Journal of Hydrology xxx (xxxx) 128828

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

High-dimensional inverse modeling of hydraulic tomography by physics informed neural network (HT-PINN)

Quan Guo [a], Yue Zhao [a], Chunhui Lu [b], Jian Luo [a,*]

[a] School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0355, USA
[b] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China

ARTICLE INFO

This manuscript was handled by Huaming Guo, Editor-in-Chief, with the assistance of Wenke Wang , Associate Editor

Keywords:
PINN
Hydraulic tomography
Large-scale
Inverse problem
Neural network

ABSTRACT

A hydraulic tomography – physics informed neural network (HT-PINN) is developed for inverting two-dimensional large-scale spatially distributed transmissivity. HT-PINN involves a neural network model of transmissivity and a series of neural network models to describe transient or steady-state sequential pumping tests. All the neural network models are jointly trained by minimizing the total loss function including data fitting errors and PDE constraints. Batch training of collocation points is used to amplify the advantage of the mesh-free property of neural networks, thereby limiting the number of collocation points per training iteration and reducing the total training time. The developed HT-PINN accurately and efficiently inverts two-dimensional Gaussian transmissivity fields with more than a million unknowns (1024 × 1024 resolution), and the inversion map accuracy exceeds 95 %. The effects of batch sampling methods, batch number and size, and data requirements for direct and indirect measurements are systematically investigated. In addition, the developed HT-PINN exhibits great scalability and structure robustness in inverting fields with different resolutions ranging from coarse (64 × 64) to fine (1024 × 1024). Specifically, data requirements do not increase with the problem dimensionality, and the computational cost of HT-PINN remains almost unchanged due to its mesh-free nature while maintaining high inversion accuracy when increasing the field resolution.

Nomenclature

AD	Automatic Differentiation
GA	Geostatistical Approach
HT	Hydraulic Tomography
HT-PINN	Hydraulic Tomography-Physics Informed Neural Network
MSE	Mean Squared Error
PINN	Physics Informed Neural Network
RGA	Reformulated Geostatistical Approach

SymbolsGeneral groundwater flow equation

S_s	Specific storage
h	Hydraulic head
Q	Water accumulation/reduction rate in control volume
$\mathbf{q}$	Specific discharge vector
T	Hydraulic transmissivity

Experimental domain

Ω	Two-dimensional (2D) spatial domain
T	Total observation time
x/y	Horizontal/vertical coordinate in 2D domain
t	Time (temporal coordinate)

Boundary and initial conditions (BC and IC)

(x_{init}, y_{init})	Grids where initial condition is applied
h_{init}	Initial hydraulic head
Γ_N	Neumann boundary
(x_N, y_N, t_N)	Grids and time that Neuman BC is applied
q_N	Flux rate at Neumann boundary
$\mathbf{n}$	Unit vector normal to Neumann BC
(x_D, y_D, t_D)	Grids and time that Dirichlet BC is applied

* Corresponding author.
E-mail address: jian.luo@ce.gatech.edu (J. Luo).

https://doi.org/10.1016/j.jhydrol.2022.128828
Received 10 August 2022; Received in revised form 3 November 2022; Accepted 10 November 2022
0022-1694/© 20XX

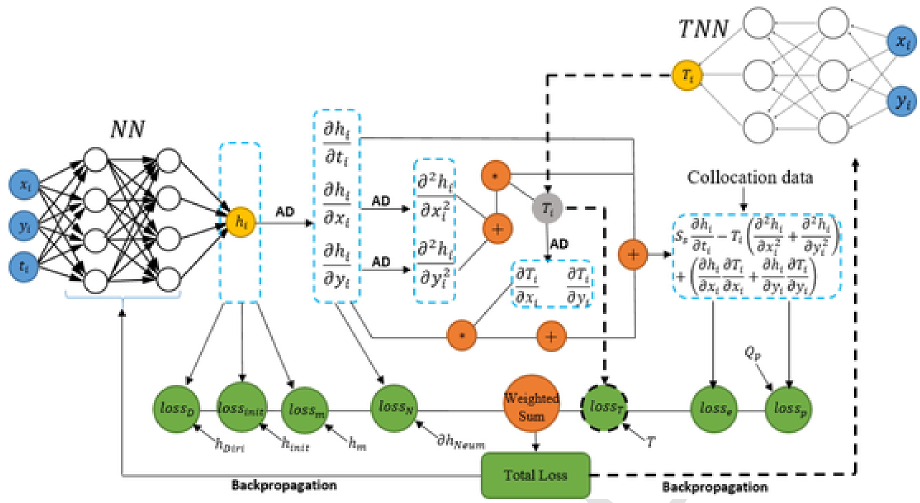

Fig. 1. Flowchart of HT-PINN: *TNN* (upper right) and *NN* (upper left) denote inverse and forward neural networks for hydraulic transmissivity and hydraulic head; T_i and h_i denote predictions from neural networks; "AD" denotes automatic differentiation used to approximate partial derivatives; Orange rounds are operators; Green rounds are loss terms; Dash line is the pathway used only by *TNN*; the solid is the pathway used by *TNN* and *NN*; Flow in two types of pathways will not convey despite the intersection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

