

# Water Resources Research®



## METHOD

10.1029/2023WR034939

### Key Points:

- RGA-FNO integrates the reduced geostatistical approach (RGA) with a Fourier neural operator (FNO) surrogate model
- A highly accurate FNO was trained to learn the solution operators of PDEs in hydraulic tomography, generating multiple hydraulic head data
- RGA-FNO can efficiently produce satisfactory estimations for random transmissivity fields with high-rank, small-scale heterogeneities

### Supporting Information:

Supporting Information may be found in the online version of this article.

### Correspondence to:

J. Luo,  
[jian.luo@ce.gatech.edu](mailto:jian.luo@ce.gatech.edu)

### Citation:

Guo, Q., He, Y., Liu, M., Zhao, Y., Liu, Y., & Luo, J. (2024). Reduced geostatistical approach with a Fourier neural operator surrogate model for inverse modeling of hydraulic tomography. *Water Resources Research*, 60, e2023WR034939. <https://doi.org/10.1029/2023WR034939>

Received 21 MAR 2023

Accepted 5 JUN 2024

### Author Contributions:

**Conceptualization:** Ming Liu, Yue Zhao, Yuzheng Liu, Jian Luo

**Formal analysis:** Quan Guo, Yu He

**Funding acquisition:** Jian Luo

**Methodology:** Quan Guo, Ming Liu,

Yue Zhao, Yuzheng Liu, Jian Luo

**Project administration:** Jian Luo

**Software:** Quan Guo

**Supervision:** Jian Luo

**Validation:** Quan Guo, Yu He

**Visualization:** Quan Guo

**Writing – original draft:** Quan Guo

**Writing – review & editing:** Yu He,

Jian Luo


© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by/4.0/) License, which permits use and

distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

## Reduced Geostatistical Approach With a Fourier Neural Operator Surrogate Model for Inverse Modeling of Hydraulic Tomography

Quan Guo<sup>1</sup> , Yu He<sup>1</sup> , Ming Liu<sup>1</sup>, Yue Zhao<sup>1</sup>, Yuzheng Liu<sup>1</sup>, and Jian Luo<sup>1</sup> 

<sup>1</sup>School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA

**Abstract** In this study, we propose a new approach for inverse modeling of hydraulic tomography (HT) using the Fourier neural operator (FNO) as a surrogate forward model. FNO is a deep learning model that directly parameterizes the integral kernel in Fourier space to learn solution operators for parametric partial differential equations (PDEs). We trained a highly accurate FNO to learn the solution operators of PDEs in HT, which can efficiently generate multiple channels of hydraulic head fields corresponding to specific pumping tests in new parameter fields through a single forward pass. The trained FNO surrogate model is integrated with the reduced geostatistical approach (RGA) to provide a powerful tool, named RGA-FNO, for inverse modeling of Gaussian random fields. RGA utilizes principal components to effectively reduce problem dimensions and encode geostatistical information. FNO benefits from deep learning automatic differentiation, enabling efficient backpropagation of gradients during inverse modeling. Our results show that RGA-FNO can efficiently produce satisfactory estimations for random transmissivity fields with an exponential covariance function, which is non-differentiable and visually non-smooth. The RGA-FNO application on steady-state multi-well HT is better than the single-well transient pumping test for inverse modeling. The data required for inverse modeling applications of RGA-FNO increases with the variance of the random transmissivity fields. FNO and other neural operator models have the potential to play an important role in groundwater modeling, especially in inverse modeling and experimental design, which often requires a large number of forward simulations.

## 1. Introduction

Hydraulic conductivity or transmissivity are key parameters that control groundwater flow through subsurface materials. However, these parameters are spatially distributed and costly to measure directly, necessitating the use of inverse modeling techniques to estimate them from hydraulic head measurements, such as hydraulic head data collected during hydraulic tomography (HT). Inverse modeling involves solving a complex and often ill-posed inverse problem, which requires specialized approaches. Kitanidis (1995) proposed the geostatistical approach (GA) in a Bayesian framework to tackle the inverse problem associated with Gaussian random fields (GRFs). In this approach, the spatial covariance function of the target GRF is often assumed to be known. GA employs gradient-based methods to maximize the likelihood of a posterior distribution of a GRF realization, taking available measurements/data and a prior into account.

The estimation of high-resolution parameter fields using GA poses computational challenges, including the high cost of matrix computation and storage, forward model simulation on high-resolution parameter fields, and the iterative forward simulations required to determine the Jacobian matrix numerically. To alleviate this computational burden, several efforts have been made, primarily focusing on accelerating matrix computation and reducing dimensionality (e.g., Ambikasaran et al., 2013; Broyden, 1965; Kitanidis & Lee, 2014; Klein et al., 2017; Lee et al., 2016; Lee & Kitanidis, 2014; Liu et al., 2013; Liu & Kitanidis, 2011; Nowak & Cirpka, 2004; Nowak et al., 2003; Saibaba et al., 2012). Zhao and Luo (2020) introduced the reduced geostatistical approach (RGA) to estimate the projections on principal components, which reduces the dimension of unknowns. They also proposed an iterative correction method to relax the requirement for precise prior information (Zhao & Luo, 2021b). Additionally, a quasi-Newton algorithm has been proposed to use a low-rank approximation of the Jacobian matrix, further reducing the number of forward simulations (Zhao & Luo, 2021a).

Despite the notable reduction in the number of forward simulations achieved through dimension reduction, the reliance on numerical forward models remains a limitation since these models lack scalability with respect to the resolution of the field. To address this challenge, several methods have been developed to accelerate forward

model computation. The method of temporal moments transforms transient forward models to steady-state, but requires long sampling times and smooth measurements, which may be challenging in the field, especially for monitoring wells with heavy-tailed measurements located in low-permeability regions (Yin & Illman, 2009). Reduced-order surrogate models aim to approximate solutions by constructing a solution subspace, which typically requires extensive forward model computations for each flow scenario, such as pumping at different wells in high-resolution fields (Boyce & Yeh, 2014; Liu et al., 2013). The Upscaled Effective Model approach (UPCIA), introduced by Zhao et al. (2022), aims to linearly upscale high-resolution fields to effective low-resolution ones and conduct forward simulation, thereby reducing the computational time for individual forward simulations. However, it should be noted that this approach may introduce additional approximation errors, particularly when dealing with fields exhibiting significant small-scale heterogeneities and employing covariance functions, such as non-differentiable exponential covariance functions. GRFs characterized by exponential covariance functions display pronounced fluctuations at small scales, typically within a range shorter than the correlation length, leading to visually non-smooth behavior. While the term “non-smoothness” describes visual characteristics rather than a mathematical property, it's important to acknowledge that estimating high-dimensional, non-smooth exponential GRFs poses challenges. This is because common upscaling methods essentially average regions within a spatial window, akin to smoothing in image processing, which might compromise the original non-smoothness and variance of the fields. Conversely, linear interpolation methods introduce smoothing effects and lose fidelity during downscaling, warranting further exploration of the approximation errors, particularly for fields with substantial variance and non-smooth features (Zhao et al., 2022).

In recent years, deep learning models have been utilized to develop surrogate models for groundwater inverse modeling. Physics-informed neural network (PINN) is an example that trains deep neural networks under physical constraints, which are mathematically represented by partial derivative equations (PDEs) (Raissi et al., 2017). An advantage of using PINN to estimate the spatial distribution of parameters is that automatic differentiation (autograd) can be used to evaluate derivatives on the domain of the function pointwise, eliminating non-scalable, matrix-wise computation in numerical models (Guo et al., 2023; Li & Tartakovsky, 2022; Tartakovsky et al., 2020; Wang et al., 2021). However, PINN is an instance-wise model that needs to be retrained when parameters change. In addition, PINN shows obvious model error when the approximated spatial distribution contains small-scale nonsmoothness, such as GRF with an exponential covariance, which may require transmissivity measurements as reference data, and the amount of data may be substantial. Another approach involves training a neural network as a surrogate for the forward model and combining it with traditional gradient-based inverse modeling methods. However, this approach demands high accuracy and generality in the surrogate model, as gradients are not directly learned by the neural network (He et al., 2020; He & Tartakovsky, 2021; Xu et al., 2021; Zhang et al., 2022; Zong et al., 2023). For example, some deep neural networks have been used to learn the transient dynamics in groundwater flow rather than solving exact parameterized PDEs (Mo et al., 2019; Tripathy & Bilonis, 2018; Zhou et al., 2022; Zhu & Zabarar, 2018).

In this study, we explore the potential of using the Fourier neural operator (FNO) as a surrogate forward model in inverse modeling of HT. FNO is effective in learning solution operators of parametric PDEs. In the PDE, the parameters and states can be continuous functions, and FNO transforms the functions through Fourier transformation and directly parameterizes the integral kernel in Fourier space (Li et al., 2021a, 2022). In groundwater modeling, we consider GRFs and their corresponding hydraulic response fields as spatial functions. These entities are interrelated within a specified groundwater PDE. We train a FNO model using GRFs and hydraulic response fields as inputs and outputs, aiming to learn the relationship between them and to approximate the solution operator of the groundwater PDE. Furthermore, our study presents a novel approach that combines FNO with RGA to efficiently inverse model the hydraulic conductivity fields characterized by significant heterogeneity. These fields are essentially realizations of GRFs with non-differentiable exponential covariance functions, whose inverse modeling is recognized as a challenging task. The RGA method incorporates Principal Component Analysis (PCA) and automatically captures the non-smoothness of exponential GRFs in principal components. Besides, the well-trained FNO is an efficient forward model surrogate and can provide fast gradient back-propagation through neural networks. Additionally, the accuracy of well-trained FNO is invariant to resolutions of input, enabling it to be trained using low-resolution data and used for high-resolution forward simulations so that the training time can be shortened leading to a competitive model runtime (Li et al., 2021a). The RGA and FNO combination can make accurate estimation of the low-dimensional parameters and high-fidelity recovery of the exponential GRF.

This note is organized as follows. First, we introduce parametric PDE in groundwater with pumping tests, the architecture of FNO, and the RGA method with dimension reduction. Next, we explain our numerical experiments, which include implementing FNO as a surrogate forward model for steady-state and transient pumping tests and integrating the surrogate model with RGA as an inverse model, referred to as RGA-FNO. We then present the results of these implementations. Additionally, we compare the RGA-FNO inverse model with a PINN inverse model. Prior to concluding, we undertake an evaluation of the RGA-FNO model's performance by manipulating the variances of the GRFs and the quantity of hydraulic head measurements. Finally, we conclude with a discussion of the results.

