sources of uncertainty. Sun and Tang (2021) employed an attention-based, deep convolutional neural network to downscale coarse-resolution satellite-based precipitation data including elevation and air temperature. To handle the missing data in gauged precipitation data due to instrumentation and data quality issues, Mital et al. (2020) developed a new sequential imputation algorithm based on a random forest technique for interpolating the missing values on time and space scale for daily precipitation records. The simulation of snow water equivalent was implement using LSTM to leverage the climate and snow water equivalent data from five observation stations (Meyal et al., 2021). They have validated the plausibility of using LSTM for large-scale operational snow water equivalent modeling and built automated prediction online system.

As for groundwater, a multilayer perceptron layer (MLP) is trained to predict three-point observations of groundwater levels using temperature, precipitation, river discharge and past groundwater data, and proves that all the time fed into the MLP system as a whole does not necessarily produce the optimal choice (Sahu et al., 2020). They also found that MLPs trained solely on temperature and historical groundwater level measurements as features were unreliable at all locations, which alluded dynamical linkage between surface and groundwater. Future sensitivity analysis well likely be accompanied by uncertainty estimates to ensure the robustness of the analysis.

In the subsurface environment, due to the extreme data scarcity, geologists usually generate synthetic data with all kinds of special characteristics for further research study. One of popular artificial techniques Generative Adversarial Networks (GAN) are becoming an alternative to Two-Point Statistics and Multiple-Point Statistics (MPS) to generate stochastic subsurface fields from training images, the main pattern or data trend can be self-learnt and captured by using optimization approach. What's more, this technique could be extended to 3D application of reconstructing subsurface architecture when it is not possible to collect exhaustive and accurate data about 3D subsurface distribution of rock types. Coiffier and Renard (2019) proposed Dimension Augmenter GAN (DiAGAN) that enables GANs to generate 3D fields from 2D training samples and this approach works well for generating complex binary subsurface media. Mital et al. (2020) use a deep neural network of the variational autoencoder to construct a parametric low-dimensional base model parameterization of complex binary geological media which was a largely unsolved challenge. This approach it more suited for probabilistic inversion

than for unconditional geostatistical simulation; and the performance is more promising than Multiple-Point Statistics (MPS) by Sequential Geostatistical Resampling (SGR).

Inspired by the rapid development of machine learning, deep learning, computer vision, etc. across many scientific disciplines, we are thinking if it is a good point to apply neural network to simulate the black box problem of inverse computation. You can think of a scenario in biology that many impulses are carried toward a body cell, transported along the axon, and finally arrive at the terminal axon. As a result, the information is carried successfully from the origin to a selected destination; the logic is repeatedly all the time in our body. Though biologists have explored the information sharing and the underlying chemical reactions for many decades, we can still employ this idea to our research. Therefore, it is likely not to rely on stochastic sampling methods such as MCMC and gradient based optimization methods such GA to address the inverse calculation at the high expensive cost of computational budget; instead, we switch to a data-driven approach to replace the inversion of a forward model and learn the relationship between indirect measurements and the large-scale parameterized random fields directly. Another significant reason is that we do not care if the forward model is linear or non-linear because many more advanced machine learning techniques such as neural network will handle the forward model in the same way. That's why in this thesis we switch from traditional geostatistical perspective to machine learning and explore its implication in hydrogeology area.

To sum up, there are multiple ways for machine learning to make advances in hydrogeology application. We can incorporate physics in ML models to improve the system interpretability, develop physics-informed neural network and provide uncertainty

analysis to model results. In this thesis, I will develop a data augmentation algorithm to create complex domains with special characteristics which will be reused in the following inverse modeling research. Neural networks will be incorporated into hydraulic physics forward models followed by uncertainty estimation with the help of Bayesian theorem. As a result, the robustness of the entire system will be enhanced on spatial and temporal scales.

CHAPTER 3. FAST CONDITIONAL IMAGE QUILTING (FCIQ)

3.1 Introduction

Subsurface heterogeneity, varying spatially and temporally, has always been a very important and challenging factor associated with most subsurface flow and transport processes and extensively considered in the management and uncertainty prediction in hydrogeological systems. Due to limited and costly field data, one urgent need is to generate realistic realizations of graphical textures with natural stochasticity and variability to represent the unknown hydrogeological system. Many geostatistical approaches have been developed to address the significant uncertainty issue in numerical modeling by characterizing the spatial structures (i.e. pores, grains, channels and fractures, etc.) to investigate important variables such as hydraulic conductivity, hydraulic head, porosity and electrical resistivity. Traditional geostatistical tools mainly rely on spatial variogram and covariance analysis, which are then used in classic kriging and indicator simulations in hydrogeological applications. However, variograms can only provide the correlation information between two points in space with strong assumptions of specific patterns such as Gaussian Random fields, and lack the flexibility to represent realistically complex textures such as channeling, which are very important in natural system, thus limiting its applicability in physical realism (Gómez-Hernández & Wen, 1998; Journel & Zhang, 2006; Mahmud et al., 2014; Sánchez-Vila et al., 1996; Schlüter et al., 2011; Western et al., 2001).

Alternatively, state-of-the-art Multiple Points Statistics (MPS) was proposed to address the incapability of two-point geostatistics based on machine learning of Training Images (TIs), from which more explicit representation of higher order statistics of

measured field data can be extracted to better formalize the spatial patterns of specific phenomena and allow the model to condition on field data (Mahmud et al., 2014; Mustapha & Dimitrakopoulos, 2010). TIs might come from real data representative of hydrology or geology under consideration, unconditional realizations from other reasonable stochastic simulation techniques, outcrops data and synthetic TIs developed from expert understanding and geologist sketching (Mahmud et al., 2014; Strebelle, 2002). Applications of MPS algorithms include reservoir modeling (Caers et al., 2003; Falivene et al., 2006; Ronayne et al., 2008), water resources modeling such as remote sensing (Boucher et al., 2008; G Mariethoz et al., 2012), fluids physics in porous media (Okabe & Blunt, 2007; Tahmasebi & Sahimi, 2013), climate modeling (Jha et al., 2013) and even medical imaging (Pham, 2012). MPS algorithms developed include SNESIM (Strebelle, 2002), SIMPAT (Arpat & Caers, 2007), FILTERSIM (Wu et al., 2008), MPPCA (Abdollahifard & Faez, 2013) and Direct Sampling (Rezaee et al., 2013). However, since large-scale spatial or temporal data of 2-D or 3-D models may contain millions of nodes, the computation of one single realization using current methods costs too much memory and time, which may become unaffordable if inverse modeling is involved. Thus, there is an urgent need of developing new methods to radically accelerate the current MPS algorithms. We should be aware that we cannot rely on the computer hardware development such as GPU graphics replacing CPU because algorithms have always been weighing more balance than hardware developments in the past decades.

In computer graphics, texture synthesis methods pursue the similar goals to MPS, i.e. generating new realistic or unrealistic graphical images with similar pattern to sample training images, as well as stochasticity and variability. Texture synthesis has been widely

used to generate large amount of images for the application of video games and animated movies(Gregoire Mariethoz & Lefebvre, 2014). An early method was proposed by (Efros & Leung, 1999), but inefficient because the simulation used pixel as the unit of synthesis, which was proceeded by exhaustive search of the entire image every time to locate all likely neighborhoods for computing the conditional distribution. Later, one patch based texture synthesis method named Image Quilting (IQ) was proposed in computer graphics literature (Efros & Freeman, 2001) to speed up the computation.

Efros and Freeman (2001) proposed a 2D unconditional IQ algorithm to synthesize new textures based on given sample training images. IQ can be compared to jigsaw puzzle since the key idea is to take all possible textures from a given TI and sew them together in a coherent manner. IQ starts from any corner of the simulation grid and assemble the patches one by one along the unilateral raster path or random raster path. In every step the algorithm selects a random patch from several best candidates which have the minimum squared Euclidean distance with the neighborhood over the overlapping area and put the corresponding patch onto the simulation grid (SG). The main parameters of this algorithm are the patch size, overlapping area size and the number of candidates for selection. Generated realizations will be different from each other by varying the parameters while maintaining the important property of stochasticity and variability in physical realism. At the same time, the patch selection will speed up significantly than traditional pixel-based algorithm and improve the performance. Mahmud et al. (2014) successfully adopted image quilting for hydrogeological applications and extended from originally unconditional simulations to conditional simulations in subsurface hydrogeology. Further improvements were made for the pattern-based simulation by using multi-scale (MS) representation of

the training image instead of traditional multi-grid strategy and accelerating the search computation in Fourier space (Tahmasebi et al., 2014).

In this chapter, we employed the pattern-based simulation approaches (Boykov & Jolly, 2001; Efros & Freeman, 2001; Kwatra et al., 2003; Mahmud et al., 2014; Tahmasebi et al., 2014) to generate conditional realizations with natural stochasticity and variability that represent the hydrogeology tomography. Specifically, multiple optimization techniques associated with IQ such as cross correlation, multiple-resolution, extended-patch for data conditioning and Fast Fourier Transform (FFT) were incorporated. As a result, conditional or unconditional realizations for fields with special characteristics are augmented that will be reused to support inverse modeling in CHAPTER 4 and CHAPTER 5.

3.2 Methodology

3.2.1 Cross Correlation

Cross Correlation (Lewis, 1995) is a measure of similarity of two series as a function the displacement of one relative to the other, which is known as a sliding dot product or sliding inner-product. It has been extensively used in applications like image-processing, signal processing, pattern recognition, single particle analysis, electron tomography, averaging, and neurophysiology. Prior to cross correlation, Euclidean distance, widely used in many areas with its simplicity and easy operation, is defined through the normal equation, which and the further expansion are given by:

$$d_{f,t}{}^2(u,v) = \sum_{x,y}[f(x+u,y+v) - t(x,y)]^2$$
$$= \sum_{x,y}[f^2(x+u,y+v) - 2f(x+u,y+v) * t(x,y) + t^2(x,y)] \quad (14)$$

where $f(x+u,y+v)$ represents TI and the sum is over x,y under the window t positioned at (u,v) with the window size of p_1, p_2

The term $\sum_{x,y}[f^2(x+u,y+v)]$ represents the energy of an image and is roughly constant, and the term $\sum_{x,y} t^2(x,y)$ from specific target patch is constant. We can approximate our original Euclidean distance equation to the term of cross correlation to simplify the similarity measurement between TI and specific feature.

$$C(u,v) = \sum_{x,y} f(x+u,y+v) * t(x,y) \quad (15)$$

In addition to simplicity, cross correlation can capture the heterogeneity information for complex image patterns and generate realizations matching with TI better. In contrast, the selection of Euclidean distance could often tend to generate realizations quite different from TI and even diverge from physical realism.

3.2.2 Extended-path

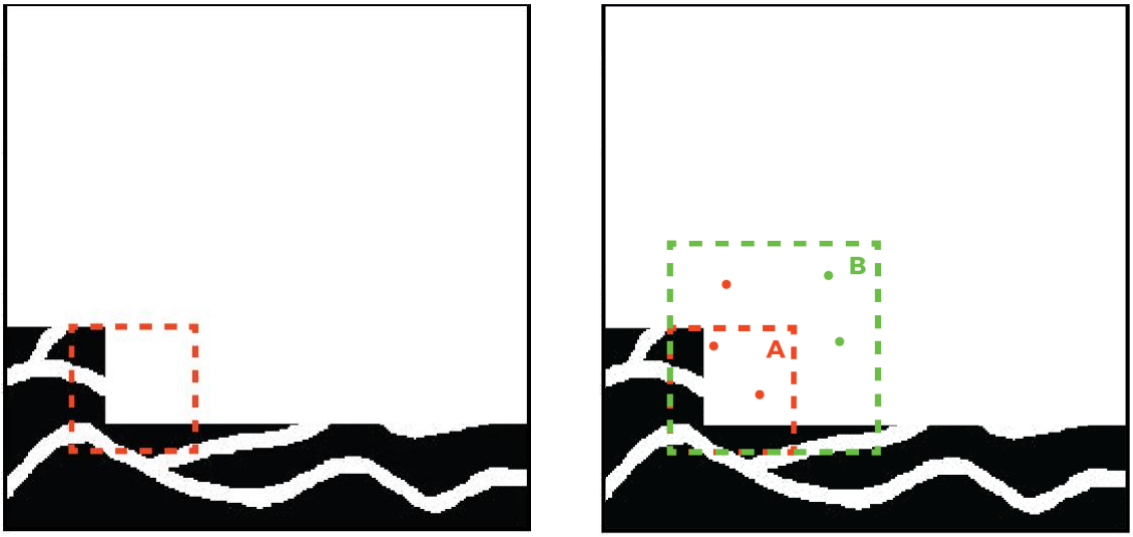
We develop the conditional image quilting by considering the hard data (i.e. the values at the randomly sampled pixel locations on the training image, which are provided as the prior information) within the target patch window. It is not only a cross correlation to determining best candidate, but also a hard data constraint within each target patch should be honored sufficiently. Limited by the patch window size, it is challenging to

consider the hard data beyond the patch window, missing the best candidate patch sometimes. Thus, extended-patch windows are adopted to better address the conditioning issue. The corresponding visualization and comparison from previous traditional method is shown in Figure 3. Also, we designed new criterion integrating the factor of cross correlation, inside hard data mismatch (red window) and outer hard data mismatch (green window – red window) as below,

$$E_t = w_1 * C(u,v) + w_2 * \frac{E_{inner\ mismatch}}{inner\ size} innerhd + w_3 * \frac{E_{outer\ mismatch}}{outer\ size} outerhd \tag{16}$$

where $w_1 + w_2 + w_3 = 1$, $E_{inner\ mismatch}$ is the error between target patch (red window) and any other patch at the hard data position, $innerhd$ the number of hard data inside red window. Similarly $E_{outer\ mismatch}$ represents the error from the outer window (green - red) and $outerhd$ is the number of outside hard data.

As we consider more constraint factors in determining the best patch candidate to match with its neighborhood, it's more reasonable and supportive for CIQ. This results in more parameters tuning of w_1, w_2 and w_3, which can be optimized in the future, although most of time we can get good realizations with the same pattern as TI.



(a) Overlap Area (b) Conditional Extended Patch

Figure 3. (a) Unilateral-path and overlapping area; (b) Extended-path for conditioning.

3.2.3 Multiple-resolution

We use a multi-resolution approach (Strebelle, 2002; Tran, 1994) to speed up the algorithm. As shown in Figure 4, given the original TI with finest resolution (1000×1000), the coarse grid with resolution of (500×500) and (250×250) can be upscaled from its upper-level finer grid by the bi-cubic interpolation method for 2-D TI. We choose three-level or two-level implementation in our test cases. The multiple-resolution algorithm starts from the lowest resolution training image ti_{g-1} and find the best solutions/patches and positions under ti_{g-1}. The appropriate positions in the next level ti_{g-2} can be easily located, then search of best solutions under ti_{g-2} can be done just around the corresponding positions by setting one searching radius. The patch size doubles from one level resolution grid to the next level and the search radius is usually as double as the patch size under the corresponding training image. This procedure can be repeated from the coarser grid to finer grid successively until the finest simulation grid is finished. Since the coarsest resolution grid is much lower compared with the finest grid, the memory cost of cross correlation is lowered, and beyond this step the cross correlation is only computed within the searching radius because we have found good solutions from the previous step. The overall computing time or the memory cost decreases significantly in this way.

In the meanwhile, we should conduct similar upscaling operations for hard data image in CIQ. For simplicity the original finest hard data in our experiments is taken randomly or uniformly from the original finest TI because our goal is to approve the feasibility and efficiency of our approach. For TI with categorical data, we do bi-cubic interpolation followed by the threshold method to turn it back to categorical values from

the interpolated continuous values. For TI with continuous data, the mean implementation is made by bi-cubic interpolation. We also design the categorical hard data resizing method from finer resolution image to its next coarser resolution image, four configurations are shown in Figure 4: (1) If there is only one hard data within the cell of coarser gird, denote the cell within the hard data; (2) When two hard data with same type are located in the cell, the cell remains the same with those hard data; (3) More than one type of hard data is found, denote the cell with the hard data whichever has higher frequency; (4) Randomly assign one type of hard value from those sharing the same frequency.

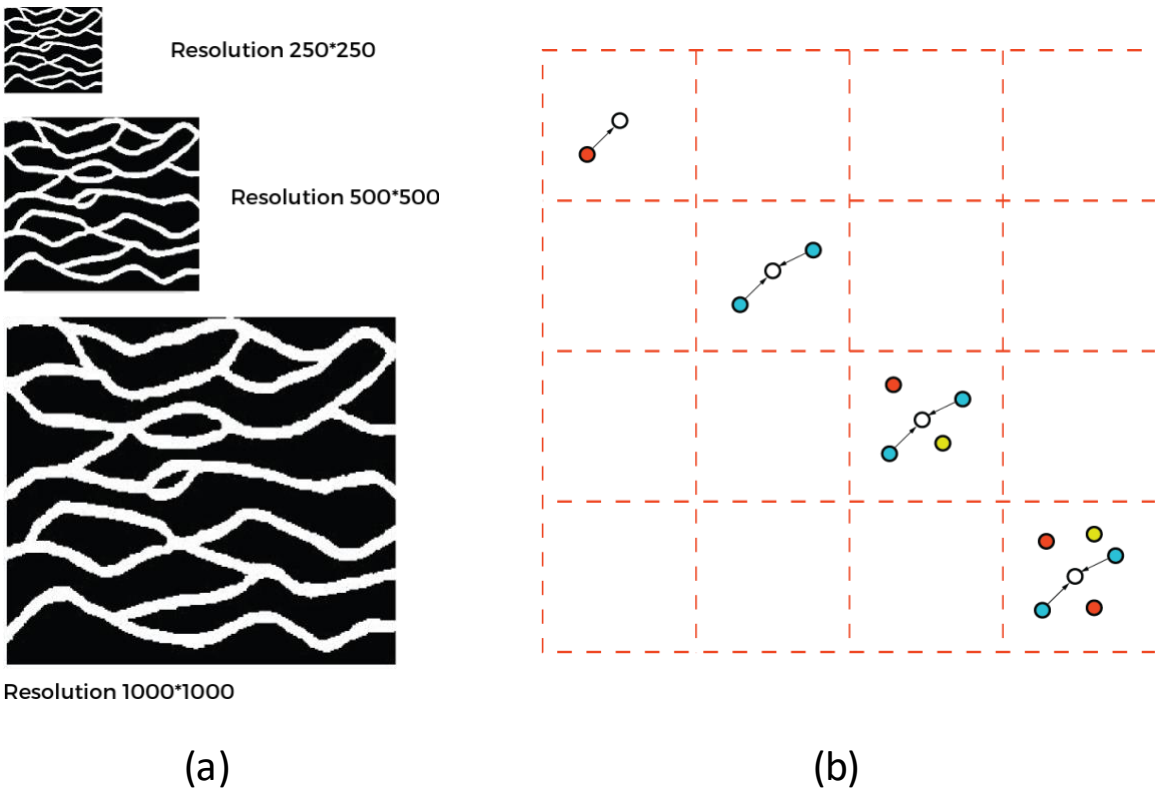


Figure 4. (a) Pyramid of multiple-resolution method; (b) Four resize configurations of categorical hard data.

3.2.4 Multiple-path

Strebelle and Zhang (2005) present another way of constraining MPS to the orientation of geological features that are different from the original TIs. They propose a series of rotation/affinity transforms to stationary training image and build a search tree to store the multiple points statistics inferred from each rotated/rescaled training image. This technique can be easily generalized to create non-stationary models using stationary training images thought to be representative of geological heterogeneity in different areas. It has been proved that the reproduction of the training patterns in non-stationary MPS models is similar to that observed in the stationary models.

The reservoir field can be divided into small regions with different orientations and size of geological features, then one can rotate or rescale the template for each region, thus this transform can be done regionally or globally. I adopt this approach to CIQ to increase the texture variabilities available for pattern reproduction as you can see in Figure 5. Given one TI, we can make horizontal flip and vertical flip separately or together getting four configurations shown in below, then by transposing each of them we can get another four configurations. Finally, all different TIs can be prepared for selection as the input of CIQ.

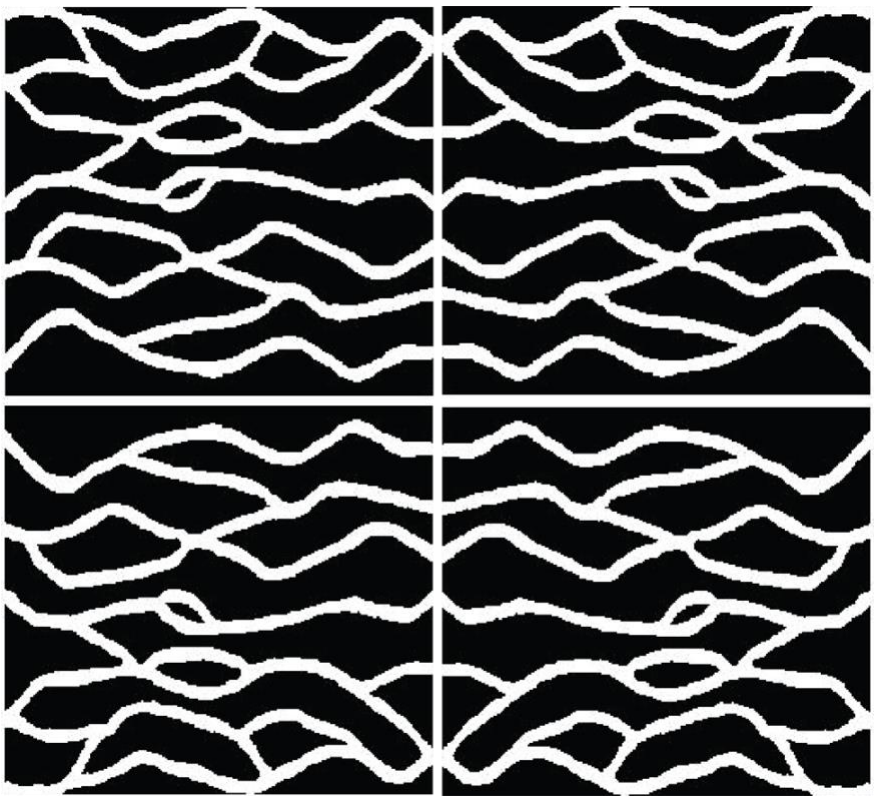


Figure 5. TI on the top left and others flip along x-axis and y-axis.

3.2.5 Fast Fourier Transform

Fast Fourier transform (FFT) was first discussed by (Cooley & Tukey, 1965), the key idea is to turn the computation from temporal or spatial domain to frequency domain. Some of significant applications of FFT includes signal processing and image processing such as filtering algorithms, solving difference equations (recurrence relations), efficient matrix-vector multiplication, fast large integer and polynomial multiplication, etc. Its acceleration comes from reducing the number of computations needed from $O(N^2)$ to $O(NlogN)$, where N is the number of points.

Since digital image is not a continuous function but a matrix of sampled data, taking a 3-D image as an example, the FFT form $F(u,v,w)$ of an image $f(x,y,z)$ with dimensions $d_x * d_y * d_z$ is defined by

$$F(u,v,w) = \sum_{x=0}^{d_x-1} \sum_{y=0}^{d_y-1} \sum_{z=0}^{d_z-1} f(x,y,z)\, e^{-i2\pi\left(\frac{ux}{d_x}+\frac{uy}{d_y}+\frac{uz}{d_z}\right)} \tag{17}$$

where (u,v,w) is the principal axes in frequency domain, $u \in 0,1,\dots,d_x-1$, $v \in 0,1,\dots,d_y-1$ and $w \in 0,1,\dots,d_z-1$.