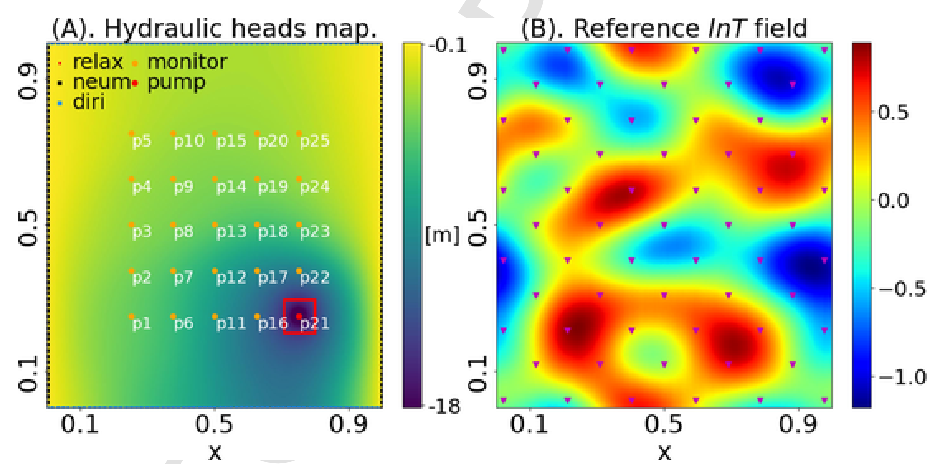

Fig. 2. Numerical experiment setup for the steady-state HT experiment. (A) the well network consists of 25 wells (p1-p25) with the pumping well, for example, located at the right bottom corner (red dot) and monitoring wells (monitor) located at orange dots. Dirichlet (diri) boundary cells are denoted by blue squares, and Neumann (neum) boundary cells are denoted by black squares. Relax region boundaries are denoted by the red line, cells outside this region but inside boundaries are under PDE constraint; (B) the reference *lnT* field used in the pumping test simulation with direct measurements of transmissivity denoted by purple triangles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

h_D	Hydraulic head at Dirichlet boundary

PDE constraint in pumping test

(x_e, y_e)	Non-pumping grids where pumping test is not conducted
(x_p, y_p)	Pumping grids where pumping test is conducted
Q_p	Pumping rate

Reference data

(x_T, y_T)	Grids where direct measurements of hydraulic transmissivity are measured
(x_m, y_m)	Grids where indirect measurements of hydraulic heads are monitored

Neural Networks

NN	Forward neural network for predicting hydraulic heads
TNN	Inverse neural network for estimating hydraulic transmissivity
$f_{NN,T}$	PDE residuals only related to forward network
$f_{NN,TNN}$	PDE residuals related to both forward and inverse networks
$Loss_e$	Loss term of PDE constraints of non-pumping grids
$Loss_p$	Loss term of PDE constraints of pumping grids
$Loss_{init}$	Loss term of initial conditions
$Loss_N$	Loss term of Neumann BC
$Loss_D$	Loss term of Dirichlet BC
$Loss_m$	Loss term of data match of indirect measurements
$Loss_T$	Loss term of data match of direct measurements
$Loss_{FP}$	Total loss of forward problem
$Loss_{IP}$	Total loss of inverse problem
λ_e/N_e	Weight and number of grids used by $Loss_e$
λ_p/N_p	Weight and number of grids used by $Loss_p$
λ_{init}/N_{init}	Weight and number of grids used by $Loss_{init}$

Table 1
Hydrogeological and geostatistical parameters for the hydraulic tomography experiment.