## 2. Models and Method

### 2.1. Parametric PDE of Groundwater Flow With Pumping

The general governing equation for groundwater flow in saturated porous media with source/sink terms is given by:

$$S_s \frac{\partial h}{\partial t} = -\nabla \cdot [K(\mathbf{x})\nabla h] + Q \quad (1)$$

where  $S_s$  ( $L^{-1}$ ) is the specific storage,  $t$  (T) is time,  $h$  (L) is hydraulic head,  $Q$  ( $T^{-1}$ ) is the water accumulation/reduction rate per volume,  $K$  ( $LT^{-1}$ ) is the hydraulic conductivity,  $\mathbf{x}$  is a vector of spatial coordinates. Assuming that the aquifer is two-dimensional, isotropic, and has constant saturated thickness and specific storage,  $\mathbf{x}$  can be represented by  $[x, y]$  and following the Dupuit-Forchheimer assumption of horizontal flow, the depth-integrated groundwater flow is:

$$S \frac{\partial h}{\partial t} = -\nabla \cdot [T(x, y)\nabla h] + W \quad (2)$$

where  $S$  (–) is the storativity,  $S = S_s b$ ;  $T$  ( $L^2 T^{-1}$ ) is the transmissivity,  $T = Kb$ ;  $W$  ( $LT^{-1}$ ) is the water accumulation/reduction rate per area,  $W = Qb$ ; and  $b$  (L) is the saturated thickness. The two-dimensional steady-state groundwater flow with pumping is given by:

$$0 = -\nabla \cdot [T(x, y)\nabla h] + W \quad (3)$$

The domain boundaries are subject to Neumann ( $\Gamma_N$ ) or Dirichlet ( $\Gamma_D$ ) boundary conditions (BCs) as:

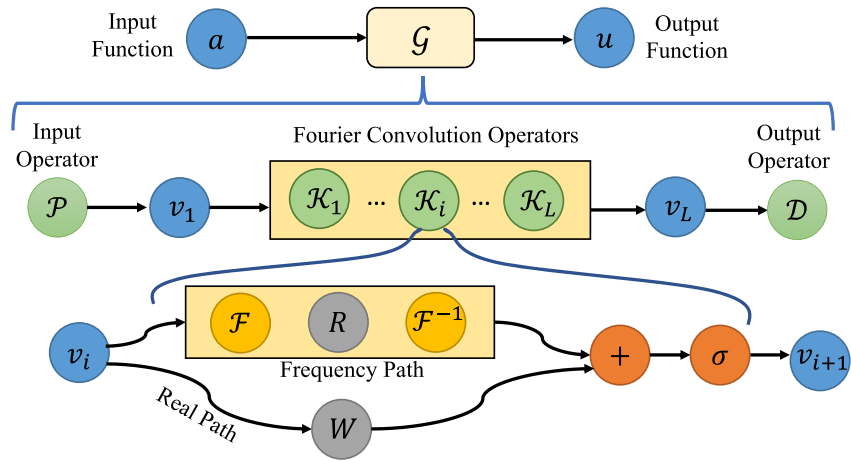
$$\mathbf{n} \cdot \nabla h(x, y) = q_N, (x, y) \in \Gamma_N \quad (4)$$

$$h(x, y) = h_D, (x, y) \in \Gamma_D \quad (5)$$

where  $\mathbf{n}$  is a unit directional vector used to denote flux direction in Neumann BC,  $q_N$  is the flux rate on Neumann boundary,  $h_D$  (L) is the hydraulic head on Dirichlet boundary. If  $T(x, y)$  is fully known, it can be solved numerically for given BCs and pumping schedules. This is known as the forward problem. A forward model surrogate is expected to take the transmissivity field  $T(x, y)$  as input and predict hydraulic head fields  $h(x, y)$  as output.

### 2.2. Fourier Neural Operator (FNO)

Fourier neural operator (FNO) is a neural network-based surrogate model proposed to approximate the solution operator of a parametric PDE (Li et al., 2021a). For example,  $\mathcal{T}(u, a) = 0$  is a parametric PDE, where  $a(x) \in \mathbb{R}^{d_a}$  and  $u(x) \in \mathbb{R}^{d_u}$  represent parameter and state functions in a real space  $x \in \mathbb{R}^{d_x}$ , and the solution operator  $\mathcal{G}$  maps  $a(x) \in \mathbb{R}^{d_a}$  to  $u(x) \in \mathbb{R}^{d_u}$ , where  $u(x) = \mathcal{G}(a)(x)$ . The basic FNO structure is illustrated in Figure 1, it is composed of an input operator  $\mathcal{P}$ ,  $L$  stacked Fourier convolution operators  $\mathcal{K}_i$ , and an output operator  $\mathcal{D}$ . During the forward pass, a batch of input functions  $a(x)$  is structured as a 4D tensor, akin to a batch of images in computer vision. The first dimension corresponds to the batch size, the second dimension represents the channel dimension, which in this case is one since we are dealing with hydraulic transmissivity as the sole parameter. It's important to note that



**Figure 1.** Flowchart of the Fourier neural operator (FNO) approach. At the top of the figure,  $a$  is the input function,  $\mathcal{G}$  is the FNO, and  $u$  is the output function. In the middle of the figure,  $\mathcal{G}$  is expanded to reveal its components and the forward pass. At the beginning, the input operator  $\mathcal{P}$  lifts the input function to a high-dimensional space as  $v_1$ . Next,  $v_1$  is processed by stacked Fourier Convolution operators  $\mathcal{K}$  and transformed to  $v_L$ . Finally,  $v_L$  is transformed by the output operator  $\mathcal{D}$  to produce the output function  $u$ . At the bottom of the figure, a Fourier Convolution operator  $\mathcal{K}_i$  is expanded. It has a frequency path, which conducts channel-wise Fourier transformation  $\mathcal{F}$ , convolution on Fourier coefficients  $\mathcal{R}$ , and inverse Fourier convolution  $\mathcal{F}^{-1}$ ; and a real path, which conducts convolution  $W$  in the real domain. The results from the two paths are added and activated by a nonlinear activation function  $\sigma$ .

the term “channel” here refers to the second dimension of the 4D tensor, much like the red, green, and blue channels in a colored image, and doesn't refer to the physical channels in hydraulic transmissivity realizations. The third and fourth dimensions correspond to the height and width, signifying the resolutions of the discretized domain. The input operator  $\mathcal{P}$  lifts the single-channel input function  $a(x)$  to a multi-channel function  $v_1(x) = (\mathcal{P}a)(x)$ . Then,  $v_1(x)$  passes through  $L$  stacked Fourier convolution operators, where each operator has two paths: a frequency path denoted as  $K$ , which performs channel-wise Fourier transformation  $\mathcal{F}$ , convolution on Fourier coefficients  $\mathcal{R}$ , and inverse Fourier transformation  $\mathcal{F}^{-1}$ ; and a real path that performs a convolution  $W$  in the real domain. The outputs of the two paths are then added and activated by a nonlinear activation function  $\sigma$ . Finally, the output operator  $\mathcal{D}$  maps  $v_L(x)$  to the output space as  $u(x) = (\mathcal{D}v_L)(x)$ , which has the same dimension as the output data. The forward pass can be expressed by:

$$u(x) = \mathcal{G}(a)(x) = (\mathcal{D}v_L)(x), x \in \mathbb{R}^{d_x} \quad (6)$$

$$\mathcal{G}(a) = (\mathcal{D} \circ \mathcal{K}_L \dots \circ \mathcal{K}_1 \circ \mathcal{P}) a \quad (7)$$

$$v_{i+1}(x) = (\mathcal{K}_i v_i)(x) = \sigma_i(W_i v_i(x) + (K_i v_i)(x)) \quad (8)$$

$$(K_i v_i)(x) = \mathcal{F}^{-1}(R_i \cdot \mathcal{F}(v_i))(x) \quad (9)$$

The transformation operators in FNO can be parameterized by linear transformations or convolutional neural networks. Additionally, if the discretized function is periodic and projected onto a unit domain of regular mesh, we can implement Fast Fourier transformation (FFT) in the Fourier operator frequency path and truncate the obtained modes to a constant number, enabling FNO to have quasi-linear complexity corresponding to the dimension of inputs. Since FNO is originally conceptualized in Fourier domain and applicable to continuous functions, numerically, all finite-dimensional approximations  $a(x) \in \mathbb{R}^{d_a}$  and  $u(x) \in \mathbb{R}^{d_u}$  originate from the same continuous function and share a common set of parameters. Therefore, FNO is resolution-invariant and can be trained with low-resolution data ( $\mathbb{R}^{d_a^{low}}$  and  $\mathbb{R}^{d_u^{low}}$ ) and evaluated with high-resolution data ( $\mathbb{R}^{d_a^{high}}$  and  $\mathbb{R}^{d_u^{high}}$ ) without performance deterioration (Kovachki et al., 2020; Li et al., 2020; Nelsen & Stuart, 2021).

FNO is a data-driven method that learns from data by minimizing the match loss ( $\mathcal{L}_{data}$ ), which can be evaluated by integrating over the domain and aggregating over independent inputs. Numerically, the integral is evaluated at discrete points where the reference data is sampled, as shown by:

$$\mathcal{L}_{\text{data}} = \frac{1}{N} \sum_{j=1}^N \int |u_j(x) - \mathcal{G}(a_j)(x)|^2 dx \approx \frac{1}{ND} \sum_{j=1}^N \sum_{i=1}^D |u_j(x_i) - \mathcal{G}(a_j)(x_i)|^2 \quad (10)$$

where  $u_j$  represents the value of the real output function at grid point  $j$  within the domain and serves as the reference data point for the FNO to learn from;  $N$  denotes the batch size, indicating the number of pairs of hydraulic transmissivity and hydraulic head fields in each batch;  $D$  signifies the number of discrete points utilized for evaluating the integral. To ensure that FNO accurately matches the output function across the entire domain, the integral must encompass the full domain. In this context,  $D$  is equal to the resolution of the output function, which has been denoted as  $\mathbb{R}^{d_u^{\text{low}}}$ .

In hydraulic tomography, multiple pumping tests are conducted at different locations, and each of them can be formulated as an individual forward problem, as described in Section 2.1. FNO can learn from all these forward problems simultaneously by using multi-channel output. For example, when simulating steady-state HT, the input of FNO is the transmissivity parameter field, and the output of FNO has multiple channels, with each channel approximating a steady-state hydraulic head field corresponding to a specific pumping test. The number of channels matches the number of pumping tests, making FNO an all-in-one surrogate for steady-state HT forward modeling. This method can be easily generalized to transient pumping tests, where each output channel approximates the hydraulic head field at a specific time step for a particular pumping test. In this case, the number of channels equals the number of pumping tests multiplied by the number of monitoring time steps, which is expected to be significantly larger than the number of channels for steady-state pumping tests.