The inverse Fourier transform is defined by

$$f(x,y,z) = \frac{1}{d_x * d_y * d_z} \sum_{u=0}^{d_x-1} \sum_{v=0}^{d_y-1} \sum_{w=0}^{d_z-1} F(u,v,w)\, e^{i2\pi\left(\frac{ux}{d_x}+\frac{uy}{d_y}+\frac{uz}{d_z}\right)} \tag{18}$$

According to the above FFT and related shift theorem of Fourier transforms, we can convert the cross correlation calculation between the overlapping area and search domain within TI or the whole TI to frequency domain and inverse it back to spatial domain, finally

we can get the appropriate pattern matching its neighborhoods with the advantage of reducing algorithm complexity. FFT will be implemented in large scale of 2-D and 3-D TIs computation for the sake of speeding up and evident performance comparison.

3.2.6 Optimal Cut Boundary Adjustment

This algorithm especially suits well for MPS methods where complex structures are considered because the cumulative minimum error is calculated along the vertical or horizontal cut direction as described in Figure 6 instead of simply cut right along the middle-line of the overlapping area, we calculate the real cut boundary in order to make the transition between patches smoother by means of a cut that maximized continuity (Boykov & Jolly, 2001; Kwatra et al., 2003).

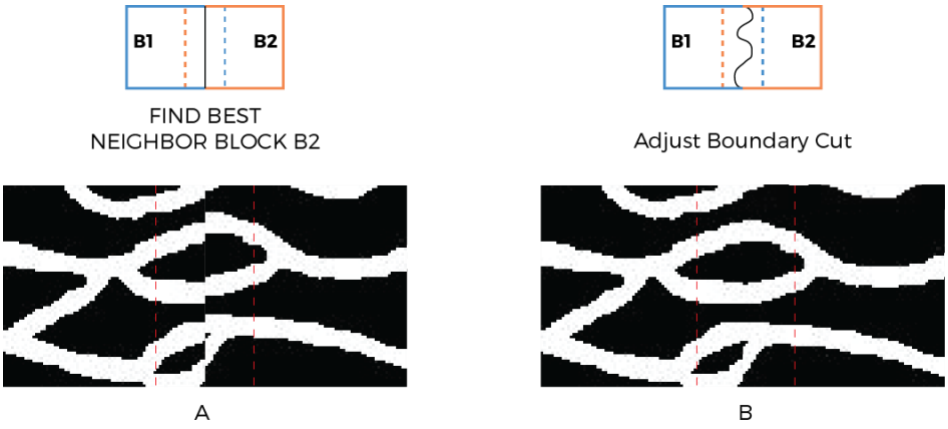


Figure 6. Left: search for best neighbor path; Right: calculate the best cut to make boundary smoother.

3.3 Experiments

We experimented the proposed algorithm on 2D Training Images (TIs) with different characteristics including binary TI shown in Figure 7, continuous TI shown in Figure 9 and extended it to complex satellite images shown in Figure 10 (Tahmasebi et al., 2014). Many experiments were conducted to determine the parameters including patch size, overlapping length that can generate high quality realizations. Here are the useful parameters in the experiment that are effective to reproduce high quality realizations: patch size accounts for $\frac{1}{10} \sim \frac{1}{6}$ of original TI; overlapping length is tuned to be $\frac{1}{4} \sim \frac{1}{3}$ of patch size.

As shown in Figure 7, we take an original TI with a resolution of 1000×1000 that represents connected channels along the EW direction. 1000 hard data (0.1%) is randomly sampled from TI as a prior information, patch size is set as 100×100 and overlapping area as 100×16, we conduct 50 times of conditional realizations and calculate the average to compare the performance of results over different strategies with or without extended-path and multiple-path. According to the performance we can conclude that extended-path and multiple-path are helpful to improve the conditional realization quality, adding them will help produce smoother images than others (bottom right in Figure 7). From quantitative perspective, we draw two different figures that represent widely used geostatistical characteristics including variogram and connectivity probability with respect to distance in Figure 8, it also indicates the result that all the simulated conditional realizations are very close to real TI (red line in each figure). The results are in line with our initial expectations that the proposed algorithm could successfully generate high quality conditional realizations with small variance.

To further testify our developed algorithm, we applied FCIQ to 2D TIs with continuous properties in Figure 9 and more complex characteristics in Figure 10 and come to the same conclusion as that of binary TI. The random selected information as a percentage of image pixels is approximately equal to 0.25% to 0.5%, which is higher than that value used in the study of binary TIs, this is because more prior information needs to be provided to overcome the difficulty of implementing the complex features existing in conditional realizations. Though in real world applications, the data collection job of getting the prior is expensive, especially if more wells need to be dug as a prerequisite, the method we developed still has great potential to generate more TIs with special characteristics. By changing the prior information, we created more conditional realizations that maintain the dominant patterns such as the channels; with the help of geologist, we can filter out noisy output, making data augmentation real. These generated conditional realizations after filtering, as well as unconditional realizations with special characteristics, will be used to in our inverse modeling study in CHAPTER 4 and CHAPTER 5.

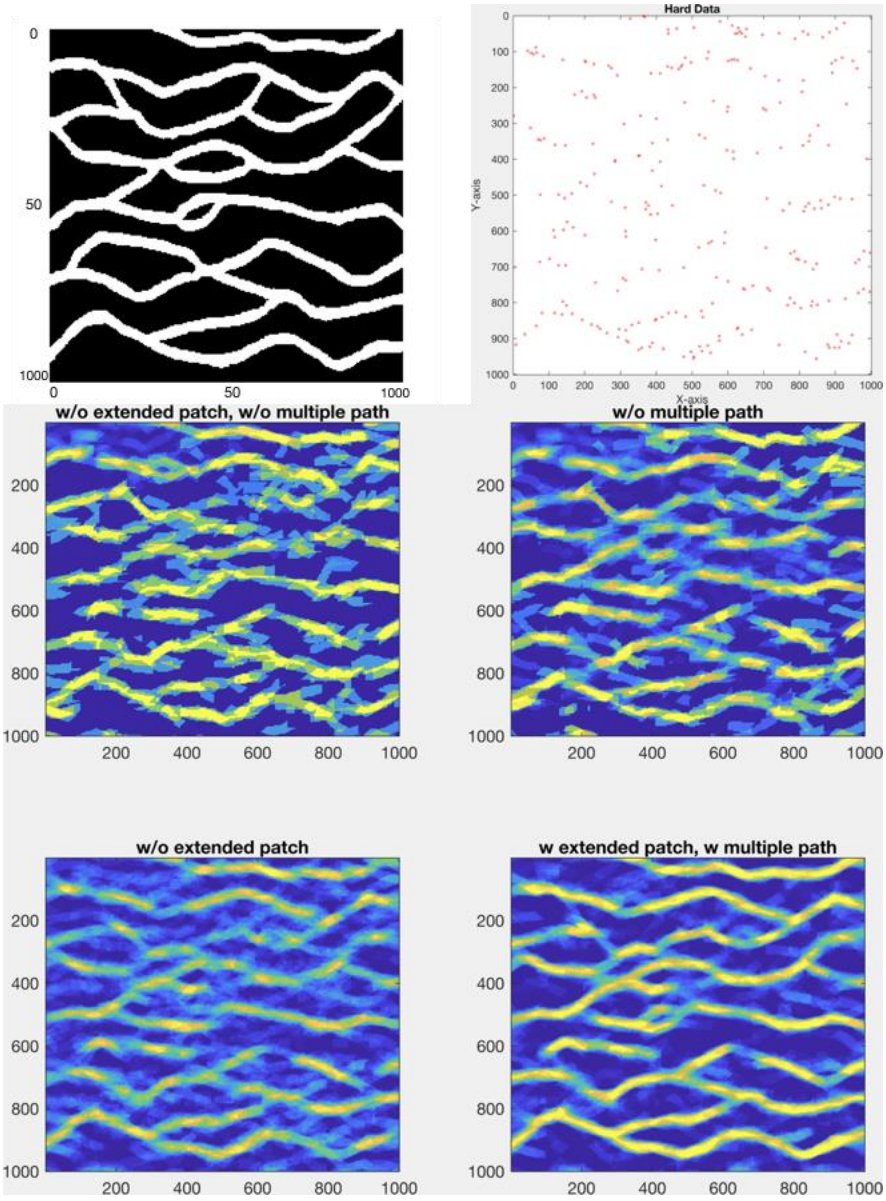


Figure 7. Performance of reconstructed channel patterns on binary TIs over different strategies with or without extended-patch and multiple-path strategy. Top right: 1000 hard data (0.1%) randomly sampled from initial TI. 2nd/3rd row: Average of 50 realizations per strategy. Parameters used in the experiments: patch size: 100×100, overlapping area: 100×16, resolution level: 2, search radius: 100.

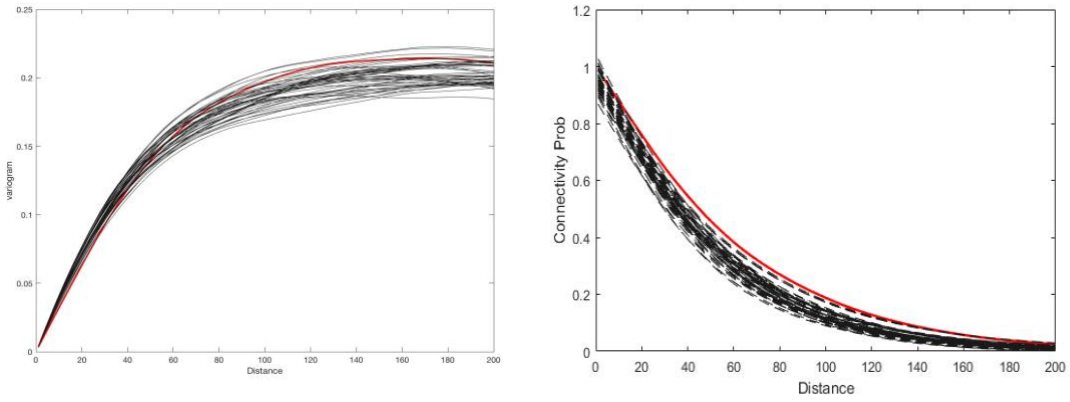


Figure 8. Variogram plot on the left; Connectivity probability distribution on the right.

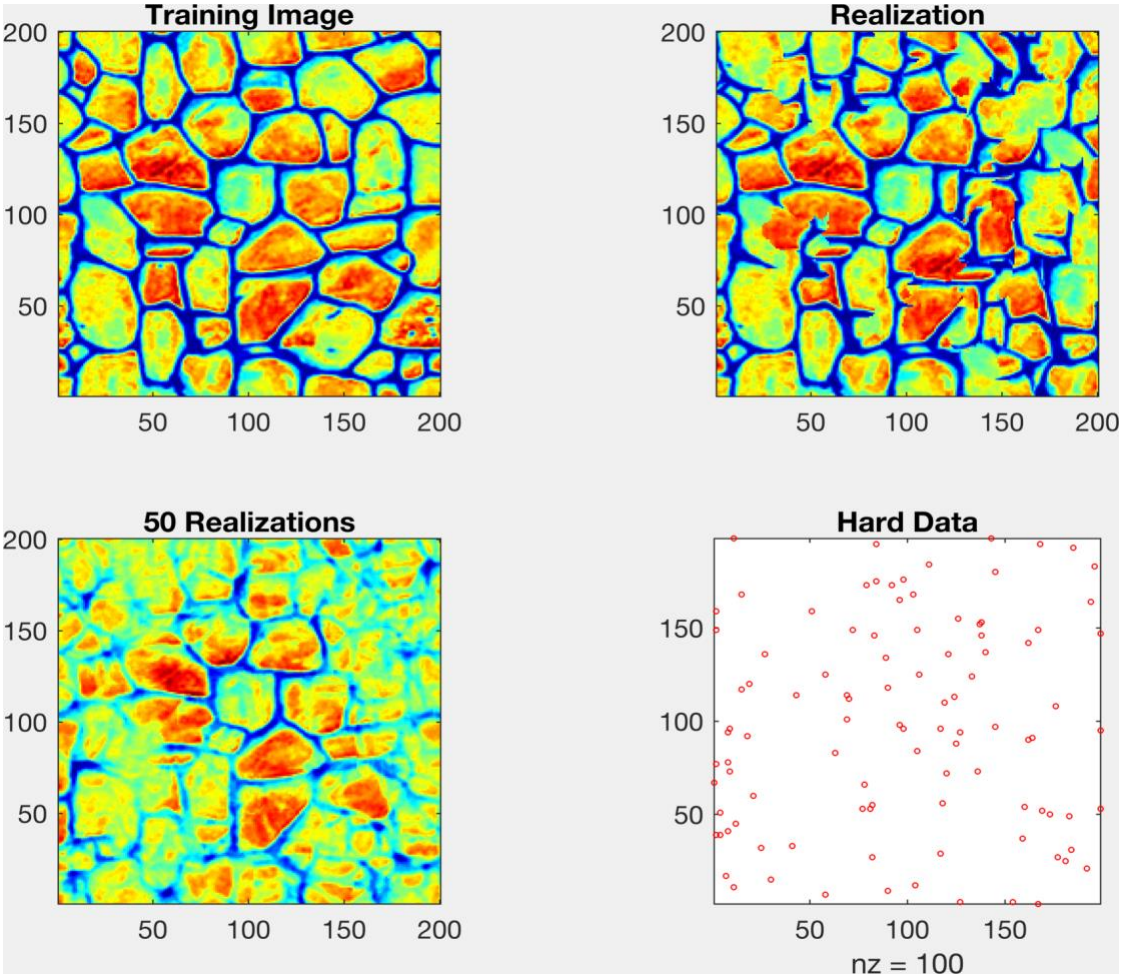


Figure 9. Conditional realizations for continuous image. Resolution: 200×200, Patch size: 36×36, Overlapping area: 36×12, Resolution level: 3, 100 hard data is used as a prior.

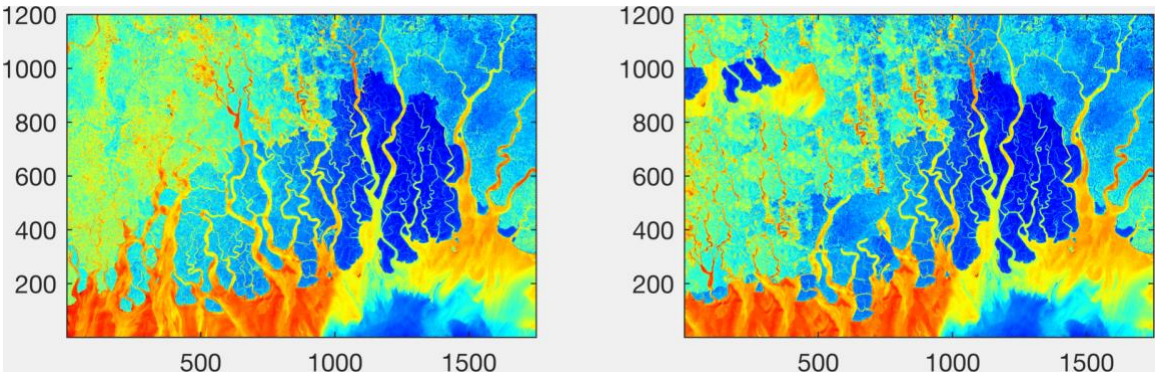


Figure 10. Conditional realization for satellite image of the Sundarbans region, Bengladesh. Resolution: 1200×1750 , Patch size: 240×320, Overlapping area: 80×240 vertically and 80×320 horizontally, Resolution level: 2.

3.4 Conclusion

We have successfully testified our data augmentation method on binary TIs, continuous TIs, and more complex images with special characteristics. The generated conditional realizations indicate that the proposed algorithm performs well in reproducing special structures including connectivity, variability and stochasticity that could not be easily reflected in traditional geostatistical methods. The strategy of using extended-patch and multiple-path proves to be useful because the former improves the algorithm for searching for good patches for its neighbors, while the later provides more usability for pattern selection. In addition, the use of multiple-resolution and Fast Fourier Transform (FFT) improves the algorithm's computational performance.

The developed method can successfully generate TIs with special characteristics such as channels, thus it could serve as an aid to the reservoir characterization application. By changing the prior information that is known as a prerequisite for the experiment, we will create more conditional realizations that maintain significant patterns of initial image; with the help of geologists, we can perform post-inspection on the output to make the data augmentation function better for our research. This checked conditional and unconditional data with special characteristics will be used to support the inverse modeling studies in CHAPTER 4 and CHAPTER 5.

CHAPTER 4. NEURAL NETWORK INVERSE (NNI) ON FULL DIMENSIONS AND NEURAL NETWORK PRINCIPAL COMPONENTS INVERSION (NNPCI)

4.1 Introduction

Inverse problems arise whenever spatially distributed parameters such as heterogeneous hydraulic conductivity are needed as inputs for simulating groundwater flow and transport. Direct measurements of hydraulic conductivity based on local-scale core samples are simply cost prohibitive for field practice. Thus, such spatially distributed parameter fields are usually estimated based on indirect measurements from field tests, such as hydraulic heads in pumping tests and concentrations in tracer tests. Hydraulic tomography is known as sequential pumping tests, which has proved to be effective to provide a large number of measurements with a high information density and relatively low costs and simple sampling techniques. However, inverse modeling of large-scale hydraulic tomography requires efficient numerical methods. So far, geostatistical approach has been the most successful for inverse modeling of hydraulic tomography.

Generally, hydrogeologic inverse problems can be formulized within a Bayesian framework. Kitanidis (1986) showed that traditional cokriging is equivalent to the estimation of the conditional mean of Gaussian processes in a Bayesian analysis. Kitanidis (1995) generalized the geostatistical approach in a rigorous Bayesian framework for solving quasilinear inverse problems to estimate spatially distributed parameters such as hydraulic conductivity. Major computational costs lie in handling auto-covariance matrices and cross-covariance matrices, including the storage and multiplication of matrices, and

iterative implementation of forward models to determine the Jacobian matrix for nonlinear problems.

Many efforts have been devoted to reduce the computational costs of the geostatistical approach. Nowak et al. (2003) proposed an efficient approach of computing the covariance matrix based on circulant embedding and the fast Fourier transform (FFT). Liu and Kitanidis (2011) proposed a sparse representation of the prior covariance matrix, which facilitates the storage of large matrices and the associated computation. Ambikasaran et al. (2013) and Saibaba et al. (2012) proposed hierarchical matrices for large-scale linear geostatistical inverse problems. For geostatistical inverse problems with massive observational data, Klein et al. (2017) proposed a preconditioned conjugate gradient algorithm, and Lin et al. (2017) proposed a randomized data reduction algorithm. Recently, Kitanidis and Lee (2014) introduced the principal component analysis (PCA) into the quasi-linear geostatistical approach and named the approach as principal component geostatistical approach (PCGA). PCGA employs a low-rank approximation of the covariance matrix, and by a so-called 'matrix-free' approach, the number of forward model runs for an explicit construction of the Jacobian matrix is reduced to almost the number of truncated principal components. We shall notice that so far all the efforts have been devoted to advance the computational implementation of the geostatistical approach so that large-dimensional inverse problems can be solved within the classical framework of the quasi-linear geostatistical approach or cokriging. Zhao and Luo (2020) proposed a Reformulated Geostatistical Approach (RGA) framework to directly estimate the principal component coefficients, which integrates the computational techniques and methodology framework.

With the rapid development of machine learning and artificial intelligence and its implication to many different areas, an increasing number of tricky questions can be resolved through neural network architecture. Today, we apply neural network for time-series predictions, anomaly detection in data, and natural language understanding, etc. More researchers are likely to investigate the so-called 'black box' problem with the help of machine learning or deep learning techniques. There has been considerable progress in developing machine learning based methodologies for many geological or geophysical applications. Baykasoglu et al. (2008) applied neural network (NN)-based approach to the strength prediction of limestone. Forghani (2020) used machine learning techniques to obtain a fast solver of shallow water equations, provided with the distribution of bathymetry derived from PCGA and achieved improved accuracy for the velocity prediction. Hojat Ghorbanidehno et al (2021) used deep learning technique for fast inference of large-scale riverine bathymetry and applied his algorithm to large dimensional riverine bathymetry problems with limited available data and computational resources.

In this chapter, this issue is approached from a different perspective. We replace traditional geostatistical methods with neural network architecture to simulate the inverse problem of hydraulic tomography, and reformulate the last layer of neural work based on the principal component analysis of the unknown parameter field, which yields new unknown parameters of principal component coefficients instead of the original parameter field. This new approach is named as neural network principal component inverse (NNPCI). Three independent numerical experiments of hydraulic tomography are presented to demonstrate the validity of the proposed approach.

4.2 Methodology

4.2.1 Inverse Problem Reformulation

The general relationship between the measurement data and unknown data is described as:

$$\mathbf{y} = \mathbf{f}(\mathbf{s}) + \boldsymbol{\epsilon} \tag{19}$$

where $\mathbf{y} \in \mathbb{R}^{n\times 1}$ means the measurement vector, $\mathbf{s} \in \mathbb{R}^{m\times 1}$ represents the unknown vector, $\mathbf{f}$ indicates the forward model that is sophisticated in most cases, and $\boldsymbol{\epsilon} \in \mathbb{R}^{n\times 1}$ is Gaussian with mean $\mathbf{0}$ and covariance $\mathbf{R} \in \mathbb{R}^{n\times n}$. For the prior information, the random field $\mathbf{s}$ is Gaussian with an unknown mean and a generalized covariance function:

$$\begin{aligned} E[\mathbf{s}] &= \mathbf{X}\boldsymbol{\beta} \\ E[(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}] &= \mathbf{C} \end{aligned} \tag{20}$$

where $\mathbf{X} \in \mathbb{R}^{m\times p}$ represents the drift of mean, $\boldsymbol{\beta} \in \mathbb{R}^{p\times 1}$ represents the coefficient vector of the drift function; typically p is a small number: $p = 1$ indicates a constant mean, and $p = 2$ represents a linear drift. $\mathbf{C} \in \mathbb{R}^{m\times m}$ is the covariance matrix, that is typically evaluated by a two-point geostatistical function in geophysical and hydrologic applications.

In typical geophysical and hydrologic application, we usually use geostatistical approaches to find the maximum a posteriori (MAP) of $\mathbf{s}$ by minimizing the negative logarithm of the posterior distribution through Bayes theorem.

$$\min_{\mathbf{s}} f(\mathbf{s}) = \min_{\mathbf{s}} \left\{ \frac{1}{2} \left(\mathbf{y} - \mathbf{f}(\mathbf{s}, \boldsymbol{\beta}) \right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{y} - \mathbf{f}(\mathbf{s}, \boldsymbol{\beta}) \right) + \frac{1}{2} \mathbf{s}^{\mathrm{T}} \mathbf{C} \mathbf{s} \right\} \tag{21}$$

For nonlinear inverse problems, the computation cost for gradient-based optimization methods is expensive due to repeated evaluations of physical model $\mathbf{f}(\mathbf{s})$ for determining Jacobian matrices in each iteration. In most cases, $\mathbf{f}$ is a numerical solver or represents a series of numerical solvers to simulate the results corresponding to the measurements, which may become problematic when the resolution of $\mathbf{s}$ goes up to high dimensions up to thousands or even millions in real world applications.