Parameter	Values
Domain size, $L_x \times L_y$	320m × 320m
Grid spacing, $\Delta x \times \Delta y$	0.3125m × 0.3125m
Spatial resolution, $n_x \times n_y$	1024 × 1024
Transmissivity, T[m^2/hr]	
Geometric mean	0
Variance of $\ln T$, $\sigma^2_{\ln T}$	1
Correlation length, $\lambda_x \times \lambda_y$	64m × 48m
Left Boundary	h = 0m
Right Boundary	h = 0m
Initial Condition	h = 0m
Pumping Rate [m^3/hr]	3.6

λ_N/N_N	Weight and number of grids used by $Loss_N$
λ_D/N_D	Weight and number of grids used by $Loss_D$
λ_m/N_m	Weight and number of grids used by $Loss_m$
λ_T/N_T	Weight and number of grids used by $Loss_T$

Evaluation Metrics

ϵ	Relative residuals
ε	Pointwise estimation error

1. Introduction

The cost of directly measuring spatially variable parameters like hydraulic conductivity or transmissivity in the field is prohibitive, leading to inverse problems of estimating these parameters through indirect measurements such as hydraulic heads or tracer concentrations. Hydraulic tomography (HT), also known as sequential pumping tests, has demonstrated great potential for aquifer characterization with relatively low cost and simple data collection techniques (Yeh and Liu, 2000). By alternatively switching pumping and monitoring wells in a well network, HT can provide larger and more informative data than a traditional single-well pumping test. The enhanced information density reduces the non-uniqueness of potential inverse solutions for the unknown parameter field like hydraulic conductivities.

A common and effective approach to solving HT inverse problems is the gradient-based geostatistical approach (GA). GA formulates the posterior distribution of target variables according to the likelihood of data matching and the prior for the smoothness regularization under the Bayesian framework (Kitanidis, 1995). The posteriori is maximized to obtain the best linear estimate and associated uncertainty. The major challenge faced by GA is its applicability when the dimension of the target variable is large for estimating a high-resolution parameter field. The cost of computing and storing the associated full-rank covariance matrix is very high, especially when using gradient-based methods such as Newton's method to evaluate the Jacobian matrix at each iteration (Ambikasaran et al., 2013; Klein et al., 2017; Liu and Kitanidis, 2011; Liu et al., 2013; Obiefuna and Eslamian, 2019). Many efforts have been made to optimize GA for large-scale inverse problems from two aspects: dimensionality reduction and efficient numerical methods. Dimensionality reduction encodes the original high-dimensional parameter field as low-dimensional random parameters, and reformulates the computations on the original field to the random parameters so that the computations are more concise and efficient. For example, Gaussian random fields can be encoded through combinations of independent projective vectors and identified with about tens of independent and identically distributed random variables. The combinations of independent projec-