### 2.3. RGA Integrated With FNO (RGA-FNO)

The HT inverse problem, which involves estimating a hydraulic transmissivity field  $T \in \mathbb{R}^{m \times 1}$ , is addressed using the RGA framework based on Bayesian analysis and principal component dimension reduction (Zhao & Luo, 2020). The low-dimensional representation of  $\ln T$  is achieved through an exponential activation function and a linear transformation  $\mathbf{A}$  with given structural parameters as:

$$T = \mathbf{A}_{\boldsymbol{\mu}, \mathbf{Z}}(\boldsymbol{\alpha}) = e^{\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\alpha}} \quad (11)$$

where the projection matrix  $\mathbf{Z} \in \mathbb{R}^{m \times k}$  is a linear combination of  $k$  scaled dominant principal components,  $\boldsymbol{\mu} \in \mathbb{R}^{m \times 1}$  is the mean parameter value which may be non-stationary and drifting in the domain, and  $\boldsymbol{\alpha} \in \mathbb{R}^{k \times 1}$  is independent, identically distributed latent variables used as the low-dimensional representation of the realization. The relation between  $\boldsymbol{\alpha}$  and observations  $\mathbf{h}$  through forward model  $\mathbf{f}(\cdot)$  can be expressed as:

$$\mathbf{h} = \mathbf{f}(T) + \boldsymbol{\epsilon} = \mathbf{f}(e^{\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\alpha}}) + \boldsymbol{\epsilon} \quad (12)$$

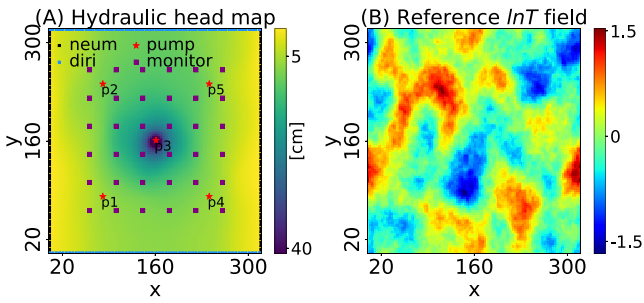
where  $\mathbf{h} \in \mathbb{R}^{n \times 1}$  corresponds to the observed hydraulic heads during pumping tests, the vector  $\boldsymbol{\epsilon} \in \mathbb{R}^{n \times 1}$  represents the measurement errors, characterized by dimension  $n$  and covariance matrix  $(\mathbf{R})$ . We operate under the assumption that these measurement errors are independent and identically distributed (i.i.d.) as Gaussian noise. Consequently,  $\mathbf{R}$  is proportional to the identity matrix.

To estimate the unknown  $\boldsymbol{\alpha}$ , RGA formulates its Bayesian posterior distribution by considering the prior normal distribution for the coefficients  $\boldsymbol{\alpha}$  and the likelihood of reference data:

$$p''(\mathbf{a}) \propto \exp\left(-\frac{1}{2}(\mathbf{h} - \mathbf{f}(e^{\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\alpha}}))^T \mathbf{R}^{-1} (\mathbf{h} - \mathbf{f}(e^{\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\alpha}})) - \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{\alpha}\right) \quad (13)$$

The negative log-likelihood of the posterior distribution is used as a loss function to be minimized:

$$\min_{\boldsymbol{\alpha}} l_f(\boldsymbol{\alpha}) = \frac{1}{2}(\mathbf{h} - \mathbf{f}(e^{\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\alpha}}))^T \mathbf{R}^{-1} (\mathbf{h} - \mathbf{f}(e^{\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\alpha}})) + \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{\alpha} \quad (14)$$



**Figure 2.** Numerical experiment setup for the HT experiment. (a) The well network consisting of 5 pumping wells (p1–p5) represented by red stars, and 36 monitoring wells located at purple dots. Dirichlet (diri) boundary cells are denoted by blue squares, and Neumann boundary (neum) cells are denoted by black squares; (b) The reference  $\ln T$  field used in the pumping test simulation.

The minimizer of Equation 13 is the maximum a posteriori (MAP) estimate, which can be found through a gradient-based Newton or quasi-Newton method. However, computing the Jacobian matrix of the loss function with respect to  $\alpha$  at each iteration can be computationally expensive using traditional finite-difference methods that iteratively perturb the components of  $\alpha$ .

A well-trained FNO is used as a surrogate for the numerical forward model in RGA:

$$\mathbf{f}(e^{\mu+Z\alpha}) \approx \text{FNO}(e^{\mu+Z\alpha}) \quad (15)$$

Following the idea of Bayesian neural network (Hernandez-Lobato & Adams, 2015), at each iteration of the inversion optimization, instead of optimizing an individual realization, we sample  $M$  realizations of  $\alpha$  from the normal distribution, and optimize the mean  $\mu_\alpha$  and variance  $\mathbf{C}_\alpha$  of the distribution. The coefficients in FNO are frozen during inversion optimization, and the loss function  $l(\alpha)$  is the mean squared error (MSE), evaluating the data match to the observations as:

$$\min_{\mu_\alpha, \mathbf{C}_\alpha} l(\alpha; \mathbf{h}, \text{FNO}) = \frac{1}{M} \sum_{i=1}^M |\mathbf{h} - \text{FNO}(e^{\mu+Z\alpha_i})|^2, \alpha \sim \text{MVN}(\mu_\alpha, \mathbf{C}_\alpha) \quad (16)$$

By implementing the FNO as the surrogate forward model in the RGA, a method we refer to as RGA-FNO, we establish a direct link between the parameter  $\alpha$  and the loss function within a unified deep learning computational graph. The integration facilitates the efficient computation of the gradient of the loss function with respect to  $\alpha$  via automatic differentiation (autograd). This approach for computing the gradient is more efficient compared to traditional explicit methods, which rely on numerically approximating the Jacobian by iteratively perturbing the components of  $\alpha$  using finite differences.

### 3. Numerical Experiments

#### 3.1. Experimental Domain and Synthetic HT Data Set

In our experiment, we model the transmissivity field as log-normally distributed. Specifically,  $\ln T$  is a realization derived from a GRF with an exponential spatial covariance function, incorporating the 50 highest-ranked principal components. This configuration of the  $\ln T$  field results in more pronounced non-smoothness, thereby rendering the inverse estimation process more challenging. In contrast, previous PINN studies, (e.g., Guo et al., 2023; Tartakovsky et al., 2020; Wang et al., 2021; Xu et al., 2021), typically employ GRFs with Gaussian covariance functions or limit themselves to no more than 20 principal components. This leads to smoother  $\ln T$  fields, making the inverse estimations comparatively less complex. The experimental domain is square with a resolution of  $128 \times 128$  and a scale of  $320 \text{ m} \times 320 \text{ m}$ . The discretization is piecewise constant and grids are square and equal-spaced, and the bottom-left corner is coordinated at  $(0 \text{ m}, 0 \text{ m})$ . Figure 2a shows the hydraulic head map and Figure 2b shows a  $\ln T$  realization. We implement HT in a domain with impermeable top and bottom boundaries (Neumann boundary) and constant head left and right boundaries (Dirichlet boundary).  $S_s$  is uniform and constant across the domain, with a value of  $0.0001 \text{ m}^{-1}$ . The geostatistical and hydrogeological parameters are listed in Table 1.

We conducted two types of HT experiments: one steady-state and the other transient. For the steady-state HT setup, we used five pumping wells (Table 2). Each pumping test involved extracting water from one of these wells at a constant rate of  $3.6 \text{ m}^3/\text{hr}$  ( $0.001 \text{ m}^3/\text{s}$ ). As a result, a transmissivity field was paired with five corresponding hydraulic head fields. In the transient HT setup, a single pumping test was conducted at well p3 (Table 2) at the same rate. The initial hydraulic head across the domain was set to 0 m. Hydraulic head observations were collected 10 times within an hour, at intervals of 0.1 hr. After 1 hr, the hydraulic head field approached a near steady state. Consequently, each input transmissivity field was paired with 10 hydraulic head fields in this experiment.

**Table 1**  
*Hydrogeological and Geostatistical Parameters for the Hydraulic Tomography Experiment*

Parameter	Values
Domain size, $L_x \times L_y$	320 m $\times$ 320 m
Grid spacing, $\Delta x \times \Delta y$	2.5 m $\times$ 2.5 m
Spatial resolution, $n_x \times n_y$	128 $\times$ 128
Thickness, $b$	1 m
Specific storage, $S_s$	0.0001 m <sup>-1</sup>
Natural logarithm of transmissivity, $\ln T$	
Mean	0 m <sup>2</sup> /s
Variance of $\ln T$ , $\sigma_{\ln T}^2$	0.25 m <sup>4</sup> /s <sup>2</sup>
Correlation length, $\lambda_x \times \lambda_y$	64 m $\times$ 48 m
Left Boundary	$h = 0$ m
Right Boundary	$h = 0$ m
Pumping Rate	0.001 m <sup>3</sup> /s

For each of the two experiments, we conduct the corresponding HT survey with 1,000 transmissivity realizations for FNO training and validation. The hydraulic head fields were solved by a self-developed Galerkin finite element method (FEM) solver, with a tolerance set at  $1 \times 10^{-8}$ . All hydraulic head values were negative. Although the forward code was developed using MATLAB 2022a, any compatible solver can be used to replicate the results. The hardware used for these computations was Intel(R) Core(TM) i9-10900 CPU 2.81 GHz.

### 3.2. Computation Implementation

#### 3.2.1. FNO Training

The input of FNO for modeling is a 4D tensor with a channel dimension of one, representing the 2D transmissivity field discretized on the  $128 \times 128$  domain. The input operator  $\mathcal{P}$  is a local, pointwise transformation parameterized by a linear layer with 128 output dimensions. The Fourier section comprises four identically structured Fourier convolution operators ( $L = 4$ ), with each frequency path truncating the 20 highest and lowest frequent modes ( $mode = 20$ ), and the convolution layers in either Fourier or real space ( $R$  and  $W$ ) are squared with  $width = 64$  and  $stride = 1$ . The nonlinear Gaussian Error Linear Unit (GELU) activation function  $\sigma$  is used after the first three Fourier operators.

Padding with  $ratio = 2$  is added and removed before and after the section. The output operator  $\mathcal{D}$  comprises three local, pointwise transformations with output dimensions of 128, 64, 5, and the GELU activation function  $\sigma$  is used after the first two layers. The channel dimension of the output is the same as the number of simulated hydraulic head fields. For steady-state FNO, the channel dimension is five where each channel corresponds to a steady-state hydraulic head field under a specific pumping test. For transient FNO, the channel dimension is 20 where each channel corresponds to a hydraulic head field at a specific time step in the single pumping test.

The numbers of HT simulations used to train and validate FNO are 900 and 100, respectively. The spatial dimensions of the input and output are upscaled to a  $64 \times 64$  resolution and pointwise normalized. During the training, the number of iterations is 500, at each iteration, FNO is trained by one batch of data, and the batch size is 10. The loss function is MSE, the optimizer is Adam with a learning rate of 0.001 (Kingma & Ba, 2017). The training is conducted on Google Colab, an online cloud computation platform equipped with a Tesla A100 GPU backend (Bisong, 2019). The training processes of steady-state and transient FNO are completed within approximately 825 and 830 s.