4.2.2 Artificial Neural Network Surrogate Model on Full Dimensions

Neural network machine learning is essential data-driven. That is, the development of a neural network requires a large amount of data, which are typically unavailable in hydrogeologic applications. For example, we cannot obtain different scenarios of hydraulic conductivity and indirect measurements including hydraulic heads and concentrations. A simple method is to develop a neural network surrogate model to replace the forward model $\mathbf{f}$ so that the optimization function becomes:

$$\min_{\mathbf{s}} f(\mathbf{s}) = \min_{\mathbf{s}} \left\{ \frac{1}{2} \left(\mathbf{y} - NN(\mathbf{s}|\boldsymbol{\theta}) \right)^{\mathrm{T}} \mathbf{R}^{-1} \left(\mathbf{y} - NN(\mathbf{s}|\boldsymbol{\theta}) \right) + \frac{1}{2} \mathbf{s}^{\mathrm{T}} \mathbf{C} \mathbf{s} \right\} \tag{22}$$

where NN is the neural network surrogate model and $\boldsymbol{\theta}$ is the hyperparameter vector in NN. The optimization procedures and techniques to estimate $\mathbf{s}$ can be exactly the same as GA and other methods by replacing $\mathbf{f}$ with NN. Since the computation cost for a trained

NN is very low, the optimization problem can be solved efficiently. The main computation cost now becomes the training of the NN. The training data can be obtained by generating random field realizations based on the geostatistics shown in Eq. (20) and conducting forward model simulations of **f**. For example, we may generate 1,000 realizations for the spatially distributed hydraulic conductivity fields, **s**, for the given spatial moments, and then simulate sequential pumping tests on each realization, $\mathbf{s}^{(i)}$, to generate the hydraulic head measurements, $\mathbf{y}_i$. Thus, NN can be determined by optimizing:

$$\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{N} \left(\mathbf{y}_i - NN\left(\boldsymbol{\theta} \middle| \mathbf{s}^{(i)}\right) \right)^{\mathrm{T}} \left(\mathbf{y}_i - NN\left(\boldsymbol{\theta} \middle| \mathbf{s}^{(i)}\right) \right) \right\} \tag{23}$$

In the present research, we further simply this procedure by directly constructing the inverse problem using neural network, named as NNI. That is, the inverse problem is described as:

$$\mathbf{s} = \mathbf{f}^{-1}(\mathbf{y}) + \boldsymbol{\epsilon}' \tag{24}$$

where $\mathbf{f}^{-1}$ is the inverse model and $\boldsymbol{\epsilon}'$ is the error vector. $\mathbf{f}^{-1}$ is typically unavailable for nonlinear problems and cannot be solved by a physics-based mechanistic model. However, it is very convenient for neural network. We only need to construct the artificial neural network that is fed with input **y** and predicts the output **s** to simulate the inverse of **f**, thus reducing the computational budget of optimization while still preserving the quality of inversion results. The neural network inverse model is obtained by optimizing:

$$\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{N} \left(\mathbf{s}^{(i)} - NNI\left(\boldsymbol{\theta} \middle| \mathbf{y}^{(i)}\right) \right)^{\mathrm{T}} \left(\mathbf{s}^{(i)} - NNI\left(\boldsymbol{\theta} \middle| \mathbf{y}^{(i)}\right) \right) \right\} \tag{25}$$

A set of measurements $\{\mathbf{y}^{(i)}\}, i = 1, \ldots, N$ are fed into the artificial neural network. Each input $\mathbf{y}^{(i)}$ is paired with a targeted random field $\mathbf{s}^{(i)}$. The goal is to learn parameters in each layer of the neural network architecture that connect the preceding layer and the following layer for each single neuron. Each input is a vector of hydraulic heads in sequential pumping tests, while the target output is a representation given by all unknown parameters of the hydraulic conductivity field. Mathematically, the objective of the system is to minimize the value of loss function, i.e. mean square error between target $\mathbf{s}^{(i)}$ and the predicted estimate $NNI\left(\boldsymbol{\theta} \middle| \mathbf{y}^{(i)}\right)$. The representation of input and output could be extended to other physical properties such as storage coefficient or tracer concentrations in different inversion applications. After NNI is determined, the best estimate of the underlying parameter field is simply given by:

$$\hat{\mathbf{s}} = NNI(\mathbf{y}) \tag{26}$$

The NNI framework is shown on the right hand side of Figure 11. Many experiments have been conducted to determine the number of hidden layers and the number of neurons in each layer and other hyperparameters such as learning rates and momentum rates for the optimization. To better handle the large computational volume, a batch normalization layer is added followed by each hidden layer to accelerate convergence of the optimization. The reason of adding batch normalization layer was in the early research study, I found the weights preceding or following hidden layers tended to explode in neural

network training because of the increasing number of neurons for NNI framework on full dimensions, so I added batch normalization after each hidden layer (i.e. 2~3 batch normalization layers were added in total). As a result, the optimization process turned out to speed up much faster than before. Adaptive moment estimation (Adam) was used as the optimization method; other machine learning tracks, such as cross validation and early stop, were applied to deal with potential overfitting issues.

4.2.3 Artificial Neural Network Surrogate Model on Principal Components

NNI is on the full dimensions because it directly focuses on the spatially distributed field, $\mathbf{s}$. Thus, we may still face the challenge of the increasing computational workload in real world applications, particularly when the number of unknown parameters reaches millions. The computation overhead is also expensive for the artificial neural network like other traditional geostatistical methods. We take advantage of principal component analysis (PCA) to further reduce the computational cost of NNI. The low-rank approximation of the symmetric covariance matrix, $\mathbf{C}$, can be written as the truncated eigen-decomposition (Lee and Kitanidis, 2014):

$$\mathbf{C} = \mathbf{V}\mathbf{D}\mathbf{V}^T \approx \mathbf{V}_k\mathbf{V}_k^{\mathrm{T}} \tag{27}$$

The summation of all eigenvalues in the diagonal matrix $\mathbf{D}$ describes the total variance at all principal components. To achieve a modest truncation number k, we simply choose a k to ensure the ratio of the selected variance to the total variance is greater than a predefined criterion, such as 0.95, which means 95 percent of variance can be described by the truncated principal components. The value of k value is selected making a trade-off

between speeding up calculation and losing how much less-importantly information in the projection. Therefore, **s** can then be approximated as:

$$\mathbf{s} \approx \mathbf{X}\boldsymbol{\beta} + \mathbf{V}_{\mathrm{k}}\mathbf{a} \tag{28}$$

In this simplified equation of **s**, $\mathbf{a} \in \mathbb{R}^{k\times 1}$ is the unknown principal component coefficient vector, quantifying that the fluctuations of **s** with respect to the mean is a linear combination of principal components. Thus, any inversion of **s** can be reduced to the inversion of **a**. As *k* is independent of the number of observations, the inversion of **a** is more efficient for cases with massive observations. Moreover, the prior probability distribution of **a** is an i.i.d. standard multi-Gaussian distribution, $\mathbf{a}\sim\mathrm{N}(\mathbf{0},\mathbf{I})$. Thus, we project the original parameter space of **s** onto the principal component space, which reduces the output size tremendously. For example, for a domain with 16,348 cells (128×128), we may only need 50 retained principal components to account for 95% of total variance produced by a Gaussian covariance function or 80% of variance produced by an exponential covariance function. We name the new approach as neural network principal component inverse (NNPCI), which is obtained by optimizing:

$$\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{N} \left(\mathbf{a}^{(i)} - NNPCI\left(\boldsymbol{\theta}|\mathbf{y}^{(i)}\right)\right)^{\mathrm{T}} \left(\mathbf{a}^{(i)} - NNPCI\left(\boldsymbol{\theta}|\mathbf{y}^{(i)}\right)\right) \right\} \tag{29}$$

The left panel in Figure 11 shows the NNPCI structure. The output is the retained principal component coefficients, **a**. The number of **a** is not a function of the size of the underlying random field, thus making the approach more scalable. Also, the NNPCI structure is simpler than the NNI structure on full dimensions of the unknown parameter

field. To generate the training data, the random realizations are generated based on Eq. (28) with random samples from $\mathbf{a}\sim\mathrm{N}(\mathbf{0},\mathbf{I})$. After $NNPCI$ is determined, the best estimate of the underlying parameter field is simply given by:

$$\hat{\mathbf{a}} = NNPCI(\mathbf{y}) \tag{30}$$

$$\hat{\mathbf{s}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{V}_{\mathrm{k}}\hat{\mathbf{a}} \tag{31}$$

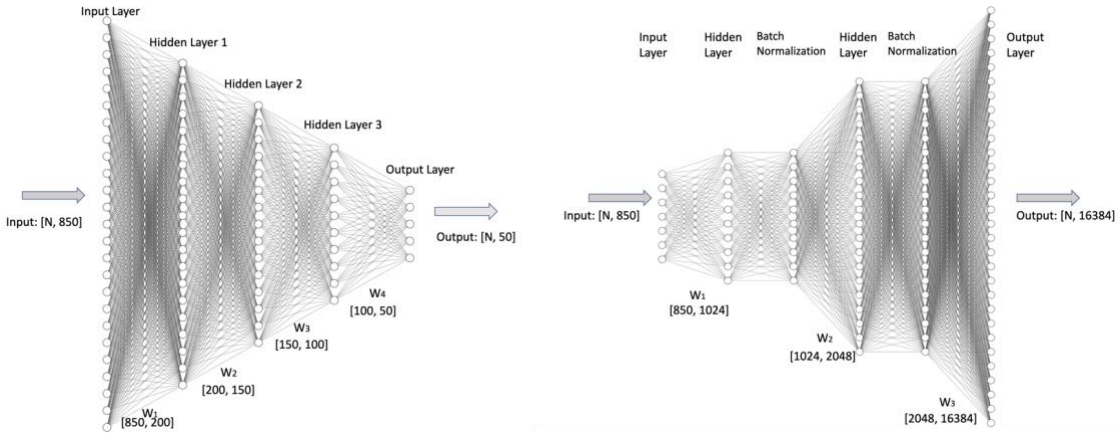


Figure 11. Neural Network Principal Components Inverse (NNPCI) and Neural Network Inverse (NNI) framework.

4.3 Numerical Experiments

4.3.1 Inverse Experiment Settings

Multiple numerical experiments of two-dimensional steady-state hydraulic tomography are presented to prove the effectiveness of the proposed NNI and NNPCI framework. Three heterogeneous fields of logarithmic hydraulic conductivity having different parameters are used in the experiments. Table 1 summarizes the geostatistical properties, hydraulic parameters, the governing equation, and associated boundary conditions. As shown in Figure 12, a random field with Gaussian covariance matrix and another random field with exponential covariance matrix are generated as the true fields of logarithmic hydraulic conductivity. There are 35 wells uniformly installed in the field, among which the black dots represent 25 dual-function (pumping and monitoring) wells, and the white dots represent 10 additional pumping wells. At every stage of hydraulic tomography, when one of dual-function wells (black dots) is used as pumping well, another 24 wells are used as monitoring wells to record the steady-state hydraulic head; when one pumping well (while dots) pumps water, the other 25 dual-function wells (black dots) record the steady-state water head. Therefore, the number of available measurement data for each experiment is 850. The observed data is also polluted by 1.5%~3%, which approximates the situation in actual application.

To generate the training data, random Gaussian field realizations with the Gaussian covariance or exponential covariance model were generated using Fast Fourier Transform. For NNPCI applications, the realizations are generated based on the retained principal components. Channelized fields were created using the data augmentation algorithm

proposed in CHAPTER 3. The purpose is to provide a direction for our algorithm to better handle inverse problem on channels or even more complex domains. The geostatistical characteristics and hydraulic parameters are summarized in Table 2; the pumping and monitoring settings shown in Figure 13 are the same as the above Gaussian field experiment.

For each of the three numerical experiments, 1000 implementations were generated and divided into 70% training, 15% validation and the remaining 15% for testing. All numerical experiments were implemented on a MacBook Pro equipped with Intel® Core i7 UHD Graphics 630 2.20 GHz processor and 16.00 GB RAM.

Table 1. Numerical experiment setup for hydraulic tomography in two heterogeneous hydraulic conductivity fields with Gaussian and exponential covariance function.

Geostatistical Properties		
Domain Scale	$100\text{m} \times 100\text{m}$	$100\text{m} \times 100\text{m}$
Field Type	Logarithmic Field	Logarithmic Field
Resolution	128×128	128×128
Covariance Model	Gaussian	exponential
Mean	$\mu = -5\text{m/s}$	$\mu = -5\text{m/s}$
Variance	$\sigma^2 = 4\text{m}^2/\text{s}^2$	$\sigma^2 = 4\text{m}^2/\text{s}^2$
Correlation length	$\text{lx} = 20\text{m}, \text{ly} = 20\text{m}$	$\text{lx} = 20\text{m}, \text{ly} = 20\text{m}$
Hydraulic Parameters		
Pumping Rate	$\text{Q} = 0.0075\text{m}^3/\text{s}$	$\text{Q} = 0.0075\text{m}^3/\text{s}$
Governing Equation	$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)\left(K\left(\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\right)\right) = 0$	
Top	Impermeable	Impermeable
Bottom	Impermeable	Impermeable
Left	$\text{h} = 0\text{m}$	$\text{h} = 0\text{m}$
Right	$\text{h} = 0\text{m}$	$\text{h} = 0\text{m}$

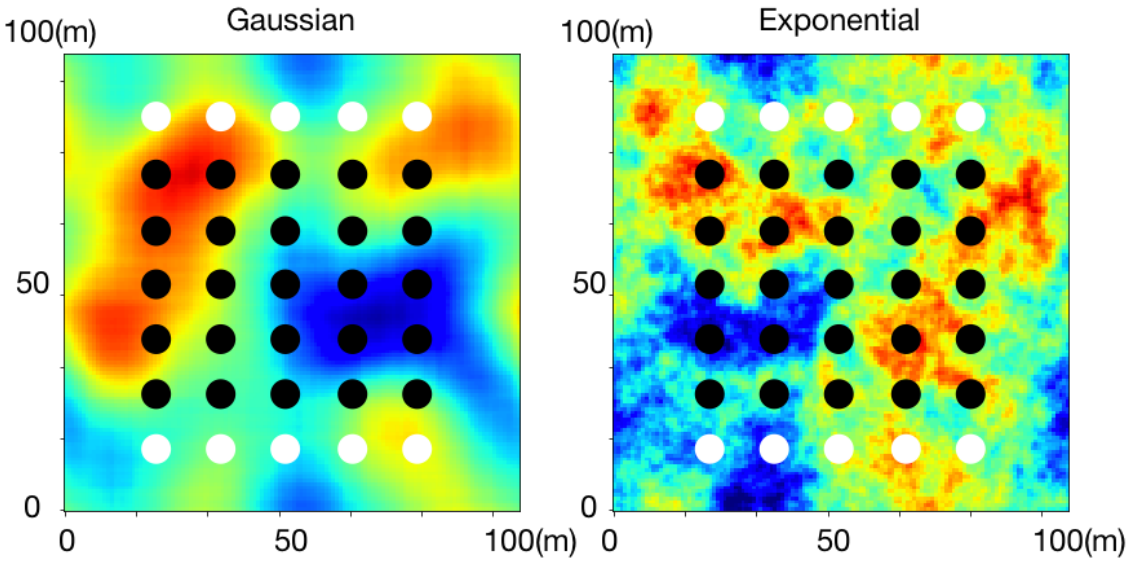


Figure 12. True random field with Gaussian (left) and exponential (right) covariance function and the well setup.

Table 2. Numerical experiment setup for hydraulic tomography in channelized fields.

Geostatistical Properties	
Domain Scale	$100\text{m} \times 100\text{m}$
Field Type	Logarithmic Field
Resolution	128×128
Mean	$\mu_1 = -3\text{m/s}$
	$\mu_2 = -5.5\text{m/s}$
Variance	$\sigma^2 = 0.5\text{m}^2/\text{s}^2$
Hydraulic Parameters	
Pumping Rate	$\text{Q} = 0.0075\text{m}^3/\text{s}$
Top	Impermeable
Bottom	Impermeable
Left	$\text{h} = 0\text{m}$
Right	$\text{h} = 0\text{m}$

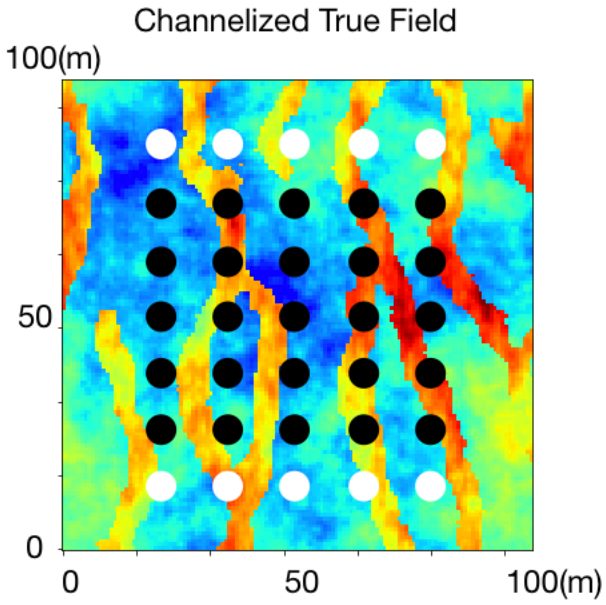


Figure 13. True channelized field and well setup.

4.3.2 Neural Network Inverse (NNI)

The results of NNI framework on the random fields with Gaussian covariance function and exponential covariance function are shown in Figure 16 and Figure 17. Three randomly sampled data were selected as comparisons listed vertically in each figure. By comparing the estimated results on the second row with the real field data on the first row, we can conclude the NNI framework produces satisfactory inversion results because the dominant patterns are well captured by the inversion and the Gaussian features are clearly reproduced in the estimation.

From a quantitative point of view, the Mapping Accuracy (MA) is used as a criterion for analyzing the inversion performance, which represents the percentage of estimated points whose absolute error is less than the threshold (Kang et al., 2017). In this section, we define 15% deviation of the absolute difference between the maximum value and minimum value of the random field as the threshold, as shown by the vertical red lines in the error histograms in Figure 16 and Figure 17. The achieved mapping accuracy is around 83%~89% for the domain with Gaussian covariance function and 70%~87% for the domain with exponential covariance function. The scatter plot density maps in the fourth row of Figure 16 and Figure 17 consolidate the good matching. It can be clearly seen that the main density points fall at 45 degrees on the coordinate, indicating a good fit between the neural network surrogate model predictions and numerical generated results. However, the correlation value of the fitting plot in Figure 16 ranging from 87% to 89%, which is higher than the range from 81% to 85% in Figure 17, because the domain with the exponential covariance function contains more details, which are typically more difficult to invert.

To further verify the effectiveness of the NNI framework, we extend the test domain from the domain with the Gaussian covariance function and exponential function to a channelized, non-Gaussian field. The result is shown in Figure 18. The mapping accuracy is 56%. The dominant channel pattern is captured in two ways: (1) The NNI framework successfully learns the main channel patch along the north-south direction; (2) Since the training image is composed of two media, i.e. channel and non-channel, the authenticity of the bimodal field characteristics is confirmed by the two-Gaussian distribution reflected in the scatter plot density map in the fourth row of Figure 18. Although the estimated performance is not as good as the above two experiments for Gaussian fields, the results are still encouraging because the NNI framework outperforms traditional stochastic sampling methods such MCMC and optimization approaches like Geostatistical Approach (GA) that cannot handle inverse problem without a spatially covariance function.

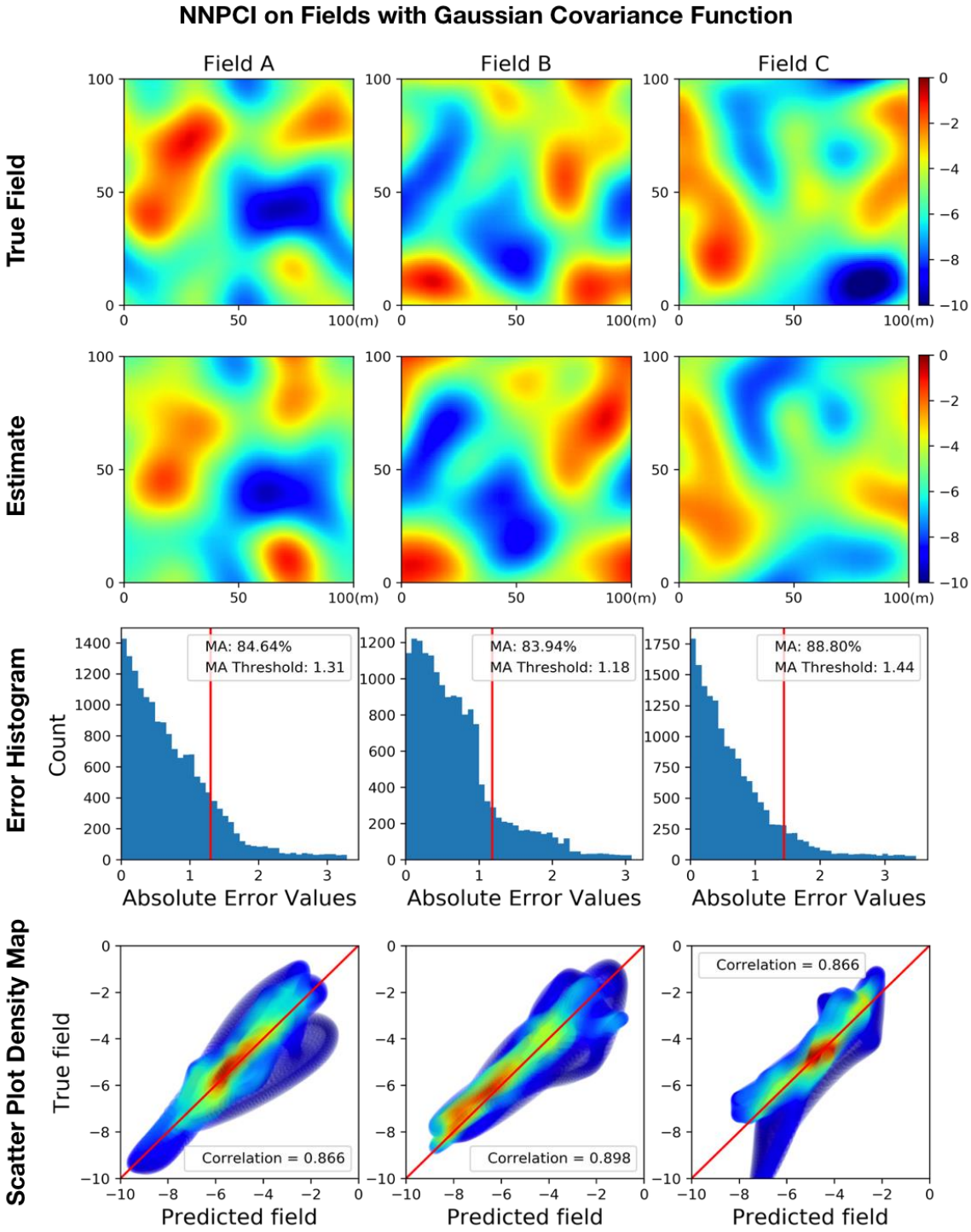


Figure 14. Inversion results with the NNPCI framework for Gaussian fields with Gaussian covariance function on retained principal components.