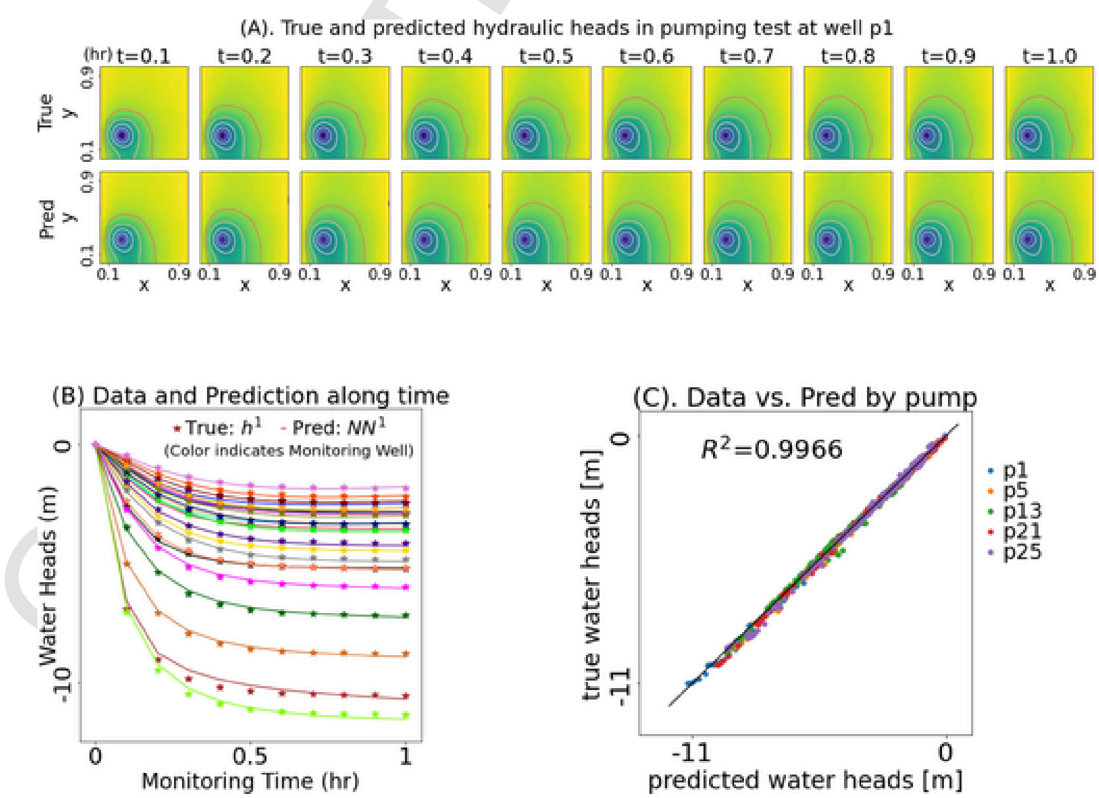

Fig. 3. Comparison of NN model with the numerical simulation for the transient pumping test. (A) hydraulic head distribution in pumping test at well p1, the first row shows $h^1(x, y, t)$ from the numerical simulation at $t = 0.1$–1.0 h, the second row shows $NN^1(x, y, t)$ prediction $t = 0.1$–1.0 h; (B) reference water heads and predicted water heads at monitoring wells in pumping test p1, each color notes an unique monitoring well, stars note reference data, solid line note model prediction; (C) reference water heads vs predicted water heads in pump tests at p1, p5, p13, p21 and p25.

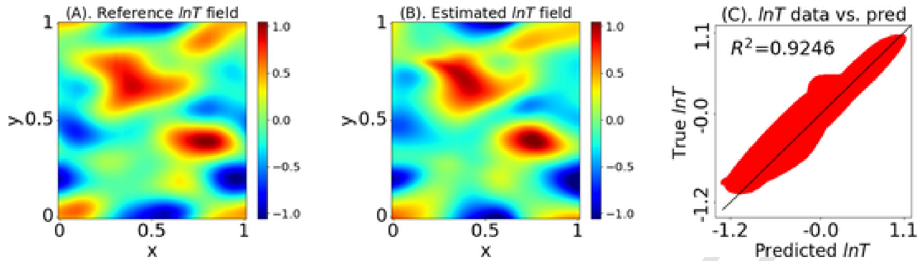

Fig. 4. Performance of TNN for inverse modeling of transient hydraulic tomography. (A) reference lnT field; (B) estimation from TNN_t; (C) true vs estimated **lnT**.

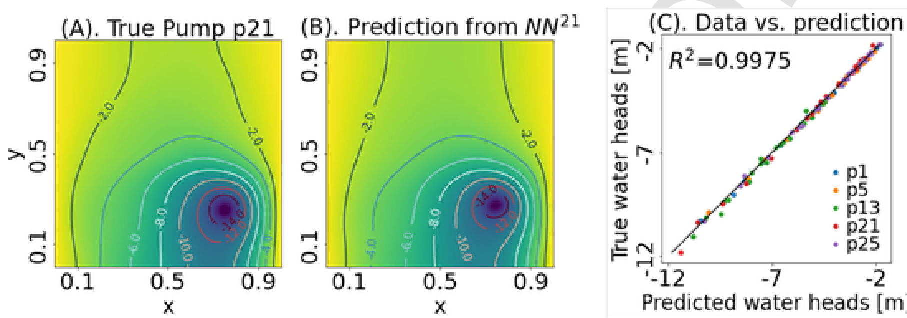

Fig. 5. Comparison of NN model with the numerical simulation for the pumping test at well p21. (A) numerical simulation of hydraulic head distribution; (B) NN^{21} prediction; (C) reference water heads vs predicted water heads for all pumping tests.

tive vectors, which is usually referred as projection matrix, can be obtained by principal component analysis (PCA) (Kitanidis and Lee, 2014; Lee and Kitanidis, 2014; Lee et al., 2016; Zhao and Luo, 2020; Zhao and Luo, 2021a). For more complicated, non-Gaussian fields, machine learning or deep learning methods can be used for dimensionality reduction (Chen et al., 2022; Chen et al., 2021; Laloy et al., 2018; Laloy et al., 2017; Pang et al., 2020). Efficiency of numerical methods can be improved through accelerating the process of solving the linear system related to forward problems such as using efficient embeddings on the covariance matrix, upscaling the parameter fields or implementing quasi-Newton methods for approximating Jacobian matrices (Broyden, 1965; Nowak and Cirpka, 2004; Nowak et al., 2003; Saibaba et al., 2012; Zhao et al., 2022; Zhao and Luo, 2021b).