#### 3.2.2. RGA-FNO Inverse Modeling

During inverse modeling, we collect pointwise hydraulic head data to estimate  $\alpha$  of the underlying transmissivity field. The data is collected from 36 monitoring wells in each hydraulic head field (points in domain). These wells are systematically distributed in a  $6 \times 6$  square pattern. The locations of wells at the four corners are listed in Table 2, with a 39.6-m separation between adjacent wells (Figure 2a). For steady-state HT, the hydraulic head measurements are taken from the 36 monitoring wells across five pumping tests, resulting in a total of 180 measurements. In transient HT, the hydraulic head measurements from the 36 monitoring wells are taken at each time step, yielding a total of 360 measurements. To account for real-world variability, these hydraulic head measurements are subject to random noise with a variance of 5% relative to the true values.

During optimization, the gradient of the loss function with respect to  $\alpha$  is evaluated by autograd, the optimizer is Stochastic Gradient Descent (SGD) optimizer, the learning rate is fixed at 1 during the entire optimization, and the optimization undergoes 500 iterations, early stop or tolerance is not used. The optimization is implemented on the same Google Colab with a Tesla A100 GPU at backend. The runtime for both the steady-state and transient

**Table 2**  
*Coordinates of Pumping Wells and Corner Monitoring Wells*

Well type	Well location (index)	Coordinates (x, y)
Pumping	Lower-left (p1)	(80 m, 80 m)
	Upper-left (p2)	(80 m, 240 m)
	Center (p3)	(160 m, 160 m)
	Lower-right (p4)	(240 m, 80 m)
	Upper-right (p5)	(240 m, 240 m)
Monitoring	Lower-left	(61 m, 61 m)
	Upper-left	(61 m, 259 m)
	Lower-right	(259 m, 61 m)
	Upper-right	(259 m, 259 m)

experiments is approximately 12.5 s. This model structure and configurations of FNO training and RGA-FNO optimization are found through trial and error.

### 3.3. Quantitative Metrics

For the forward model surrogate, the prediction error is quantified by the forward relative residual metric, which is formulated as:

$$\epsilon_h = \frac{\|h_{\text{TRUE}} - h_{\text{FNO}}\|_2^2}{\|h_{\text{TRUE}}\|_2^2}, \forall (x, y) \in \Omega \quad (17)$$

where  $h_{\text{TRUE}}$  and  $h_{\text{FNO}}$  represent the true and FNO-approximated hydraulic heads, respectively. Similarly, the inverse relative residual  $\epsilon_T$  is evaluated by:

$$\epsilon_T = \frac{\|\hat{T}(x, y) - T(x, y)\|_2^2}{\|T(x, y)\|_2^2}, \forall (x, y) \in \Omega \quad (18)$$

where  $T(x, y)$  and  $\hat{T}(x, y)$  are the true and estimated transmissivity vectors. Moreover, the inverse result is assessed by map accuracy, which is defined as the percentage of the grid that has transmissivity correctly inverted (Kang et al., 2017). The condition for correct inversion is that pointwise estimation error  $\epsilon(x, y)$  is less than a predefined threshold, set to 10% in this study:

$$\epsilon(x, y) = \frac{|\hat{T}(x, y) - T(x, y)|}{T^{\max} - T^{\min}}, \forall (x, y) \in \Omega \quad (19)$$

## 4. Results and Discussion

### 4.1. FNO for Steady-State HT

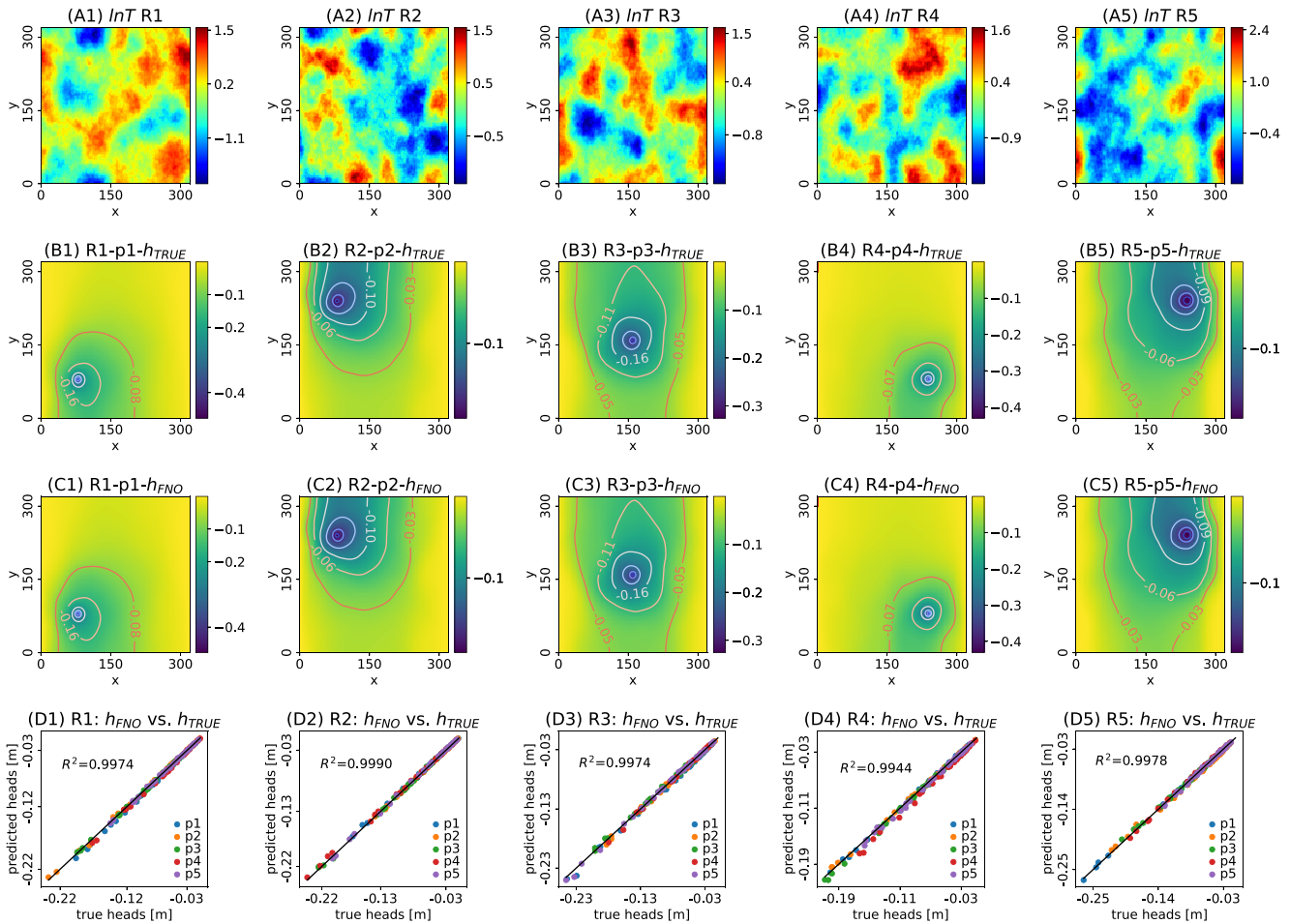
#### 4.1.1. Evaluation of Steady-State Forward Model Surrogate

To validate the FNO as an effective surrogate forward model of steady-state HT, we conducted simulations of the full steady-state HT surveys (p1–p5) using 100 different transmissivity realizations. Each of these transmissivity fields is set at a resolution of  $128 \times 128$ . These fields serve as inputs for the FNO, which is then tasked with predicting the corresponding steady-state hydraulic heads for the five pumping tests (p1–p5). The accuracy of these predictions is assessed by comparing them with reference data derived from FEM simulations. Visualized in Figure 3 are five illustrative validation cases, where each column represents a set of input, output, reference, and comparison. For example, Figure 3a1 presents a  $lnT$  field named R1, Figures 3b1 and 3c1 depict the hydraulic heads simulated by FEM (reference,  $h_{\text{TRUE}}$ ) and FNO (prediction,  $h_{\text{FNO}}$ ), respectively, under a pumping test at p1 in the R1 field. Figure 3d1 plots sampled points from  $h_{\text{FNO}}$  and  $h_{\text{TRUE}}$  across all five pumping tests in R1, showcasing their strong alignment with a  $R^2$  score of 0.9974, indicating a well-simulated steady-state HT test by FNO.

Similarly, Figures 3a2–3a5 present four other randomly chosen realizations (R2–R5) from the validation set. Corresponding, Figures 3b2–3b5 show the reference solutions from FEM for distinct realizations R2–R5 and pumping tests p2–p5, Figures 3c2–3c5 display FNO-predicted hydraulic heads. Finally, Figures 3d2–3d5 compare the sampled points. The agreement between predicted and reference hydraulic maps and the alignment of sampled points ( $R^2$  scores all above 0.99) demonstrate FNO's robustness and generality, confirming its accuracy in this validation exercise.

Quantitative evaluation is performed independently on each pumping test using the relative error as a metric. The means and standard deviations of the relative errors for pumping tests p1–p5 are  $3.43\% \pm 1.60\%$ ,  $3.75\% \pm 1.87\%$ ,  $2.75\% \pm 1.06\%$ ,  $3.59\% \pm 1.54\%$ ,  $3.28\% \pm 1.17\%$ , respectively. The results show that FNO successfully learns the solution operator of the groundwater flow equation and can generalize to unseen transmissivity fields generated from the same GRF, producing accurate predictions of hydraulic heads for the configured HT. Furthermore, FNO is resolution-invariant within certain range considering that it is trained with low-resolution data ( $64 \times 64$ ) and