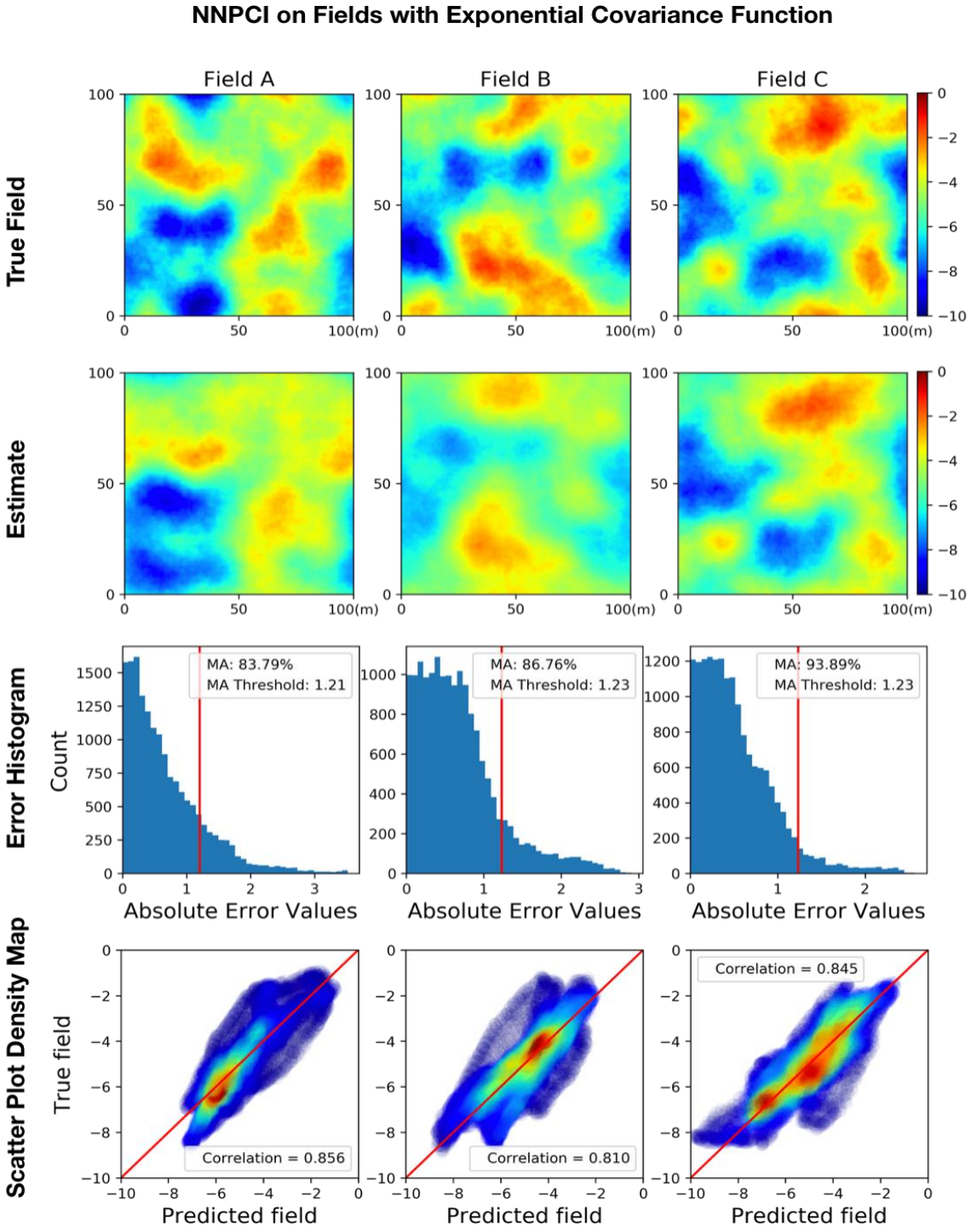


Figure 15. Inversion results with the NNPCI framework for Gaussian fields with exponential covariance function on retained principal components.

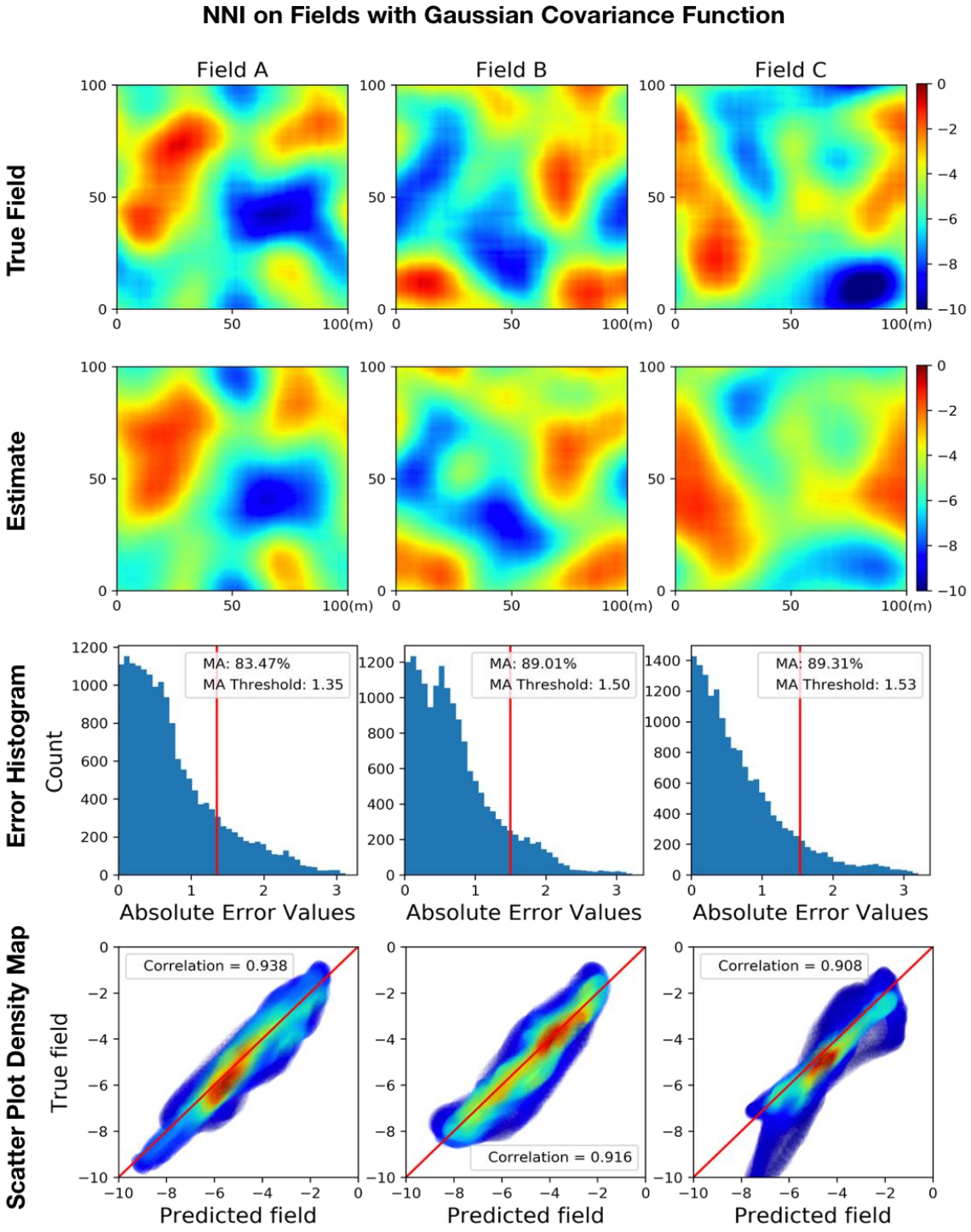


Figure 16. Inversion results with the NNI framework for Gaussian fields with Gaussian covariance function on full dimensions.

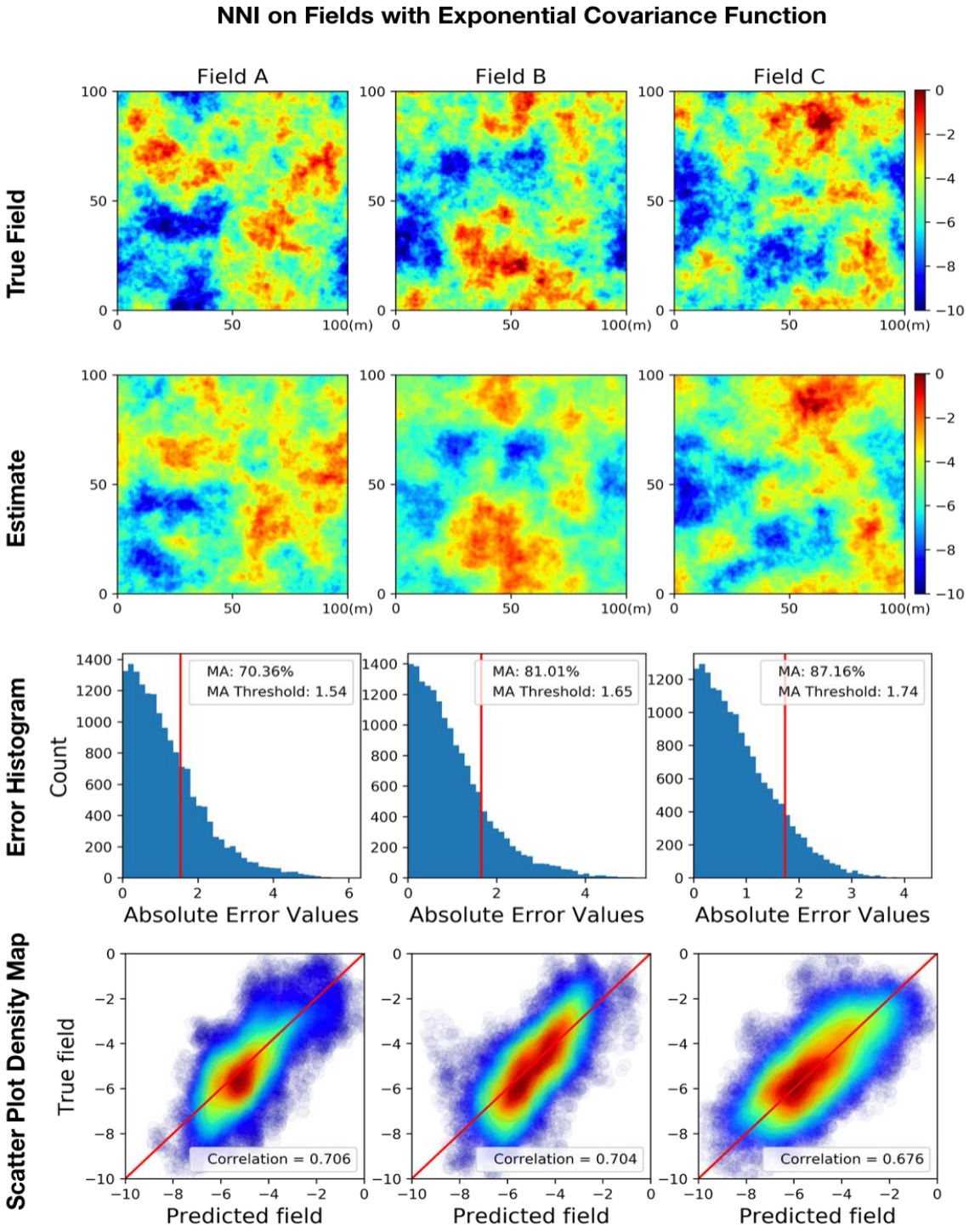


Figure 17. Inversion results with the NNI framework for Gaussian fields with exponential covariance function on full dimensions.

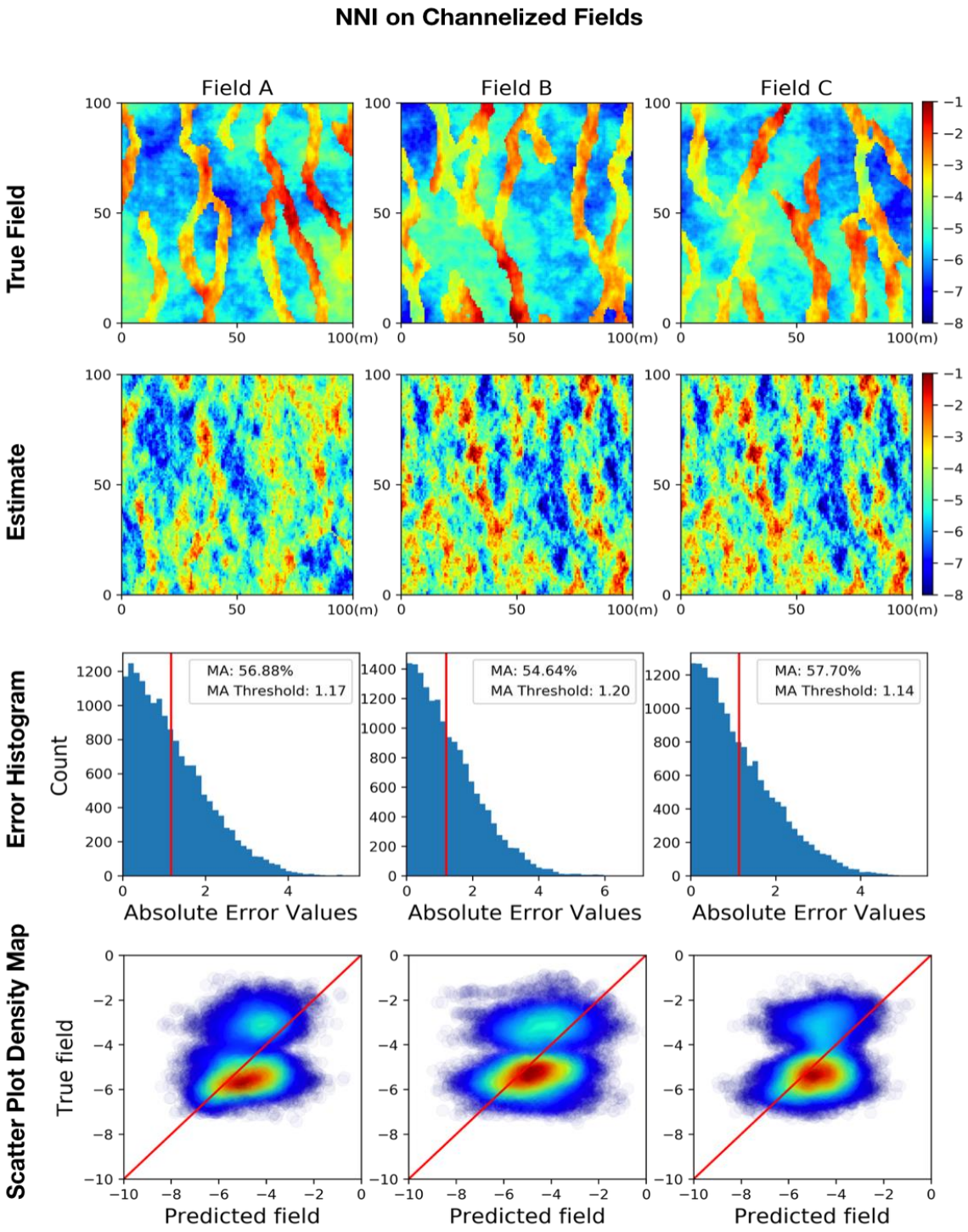


Figure 18. Inversion results of channelized random fields obtained using the NNI framework.

4.3.3 Neural Network Principal Component Inversion (NNPCI)

In hydrogeological inverse problems, the spatial range of the random fields of interest ranges from hundreds to millions, thus slowing down the inversion process. The inverse problems can be solved in the reduced space rather than in the initial space and converted back into the original space. Inspired by Reformulated Geostatistical Approach (RGA) proposed by Zhao and Luo (2020), we replaced the number of neurons in the last layer of the NNI framework with the retained principal components, and converted back to the real estimated image after inversion was completed, which greatly speeds up the algorithm and makes the framework better handle larger inverse problems in hydrogeological application.

We verified the effectiveness of NNPCI on the domains with the Gaussian covariance function and exponential covariance function, and the results are shown in Figure 14 and Figure 15. The NNPCI framework also produces satisfactory inversion results with the dominant patterns being well captured and the features being clearly reproduced in the estimation. The achieved mapping accuracy is improved to 85%~95% for the domain with Gaussian covariance function and the domain reflected by exponential covariance function; the correlation coefficient of the two domains is about 85%. Therefore, we can conclude the NNPCI framework can produce convincing inversion results. In the next section we will discuss that NNPCI outperforms NNI in terms of computational budget.

4.3.4 Effect of Training Data

Since preparing the training dataset for the NNI and NNPCI models requires multiple forward model runs, we test the effect of the amount of training data on the performance of the inverse modelling. We conducted a set of parallel experiments on Gaussian random fields with a Gaussian covariance matrix by varying the training data size from 200 to 800 realizations. We use the mapping accuracy and correlation coefficient as the metrics and take the average value on 100 testing dataset, which is consistent in all the parallel experiments.

Figure 19 summarizes the comparisons between a random Gaussian field with a Gaussian covariance matrix and the corresponding inversion results. The inversion results achieve good performance with a mapping accuracy of 82% and a correlation coefficient of 0.83 when the neural network is trained with only 200 datasets; and the scatter plot density map indicates a good fit between the estimated results and the true field. By feeding more datasets into the neural network surrogate model such as 800, the model only has slight improvement. This conclusion is further validated by the mapping accuracy average value and correlation coefficient average value of 100 testing data sets, as shown in Table 3. Thus, we can conclude that our neural network surrogate model can be further accelerated to outperform traditional geostatistical methods.

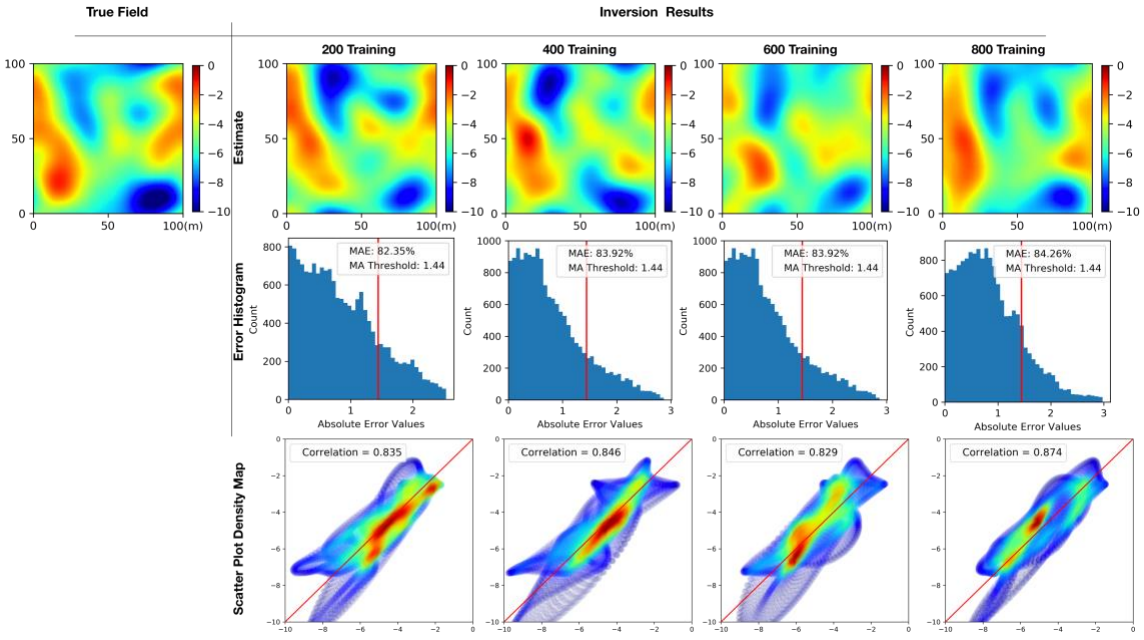


Figure 19. Inversion results of random Gaussian fields with Gaussian covariance matrix obtained using different sizes of training data.

Table 3. Averages of mapping accuracy and correlation coefficient on 100 testing Gaussian fields with Gaussian covariance function. 200, 400, 600 and 800 training dataset are used in each experiment.

Training Dataset	Average MA	Average Correlation
200	0.721	0.768
400	0.739	0.798
600	0.725	0.783
800	0.77	0.81

4.3.5 Time Analysis

At the training stage, it is worthwhile to mention that for the NNPCI framework, the inverse problems are solved in the reduced space, and converted back into the original space. The scalability of the system is enhanced to handle high-dimensional inverse problems, regardless of the unknown parameters in the initial parameter field. Figure 20 illustrates the trend of computational cost relative to the unknowns in the field. When the number of unknowns increases from 256 to 262144, the training time almost remains unchanged. Therefore, the NNPCI framework has better scalability than NNI framework when the computational demand is very large. Figure 21 summarizes the time comparison between NNPCI (including the time of generating training data) and RGA with respect to the number of unknowns of the parameter field, the running time of the two algorithms is similar when there are not many unknowns; however, NNPCI is much more efficient with around 3800s total time compared with 6000s for RGA algorithm. In addition, running time and mapping accuracy performance is further tested using a specific case on 128×128 resolution. It is clear to see NNPCI is more computationally efficient while achieving competitive mapping accuracy compared with BSGN (MCMC) and PCGA. Thus, we can conclude our proposed NNPCI can save more computational budget than many traditional MCMC algorithms and geostatistical approaches in tackling large scaled inverse problem in hydrogeological area.

At the testing stage, both frameworks could generate estimates faster than traditional geostatistical methods. Specifically, taking the Reformulated Geostatistical Approach (RGA) as an example, this method needs to solve multiple equations, and each iteration requires running multiple forward models to determine the Jacobian matrix. In contrast,

when the input is fed into the well-trained NNI or NNPCI framework, the conditional realization will be generated quickly, completely dependent on the trained neuron weights, and multiple realizations can be completed in parallel in a short time. Therefore, in scenarios where multiple estimates are to be generated, the NNI and NNPCI framework are much more efficient than RGA by taking advantage of offline training.

The calculations of forward model are time-consuming in traditionally geostatistical approach. We include this part and compare the computational budget of NNPCI with that of RGA. Assume a scenario where we retain the first k principal components, and it takes i iterations for RGA to converge to the best estimate. The total number of forward model runs is computed as $i \times k$, and total computational time is $i \times k \times f$, where f is the defined as the time consumed in each forward model. Assume there are p available processors, ideally, the time complexity is $O(\frac{i \times k \times f}{p})$. For the NNI or NNPCI framework, assume N training data is generated, the total computational time is $N \times f$. By leveraging parallel computing of p processors, the time complexity reduces to $O(\frac{N \times f}{p})$. To obtain a more concrete understanding of the computational time comparison, a few numbers are plunged into the equation. The results are shown in Table 4. In a scenario where the number of retained principal components is 50, for example, the iteration number for RGA is 20, each forward model consumes 1s, and assume 300 training data is generated. Initially the two algorithms consume similar time, with the increase of processors used in parallel computing, NNPCI will be more efficient in producing the training data.

Table 4. RGA and NNPCI speed comparison in parallel computing. In scenario where the number of retained principal components is 50, iteration number for RGA is 10, each forward model consumes 1, 300 training data is generated.