In the present research, we aim to explore a new method, physics informed neural network (PINN), for solving high-dimensional HT inverse problems. PINN has demonstrated good potential in inverse modeling due to its pointwise computation and mesh-free property. PINN is a deep neural network (DNN) in structure, but unlike DNN which only contains the loss function of the residuals of data matching, PINN integrates additional constraints from prior physics information in the loss function to overcome the inability of traditional DNNs for absorbing prior physics knowledge (Raissi et al., 2017a; Raissi et al., 2017b; Raissi et al., 2017c; Raissi et al., 2017d). These constraints are mathematically expressed in partial differential equations (PDEs) and derivative values approximated by automatic differentiation (AD) (Griewank, 2003). After optimizing such a loss function, PINN can make effective predictions in accordance with the governing PDEs in addition to the observed data (Jagtap et al., 2020; Karniadakis, 2019; Kharazmi et al., 2021; Raissi and Karniadakis, 2018). For groundwater inverse problems, Tartakovsky et al. (2020) used two PINNs to jointly solve the forward and inverse groundwater flow problems: one was used as the forward model for approximating hydraulic heads, and the other as the inverse model for estimating heterogeneous hydraulic conductivities. The study was extended to groundwater transport problems by assimilating tracer concentration and hydraulic heads data together for inversing the hydraulic conductivity field (He et al., 2020; He and Tartakovsky, 2021). Wang et al. (2020) developed a theory-guided neural network (TgNN), which was incorporated with weak form PDE constraints and used to infer the inverse solution of a Gaussian hydraulic conductivity field with known spatial covariance information (Wang et al., 2021a). PINN has also been applied to inverse problems in unsaturated groundwater flow (Depina et al., 2022).

Compared with gradient-based GA, PINN transformed the inverse problem into a predictive task and solved it directly by a continuous function on mesh coordinates (Bottou and Bousquet, 2008; Zhu et al., 2019). The required gradients, whether in the variable space or the model coefficient space, were evaluated using AD, which was mesh-free and much faster than implementing groundwater flow simulation to determine the Jacobian matrix (Yang et al., 2021; Yang et al., 2020; Yang and Perdikaris, 2019). Although PINN theoretically has the potential to deal with large-scale inverse problems due to its pointwise computation and mesh-free property, it has not been tested for estimating high-resolution parameter fields. For example, in the above-mentioned groundwater application literature, the number of hydraulic conductivity to be estimated is on the order of 10^3 –10^4 for PINN, while GA has been successfully applied to estimate high-resolution hydraulic conductivity fields with millions of unknowns (Lee and Kitanidis, 2014; Zhao and Luo, 2020). Tartakovsky et al. (2020) investigated the effect of the number of collocation points with PDE regularization and demonstrated that approximately 10 % of the total grid can provide inverse results comparable to the full grid. However, their problem dimension is only 1,024, i.e., a 32 × 32 field. For a high-resolution field with millions of unknowns, 10 % of the total collocation points for PDE regularization are computationally unaffordable.

In this study, we develop a hydraulic tomography-PINN (HT-PINN) to jointly solve forward and inverse problems for two-dimensional large-scale hydraulic tomography. To the best of our knowledge, this is the first PINN for groundwater inverse problems involving pumping tests, especially multiple pumping tests. We extend the application of PINN by incorporating the batch training technique to solve large-scale inverse problems for estimating high-resolution parameter fields with over millions of unknowns. Unlike traditional batch training techniques that divide big data into subsets, we divide collocation points with PDE regularization into subsets while keeping all measurement data as one batch to train the network sequentially. We investigate the collocation point batch training for the developed HT-PINN to demonstrate high computational efficiency based on pointwise computation and mesh-free property. Furthermore, we compare the performance of HT-PINN

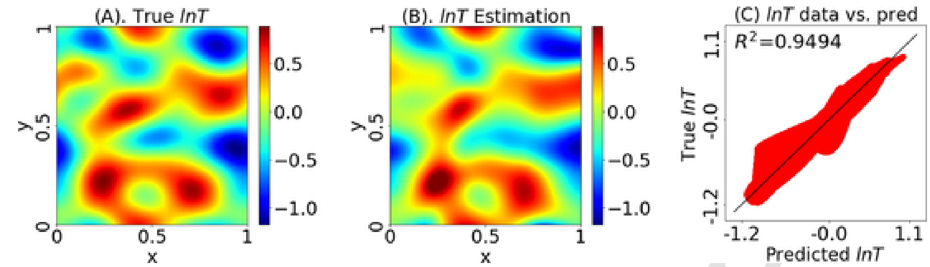

Fig. 6. Performance of *TNN* for inverse modeling of hydraulic tomography. (A) reference lnT field; (B) estimation from *TNN*; (C) true vs estimated *lnT*;