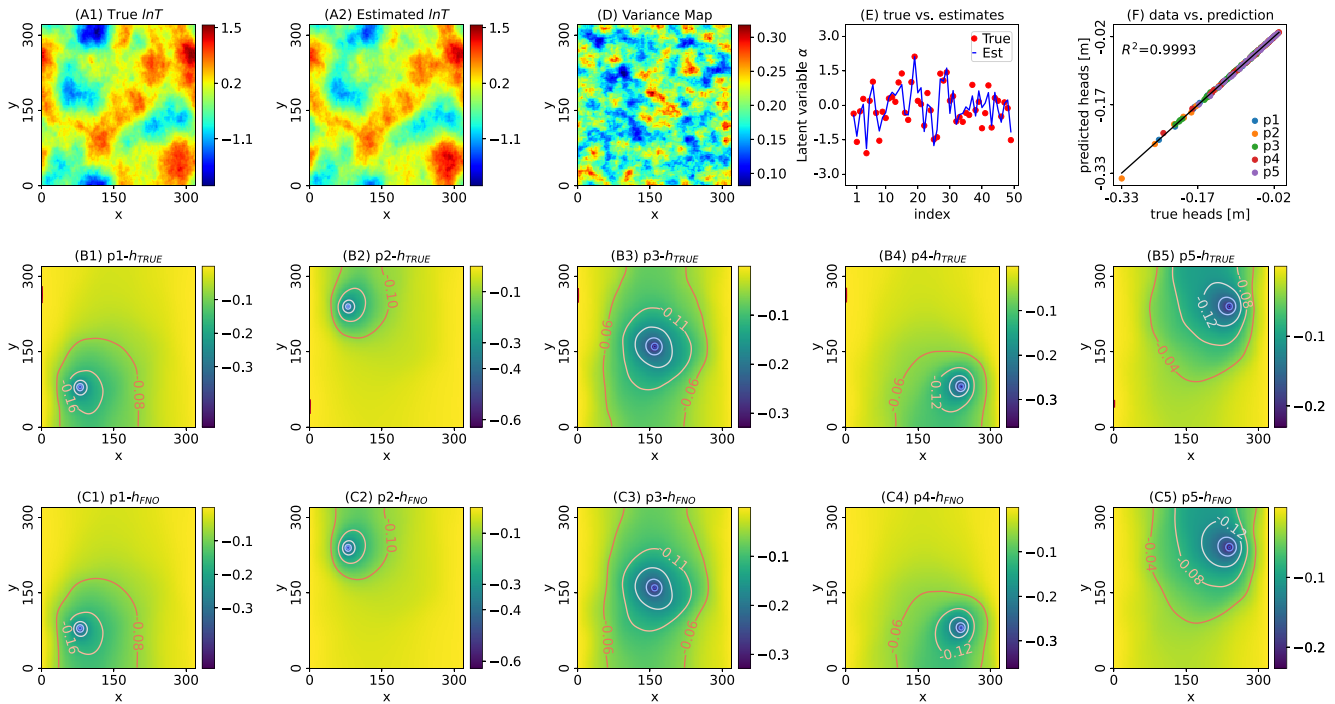


**Figure 3.** Steady-state FNO surrogate forward model results of different realizations and pumping tests. (a1–a5)  $\ln T$  realizations, named as R1–R5; (b1–b5) Reference or true hydraulic heads with respect to different realizations and pumping locations ( $h_{TRUE}$ ), for example, R1-p1- $h_{TRUE}$  means the FEM-solved or true hydraulic heads in pumping test p1 at  $\ln T$  realization R1; (c1–c5) Hydraulic heads predicted by FNO with respect to different realizations and pumping locations ( $h_{FNO}$ ), for example, R1-p1- $h_{FNO}$  means the FNO-predicted hydraulic heads in pumping test p1 at  $\ln T$  realization R1; (d1–d5)  $h_{TRUE}$  versus  $h_{FNO}$  for each realizations at monitoring well locations,  $R^2$  coefficient greater than 0.99 for all five realizations.

used for high-resolution evaluation ( $128 \times 128$ ). We also train FNO with data of lower resolution ( $32 \times 32$ ), but the accuracy of the model is not satisfactory. The resolution-invariant property and limitation of the FNO depend on the model structure and need further investigation, which will not be discussed in this study. The major advantage of FNO is its computational speed. For example, a single forward pass of FNO (a steady-state HT simulation of pumping tests at all five pumping wells) takes only 0.01 s, compared with the HT simulation of FEM model, which takes about 0.48 s, demonstrating the improvement in computational efficiency provided by the FNO. We also train FNO to learn the Darcy flow PDE without sinks, and the results are shown in the Supporting Information S1.

#### 4.1.2. Evaluation of RGA-FNO Inverse Modeling

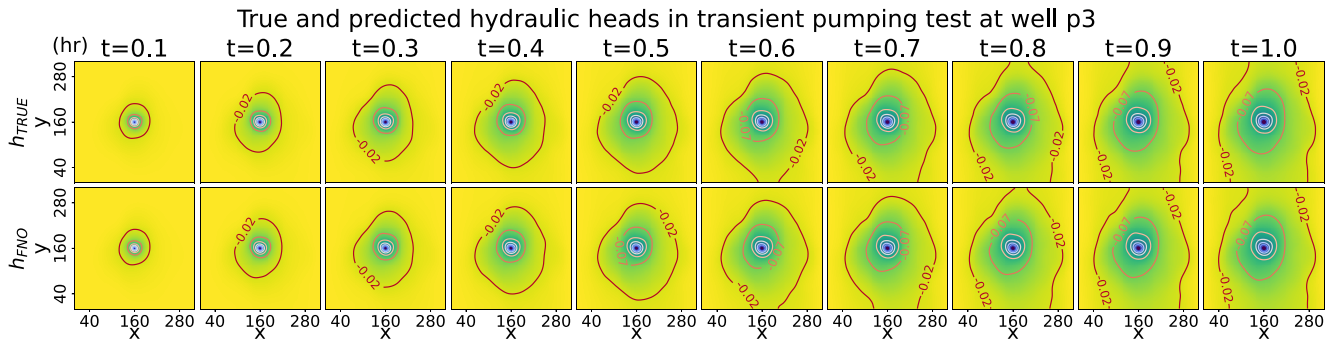
The outcomes of the RGA-FNO inverse modeling utilizing the steady-state FNO forward model surrogate are depicted in Figure 4. Specifically, Figures 4a1 and 4a2 illustrate the  $\ln T$  field alongside the best estimation, marked by a relative error of 15.07% and a map accuracy of 94.78%. In Figures 4b1–4b5, the reference hydraulic heads produced using the FEM forward model and the true transmissivity field are shown, while Figures 4c1–4c5 exhibit the hydraulic head fields for the five pumping tests, predicted by FNO with the best-estimated transmissivity field. The predictions correspond well to the references in both magnitude and contours, with relative errors for pumping tests p1–p5 at 2.47%, 4.37%, 2.04%, 2.58%, and 3.27%. The errors arise from a combination of FNO model inaccuracies and disparities between the true and estimated transmissivity fields. Figure 4d illustrates



**Figure 4.** RGA-FNO inverse results with the trained steady-state FNO surrogate forward model. (a1) True  $\ln T$  realization; (a2) Best estimation with a relative error of 15.07% and map accuracy of 94.78%; (b1)–(b5) Reference hydraulic heads of each pumping test that simulated by FEM with the true transmissivity field ( $h_{TRUE}$ ); (c1–c5) Predicted hydraulic heads of each pumping test that from FNO with the estimated transmissivity field ( $h_{FNO}$ ); (d) Variance of the estimated  $\ln T$ , evaluated by the Monte Carlo method; (e) True latent variables and best estimates (red dots denote true, and the blue line denotes best estimates); (f) Observed (reference data) versus predicted hydraulic heads.

the pointwise estimation variance, assessed using a Monte Carlo technique with 200 samples drawn from the normal distribution with optimized mean and variance, demonstrating low uncertainty in transmissivity estimates. Comparisons between true and best-estimated latent random variables  $\alpha$  are presented in Figure 4e, while Figure 4f displays a cross-plot of observed (reference data) and predicted hydraulic heads, with an  $R^2$  coefficient exceeding 0.99. The MSE at the last iteration of the optimization is 0.003 which can be seen as the error (squared error) of fitting the measurements.

The estimated and actual  $\alpha$  values, represented by blue lines and red dots in Figure 4e, exhibit strong alignment, as do the predicted and observed hydraulic heads. This alignment underscores the effectiveness of autograd-enabled gradient backpropagation in achieving desired convergence ( $\alpha$  alignment) while minimizing the objective function (hydraulic head alignment). The inverse model produced accurate estimations without utilizing transmissivity measurements, and it optimizes  $\alpha$  for 500 iterations with a total computational time of approximately 12.5 s. In comparison, the RGA with an FEM solver (RGA-FEM) optimizes  $\alpha$  for 6 iterations and achieves a map accuracy of 95.31% and a relative error of 19.69%, and the total computational time is about 161 s. RGA-FNO demonstrated comparable accuracy to RGA-FEM but with faster performance. This enhanced efficiency can be attributed to RGA-FNO's employment of automatic differentiation (autograd) for efficiently computing the gradient values of the loss function with respect to latent variables. This process is integrated within the deep learning computational graph, with each iteration requiring only 0.025 s. In contrast, RGA + FEM calculates the Jacobian matrix through finite difference perturbation. Specifically, 50 entities of  $\alpha$  are sequentially perturbed by  $1e-6$  to derive the Jacobian matrix, requiring 50 forward simulations over 24 s. Such a comparison distinctly highlights the efficiency and speed advantages of employing a neural network-based surrogate forward model in inversion processes. It is noteworthy that the RGA-FEM model, as originally proposed by Zhao and Luo (2020), relies solely on the Jacobian matrix for optimization, without incorporating any adjoint-state analysis. They assert that utilizing the Jacobian matrix alone is adequate for achieving satisfactory results. Although implementing adjoint-state analysis in RGA-FEM could potentially enhance optimization, it would require additional forward simulations, leading to an increase in computation time.

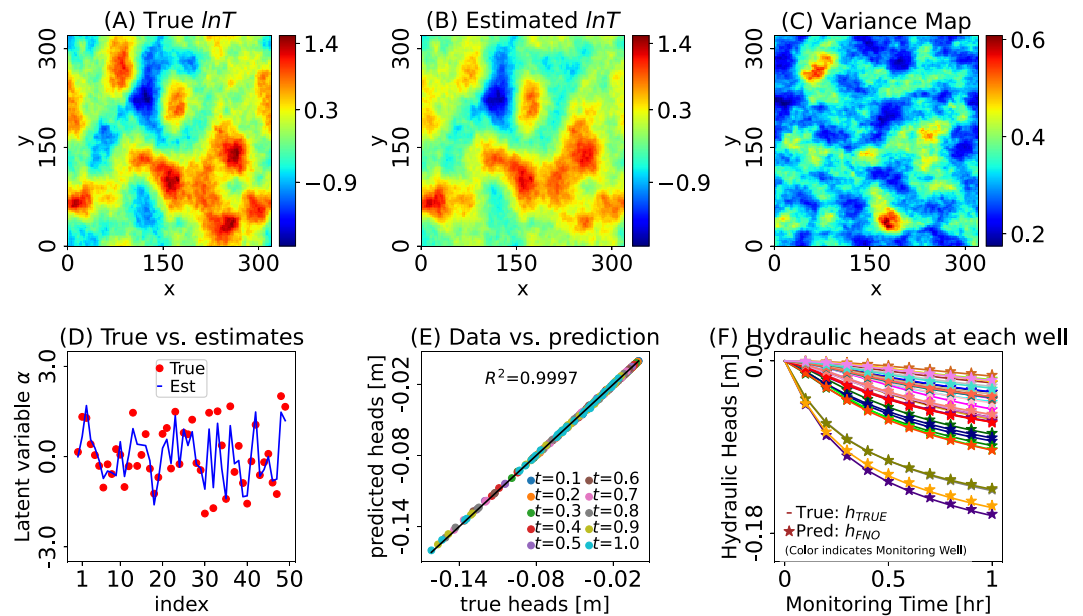


**Figure 5.** Comparison of transient FNO model ( $h_{FNO}$ ) with the numerical simulation ( $h_{TRUE}$ ) for the transient pumping test, hydraulic head distribution in pumping test at well p3 which is located at the center of the domain, first row shows  $h_{TRUE}$  from numerical simulation at  $t = 0.1-1.0$  hr; second row shows  $h_{FNO}$  prediction; each column shows a time step, from left to right,  $t = 0.1, 0.2, \dots, 1.0$  hr. The relative errors between  $h_{FNO}$  and  $h_{TRUE}$ ,  $\epsilon_h^{t=t_i}$ , in time order are 2.20%, 2.11%, 1.99%, 1.87%, 1.78%, 1.69%, 1.64%, 1.60%, 1.56%, 1.54%.