	RGA	**NNPCI**
Given $p = 1$	500	300
Given $p = 2$	250	150
Given $p = 5$	100	60
Given $p = 20$	25	15
Given $p = 50$	10	6
Given $p = 100$	10	3
Given $p = 200$	10	2

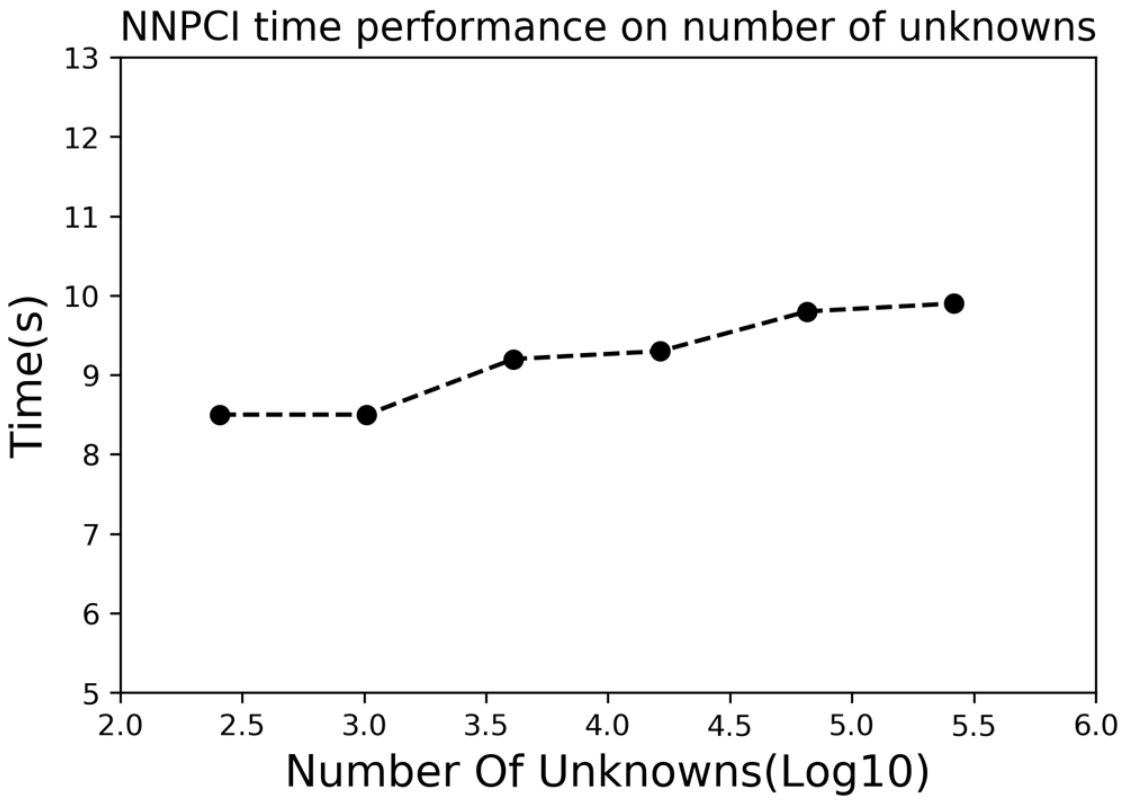


Figure 20. NNPCI time performance on the number of unknowns.

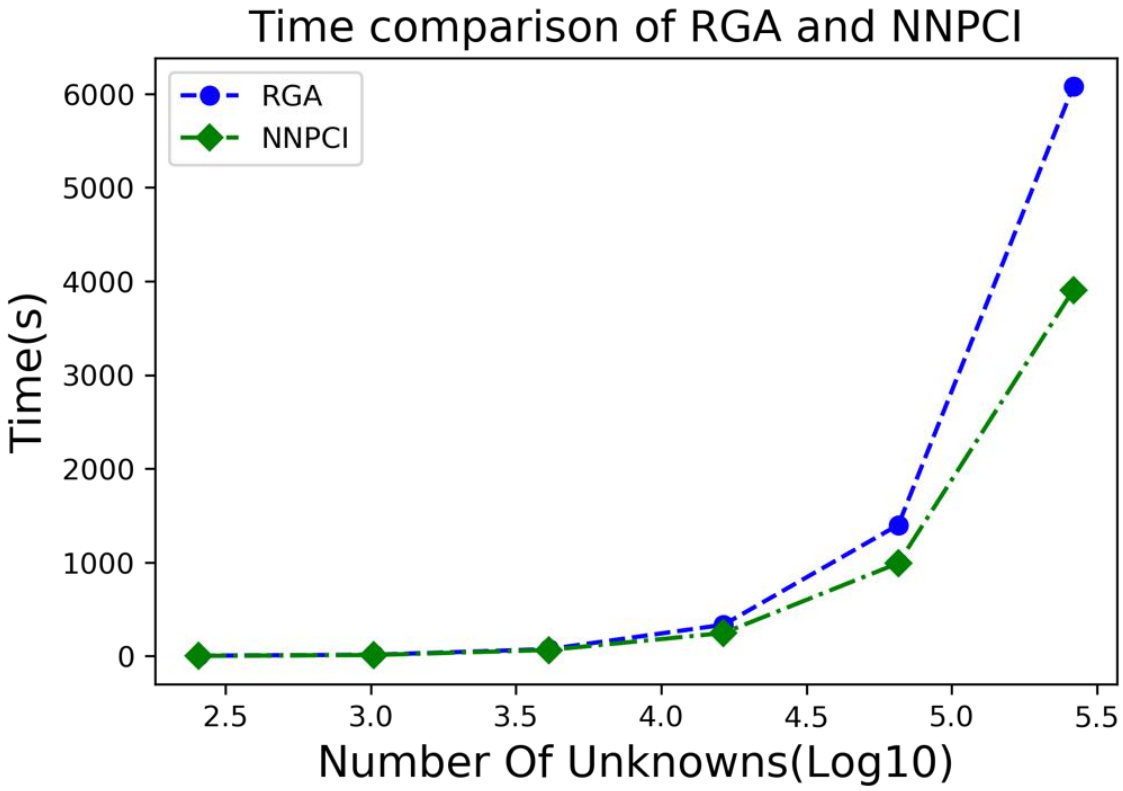


Figure 21. Time comparison of RGA and NNPCI.

Table 5. Performance comparison between NNPCI, BSGN and PCGA on a field case with 128×128 resolution.

	NNPCI	**BSGN**	**PCGA**
Field Type	Gaussian field with exponential covariance matrix	Gaussian field with exponential covariance matrix	Gaussian field with Gaussian covariance matrix
Field Size	128×128	128×128	128×128
Total time(s)	246	1353	845
Mapping Accuracy	86.76%	83.37%	88%

4.4 Conclusion

In this chapter, we proposed Neural Network Inverse (NNI) framework on full dimensions and Neural Network Principal Component Inverse (NNPCI) framework on reduced dimensions by PCA to handle the inverse problems of hydraulic tomography in heterogeneous hydraulic conductivity fields. Our neural network surrogate models directly describe the inverse relationship between the indirect measurements, i.e. hydraulic heads and the underlying parameter fields, hydraulic conductivity, or the retained principal component coefficients. Thus, the inverse estimates can be directly obtained through the trained NNI and NNPCI, and there is no need to use the trained model for another optimization.

Numerical experiments show that both neural network architectures result in satisfactory performance of inversion results. In particular, NNPCI is more efficient for Gaussian random fields by reducing the size of unknown parameters to that of retained principal components without jeopardizing the quality of inversion results. Moreover, for non-Gaussian random fields, such as the channelized heterogeneous fields, NNI also provides good inverse results. Because such fields cannot be described by the two-point spatial covariance, geostatistical approach is not applicable.

In terms of computational complexity and efficiency, both frameworks could provide the estimates faster than traditional geostatistical results when the models are trained. The major computational cost is the generation of the training data and model training. In the present study, we generated 1,000 random realizations for each field, and correspondingly implemented 1,000 times of forward model simulations. The number of the forward model

run is comparable to the traditional geostatistical approach. The recent RGA framework may reduce the number of the forward model runs by incorporating the quasi-Newton algorithms. However, the geostatistical approach needs to implement the forward model simulations iteratively during the optimization for determining the Jacobian matrix. In contrast, all the forward model simulations in our neural network models can be generated independently offline. That is, NNI and NNPCI do not rely on any iteration and gradient evaluation, and the forward model runs are completely independent to each other. Thus, generating the training data can be done in advance and independently by taking the advantage of parallel computing. Moreover, for established NNI and NNPCI, when an input is fed into well-trained NNI or NNPCI framework, an estimate can be fast generated purely replying on the weights of neurons.

CHAPTER 5. VARIATIONAL BAYESIAN NEURAL INVERSE (VBNI) ON FULL DIMENSIONS AND VARIATIONAL BAYESIAN NEURAL PRINCIPAL COMPONENTS INVERSION (VBNPCI)

5.1 Introduction

Comprehension and modeling of flow and transport behavior in groundwater systems require accurate and efficient characterization of the underlying spatially distributed hydraulic conductivity field. Local-scale estimates of hydraulic conductivities can be obtained by conducting laboratory experiments on collected core samples from drilled boreholes (A. Klute & R. C. Dinauer, 1986; Zha et al., 2018). However, high costs of well drilling hinder this approach from mapping field-scale hydraulic conductivity distributions onto a fine resolution. Therefore, field-scale characterization is usually achieved by inverse approaches to infer hydraulic conductivities from measurements of aquifer tests, such as pumping tests, slug tests, constant-head tests, and tracer tests etc. (Cardiff et al., 2009; Cirpka et al., 2007; Fienen et al., 2006; Liao & Cirpka, 2011; Yeh & Liu, 2000; Z. Zhang et al., 2014; Zhao et al., 2018).

Various numerical approaches have been developed to solve Bayesian inverse problems. Ensemble-based approaches include Markov chain Monte Carlo (MCMC) that simulates a large number of the posterior samples by running a Markov chain (Cui et al., 2014; Geman & Geman, 1984; Hastings, 1970; Martin et al., 2012; Vrugt et al., 2009), ensemble Kalman filter and ensemble smoother approaches that simulates samples from a Gaussian approximation of the posterior distribution (Jiangjiang Zhang et al., 2018; Zhou et al., 2012). However, for highly-parameterized Bayesian inverse problems that admit

high-resolution hydraulic conductivity fields, the geostatistical approach still requires efficient implementations to facilitate the computation (Lee & Kitanidis, 2014; Liu & Kitanidis, 2011). The reasons lie in two aspects: numerical solver needs to be performed on a high-resolution grid for multiple times in nonlinear problems (Kitanidis, 1995); and the large number of unknowns increases the cost of matrix computation during the inversion.

Many efforts have been dedicated to reduce the computational cost of the geostatistical approach. Nowak et al. (2003) proposed an efficient approach of computing the cross-covariance matrix based on circulant embedding and the fast Fourier transform, which was successfully applied to a sandbox inverse problem (Nowak & Cirpka, 2006). Liu and Kitanidis (2011) proposed a sparse representation of the prior covariance matrix that facilitates the storage and the associated computation, and proved its effectiveness in an inverse problem of sandbox hydraulic tomography. For large-scale linear geostatistical inverse problems, Ambikasaran et al. (2013) and Saibaba et al. (2012) proposed hierarchical matrices to speed up the solving process. For geostatistical inverse problems with huge volume of observational data, Klein et al. (2017) proposed a preconditioned conjugate gradient algorithm, and Lin et al. (2017) proposed a randomized data reduction algorithm. For inversion of pumping and tracer tests, temporal moments was applied to transform transient forward models to steady-state models for faster model evaluations (Li et al., 2005; Nowak & Cirpka, 2006; Pollock & Cirpka, 2008; Yin & Illman, 2009; Zhu & Yeh, 2006).

Recently, Kitanidis and Lee (2014) introduced the Principal Component Analysis (PCA) in the quasi-linear geostatistical approach to cut down the computational cost of

Bayesian inverse problems, named as principal component geostatistical approach (PCGA), and demonstrated its effectiveness and efficiency in several applications (Kang et al., 2017; Lee & Kitanidis, 2014; Lee et al., 2016). PCGA decreases the computational overhead associated with the covariance matrix and improves the computational efficiency for the Jacobian matrices by a 'matrix-free' approach. A similar idea was also incorporated into the successive co-kriging estimator for hydraulic tomography applications (Zha et al., 2018).

Under such a background of inversing modeling, in CHAPTER 4, we have developed NNPCI from the machine learning perspective and generate a new system that simulates the relationship between hydraulic conductivity and hydraulic head. The proposed framework could easily extend to other disciplines like earth sciences, environmental science and engineering, etc. However, we cannot guarantee the predicted conditional realization output from this system is always reliable because of different sources of errors such as error in input data, model structure, model parameters. These errors will consistently cause uncertainty issues to the system, making the predicted result not that accurate as we always expect. It is therefore worthwhile to put more efforts into uncertainty analysis when the outcome of the model is significant for research work or industrial financial purposes. We always expect the system to perform stably and be consistent at varying spatial and timing scales though sometimes input data is contaminated and further generalize the system or model well to other implications.

So far, much research work on algorithm investigations simulating the uncertainty and metrics to quantify uncertainty analysis have been evolved fast in hydrogeological areas in recent decades. Rafiei Emam et al. (2018) conducted uncertainty analysis for a

semi-distributed hydrological model based on four algorithms including Generalized Likelihood Uncertainty Estimation (GLUE), Parameter Solution Method (ParaSol), Sequential Uncertainty Fitting (SUFI) and Particle Swarm Optimization (PSO); and compared the performance of the algorithms using P-factor and R-factor and coefficient determination. JL Zhang et al. (2016) proposed a Bayesian framework for investigating uncertainties in input data such as temperature and precipitation and model parameters as well as their effects on the runoff response, which successfully enhances the capacity of the hydrologic system for predicting water resources during different seasons. Bayesian model averaging (BMA) method was employed on different hydrologic models and proved that BMA can provide a statistical distribution of the quantity to be forecasted (Dong et al., 2013).

Unlike the BMA method that requires multiple hydrological models in forecasting, we want to simplify the system and inject the so-called "uncertainty" to the weights of neural network architecture. In this chapter, we employed the idea of Bayes by backprop (Blundell et al., 2015a) to learn a probability distribution on the weights of a neural network. As a result, Bayesian theorem was incorporated into the NNPCI framework to add uncertainty for the whole system. The new inverse framework is named as Variational Bayesian Neural Principal Component Inversion (VBNPCI). Three numerical experiments with Gaussian and non-Gaussian dataset were tested on the model to validate the stableness of the system.

5.2 Variational Bayesian Neural Principal Component Inverse (VBNPCI)

In this chapter we introduced NNPCI in which the weights connecting the preceding layer and the next layer are constants. To increase uncertainty for the generations we need to add "variations" into the architecture, so Bayesian theorem was employed to NNPCI and the modified framework become really variational. In VBNPCI, every weight follows a distribution with a prior added. The weights w^{MAP} could be expressed as:

$$w^{MAP} = \arg\max_{\mathrm{w}} \log P(w|x,y) = \arg\max_{\mathrm{w}} \log P(x,y|w) + \log P\,(w) = \\ = \arg\max_{\mathrm{w}} \sum_{i} \log P(y^{(i)}|x^{(i)},w) + \log P\,(w) \tag{32}$$

In this way, the learnt distribution could answer predictive queries upon independent testing unknown data $\hat{x}$ with the help of one possible configuration of the weights sampled from the posterior distribution and make an appropriate decision, which could be expressed as:

$$P(\hat{y}|\hat{x}) = \mathbb{E}_{p(w|\mathcal{D})} P(\hat{y}|\hat{x},w) \tag{33}$$

We assume each weight in the variational neural network follows a Gaussian distribution and apply Kullback-Leibler (Van Erven & Harremos, 2014) between the variational distribution $q(w|\theta)$ and the true posterior $P(w|\mathcal{D})$. By Bayesian theorem we can easily derive:

$$
\begin{aligned}
&KL\big(q(w|\theta) \parallel P(w|\mathcal{D})\big) \\
&\quad = \mathbb{E}_{q(w|\theta)} \log \frac{q(w|\theta)}{P(x,y|w)P(w)} P(x,y) \\
&\quad = \mathbb{E}_{q(w|\theta)}[\log q(w|\theta) - \log P(x,y|w) \\
&\quad - \log P(w) + \log P(x,y)] \\
&\quad = KL\big(q(w|\theta) \parallel P(w)\big) - \mathbb{E}_{q(w|\theta)} \log P(x,y|w) + \log P(x,y)
\end{aligned}
\tag{34}
$$

The first two items together are also known as variational free energy $\mathcal{F}(x, y, \theta)$ and the last term $\log P(x, y)$ is log marginal likelihood which have no relationship with w, so we can simplify our learning objective function as:

$$
\mathcal{F}(x, y, \theta) = KL\big(q(w|\theta) \parallel P(w)\big) - \mathbb{E}_{q(w|\theta)} \log P(x, y|w)
\tag{35}
$$

By using Monto Carlo sampling (Blundell et al., 2015b), we can further approximate the learning objective as:

$$
\mathcal{F}(x, y, \theta) \approx \frac{1}{N} \sum_{i=1}^{N} [log\, q(w^{(i)}|\theta) - p\big(w^{(i)}\big) - log\, p\big(x^{(i)}, y^{(i)}\big|w^{(i)}\big)]
\tag{36}
$$

Where $w^{(i)}$ represents one single Monte Carlo weight configuration sampled from the variational posterior $q(w^{(i)}|\theta)$, which also correspond to each input that is fed into the model. The first term on the right hand side is called the KL-divergence between variational distribution $q(w|\theta)$ and the prior $p(w)$, and the second term is the expectation of the likelihood $P(x, y|w)$ with respect to the variational distribution $q(w|\theta)$ therefore is also called the likelihood cost.

In this way, we sample the variational distribution using a Gaussian distribution which consists of parameters (μ, σ) where μ is the mean vector of the distribution and σ represents the standard deviation vector and the uncertainty of this framework could be parameterized by μ and σ. A training iteration consists of a forward pass and backward pass. During a forward pass, a single sample is drawn from the variational posterior distribution, which is used to evaluate the approximated loss function. During the backward pass, the gradients of μ and σ are calculated via backpropagation so that their values can be updated by an optimizer. It's worthwhile to mention that for numeric stability we parameterize the network with ρ instead of σ directly and transform ρ with the softplus function to obtain $\sigma = log\ (1 + \exp\ (\rho))$, which can ensure σ is always positive thus improving training process.

As denoted in Figure 22, the input size of VBNPCI is proportional to the size of measurements and output size is determined by the number of principal components. Each weight is sampled following a Gaussian distribution, which finally results in the reasonable uncertainty for the result. We use mean square error as the loss function to tune the format of output and maintain the dominant patterns not to be jeopardized by the added variations.

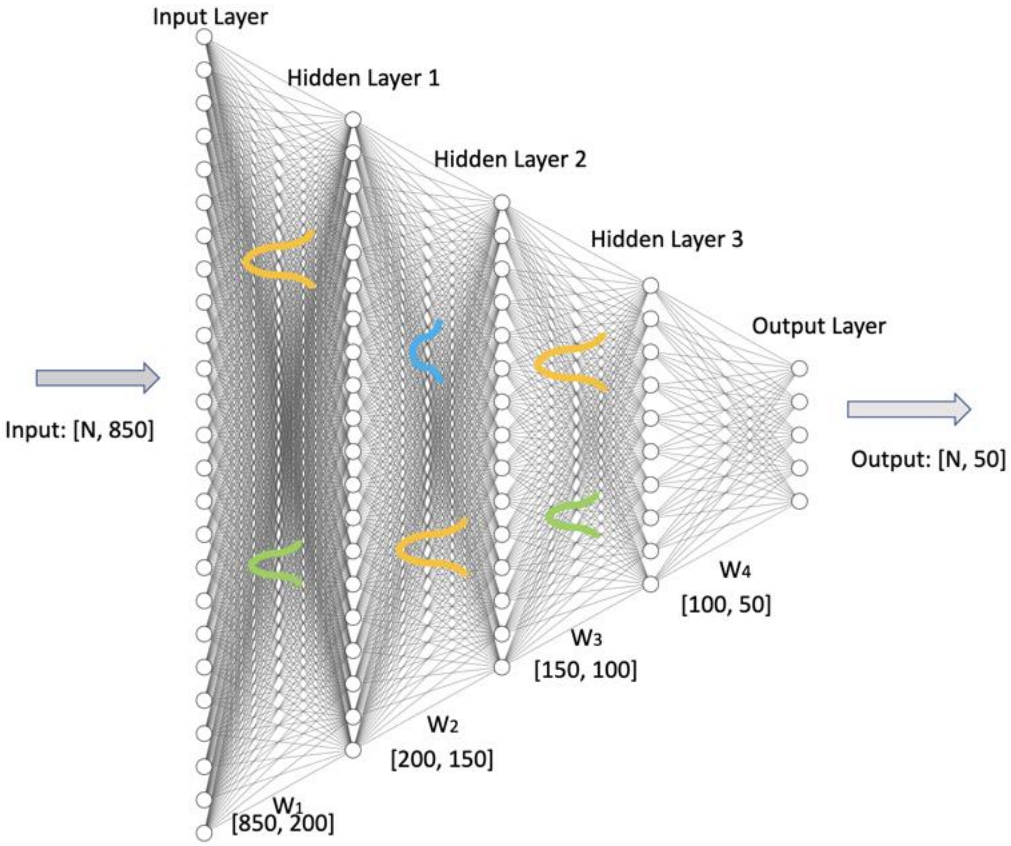


Figure 22. Variational Bayesian Neural Principal Component Inverse (VBNPCI) Architecture.

5.3 Numerical Experiments

5.3.1 Inverse Experiment Settings

Multiple numerical experiments of two-dimensional steady-state hydraulic tomography are presented to testify the uncertainty performance of the added Bayesian theorem on top of the proposed NNI and NNPCI framework in CHAPTER 4. For the consistency of the experiment, we use the same experimental parameters and well setups as in CHAPTER 4. Specifically, three heterogeneous fields of logarithmic hydraulic conductivity with different parameters are used in the experiments. Table 1 summarizes the geostatistical properties, hydraulic parameters, the governing equation, and associated boundary conditions. Figure 12 depicts a random field with Gaussian covariance matrix and another random field with exponential covariance matrix as the true fields of logarithmic hydraulic conductivity. There are 35 wells uniformly installed in the field, among which the black dots represent 25 dual-function (pumping and monitoring) wells, and the white dots represent 10 additional pumping wells. At every stage of hydraulic tomography, when one of dual-function wells (black dots) is used as pumping well, another 24 wells are used as monitoring wells to record the steady-state hydraulic heads; when one pumping well (white dots) pumps water, the other 25 dual-function wells (black dots) record the steady-state water head. Therefore, the number of available measurement data for each experiment is 850 in total. The observed data is also polluted by 1.5%~3%, which approximates the situation in actual application.

To generate the training data, random Gaussian field realizations with the Gaussian covariance or exponential covariance model were generated using Fast Fourier Transform.

For VBNPCI applications, the realizations are generated based on the retained principal components. Channelized fields were created using the data augmentation algorithm proposed in CHAPTER 3 in order to further substantiate the inverse modeling research. The aim is to prove the effectiveness of the proposed algorithm for generalization and uncertainty analysis in channels or even more complex fields. The geostatistical characteristics and hydraulic parameters are summarized in Table 2; the well setup shown in Figure 13 is the same as the above Gaussian random field experiment.