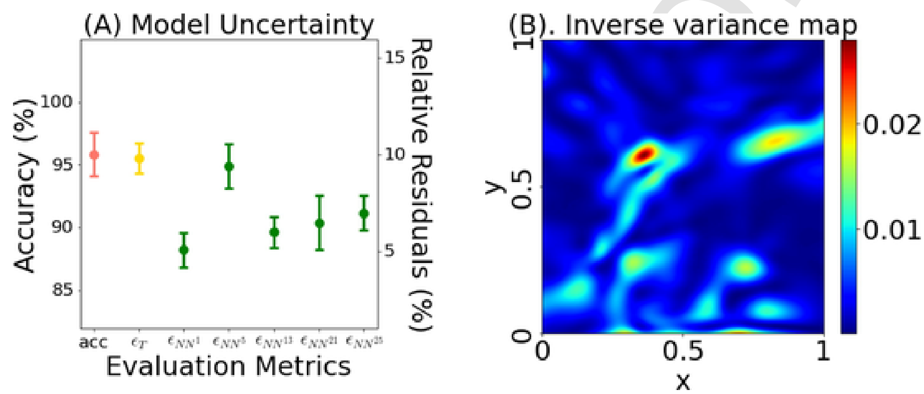

Fig. 7. Uncertainty quantification of HT-PINN by repeating the inversion with different initial guess and batch generation. (A) uncertainty of evaluation metrics. The orange plot (acc) is the map accuracy, the gold plot is the inverse relative residuals (ϵ_T), and the green plots are forward relative residuals of 5 simulated pumping tests (ϵ_{NN^1}, ϵ_{NN^5}, $\epsilon_{NN^{13}}$, $\epsilon_{NN^{21}}$, $\epsilon_{NN^{25}}$); (B) variance of inverse estimation of transmissivity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with GA for estimating parameter fields at different resolutions and discuss the measurement data required for the developed HT-PINN.

2. Models and method

2.1. Groundwater flow with pumping

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

$$S_s \frac{\partial h}{\partial t} = -\nabla \bullet \mathbf{q} + Q \tag{1}$$

where S_s is specific storage, t is time, h is hydraulic head, Q is the water accumulation/reduction rate in the selected control volume, and $\mathbf{q}$ is specific discharge vector. For simplicity, we consider hydraulic tomography in a confined, isotropic, heterogeneous aquifer in a 2D spatial domain Ω in a time window $[0, \mathrm{T}]$. Thus, at a location or grid (x_e, y_e) in the domain with no source/sink ($Q = 0$), the specific governing PDE becomes:

$$S_s \frac{\partial h(x_e, y_e, t_e)}{\partial t} - \nabla \bullet [T(x_e, y_e) \nabla h(x_e, y_e, t_e)] = 0, (x_e, y_e) \in \Omega, t_e \in (0, \mathrm{T}] \tag{2}$$

where T is isotropic, heterogeneous hydraulic transmissivity. At the initial condition (IC), $t = 0$, we have background hydraulic heads:

$$h(x_{init}, y_{init}, 0) = h_{init}, (x_{init}, y_{init}) \in \Omega, \tag{3}$$

To satisfy Neumann (Γ_N) and Dirichlet (Γ_D) boundary conditions (BCs) of the domain, we have:

$$\mathbf{n} \bullet \nabla h(x_N, y_N, t_N) = q_N, (x_N, y_N) \in \Gamma_N, t_N \in (0, \mathrm{T}] \tag{4}$$

$$h(x_D, y_D, t_D) = h_D, (x_D, y_D) \in \Gamma_D, t_D \in (0, \mathrm{T}] \tag{5}$$

where $\mathbf{n}$ is the unit vector normal to Neumann BC. If water is extracted from a pumping well located in a specific grid (x_p, y_p) at a constant flow rate, Q_p, the governing equation for this grid is:

$$S_s \frac{\partial h(x_p, y_p, t_p)}{\partial t} - \nabla \bullet [T(x_p, y_p) \nabla h(x_p, y_p, t_p)] = Q_p, (x_p, y_p) \in \Omega, t_p \in (0, \mathrm{T}] \tag{6}$$

Eqs. (2) – (6) complete a PDE system for groundwater flow in a pumping test. If $T(x, y)$ is fully characterized, $h(x, y, t)$ can be solved with numerical solvers for given initial and boundary conditions and pumping schedules. This is known as a forward problem. Reversely, estimating spatially variable $T(x, y)$ with measurements of $h(x, y, t)$ at specific monitoring locations and time and limited local measurements of hydraulic transmissivity is an inverse problem, which is the focus of the present study. In hydraulic tomography, multiple sets of $h(x, y, t)$ can be collected by conducting pumping tests sequentially at different pumping wells.