## 4.2. FNO for Transient HT

### 4.2.1. Evaluation of Transient Forward Model Surrogate

Illustrated in Figure 5, the transient FNO's performance is shown. The top row displays true hydraulic heads ( $h_{TRUE}$ ), while the second row exhibits FNO predictions ( $h_{FNO}$ ) for time steps ranging from 0.1 to 1 hr in a pumping test centered around the domain's midpoint (p3). The parameter field is depicted in Figure 6a. The color map and contour lines affirm that the FNO's predictions closely align with the outcomes of numerical simulations at each time step. Additionally, for the time interval  $t = 0.8-1.0$  hr, the contour lines assume a perpendicular orientation with respect to the top and bottom boundaries, consistent with the impermeable Neumann BCs.



**Figure 6.** RGA-FNO inverse results with the trained transient FNO surrogate forward model. (a1) True  $\ln T$  realization; (b) Best estimation with a relative error of 24.15% and map accuracy of 83.55%; (c) Variance of the estimated  $\ln T$  from RGA-FNO, evaluated by the Monte Carlo method; (d) True latent variables and best estimates (red dots denote true, and the blue line denotes best estimates); (e) Observed (reference data) versus predicted hydraulic heads, observed data is corrupted with 5% noise; (f) reference water heads and predicted water heads at monitoring wells in pumping test p3, each color notes a unique monitoring well, solid lines note reference data, it includes the zero-head initial condition at  $t = 0$ , stars note model prediction, it starts from  $t = 0.1$  hr.

Among the 100 validation pairs, the mean and standard deviations of the relative errors in hydraulic heads at each time step are as follows:  $3.89\% \pm 3.40\%$ ,  $3.55\% \pm 2.16\%$ ,  $3.33\% \pm 2.02\%$ ,  $3.17\% \pm 1.91\%$ ,  $3.04\% \pm 1.81\%$ ,  $2.95\% \pm 1.73\%$ ,  $2.87\% \pm 1.67\%$ ,  $2.81\% \pm 1.61\%$ ,  $2.77\% \pm 1.56\%$ ,  $2.73\% \pm 1.53\%$ . It is noteworthy that the relative errors tend to be higher and more variable at the earlier time steps. This trend can be attributed to the smaller magnitude of hydraulic heads during the initial time steps, stemming from the zero-head initial condition. Conversely, as time progresses and reaches a point beyond 0.8 hr, the hydraulic heads attain a steady state, leading to more stable relative errors.

#### 4.2.2. Evaluation of RGA-FNO Inverse Modeling

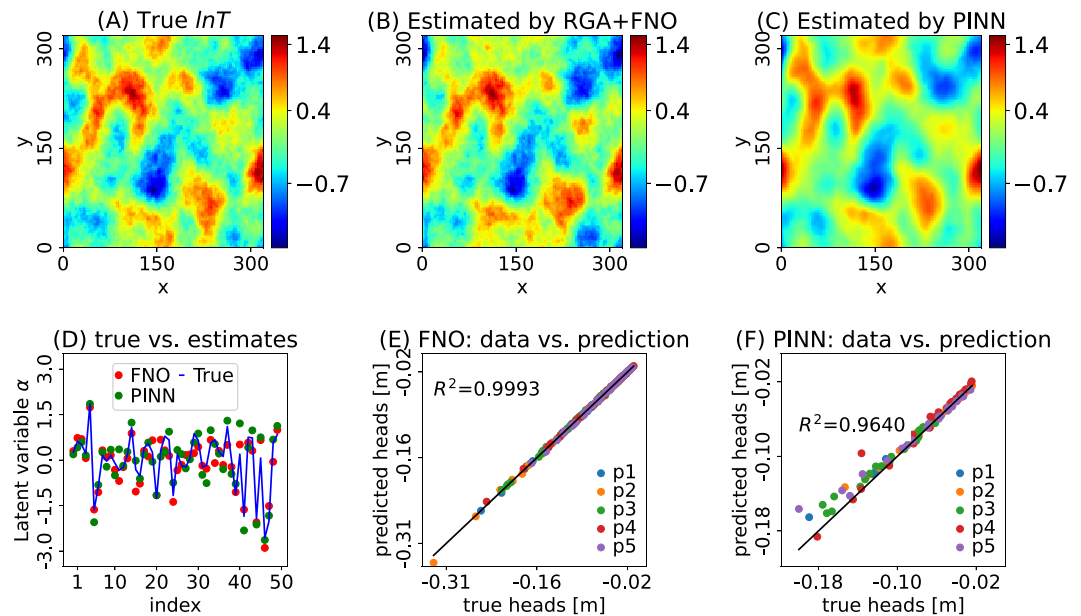
Figure 6 displays the outcomes of the RGA-FNO inverse modeling technique employing the surrogate forward model of transient FNO. Figures 6a and 6b display the true  $\ln T$  field and the corresponding best estimation, yielding a relative error  $\epsilon_T$  of 24.15% and a map accuracy of 83.55%. Comparatively, the RGA-FNO approach, despite utilizing more reference data in the form of 360 hydraulic head measurements from a single-well transient pumping test with 10 time steps, is less accurate than the steady-state case. However, the inverse estimation remains representative, effectively capturing the primary high and low transmissivity regions within the domain. Figure 6c offers insight into the pointwise variance of the estimation, computed based on 200 samples drawn from the optimized mean and variance of the Multivariate Normal distribution, which shows larger uncertainty than the steady-state case at most of the grids. Figure 6d visually contrasts the true and best estimated latent random variables  $\alpha$ . Utilizing the estimated  $\ln T$  field and transient FNO model, we approximate the hydraulic heads. The relative errors at each time step ( $t = 0.1, 0.2, \dots, 1.0$  hr) are computed as 10.54%, 9.41%, 8.30%, 7.51%, 6.93%, 6.47%, 6.10%, 5.85%, 5.62%, and 5.43%, respectively. Notably, these errors are higher than those associated with the transient FNO, primarily attributed to variations between the true and estimated transmissivity fields. Figure 6e displays observed (reference data) and predicted hydraulic heads at each time step, with distinctive colors marking different time steps. The regression analysis yields an  $R^2$  coefficient surpassing 0.99. Figure 6f provides a visualization of reference and predicted hydraulic heads at monitoring wells in pumping test p3. Each color denotes a distinct monitoring well. Solid lines represent reference data, incorporating the zero-head initial condition at  $t = 0$ , while stars depict model predictions, commencing from  $t = 0.1$  hr. The plot confirms that the transient FNO predictions closely align with reference data.

Despite its impressive performance, the transient FNO does possess inherent limitations. Firstly, it lacks autoregressive modeling capability. Transient pumping tests inherently involve an autoregressive structure, where subsequent time steps rely on previous ones. However, the transient FNO employs individual channels to represent time steps, lacking explicit autoregressive dependencies among these channels. Secondly, the transient FNO exhibits reduced accuracy, particularly in the prediction of hydraulic heads during early time steps. This limitation has been elucidated in Section 4.2.1.

The complexity of the FNO model is contingent upon the number of hydraulic head fields derived from distinct pumping tests and varying time steps within a single test. This complexity translates to an augmented number of parameters within the model. If hydraulic heads are monitored across multiple time steps, the output operator must be expanded to incorporate additional output channels. This complexity amplifies in scenarios involving multiple transient pumping tests, necessitating a doubling of the output's channel dimension. In light of these considerations, the recommendation leans toward employing the steady-state FNO. This choice stems from the transient FNO's superior accuracy in predicting steady-state hydraulic heads. Furthermore, the utilization of HT pumping tests conducted at different locations serves as a more effective strategy for estimating local transmissivity.

#### 4.3. RGA-FNO Versus PINN

In this section, we compare RGA-FNO with PINN for inverse modeling of a steady-state HT. The setup of HT and RGA-FNO is the same as in the previous section. The PINN model is newly proposed based on HT-PINN (Guo et al., 2023) and TgNN-Geo (Wang et al., 2021). It consists of five forward nets and an inverse net. Each forward net approximates the hydraulic heads for a specific pumping test, with spatial coordinates serving as input and hydraulic heads at the location being predicted as output. The inverse net, *GANet*, is pretrained as a pointwise stochastic model that approximates the spatial distribution of the GRF, for example,  $T(x, y; \alpha) \approx \text{GANet}(x, y; \alpha)$ ,  $x, y \in \mathbb{R}^1$ ,  $\alpha \in \mathbb{R}^k$ ,  $T(x, y; \alpha) \in \mathbb{R}^1$ . The architecture of *GANet* is based on the scaled dot-product attention



**Figure 7.** Inverse results of RGA-FNO and HT-PINN. (a) The reference field, in which purple triangles denote transmissivity measurements used by HT-PINN; (b) Estimated field by RGA-FNO; (c) Estimated field by HT-PINN; (d) True and estimated latent variables  $\alpha$  from two inverse models, blue line denotes the true  $\alpha$ , red dot denotes the estimation from RGA-FNO, green dot denotes the estimation from HT-PINN; (e) Observed hydraulic heads (reference data) versus FNO predictions; (f) Observed hydraulic heads (reference data) versus PINN predictions.

mechanism (Vaswani et al., 2017). It takes spatial coordinates  $(x, y)$  and latent random variables  $(\alpha)$  as input and predicts transmissivity as output. *GANet* is pretrained using low-resolution realizations from the GRF with exponential covariance. After training, *GANet* learns the map of the principal components. It incorporates spatial coordinates and leverages autograd to evaluate spatial derivatives as well as PDE constraints.