For the consistency of the experiment, in each of the three numerical experiments, we used the same input data as that in CHAPTER 4, and divided it into 70% training, 15% validation and the remaining 15% for testing. All numerical experiments were implemented on a MacBook Pro equipped with Intel® Core i7 UHD Graphics 630 2.20 GHz processor and 16.00 GB RAM.

5.3.2 Results and Discussion

In CHAPTER 4, we have proved the effectiveness and efficiency of the NNI and NNPCI framework on the random fields with Gaussian covariance function or exponential covariance function and even more complex fields with special characteristics (such as channels). Mapping accuracy and correlation coefficient were used to quantitatively confirm our observations. However, in practical applications, models or systems are often affected by errors from various sources, including input errors, model structure errors, model parameter errors, and so on. In other words, we cannot guarantee our predictions from the NNI or NNPCI framework to be always reliable. That is why we need to consider these errors and perform the following research study with the help of VBNI and VBNPCI.

Taking the VBNPCI framework as an example, the field experiments of Gaussian covariance and exponential covariance matrix are shown in Figure 23 and Figure 24 respectively, and similar results have been obtained. In the second row of each figure, the three predictions show promising results and provide detailed difference, especially around areas with very high or very low hydraulic conductivity values. We take the average of 100 predictions in each experiment, and the results show that the main patterns are well captured. According to the variance map, even in the boundary area where we did not collect enough measurements, the variance is very low. This means that how to install wells has little effect on the inverse estimation by using our proposed algorithm, and it is worthwhile to conduct more experiments in the future. As more conditional realizations emerge, more possibilities in each specific field would provide more guidance for researchers or field workers. What's more, taking the average of all realizations will increase the reliability of the output, so that the entire system or model becomes more stable and consistent on varying spatial and temporal scales.

In terms of pattern capture performance, the mapping accuracy and correlation coefficient were quantitatively analyzed according to same standards in CHAPTER 4, and the results are promising. The principal components fitting graph is unique for VBNPCI, as illustrated in the bottom sub-figures of Figure 23 and Figure 24. Comparing the true principal component coefficients with the estimated coefficients, we will see that for both cases, the true component coefficients are well contained in the domain formed by the averaged principal coefficients.

In addition, experiments of the VBNI framework were conducted on random with Gaussian covariance function in Figure 25, field with exponential covariance function in

Figure 26, and channelized field shown in Figure 27, and results were similar to the above-mentioned analysis with the VBNPCI framework. Therefore, we can conclude that VBNI and VBNPCI can provide a new perspective for uncertainty study to better understand the aquifer characterizations in hydrogeology area, and also could be extended to other scientific disciplines, i.e. environmental engineering, geophysics, earth science, and so on.

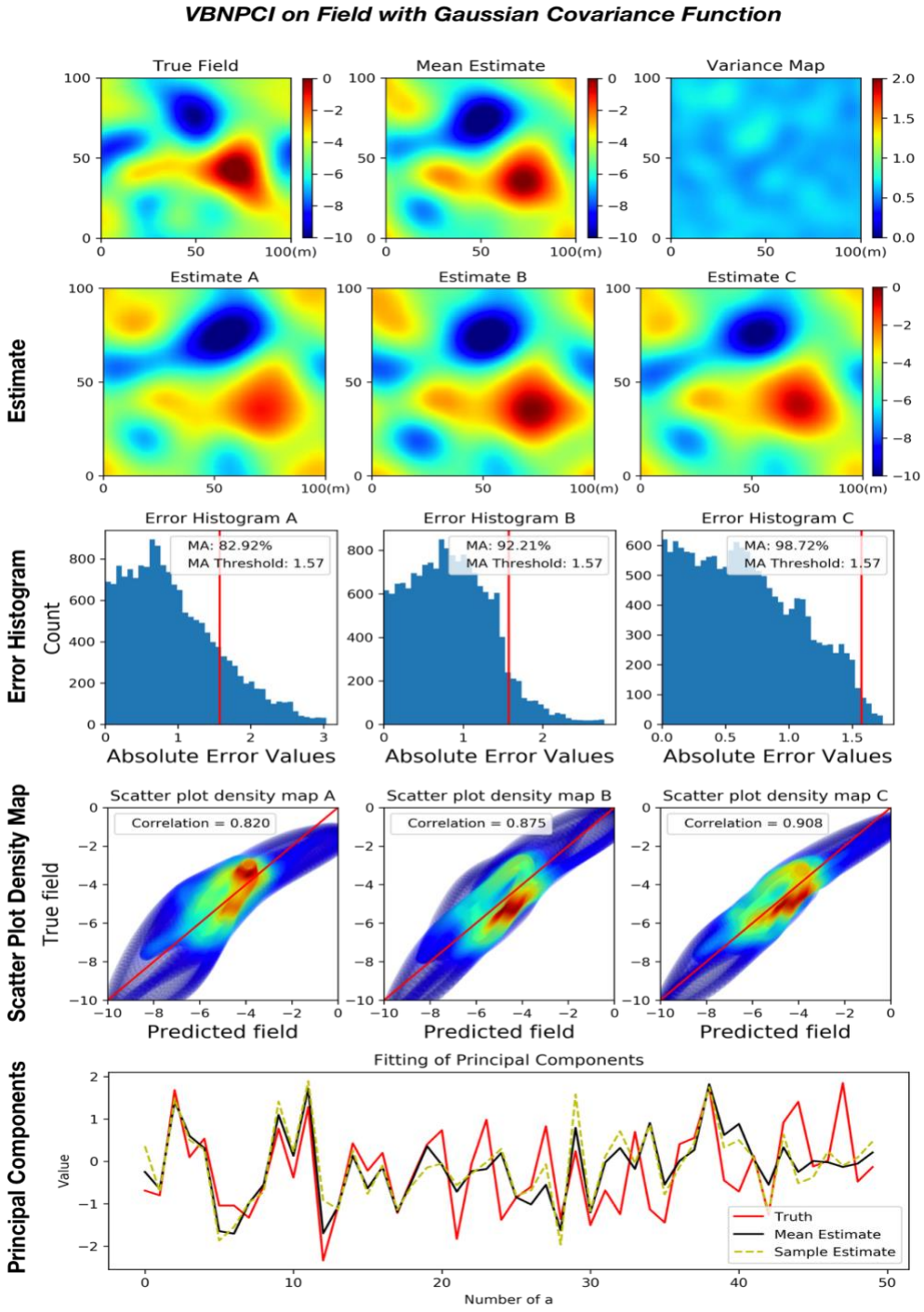


Figure 23. Inversion results with the VBNPCI framework for Gaussian random fields with Gaussian covariance function on retained principal components. The first row: true field (left), average estimate of 100 predictions (middle), and variance map (right); the second row: three predictions; the third row: scatter plot density map for each prediction; the fourth row: principal component fitting.

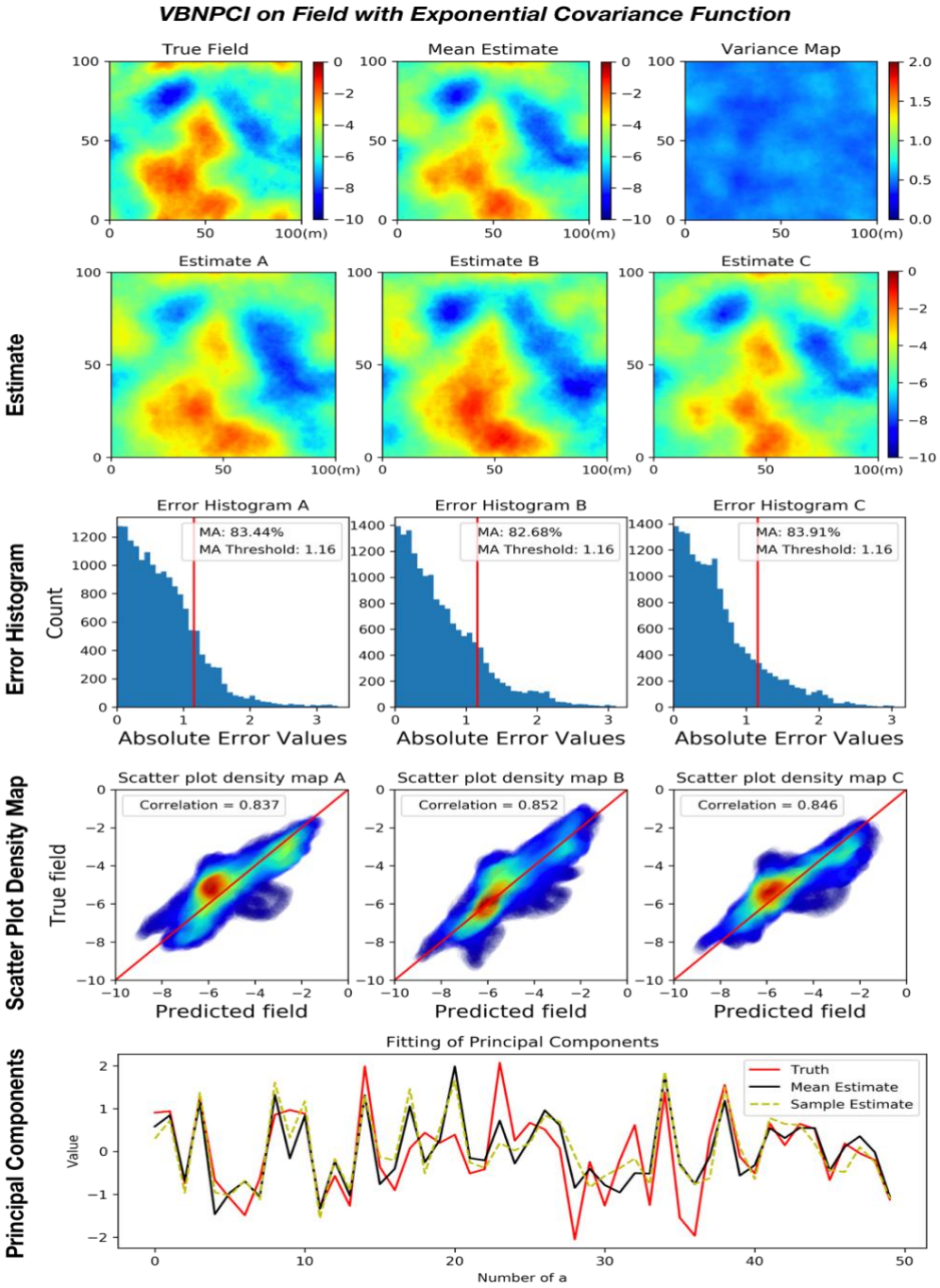


Figure 24. Inversion results with the VBNPCI framework for Gaussian random fields with exponential covariance function on retained principal components. The first row: true field (left), average estimate of 100 predictions (middle), and variance map (right); the second row: three predictions; the third row: absolute error histogram; the fourth row: scatter plot density map for each prediction; the fifth row: principal component fitting.

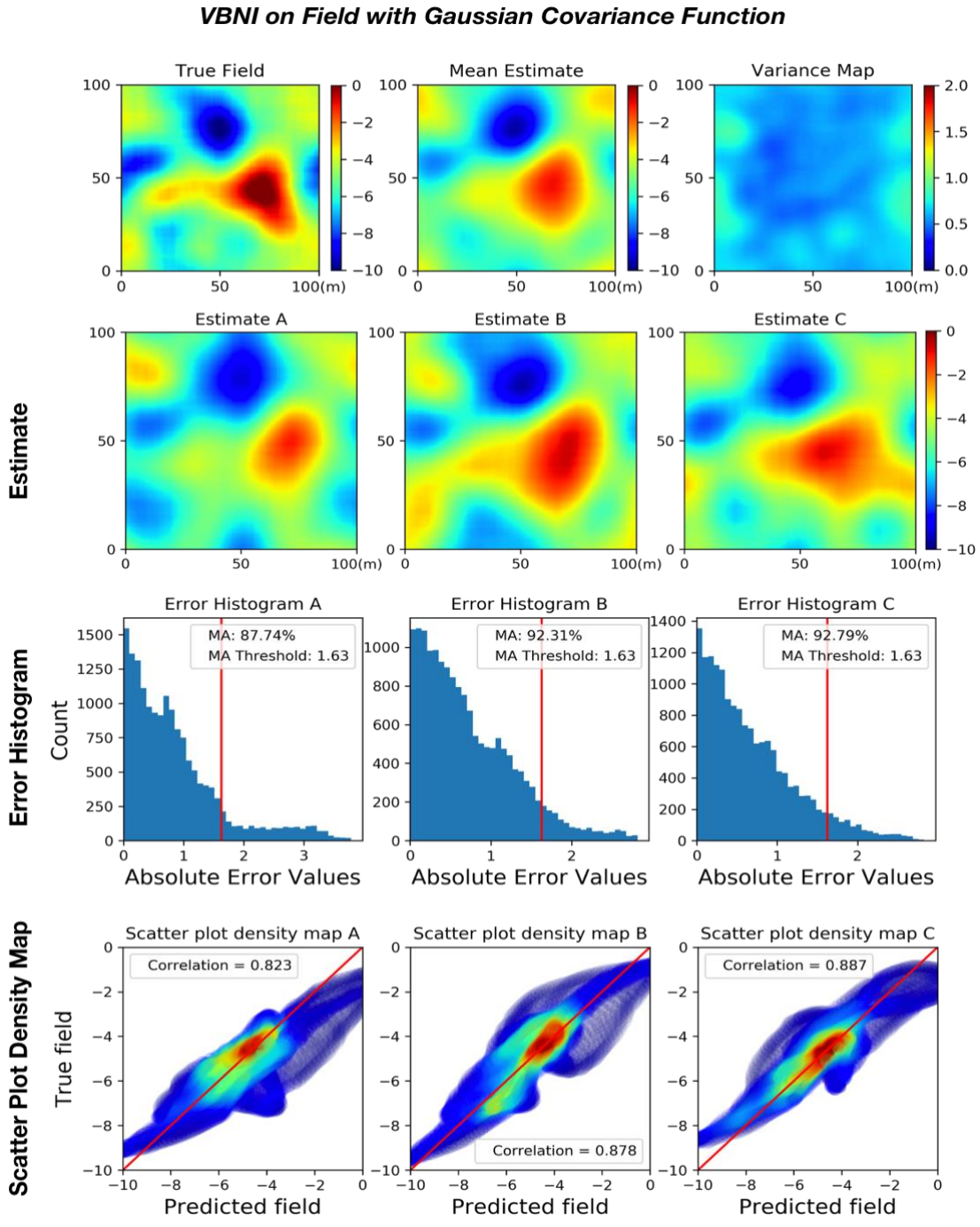


Figure 25. Inversion results with the VBNI framework for Gaussian random fields with Gaussian covariance function on full dimensions. The first row: true field (left), average estimate of 100 predictions (middle), and variance map (right); the second row: three predictions; the third row: absolute error histogram; the fourth row: scatter plot density map for each prediction.

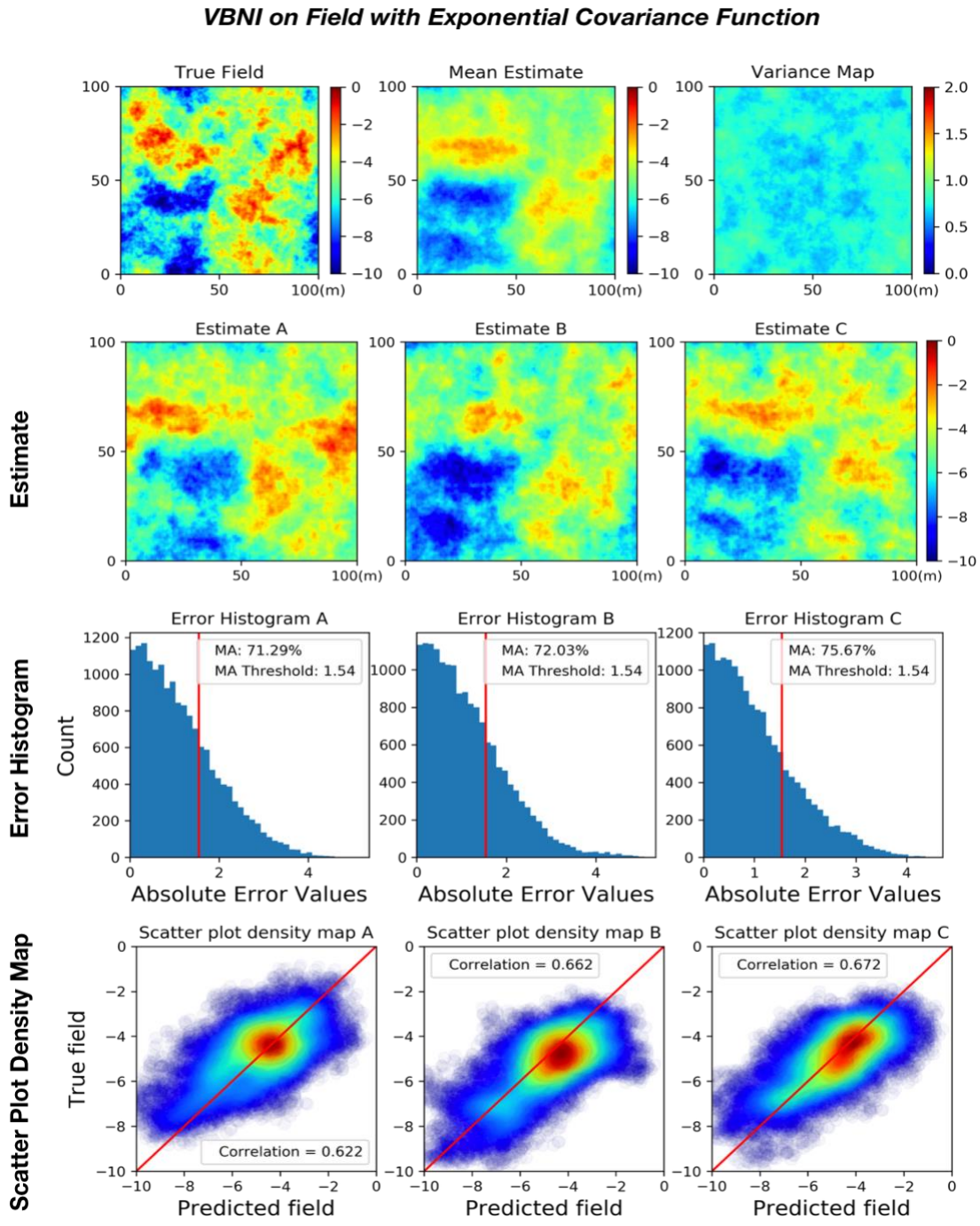


Figure 26. Inversion results with the VBNI framework for Gaussian random fields with exponential covariance function on full dimensions. The first row: True field, average estimate of 100 predictions, variance map; the second row: three predictions; the third row: absolute error histogram; the fourth row: scatter plot density map for each prediction.

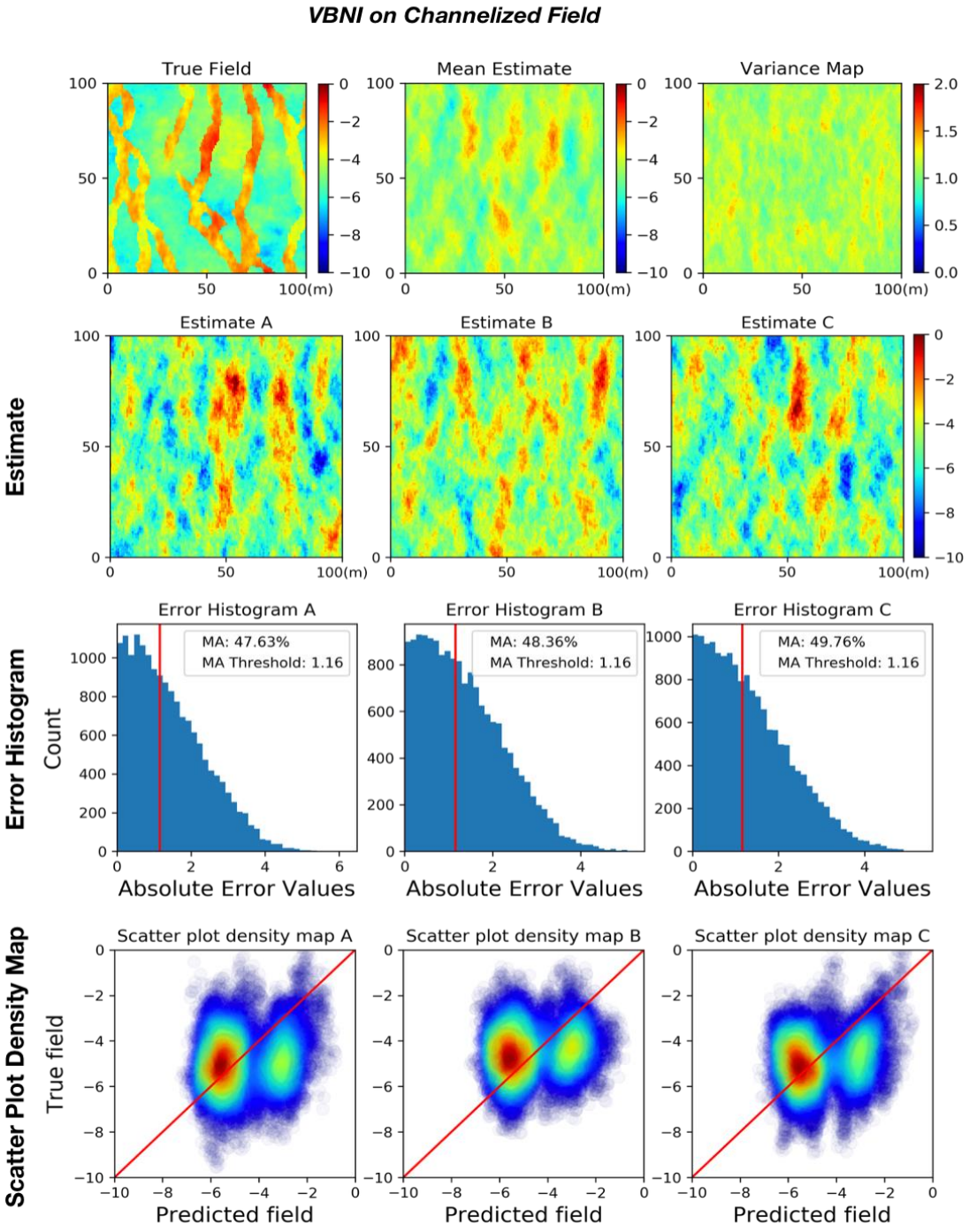


Figure 27. Inversion results with the VBNI framework for channelized random fields on full dimensions. The first row: True field, average estimate of 100 predictions, variance map; the second row: three predictions; the third row: absolute error histogram; the fourth row: scatter plot density map for each prediction.

5.4 Conclusion

In this chapter, we develop a Variational Bayesian Neural Inverse (VBNI) on full dimensions and Variational Bayesian Neural Principal Component Inverse (VBNPCI) framework on retained principal components to handle the uncertainty estimates of hydraulic tomography in heterogeneous hydraulic conductivity fields. This proposed approach is composed of NNI or NNPCI developed in CHAPTER 4 and Bayesian theorem, and provide uncertainty analysis without jeopardizing the entire trend of the field. Three sets of numerical experiments were conducted on fields with different patterns and hydraulic characteristics and a satisfactory performance was achieved in capturing the main characteristics for each field.