2.2. PINN forward model

To construct a PINN forward model $NN(x, y, t)$ for hydraulic responses $h(x, y, t)$ in a pumping test, the loss function is composed of residuals from data matching, IC constraint, BC constraints and PDE constraints. These residuals should be evaluated at separate time steps in the time window $(0, \mathrm{T}]$, except for the residual of IC constraint which is determined at $t = 0$. The PDE constraint residual is the left-hand side of Eq. (2):

Table 2
Model performance with different batch sizes, number, and generation methods.

Batch			Accuracy (%)		ϵ_T(%)		Training time (s)
Type	Number (N_B)	Size (M_B)	Mean	s.t.d	Mean	s.t.d	
random + recurrent	10	0	73.83	9.37	21.74	6.87	315
	10	100	95.67	1.09	9.13	0.67	1200
	10	200	95.58	1.35	9.28	0.78	1475
	10	300	96.35	1.43	9.28	0.65	1719
	10	400	95.46	1.00	8.98	0.68	1953
	10	600	96.05	1.06	9.28	0.48	2468
	10	900	96.40	1.19	9.15	0.84	2768
	10	1200	96.15	1.62	9.95	1.08	3141
	20	100	96.80	–	9.02	–	2400
	20	300	95.61	–	8.98	–	3450
	20	600	96.59	–	9.76	–	4890
	1	3000	86.77	–	10.27	–	34,986
	1	9000	84.69	–	8.74	–	103,380
random + non-recurrent	–	100	95.05	–	9.94	–	1223
	–	300	96.45	–	9.26	–	1838
	–	600	94.75	–	9.81	–	2618
uniform + recurrent	10	100	95.11	–	10.22	–	1212
	10	300	94.91	–	9.92	–	1629
	10	600	95.90	–	9.88	–	2482

$$f_{NN,T}(x,y,t) = S_s \frac{\partial NN(x,y,t)}{\partial t} - \nabla \bullet [T(x,y) \nabla NN(x,y,t)] \tag{7}$$

where $f_{NN,T}$ means that residual is evaluated with $NN(x,y,t)$ and $T(x,y)$. The partial derivatives and gradients on the right-hand side are approximated by AD. The residuals are evaluated for two types of grids pumping and non-pumping grids. For a non-pumping grid (x_e, y_e), i.e., no pumping well is located in the grid, using mean squared error (MSE) as a metric, the PDE loss ($Loss_e$) is expressed as:

$$Loss_e = \frac{1}{N_e} \sum_{j=1}^{N_e} \left| f_{NN,T}\left(x_e^j, y_e^j, t_e^j\right) \right|^2 \tag{8}$$

For simplicity, the total number of non-pumping grids, N_e may be the total grid number if no pumping is applied in the domain. However, N_e can be selected and does not necessarily depend on the domain resolution.

For a pumping grid (x_p, y_p), the pumping PDE loss ($Loss_p$) in MSE is given by:

$$Loss_p = \frac{1}{N_p} \sum_{j=1}^{N_p} \left| f_{NN,T}\left(x_p^j, y_p^j, t_p^j\right) - Q_p \right|^2 \tag{9}$$

In a pumping test, if water is drawn from a single well located in a grid, $N_p = 1$ and (x_p, y_p) are the coordinates of the grid where the pumping well is located.

The IC constraint from Eq. (3) is applied to IC grids (x_{init}, y_{init}), which can be anywhere in the spatial domain. The resulted MSE loss ($Loss_{init}$) is expressed as:

$$Loss_{init} = \frac{1}{N_{init}} \sum_{j=1}^{N_{init}} \left| NN\left(x_{init}, y_{init}, 0\right) - h\left(x_{init}, y_{init}, 0\right) \right|^2 \tag{10}$$

BC constraint residuals are evaluated for two types of domain boundary grids: Neumann and Dirichlet boundary grids, described by Eqs. (4) and (5), respectively. For Neumann BC grids (x_N, y_N), the MSE loss ($Loss_N$) evaluation requires the partial derivatives on $NN(x,y,t)$.

$$Loss_N = \frac{1}{N_N} \sum_{j=1}^{N_N} \left| \mathbf{n} \bullet \nabla NN\left(x_N, y_N, t_N\right) - q_N \right|^2 \tag{11}$$

The partial derivatives should be determined across the boundary. For example, for a rectangular domain, if Neumann BC is on the left or right boundary, then $\mathbf{n} \bullet \nabla NN(x_N, y_N, t_N) = \frac{\partial NN(x_N, y_N)}{\partial x_N}$, and $\mathbf{n} \bullet \nabla NN(x_N, y_N, t_N) = \frac{\partial NN(x_N, y_N, t_N)}{\partial y_N}$ is used for the top or bottom boundary. The partial derivatives in Eq. (11) are approximated by AD.