During inverse modeling, both the inverse and forward nets are trained together to estimate the transmissivity field and hydraulic heads. The training requires reference data for data match loss and collocation data for PDE loss. The reference data for the forward nets are observed hydraulic heads, which are the same as the reference data in RGA-FNO. The reference data for the inverse net are 36 transmissivity measurements, which are evenly collected in the domain and indicated by purple triangles in Figure 7a. More details about HT-PINN structure, *GANet* pretraining, collocation points, PDE constraints, and inverse modeling procedures can be found in Supporting Information S1 and Guo et al. (2023). The pretraining and inverse modeling are conducted on the Google Colab Platform with Tesla A 100 GPU, the pretraining takes 489 s, and the inverse modeling takes about 8,937 s.

The performance of RGA-FNO and HT-PINN in inverse modeling is presented in Figure 7. Figure 7a displays the true  $\ln T$  field. Figures 7b and 7c display the estimations from RGA-FNO and HT-PINN, Figure 7d plots the true and estimated latent random variables where blue line denotes the true  $\alpha$ , red dot denotes the estimation from RGA-FNO, green dot denotes the estimation from HT-PINN. Both models produce estimations in good agreement with the true realization. However, RGA-FNO captures small-scale non-smoothness better than HT-PINN. It is not only found from the visual appearance in Figures 7b and 7c, but also from the fit of  $\alpha$  at indices from 30 to 50, which represents the high-frequency patterns as well as small-scale heterogeneity and non-smoothness. A detailed look at Figure 7d, we can find that, while index is over 30, the red dots are closer to the blue line than green dots, suggesting that the estimation from RGA-FNO has a more similar high-frequency patterns to the true field than HT-PINN.

Table 3 provides a comprehensive evaluation of the models' performance, including map accuracy, relative error, and runtime. For HT-PINN, the number of transmissivity measurements ( $N_K$ ) is 36, and the map accuracy is

**Table 3**  
*Performance of Different Inverse Models*

Model	$N_K$	Map accuracy (%)	Relative error (%)						Runtime (s)
			Inverse	Forward					
				p1	p2	p3	p4	p5	
HT-PINN	36	90.23	18.07	21.41	12.26	13.78	31.31	20.06	8,937
RGA-FNO	0	91.74	16.74	3.49	2.14	2.01	2.21	3.23	70

90.23%. The relative error for forward simulations ranges from 12.26% to 31.31%, and the runtime is 8,937 s. For RGA-FNO, no transmissivity measurements are required, and the Map accuracy is 91.74%. The relative error for forward simulations ranges from 2.01% to 3.49%, and the runtime is only 15 s. Overall, RGA-FNO outperforms HT-PINN in all evaluation metrics, especially in terms of data requirement, forward model relative error, and runtime. HT-PINN needs transmissivity measurements for accurate inversion since the plain neural networks with a large number of coefficients (up to millions) may easily overfit the spatial function of  $\ln T$  (Guo et al., 2023; Wang et al., 2021); however, FNO does not require any transmissivity measurements, and the forward simulations are more accurate. Additionally, the runtime of RGA-FNO is much shorter than that of HT-PINN. The runtime for both methods only considers the time taken for inverse estimation and does not include the pretraining time for the FNO and inverse network in HT-PINN, as they only need to be trained once. However, HT-PINN is not easily generalizable and needs to be retrained for each new realization even with the same geostatistical parameters, as a contrast, FNO only needs to be retrained when the geostatistical parameters of the GRF change. During the estimation of  $\alpha$ , the combination of RGA and FNO is mutually beneficial as the principal components effectively utilize the geostatistical information, and FNO provides efficient and precise forward simulations. In contrast, HT-PINN relies on the inverse net, which even though pre-trained, still has model error, and the initialized forward nets are biased and require training for co-convergence, which can be a time-consuming process. Without accurate geostatistical information and forward models, transmissivity measurements play a crucial role in HT-PINN during inverse modeling.

#### 4.4. Data Demands for RGA-FNO

We utilize RGA-FNO to model GRFs across a range of parameter variances and different volumes of hydraulic head measurements. Through these applications, we aim to demonstrate both the versatility and constraints of the RGA-FNO methodology. The variance of  $\ln T$ , denoted as  $\sigma_{\ln T}^2$ , is systematically increased from 0.25 to 1.0, 2.0, 4.0, and 8.0. This alteration augments the magnitude of the fluctuation of  $\ln T$ . All other domain properties remain consistent with previous experiments (Table 1). The study employs HT involving five steady-state pumping tests. The FNO's architecture and training configurations remain identical to those of the previous experiment. Following training, the trained FNO is combined with RGA for inference, employing known measurements—the steady-state hydraulic heads observed from the monitoring wells. In previous experiments, the domain features 36 monitoring wells (Figure 2a). However, in this section, the distribution of monitoring wells is adjusted to square configurations of  $4 \times 4$  and  $8 \times 8$ . While the locations of the monitoring wells at the four corners remain unchanged, the remaining wells—both at the square boundary and within the square—are uniformly distributed. This manipulation yields two sets of monitoring well configurations, each containing 16 and 64 wells, respectively, offering differing volumes of measurements for inference. All optimization settings mirror those of previous experiments, with the exception of the number of hydraulic head measurements. The assessment metrics encompass the accuracy of the estimated  $\ln T$ , the relative error of the estimated  $T$ , and the relative errors of the estimated  $h_{\text{FNO}}$  in each pumping test.

Table 4 presents the inversion metrics for random fields with increasing transmissivity variances ( $\sigma_{\ln T}^2$ ) in different scenarios. The table is structured to show accuracy (%) and relative error values for both hydraulic head ( $\epsilon_h$ ) and transmissivity ( $\epsilon_T$ ) predictions under various conditions. Each section of the table corresponds to different  $\sigma_{\ln T}^2$  values: 0.25, 1.0, 2.0, 4.0, and 8.0. For each  $\sigma_{\ln T}^2$  value, the table displays accuracy percentages for different numbers of monitoring wells ( $N_h$ ), specifically 16, 36, and 64. Additionally,  $\epsilon_h$  across all five pumping tests (p1,

**Table 4**  
Inversion Metrics for Random Fields With Different Variances

$\sigma_{lnT}^2$	$N_h$	Accuracy (%)	$\epsilon_T$ (%)	$\epsilon_h$ (%)				
				p1	p2	p3	p4	p5
0.25	16	91.24	15.65	5.06	4.67	2.76	3.91	3.69
	36	98.89	8.65	4.24	2.57	1.92	2.91	3.15
	64	97.27	12.96	1.99	2.29	1.85	2.73	2.32
1.0	16	82.13	28.33	9.29	11.40	5.80	13.55	9.53
	36	93.42	21.39	5.08	4.08	3.55	8.40	3.55
	64	92.68	30.94	5.60	3.00	3.35	4.79	2.75
2.0	16	78.55	47.79	8.94	8.85	6.14	8.85	7.22
	36	90.06	43.59	6.61	8.49	4.88	4.79	6.67
	64	87.71	47.83	5.33	8.66	4.68	4.90	8.19
4.0	16	63.44	73.30	11.92	11.33	15.03	10.68	16.10
	36	86.24	44.82	11.04	7.35	5.78	9.49	7.83
	64	87.07	35.77	6.31	13.45	6.34	5.88	9.09
8.0	16	46.31	97.94	16.52	16.51	18.80	33.72	17.65
	36	80.89	58.40	9.49	9.61	7.77	19.26	8.89
	64	81.65	69.53	10.62	8.28	8.03	8.04	7.56

p2, p3, p4, p5) are provided. These results illustrate the varying performance of the inversion approach in different scenarios of random fields with different variances.

The analysis of accuracy in relation to different variances and numbers of monitoring wells reveals significant insights. Firstly, when considering the effect of  $\sigma_{lnT}^2$  on accuracy, it's evident that as the variance increases, the accuracy of the inverse modeling decreases. This trend is consistent across all three scenarios of monitoring well numbers ( $N_h = 16, 36, 64$ ). This observation implies that higher variance fields pose a greater challenge to accurate inverse modeling, resulting in decreased accuracy in predicting both hydraulic heads and transmissivity. Secondly, the impact of the number of monitoring wells is notable. Generally, as the number of monitoring wells increases, the accuracy of the inversion process improves. This trend can be observed across different variance levels. Higher numbers of monitoring wells ( $N_h = 36, 64$ ) consistently lead to higher accuracy percentages and lower error values for both  $\epsilon_h$  and  $\epsilon_T$ , compared to the case with fewer monitoring wells ( $N_h = 16$ ). In summary, the findings suggest that accurate inverse modeling is affected by both the variance of the random field and the number of monitoring wells. While higher variance random fields challenge accurate predictions, having a greater number of monitoring wells enhances the accuracy of the inversion process. These observations emphasize the importance of considering these factors when conducting hydraulic tomography and inverse modeling in real-world scenarios.

### 5. Conclusion

In this study, we proposed a new method called RGA-FNO and tested it on examples of inverse modeling of HT, which utilizes FNO as a surrogate forward model. As a data-driven model, FNO can learn the solution operator of PDEs in HT from data without the need to impose any PDE or BC constraints explicitly in the loss function during training. The well-trained FNO is accurate and generalizable and can solve parametric PDEs with new parameter instances through a single forward pass, providing an efficient alternative to traditional forward models solved by numerical methods. By replacing the traditional numerical solver in the RGA, the new RGA-FNO model achieves the same level of accuracy as the traditional RGA model, but with greater efficiency. This is because FNO, as a deep learning forward model, can backpropagate gradients through autograd efficiently. RGA-FNO is a powerful approach for inverse modeling of heterogeneous, non-smooth GRFs. RGA effectively reduces the dimension of unknowns and encodes non-smoothness in the prior distribution, while FNO is flexible enough to incorporate upscaling and provide reliable forward simulations. Together, RGA-FNO accurately estimates the parameters of the GRF, making it a valuable tool for a variety of applications in geostatistics and beyond. While we employ FNO for simulating both steady-state and transient HT, the performance of transient FNO falls short compared to steady-state FNO, despite its increased complexity and number of measurements. The integration of steady-state HT with multiple pumping wells at different locations emerges as a more effective strategy for enhancing the performance of the RGA-FNO model. Compared to the PINN inverse model, RGA-FNO does not require transmissivity measurements and produces more accurate estimations. Additionally, the inverse modeling of RGA-FNO is more efficient than PINN since FNO is trained for a GRF with specific geostatistical parameters and can be used to model the forward problems of all realizations from the GRF, however, PINN needs to retrain forward neural networks for every new realization even though it is from the same GRF. During the RGA-FNO application, higher variances of the underlying  $lnT$  are associated with diminished accuracy, whereas higher monitoring well numbers consistently contribute to improved accuracy. These findings underscore the crucial role of these factors in practical hydraulic tomography and inverse modeling scenarios.