As more conditional realizations emerge, uncertainties in each specific field would provide more guidance for researchers or field workers. At the same time, the average value of all realizations will increase the reliability of the output, so that the entire system or model behaves more stable and consistent on varying spatial and temporal scales. Therefore, we can conclude that VBNI and VBNPCI can provide a new uncertainty research perspective for better characterization of underground aquifers in the field of hydrogeology, and may also be extended to other scientific disciplines, namely environmental engineering, geophysics, earth science, and so on.

CHAPTER 6. APPLICATION OF NNI AND NNPCI TO REAL WORLD DATA

6.1 Introduction

In this chapter, the proposed neural network principal component inverse (NNPCI) is applied to investigate the hydraulic conductivity field and storage coefficient field of a deep lime stone aquifer in Xingdong coalmine site located in China. The inversion results provide more quantitative geological information of the coalmine site and have the potential to provide guidance for coalmine operations in the future to evade incidents.

There are seven observation wells installed around the coalmine boundary to monitor the water level changes, which functions to provide early warning when incidents such as water inrush happen. These wells are shown in Figure 28 with names of D1, D3, D4, D5, D6, D9 and D11. Out of safety considerations, no monitoring well is installed in the mining area.

To characterize hydraulic conductivity and storage coefficient of the coalmine area, a pumping test was conducted. As shown in Figure 29, during this pumping test, in which D5 acts as the pumping well while the rest six wells act as monitoring wells, there are 14 water level readings recorded for each monitoring well. The 14 readings of each well record water heads in 7 days for every half a day. In addition to the pumping test, Figure 29 also records the drawdown curves for a water inrush incident happened in W1, which can be regarded as another pumping test and complete the sequential pumping test. The data of the water inrush incident is also recorded on the same time resolution. It worth noting that during the water inrush, drawdown in D9 is much larger than those in the other wells.

Therefore, another separated axis is adopted on the left of the figure to illustrate the range of the drawdown curve in D9, relevant experiment results regarding D9 will also be discussed in later section of this chapter.

To estimate the hydraulic conductivity and the storage coefficient, we treat these two unknown fields independently and separately, and apply a joint neural network inversion approach to the sequential pumping test data. In this implication, storage coefficient is also unknown like hydraulic conductivity. We build two relationships between hydraulic conductivity (or storage coefficient) with drawdown respectively to invert the two unknown fields on full dimensions or retained principal components. Thus, a premium advantage of the joint neural network inverse model is that we treat the two inverse models not dependent with each other, making the training process simpler than traditional geostatistical methods which requires computing Jacobian matrix iteratively. Taking the NNI framework as an example, the relationship between the hydraulic conductivity $\boldsymbol{k}$ or the storage coefficient $\boldsymbol{s}$ and the drawdown values $\boldsymbol{y}$ can be expressed in Equation. (37):

$$\begin{aligned} \boldsymbol{s} &= NNI(\boldsymbol{y}) \\ \boldsymbol{k} &= NNI\_joint(\boldsymbol{y}) \end{aligned} \tag{37}$$

Therefore, the reformulated forward model is shown in Equation. (38):

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{s}, \boldsymbol{k}) + \boldsymbol{\epsilon} \tag{38}$$

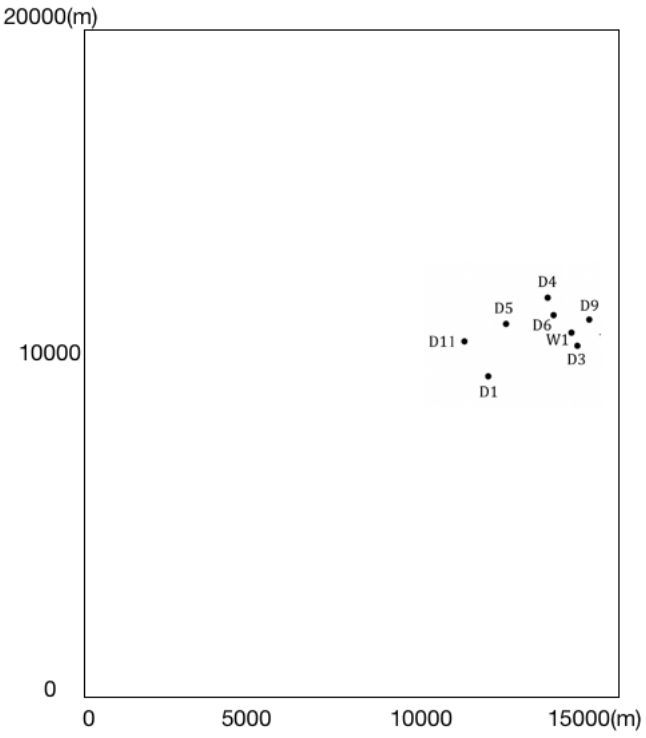


Figure 28. Setup of installed wells in the coalmine area.

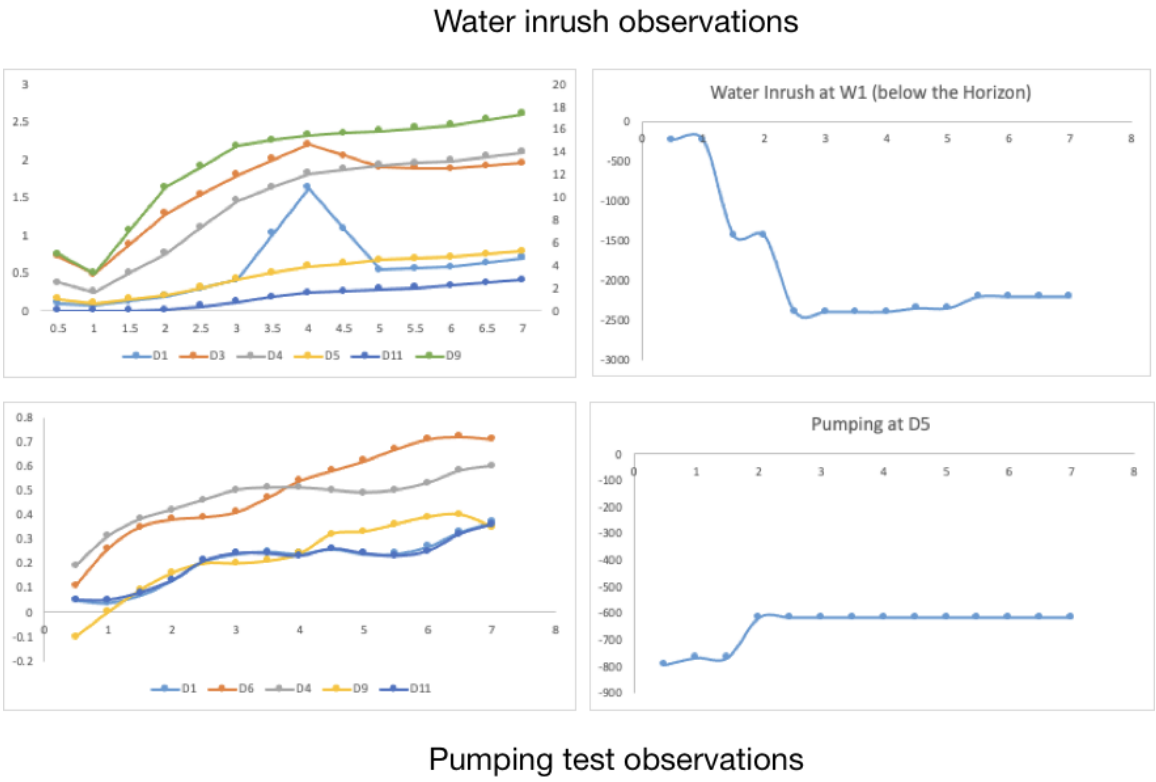


Figure 29. Draw down curves for the water inrush test and the pumping test in the coalmine area.

6.2 Real World Data Inversion

6.2.1 Training Data

The inversion of real-world data is always based on a prior guess of the field characteristics governed by structural parameters of the field in hydrogeology and geostatistical area (Zhao, 2020). The initialization of the geostatistical parameters is shown in Table 6, which reflects our prior belief in the hydraulic conductivity field and the storage coefficient field. Other values of the parameters are also explored in data inversion; however, the inversion of the data does not produce significant differences for different prior setups. This is because a similar set of geostatistical parameters can provide almost the same inversion results.

The studied domain stretches $15000\,m \times 20000\,m$ with a discretized resolution 120×160. 1000 synthetic realizations were generated for fields of hydraulic conductivity and storage coefficient respectively; each realization contains two parameterized fields that share the same drawdown value. Thus, we divided the whole dataset into 70% training, 15% validation and the remaining 15% for testing.

All numerical experiments are implemented on a MacBook Pro equipped with Intel® Core i7 UHD Graphics 630 2.20 GHz processor and 16.00 GB RAM.

Table 6. Geostatistical and hydraulic parameters in real-world data inverse modeling.

Geostatistical Properties	
Domain Scale	$15000\text{m} \times 20000\text{m}$
Field Type	Logarithmic Field
Resolution	750×1000
Covariance Model	exponential model
Mean	$\mu = 1.2\text{m/s}$ $\mu = -9.0$
Variance	$\sigma^2 = 1.5\text{m}^2/\text{s}^2$ $\sigma^2 = 1.5$
Correlation length	$1000\text{m} \times 1000\text{m}$
Governing Equation	
$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)\left(K\left(\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\right)\right) = Q + S\frac{\partial h}{\partial t}$	
North	$h = 950\text{m}$
South	$h = 950\text{m}$
West	$h = 950\text{m}$
East	Impermeable

6.2.2 Inversion of Real Data and Cross Validation

The inversion is solved by NNI and NNPCI for the two fields of hydraulic conductivity and storage coefficient respectively. The reversion results from these two frameworks are similar, which validate our research in Chapter 4. For illustration purpose, we displayed the estimates of the logarithmic hydraulic conductivity and the logarithmic storage coefficient obtained using the NNI framework. In order to better take advantage of the limited and precious real world data, we applied cross validation strategy to verify if the frameworks perform well in the specific implication. Specifically, we conducted two parallel experiments as follows.

Cross Validation 1: Since part of water inrush observations are not accurate, e.g. outliers might be existing in the observations at D1 shown in Figure 29. The measurements at D9 are much larger than other monitoring wells, which means some unexpected phenomena might occur during the observation period. Thus, we decide to treat them independently and use pumping test realizations only in the training of first inverse model and invert the predicted fields of hydraulic conductivity and storage coefficient, which is followed by cross validation using the remaining six wells in the water inrush test. The inverse estimations of hydraulic conductivity and storage coefficient are shown in Figure 30. Clearly, the result indicates that the studied area possesses highly heterogeneous behaviors in both fields. The magnitude of the hydraulic conductivity **K** ranges from 10^{-3} to 10^{5}, while the storage coefficient **S** ranges from 10^{-18} to 10^{-2}. The result also reveals the spatial patterns of conductivity and storage coefficient in the studied area. However, we cannot fully guarantee these two approximate the situation in actual application because it is impossible to get the real data in hand for this domain. Thus, we investigate the

performance of NNI from the perspective of reproducing the measurements as well as cross validation on the remaining data. For example, in experiment 1, the pumping test data is used for reproducing measurements and water inrush data is used for cross validation.

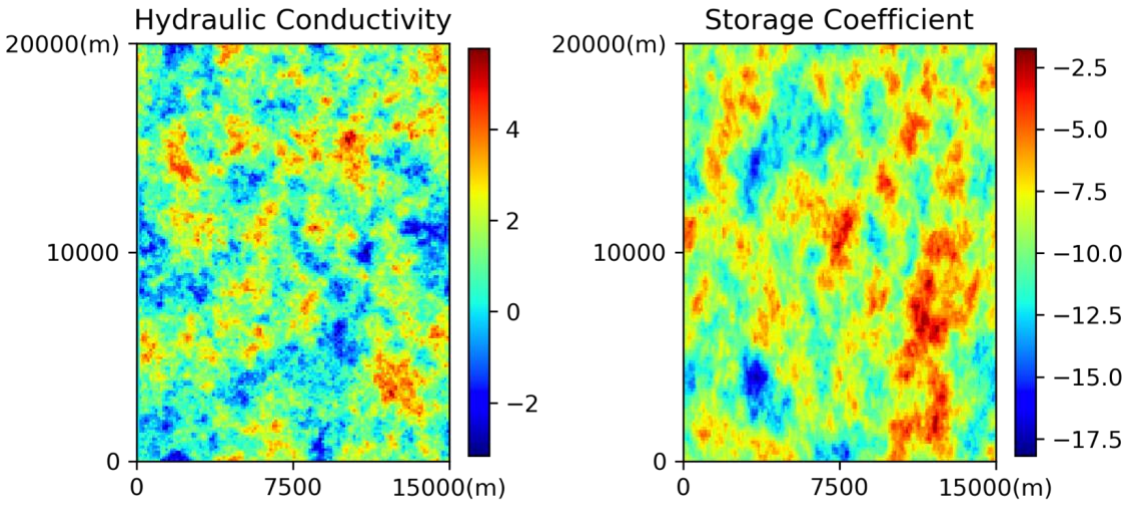


Figure 30. Estimates of logarithmic hydraulic conductivity field and logarithmic storage field.

The results indicate that the proposed approach successfully reproduces the drawdown curves of the pumping test as shown in Figure 31, which is conducted under a normal pumping rate. This matches with our initial expectation because we applied realizations from the pumping test wells to train the NNI framework. However, the cross validation on the water inrush incident (the remaining 6 wells) does not perform as well as the reproduced measurements in the pumping test, especially in D3, D4 and D9 shown in Figure 32. There might be multiple reasons for this. Firstly, unlike a pumping test that is conducted by researchers and engineers with deliberately designed setup, an extreme flow rate might occur in the water inrush incident, which potentially leads to unexpected hydraulic behaviors during the incident (Mao et al., 2018; Zhao, 2020). Secondly, the pumping of water from the coalmine is not simultaneous when the water inrush happens. Lastly, the provided data may be badly contaminated for some reasons, especially the measurements from water inrush test.

Cross Validation 2: In the second validation, we use D11 well as the testing well and the data from all other wells except D9 to train the NNI framework because based on experiment 1 we guess data from D9 in water inrush test was contaminated for unknown reasons. The performance of NNI in inverting field data is also investigated from the perspective of reproducing the measurement as shown in Figure 33, which demonstrates the NNI framework works well overall, particularly D3 and D4 data from the water inrush wells fit better than that in the first experiment. In addition, Figure 34 shows the cross validation performance of this experiment; the good fitting between real data in water inrush well D11 and the predicted value using the inverted hydraulic conductivity and storage coefficient is encouraging for us. Thus, we can conclude that our model perform

well overall and cross validation works well in the real-world application. For future study, if there are more data trunks collected, the NNI framework will assimilate the new data and further calibrate the inverse model that will provide more guidance for this coalmine field.

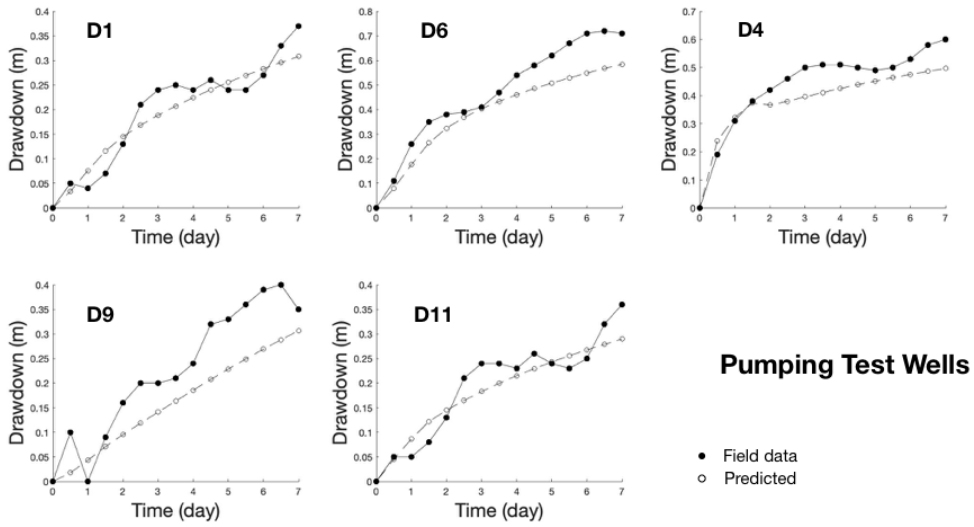


Figure 31. Experiment 1: fitting of measurements in five pumping wells.

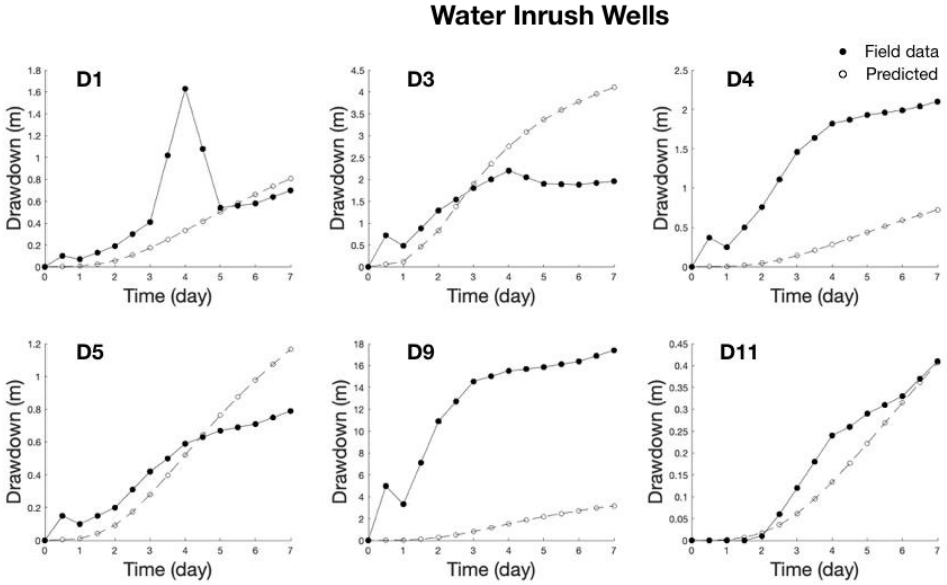


Figure 32. Experiment 1: cross validation of estimated parameter fields by measurements in the remaining six water inrush wells.

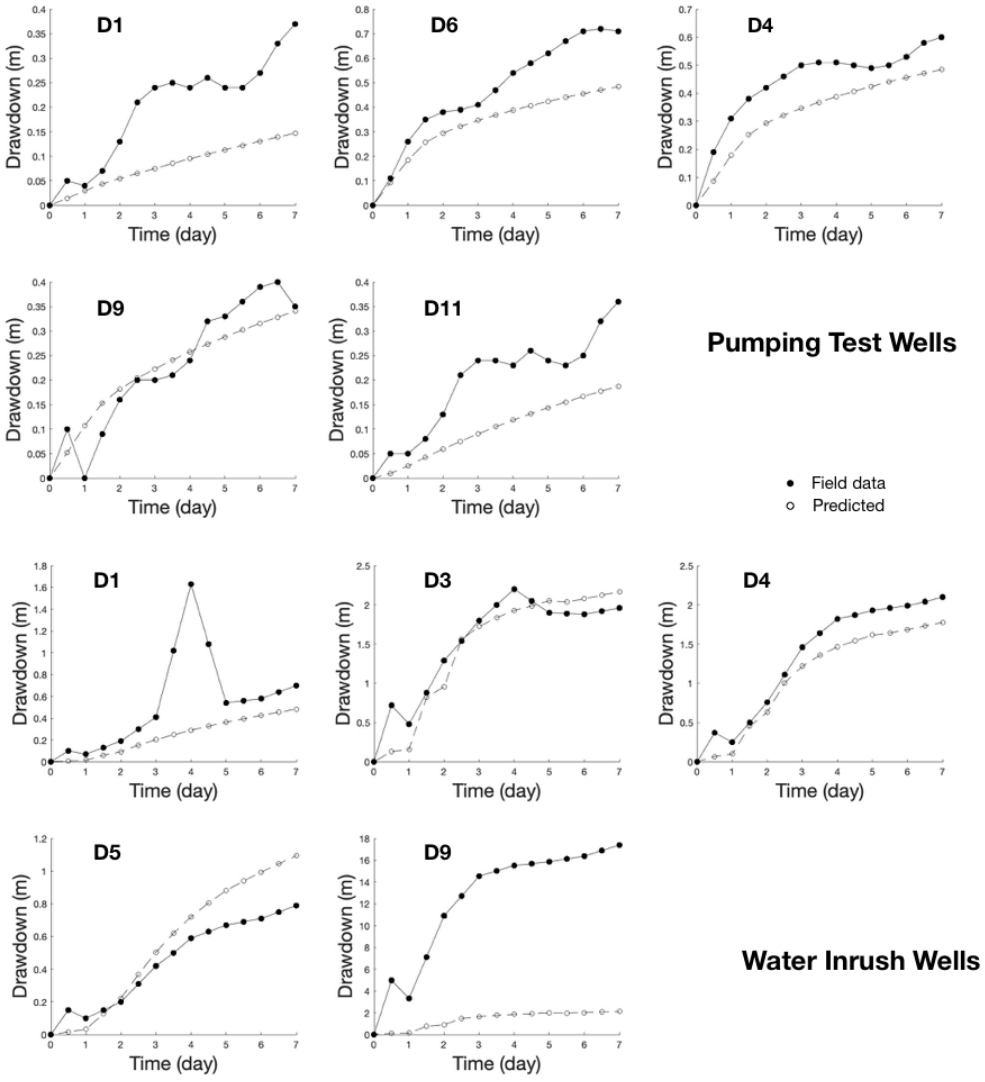


Figure 33. Experiment 2: fitting of measurements in all pumping wells and five water inrush wells.

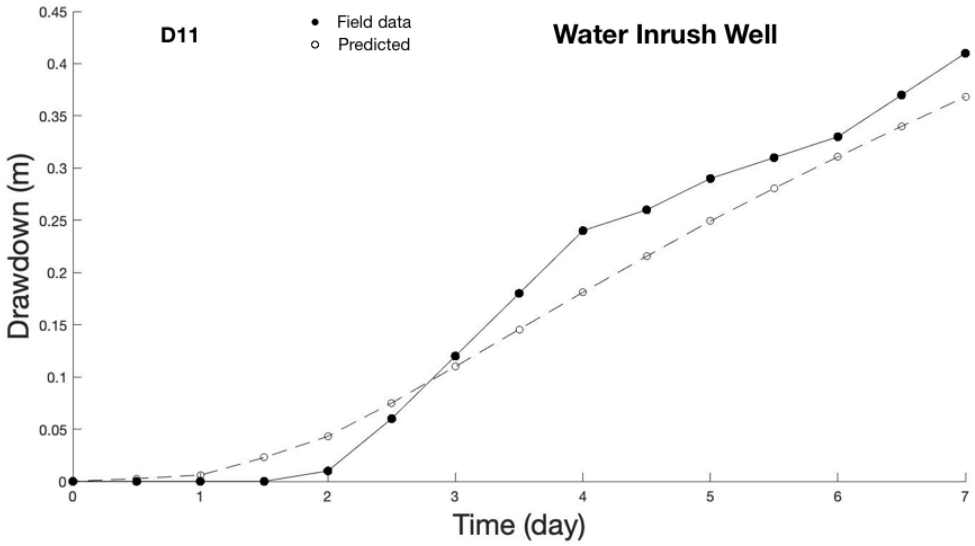


Figure 34. Experiment 2: cross validation of estimated parameter fields by measurements in D11 (water inrush well).

6.3 Conclusion

In this chapter, we investigate the performance of the proposed NNI inverse algorithm with a real-world application, i.e., a large-scale inverse problem with transient pumping data in Xingdong coalmine site located in China. The inversion result reveals heterogeneous fields of high or low hydraulic conductivity and storage coefficients. Our NNI framework performs well overall; the fitting of measurements is aligned with that in the cross validation of estimated parameter field by measurements in the testing well; this will continue to contribute to the future study of this coalmine field. For future study, if there are more trunk data collected, the NNI framework will assimilate the new data and better calibrate the inverse model and provide more guidance for this coalmine field.

CHAPTER 7. CONCLUSION AND RECOMMENDATION

7.1 Summary and Conclusions

In this thesis, we improve the hydrogeological inverse modeling with modern machine learning algorithms and demonstrate their applications in estimating large-dimensional, highly-heterogeneous hydrogeological parameter fields. Specifically, we develop a channelized field generation approach and bridge the connection between hydraulic tomography inverse modeling and the state-of-the-art algorithm neural network, providing new methods for better aquifer characterization.

In CHAPTER 3, based on the texture synthesis and the image quilting approach, a Fast Conditional Image Quilting (FCIQ) method was proposed. The efficiency and effectiveness have been successfully verified on binary TIs, continuous TIs, and more complex images with special characteristics. This algorithm was used to generate channelized random fields to support the inverse modeling research study in CHAPTER 4 and CHAPTER 5.

In CHAPTER 4, we proposed Neural Network Inverse (NNI) framework on full dimensions and Neural Network Principal Component Inverse (NNPCI) framework on reduced dimensions by PCA to handle the inverse problems of hydraulic tomography in heterogeneous hydraulic conductivity fields. Our neural network surrogate models directly describe the inverse relationship between the indirect measurements, i.e. hydraulic heads and the underlying parameter fields, hydraulic conductivity, or the retained principal component coefficients. The efficiency of the NNPCI is higher than NNI because the

retained principal components coefficients improve the scalability for the entire system. In contrast, the NNI framework could be generalized well to handle channelized random fields, which cannot be reflected by a covariance function and thus cannot be solved by traditional geostatistical approaches. In addition, one of the major computational advantages for NNI and NNPCI is that the training data can be generated by independent forward model simulations that can be done efficiently using parallel computing.

In CHAPTER 5, we develop a Variational Bayesian Neural Inverse (VBNI) on full dimensions and Variational Bayesian Neural Principal Component Inverse (VBNPCI) framework on retained principal components to handle the uncertainty estimates of hydraulic tomography in heterogeneous hydraulic conductivity fields. This proposed approach is composed of NNI or NNPCI developed in CHAPTER 4 and Bayesian theorem, and provide uncertainty analysis without jeopardizing the entire trend of the field. The system behaves more stably and consistently on varying spatial and temporal scales. Multiple numerical experiments were conducted on fields with different patterns and heterogeneous hydraulic characteristics. The main characteristics for each field was well captured with minor difference in details.

In CHAPTER 6, the performance of NNPCI is investigated by a field data case. The dataset is collected by conducting a pumping test and a water inrush incident in a coalmine site. The proposed approach inverts the field with alleviated computational cost, while simultaneously identifies the dominant patterns of hydraulic conductivity and storage coefficient in the studied area. The uncovered spatial patterns are aligned with what has been found by other researchers (Mao et al., 2018), and contribute to future exploration operation of this coalmine field.

7.2 Future Research Recommendations

(1) Uncertainty of mean and spatial covariance for NNPCI and DNN

Our NNPCI assumes that the mean and spatial covariance are known. Research is needed to account for the uncertainties from these parameters. For example, GA and RGA methods can use iterative procedures to gradually correct the spatial covariance. For neural network models, one possible solution is to include these parameters as the inputs, i.e., part of the training data, to the neural network models and develop new structures that can reflect their impacts. We have conducted some preliminary work in this area, and will continue to work on it. With uncertainties of mean and spatial covariance included in the machine learning, a new advanced model, DNN, is a worthwhile strategy to explore in the future.

(2) New applications of developed models

Most of our applications are hydraulic tomography because of its effectiveness in aquifer characterization. Our proposed approaches have the potential to be extended to other inversion applications such as seismic wave inversion for earthquake researcher and tracer data inversion.

(3) New machine learning algorithms

With the fast development of machine learning algorithms, we will continue to develop inverse modeling by incorporating the state-of-art machine learning algorithms. For example, we can develop Generative Adversarial Networks (GAN)-based algorithm to

create high quality complex field, generalize to 3D reproduction, and finally support the inverse modeling development in hydrogeology area.

(4) Physics-informed neural network (PINN)

In recent years, PINN has been proposed for inverse modeling. PINN includes the governing equations, boundary conditions, and others in the objective function to optimize. Thus, PINN can learn the actual spatially distributed solutions, while we only establish the relationships between the measurements and the underlying parameter field. It is worth exploring the potential to integrate our models with PINN.

APPENDIX.

REFERENCES

Abdollahifard, M. J., & Faez, K. (2013). Stochastic simulation of patterns using Bayesian pattern modeling. *Computational Geosciences, 17*(1), 99-116.

Ambikasaran, S., Li, J. Y., Kitanidis, P. K., & Darve, E. (2013). Large-scale stochastic linear inversion using hierarchical matrices. *Computational Geosciences, 17*(6), 913-927.

Arpat, G. B., & Caers, J. (2007). Conditional simulation with patterns. *Mathematical geology, 39*(2), 177-203.

Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015a). *Weight uncertainty in neural network.* Paper presented at the International Conference on Machine Learning.

Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015b). Weight uncertainty in neural networks. *arXiv preprint arXiv:1505.05424*.

Boucher, A., Kyriakidis, P. C., & Cronkite-Ratcliff, C. (2008). Geostatistical solutions for super-resolution land cover mapping. *IEEE Transactions on Geoscience and Remote Sensing, 46*(1), 272-283.

Boykov, Y. Y., & Jolly, M.-P. (2001). *Interactive graph cuts for optimal boundary & region segmentation of objects in ND images.* Paper presented at the Proceedings eighth IEEE international conference on computer vision. ICCV 2001.

Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G., & Wilcox, L. C. (2012). *Extreme-scale UQ for Bayesian inverse problems governed by PDEs.* Paper presented at the Proceedings of the international conference on high performance computing, networking, storage and analysis.

Caers, J., Strebelle, S., & Payrazyan, K. (2003). Stochastic integration of seismic data and geologic scenarios: A West Africa submarine channel saga. *The Leading Edge, 22*(3), 192-196.

Cardiff, M., Barrash, W., Kitanidis, P., Malama, B., Revil, A., Straface, S., & Rizzo, E. (2009). A potential ‐ based inversion of unconfined steady ‐ state hydraulic tomography. *Groundwater, 47*(2), 259-270.

Carrera, J., & Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Water Resources Research, 22*(2), 199-210.

Chen, J., Kemna, A., & Hubbard, S. S. (2008). A comparison between Gauss-Newton and Markov-chain Monte Carlo–based methods for inverting spectral induced-polarization data for Cole-Cole parameters. *Geophysics, 73*(6), F247-F259.

Cirpka, O. A., Fienen, M. N., Hofer, M., Hoehn, E., Tessarini, A., Kipfer, R., & Kitanidis, P. K. (2007). Analyzing bank filtration by deconvoluting time series of electric conductivity. *Groundwater, 45*(3), 318-328.

Coiffier, G., & Renard, P. (2019). *3d geological image synthesis from 2d examples using generative adversarial networks.* Paper presented at the Petroleum Geostatistics 2019.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation, 19*(90), 297-301.

Cui, T., Martin, J., Marzouk, Y. M., Solonen, A., & Spantini, A. (2014). Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems, 30*(11), 114015.

Dong, L., Xiong, L., & Yu, K.-x. (2013). Uncertainty analysis of multiple hydrologic models using the Bayesian model averaging method. *Journal of Applied Mathematics, 2013*.

Efros, A. A., & Freeman, W. T. (2001). *Image quilting for texture synthesis and transfer.* Paper presented at the Proceedings of the 28th annual conference on Computer graphics and interactive techniques.

Efros, A. A., & Leung, T. K. (1999). *Texture synthesis by non-parametric sampling.* Paper presented at the Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on.

Falivene, O., Arbus, P., Gardiner, A., Pickup, G., Muoz, J. A., & Cabrera, L. (2006). Best practice stochastic facies modeling from a channel-fill turbidite sandstone analog (the Quarry outcrop, Eocene Ainsa basin, northeast Spain). *AAPG bulletin, 90*(7), 1003-1029.

Fang, K., Shen, C., Kifer, D., & Yang, X. (2017). Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. *Geophysical Research Letters, 44*(21), 11,030-011,039.

Fienen, M. N., Luo, J., & Kitanidis, P. K. (2006). A Bayesian geostatistical transfer function approach to tracer test analysis. *Water Resources Research, 42*(7).

Fogg, G. E. (1986). Groundwater flow and sand body interconnectedness in a thick, multiple ‐ aquifer system. *Water Resources Research, 22*(5), 679-694.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(6), 721-741.

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

Gong, R. (2013). *Mixing-controlled reactive transport in connected heterogeneous domains.* Georgia Institute of Technology,

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

Hoeksema, R. J., & Kitanidis, P. K. (1984). An application of the geostatistical approach to the inverse problem in two‐dimensional groundwater modeling. *20*(7), 1003-1020.

Hu, W., Abubakar, A., Habashy, T., & Liu, J. (2011). Preconditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion. *Geophysical Prospecting, 59*(3), 477-491.

Jha, S. K., Mariethoz, G., Evans, J. P., & McCabe, M. F. (2013). Demonstration of a geostatistical approach to physically consistent downscaling of climate modeling simulations. *Water Resources Research, 49*(1), 245-259.

Journel, A., & Zhang, T. (2006). The necessity of a multiple-point prior model. *Mathematical geology, 38*(5), 591-610.

Kang, P. K., Lee, J., Fu, X., Lee, S., Kitanidis, P. K., & Juanes, R. (2017). Improved characterization of heterogeneous permeability in saline aquifers from transient pressure data during freshwater injection. *Water Resources Research, 53*(5), 4444-4458.

Kitanidis, & Vomvoris, E. G. (1983). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one‐dimensional simulations. *19*(3), 677-690.

Kitanidis, P. K. (1986). Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Research, 22*(4), 499-507.

Kitanidis, P. K. (1995). Quasi‐linear geostatistical theory for inversing. *Water Resources Research, 31*(10), 2411-2419.

Kitanidis, P. K. (1996). On the geostatistical approach to the inverse problem. *Advances in Water Resources, 19*(6), 333-342.

Kitanidis, P. K. (1997). *Introduction to geostatistics: applications in hydrogeology*: Cambridge University Press.

Kitanidis, P. K., & Lee, J. (2014). Principal Component Geostatistical Approach for large‐dimensional inverse problems. *Water Resources Research, 50*(7), 5428-5443.

Klein, O., Cirpka, O. A., Bastian, P., & Ippisch, O. (2017). Efficient geostatistical inversion of transient groundwater flow using preconditioned nonlinear conjugate gradients. *Advances in Water Resources, 102*, 161-177.

Klute, A., & Dinauer, R. C. (1986). Physical and mineralogical methods. *Planning, 8*, 79.

Klute, A., & Dinauer, R. C. J. P. (1986). Physical and mineralogical methods. *8*, 79.

Kwatra, V., Schödl, A., Essa, I., Turk, G., & Bobick, A. (2003). Graphcut textures: Image and video synthesis using graph cuts. *Acm transactions on graphics (tog), 22*(3), 277-286.

Lee, J., & Kitanidis, P. K. (2014). Large‐scale hydraulic tomography and joint inversion of head and tracer data using the Principal Component Geostatistical Approach (PCGA). *Water Resources Research, 50*(7), 5410-5427.

Lee, J., Kokkinaki, A., & Kitanidis, P. K. (2018). Fast large-scale joint inversion for deep aquifer characterization using pressure and heat tracer measurements. *Transport in Porous Media, 123*(3), 533-543.

Lee, J., Yoon, H., Kitanidis, P. K., Werth, C. J., & Valocchi, A. J. (2016). Scalable subsurface inverse modeling of huge data sets with an application to tracer concentration breakthrough data from magnetic resonance imaging. *Water Resources Research, 52*(7), 5213-5231.

Lewis, J. P. (1995). *Fast normalized cross-correlation.* Paper presented at the Vision interface.

Li, W., Nowak, W., & Cirpka, O. A. (2005). Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown. *Water Resources Research, 41*(8).

Liao, Z., & Cirpka, O. A. (2011). Shape‐free inference of hyporheic traveltime distributions from synthetic conservative and “smart” tracer tests in streams. *Water Resources Research, 47*(7).

Lin, Y., Le, E. B., O'Malley, D., Vesselinov, V. V., & Bui‐Thanh, T. J. W. R. R. (2017). Large‐scale inverse model analyses employing fast randomized data reduction. *Water Resources Research, 53*(8), 6784-6801.

Liu, S., Yeh, T. C. J., & Gardiner, R. (2002). Effectiveness of hydraulic tomography: Sandbox experiments. *Water Resources Research, 38*(4), 5-1-5-9.

Liu, X., Illman, W., Craig, A., Zhu, J., & Yeh, T. C. J. W. R. R. (2007). Laboratory sandbox validation of transient hydraulic tomography. *43*(5).

Liu, X., & Kitanidis, P. (2011). Large ‐ scale inverse modeling with an application in hydraulic tomography. *Water Resources Research, 47*(2).

Luo, J., & Cirpka, O. A. (2011). How well do mean breakthrough curves predict mixing ‐ controlled reactive transport? *Water Resources Research, 47*(2).

Mahmud, K., Mariethoz, G., Caers, J., Tahmasebi, P., & Baker, A. (2014). Simulation of Earth textures by conditional image quilting. *Water Resources Research*, *50*(4), 3088-3107.

Mao, D., Liu, Z., Wang, W., Li, S., Gao, Y., Xu, Z., & Zhang, C. (2018). An application of hydraulic tomography to a deep coal mine: combining traditional pumping tests with water inrush incidents. *Journal of Hydrology, 567*, 1-11.

Mariethoz, G., & Lefebvre, S. (2014). Bridges between multiple-point geostatistics and texture synthesis: Review and guidelines for future research. *Computers & geosciences, 66*, 66-80.

Mariethoz, G., McCabe, M., & Renard, P. (2012). Multivariate spatio-temporal reconstruction of gaps for spatially continuous satellite based retrievals. *Water Resour. Res, 48*, W10507.

Martin, J., Wilcox, L. C., Burstedde, C., & Ghattas, O. (2012). A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing, 34*(3), A1460-A1487.

Matheron, G. (1963). Principles of geostatistics. *Economic geology, 58*(8), 1246-1266.

Meyal, A. Y., Versteeg, R., Alper, E., Johnson, D., Rodzianko, A., Franklin, M., & Wainwright, H. (2021). Automated cloud based Long Short-Term Memory neural network based SWE prediction. *Broadening the Use of Machine Learning in Hydrology*.

Mital, U., Dwivedi, D., Brown, J. B., Faybishenko, B., Painter, S. L., & Steefel, C. I. (2020). Sequential imputation of missing spatio-temporal precipitation data using random forests. *Frontiers in Water, 2*, 20.

Mustapha, H., & Dimitrakopoulos, R. (2010). High-order stochastic simulation of complex spatially distributed natural phenomena. *Mathematical Geosciences*, *42*(5), 457-485.

Neuweiler, I., & Cirpka, O. A. (2005). Homogenization of Richards equation in permeability fields with different connectivities. *Water Resources Research, 41*(2).

Nowak, W., & Cirpka, O. A. (2004). A modified Levenberg–Marquardt algorithm for quasi-linear geostatistical inversing. *Advances in Water Resources, 27*(7), 737-750.

Nowak, W., & Cirpka, O. A. (2006). Geostatistical inference of hydraulic conductivity and dispersivities from hydraulic heads and tracer data. *42*(8).

Nowak, W., Tenkleve, S., & Cirpka, O. A. (2003). Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities. *Mathematical Geology, 35*(1), 53-66.

Okabe, H., & Blunt, M. J. (2007). Pore space reconstruction of vuggy carbonates using microtomography and multiple ‐ point statistics. *Water Resources Research, 43*(12).

Pham, T. D. (2012). Supervised restoration of degraded medical images using multiple-point geostatistics. *Computer methods and programs in biomedicine, 106*(3), 201-209.

Pollock, D., & Cirpka, O. A. (2008). Temporal moments in geoelectrical monitoring of salt tracer experiments. *Water Resources Research, 44*(12).

Rafiei Emam, A., Kappas, M., Fassnacht, S., & Linh, N. H. K. (2018). Uncertainty analysis of hydrological modeling in a tropical area using different algorithms. *Frontiers of Earth Science, 12*(4), 661-671. https://doi.org/10.1007/s11707-018-0695-y

Rezaee, H., Mariethoz, G., Koneshloo, M., & Asghari, O. (2013). Multiple-point geostatistical simulation using the bunch-pasting direct sampling method. *Computers & geosciences, 54*, 293-308.

Ronayne, M. J., Gorelick, S. M., & Caers, J. (2008). Identifying discrete geologic structures that produce anomalous hydraulic response: An inverse modeling approach. *Water Resources Research, 44*(8).

Sahu, R. K., Müller, J., Park, J., Varadharajan, C., Arora, B., Faybishenko, B., & Agarwal, D. (2020). Impact of Input Feature Selection on Groundwater Level Prediction From a Multi-Layer Perceptron Neural Network. *Frontiers in Water, 2*, 46.

Saibaba, A. K., Ambikasaran, S., Li, J. Y., Kitanidis, P. K., & Darve, E. F. (2012). Application of hierarchical matrices to linear inverse problems in geostatistics. *Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles, 67*(5), 857-875.

Sánchez-Vila, X., Carrera, J., & Girardi, J. P. (1996). Scale effects in transmissivity. *Journal of Hydrology, 183*(1-2), 1-22.

Schlüter, S., Weller, U., & Vogel, H. J. (2011). Soil ‐ structure development including seasonal dynamics in a long ‐ term fertilization experiment. *Journal of Plant Nutrition and Soil Science, 174*(3), 395-403.

Shen, C., Chen, X., & Laloy, E. (2021). Broadening the Use of Machine Learning in Hydrology. *Frontiers in Water, 3*, 38.

Silliman, S., & Wright, A. (1988). Stochastic analysis of paths of high hydraulic conductivity in porous media. *Water Resources Research, 24*(11), 1901-1910.

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

Strebelle, S., & Zhang, T. (2005). Non-stationary multiple-point geostatistical models. In *Geostatistics Banff 2004* (pp. 235-244): Springer.

Sun, A. Y., & Tang, G. (2021). Downscaling satellite and reanalysis precipitation products using attention-based deep convolutional neural nets. *Broadening the Use of Machine Learning in Hydrology*.

Tahmasebi, P., & Sahimi, M. (2013). Cross-correlation function for accurate reconstruction of heterogeneous media. *Physical review letters, 110*(7), 078002.

Tahmasebi, P., Sahimi, M., & Caers, J. (2014). MS-CCSIM: accelerating pattern-based geostatistical simulation of categorical variables using a multi-scale search in Fourier space. *Computers & Geosciences, 67*, 75-88.

Torquato, S., Beasley, J., & Chiew, Y. (1988). Two‐point cluster function for continuum percolation. *The Journal of chemical physics, 88*(10), 6540-6547.

Tran, T. T. (1994). Improving variogram reproduction on dense simulation grids. *Computers & geosciences, 20*(7-8), 1161-1168.

Van Erven, T., & Harremos, P. (2014). Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory, 60*(7), 3797-3820.

Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation, 10*(3), 273-290.

Western, A. W., Blöschl, G., & Grayson, R. B. (2001). Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research, 37*(1), 83-97.

Wu, J., Zhang, T., & Journel, A. (2008). Fast FILTERSIM simulation with score-based distance. *Mathematical Geosciences, 40*(7), 773-788.

Yeh, T. C. J., Gutjahr, A. L., & Jin, M. (1995). An iterative cokriging‐like technique for ground‐water flow modeling. *Groundwater, 33*(1), 33-41.

Yeh, T. C. J., Jin, M., & Hanna, S. (1996). An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields. *Water Resources Research, 32*(1), 85-92.

Yeh, T. C. J., & Liu, S. (2000). Hydraulic tomography: Development of a new aquifer test method. *Water Resources Research, 36*(8), 2095-2105.

Yin, D., & Illman, W. A. (2009). Hydraulic tomography using temporal moments of drawdown recovery data: A laboratory sandbox study. *Water Resources Research, 45*(1).

Zha, Y., Yeh, T. C. J., Illman, W. A., Zeng, W., Zhang, Y., Sun, F., & Shi, L. J. W. R. R. (2018). A reduced‐order successive linear estimator for geostatistical inversion and its application in hydraulic tomography. *Water Resources Research, 54*(3), 1616-1632.

Zhang, J., Li, Y., Huang, G., Wang, C., & Cheng, G. (2016). Evaluation of uncertainties in input data and parameters of a hydrological model using a Bayesian Framework: A case study of a snowmelt–precipitation-driven watershed. *Journal of Hydrometeorology, 17*(8), 2333-2350.

Zhang, J., Lin, G., Li, W., Wu, L., & Zeng, L. J. W. r. r. (2018). An iterative local updating ensemble smoother for estimation and uncertainty assessment of hydrologic model parameters with multimodal distributions. *54*(3), 1716-1733.

Zhang, Z., Jafarpour, B., & Li, L. (2014). Inference of permeability heterogeneity from joint inversion of transient flow and temperature data. *Water Resources Research, 50*(6), 4710-4725.

Zhao, Y. (2020). *BAYESIAN INVERSE MODELING: ALGORITHMS AND APPLICATIONS IN GROUND WATER.* Georgia Institute of Technology,

Zhao, Y., & Luo, J. (2020). Reformulation of Bayesian geostatistical approach on principal components. *Water Resources Research, 56*(4), e2019WR026732.

Zhao, Y., Rathore, S. S., Liu, M., & Luo, J. (2018). Joint Bayesian Inversion for Analyzing Conservative and Reactive Breakthrough Curves. *Journal of Hydrology*.

Zhou, H., Li, L., Franssen, H.-J. H., & Gómez-Hernández, J. J. J. M. G. (2012). Pattern recognition in a bimodal aquifer using the normal-score ensemble Kalman filter. *44*(2), 169-185.

Zhu, J., & Yeh, T.-C. J. (2005). Characterization of aquifer heterogeneity using transient hydraulic tomography. *Water Resources Research, 41*(7).

Zhu, J., & Yeh, T. C. J. (2006). Analysis of hydraulic tomography using temporal moments of drawdown recovery data. *Water Resources Research, 42*(2).

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