In the current study, we assume that the prior information of the GRF is known and that FNO only learns from data generated from the GRF. However, it has been shown that neural operators can learn the mapping from the GRF of different structural parameters to the corresponding Darcy flow solution (Lu et al., 2021, 2022). This type of neural operator learning requires neural networks with different architectures and training strategies, which are

beyond the scope of this study. In the future, we will expand the use of FNO as a surrogate forward model by combining it with generative deep learning models, such as variational auto-encoders (VAE) and generative adversarial networks (GAN) for inverse modeling of HT in non-Gaussian, complex random fields. The efficiency of FNO can also be beneficial for experimental design of monitoring wells given fixed pumping wells, which often requires a large number of forward simulations. Furthermore, we will explore the feasibility of using FNO to directly learn the solution operator for the inverse problem. Although the potential of FNO as an inverse operator has been demonstrated with structured, full-map inputs and outputs, we aim to extend its application to predictive tasks with unstructured and sparse measurements. We are optimistic that FNO, along with other neural operator models, may play an important role in groundwater modeling.

## Data Availability Statement

Data sets, computational notebooks, and saved models that support the findings of this study are available at (Guo, 2023). Open source software includes Fourier neural operator (Kovachki et al., 2021; Li et al., 2021b) and PyTorch (Paszke et al., 2019).

## Acknowledgments

Authors express gratitude to Li, Zongyi et al. making their code and data open source.

## References

- Ambikasaran, S., Li, J. Y., Kitanidis, P. K., & Darve, E. (2013). Large-scale stochastic linear inversion using hierarchical matrices. *Computers & Geosciences*, 17(6), 913–927. <https://doi.org/10.1007/s10596-013-9364-0>
- Bisong, E. (2019). Google colabouratory In. *Building machine learning and deep learning models on Google cloud platform: A comprehensive guide for beginners* (pp. 59–64). Apress. [https://doi.org/10.1007/978-1-4842-4470-8\\_7](https://doi.org/10.1007/978-1-4842-4470-8_7)
- Boyce, S. E., & Yeh, W. W. G. (2014). Parameter-independent model reduction of transient groundwater flow models: Application to inverse problems. *Advances in Water Resources*, 69, 168–180. <https://doi.org/10.1016/j.advwatres.2014.04.009>
- Broyden, C. G. (1965). A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, 19(92), 577–593. <https://doi.org/10.1090/s0025-5718-1965-0198670-6>
- Guo, Q. (2023). QuanGuo/RGA-FNO-HT-INV [Data & Software]. <https://doi.org/10.5281/zenodo.8052496>
- Guo, Q., Zhao, Y., Lu, C., & Luo, J. (2023). High-dimensional inverse modeling of hydraulic tomography by physics informed neural network (HT-PINN). *Journal of Hydrology*, 616, 128828. <https://doi.org/10.1016/j.jhydrol.2022.128828>
- He, Q., Barajas-Solano, D., Tartakovsky, G., & Tartakovsky, A. M. (2020). Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Advances in Water Resources*, 141, 103610. <https://doi.org/10.1016/j.advwatres.2020.103610>
- He, Q., & Tartakovsky, A. M. (2021). Physics-informed neural network method for forward and backward Advection-Dispersion Equations. *Water Resources Research*, 57(7), e2020WR029479. <https://doi.org/10.1029/2020WR029479>
- Hernandez-Lobato, J. M., & Adams, R. (2015). Probabilistic backpropagation for scalable learning of Bayesian neural networks. *arXiv*. <https://doi.org/10.48550/arXiv.1502.05336>
- 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. <https://doi.org/10.1002/2016WR020089>
- Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. *arXiv*. <https://doi.org/10.48550/arXiv.1412.6980>
- Kitanidis, P., & Lee, J. (2014). Principal component geostatistical approach for large-dimensional inverse problems. *Water Resources Research*, 50(7), 5428–5443. <https://doi.org/10.1002/2013WR014630>
- Kitanidis, P. K. (1995). Quasi-Linear Geostatistical theory for inversing. *Water Resources Research*, 31(10), 2411–2419. <https://doi.org/10.1029/95WR01945>
- 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. <https://doi.org/10.1016/j.advwatres.2016.12.006>
- Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Neural operator: Graph kernel network for partial differential equations. *arXiv*. <https://doi.org/10.48550/arXiv.2003.03485>
- Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2021). Neural operator: Learning maps between function spaces with applications to PDEs [Software]. <https://doi.org/10.48550/arXiv.2108.08481>
- Lee, J., & Kitanidis, P. (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. <https://doi.org/10.1002/2014WR015483>
- 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. <https://doi.org/10.1002/2015WR018483>
- Li, J., & Tartakovsky, A. M. (2022). Physics-informed Karhunen-Loève and neural network approximations for solving inverse differential equation problems. *Journal of Computational Physics*, 462, 111230. <https://doi.org/10.1016/j.jcp.2022.111230>
- Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Multipole graph neural operator for parametric partial differential equations. Paper presented at. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Curran Associates Inc.
- Li, Z., Kovachki, N. B., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2021a). Fourier neural operator for parametric partial differential equations. Paper presented at. In *International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.2010.08895>
- Li, Z., Kovachki, N. B., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2021b). Fourier neural operator for parametric partial differential equations (Version 0.3.0) [Software]. *arXiv*. <https://doi.org/10.48550/arXiv.2010.08895>
- Li, Z., Zheng, H., Kovachki, N., Jin, D., Chen, H., Liu, B., et al. (2022). Physics-informed neural operator for learning partial differential equations. *arXiv*. <https://doi.org/10.48550/arXiv.2111.03794>



- Liu, X., & Kitanidis, P. (2011). Large-scale inverse modeling with an application in hydraulic tomography. *Water Resources Research*, 47(2), W02501. <https://doi.org/10.1029/2010WR009144>
- Liu, X., Zhou, Q., Birkholzer, J., & Illman, W. A. (2013). Geostatistical reduced-order models in underdetermined inverse problems. *Water Resources Research*, 49(10), 6587–6600. <https://doi.org/10.1002/wrcr.20489>
- Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3), 218–229. <https://doi.org/10.1038/s42256-021-00302-5>
- Lu, L., Meng, X., Cai, S., Mao, Z., Goswami, S., Zhang, Z., & Karniadakis, G. E. (2022). A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data. *Computer Methods in Applied Mechanics and Engineering*, 393, 114778. <https://doi.org/10.1016/j.cma.2022.114778>
- Mo, S., Zabarar, N., Shi, X., & Wu, J. (2019). Deep autoregressive neural networks for high-dimensional inverse problems in groundwater contaminant source identification. *Water Resources Research*, 55(5), 3856–3881. <https://doi.org/10.1029/2018WR024638>
- Nelsen, N. H., & Stuart, A. M. (2021). The random feature model for input-output maps between Banach spaces. *SIAM Journal on Scientific Computing*, 43(5), A3212–A3243. <https://doi.org/10.1137/20m133957x>
- Nowak, W., & Cirpka, O. A. (2004). A modified Levenberg–Marquardt algorithm for quasi-linear geostatistical inverting. *Advances in Water Resources*, 27(7), 737–750. <https://doi.org/10.1016/j.advwatres.2004.03.004>
- Nowak, W., Tenkve, S., & Cirpka, O. A. (2003). Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities. *Mathematical Geology*, 35(1), 53–66. <https://doi.org/10.1023/A:1022365112368>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library (version 2.1). [Software]. *ACM Digital Library*. <https://doi.org/10.5555/3454287.3455008>
- Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). Machine learning of linear differential equations using Gaussian processes. *Journal of Computational Physics*, 348, 683–693. <https://doi.org/10.1016/j.jcp.2017.07.050>
- 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 and Gas Science and Technology*, 67(5), 857–875. <https://doi.org/10.2516/ogst/2012064>
- Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., & Barajas-Solano, D. (2020). Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems. *Water Resources Research*, 56(5), e2019WR026731. <https://doi.org/10.1029/2019WR026731>
- Tripathy, R. K., & Bilionis, I. (2018). Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification. *Journal of Computational Physics*, 375, 565–588. <https://doi.org/10.1016/j.jcp.2018.08.036>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. Paper presented at. In *Neural information processing systems*. Curran Associates, Inc. <https://doi.org/10.48550/arXiv.1706.03762>
- Wang, N., Chang, H., & Zhang, D. (2021). Deep-learning-based inverse modeling approaches: A subsurface flow example. *Journal of Geophysical Research: Solid Earth*, 126(2), e2020JB020549. <https://doi.org/10.1029/2020JB020549>
- Xu, R., Wang, N., & Zhang, D. (2021). Solution of diffusivity equations with local sources/sinks and surrogate modeling using weak form Theory-guided Neural Network. *Advances in Water Resources*, 153, 103941. <https://doi.org/10.1016/j.advwatres.2021.103941>
- Yin, D., & Illman, W. A. (2009). Hydraulic tomography using temporal moments of drawdown recovery data: A laboratory sandbox study. *Water Resources Research*, 45(1), W01502. <https://doi.org/10.1029/2007WR006623>
- Zhang, X., Zhu, Y., Wang, J., Ju, L., Qian, Y., Ye, M., & Yang, J. (2022). GW-PINN: A deep learning algorithm for solving groundwater flow equations. *Advances in Water Resources*, 165, 104243. <https://doi.org/10.1016/j.advwatres.2022.104243>
- Zhao, Y., Guo, Q., Lu, C., & Luo, J. (2022). High-dimensional groundwater flow inverse modeling by upscaled effective model on principal components. *Water Resources Research*, 58(7), e2022WR032610. <https://doi.org/10.1029/2022WR032610>
- Zhao, Y., & Luo, J. (2020). Reformulation of Bayesian geostatistical approach on principal components. *Water Resources Research*, 56(4), e2019WR026732. <https://doi.org/10.1029/2019WR026732>
- Zhao, Y., & Luo, J. (2021a). A Quasi-Newton reformulated geostatistical approach on reduced dimensions for large-dimensional inverse problems. *Water Resources Research*, 57(1), e2020WR028399. <https://doi.org/10.1029/2020WR028399>
- Zhao, Y., & Luo, J. (2021b). Bayesian inverse modeling of large-scale spatial fields on iteratively corrected principal components. *Advances in Water Resources*, 151, 103913. <https://doi.org/10.1016/j.advwatres.2021.103913>
- Zhou, Z., Zabarar, N., & Tartakovsky, D. M. (2022). Deep learning for simultaneous inference of hydraulic and transport properties. *Water Resources Research*, 58(10), e2021WR031438. <https://doi.org/10.1029/2021WR031438>
- Zhu, Y., & Zabarar, N. (2018). Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. *Journal of Computational Physics*, 366(C), 415–447. <https://doi.org/10.1016/j.jcp.2018.04.018>
- Zong, Y., He, Q., & Tartakovsky, A. M. (2023). Improved training of physics-informed neural networks for parabolic differential equations with sharply perturbed initial conditions. *Computer Methods in Applied Mechanics and Engineering*, 414, 116125. <https://doi.org/10.1016/j.cma.2023.116125>