For Dirichlet BC grids (x_D, y_D), the MSE loss ($Loss_D$) is:

$$Loss_D = \frac{1}{N_D} \sum_{j=1}^{N_D} \left| NN\left(x_D^j, y_D^j, t_D^j\right) - h\left(x_D^j, y_D^j, t_D^j\right) \right|^2 \tag{12}$$

The numbers of grids, N_N and N_D, and the spatial coordinates (x_N, y_N), (x_D, y_D) of Neumann and Dirichlet BC grids depend on the domain resolution and boundary setup.

Data matching residuals are evaluated for grids with monitoring wells and hydraulic head measurements. For monitoring grids (x_m, y_m) and measured hydraulic heads $h(x_m, y_m, t_m)$, the MSE loss ($Loss_m$) has the same form as $Loss_D$:

$$Loss_m = \frac{1}{N_m} \sum_{j=1}^{N_m} \left| NN\left(x_m^j, y_m^j, t_m^j\right) - h\left(x_m^j, y_m^j, t_m^j\right) \right|^2 \tag{13}$$

N_m is the number of measurements, and (x_m, y_m) are the coordinates of monitoring wells.

The total loss of the forward problem ($Loss_{FP}$) is the sum of the weighted loss terms:

$$\begin{aligned} Loss_{FP} = {} & \lambda_m Loss_m + \lambda_e Loss_e + \lambda_N Loss_N \\ & + \lambda_D Loss_D + \lambda_p Loss_p + \lambda_{init} Loss_{init} \end{aligned} \tag{14}$$

The weights of the loss terms, (λ_m, λ_e, λ_p, λ_N, λ_D), are hyperparameters.

It should be noted that in different pumping tests, the coordinates of the pumping well, (x_p, y_p), are different, and the coordinates of other types of grids other than the domain boundary have also changed. Consequently, different $NN(x,y,t)$ need to be trained as forward models corresponding to different pumping wells. Therefore, to simulate a hydraulic tomography consisting of n sequential pumping tests at different locations, n PINN forward models, $NN^i(x,y,t)$, $i = 1, 2 \cdots n$, are needed to serve in ensemble as the surrogate model, and the total loss function is the sum of Eq. (18) for all individual pumping tests.

2.3. PINN inverse model

To estimate the hydraulic transmissivity field $T(x,y)$, we develop an inverse PINN model, $TNN(x,y)$. Its data matching loss $Loss_T$ is given by:

$$Loss_T = \frac{1}{N_T} \sum_{j=1}^{N_T} \left| TNN\left(x_T^j, y_T^j\right) - T\left(x_T^j, y_T^j\right) \right|^2 \tag{15}$$

where (x_T, y_T) are the coordinates of the grids with direct measurements of hydraulic transmissivity, and N_T is the number. These measurements are directly used to constrain the hydraulic transmissivity distribution, and no more information or assumptions such as spatial covariance are needed. The BC and IC constraints do not apply to $TNN(x,y)$, but the PDE constraints are involved. Substituting $T(x,y)$ in Eq. (11) with $TNN(x,y)$, the PDE residual function of the pumping test i is expressed as:

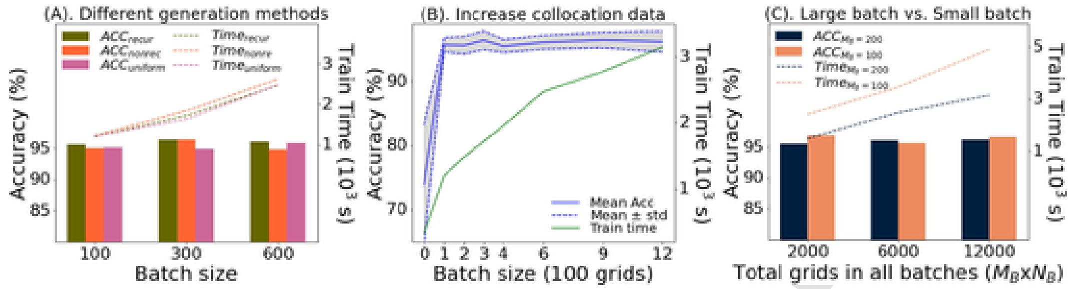

Fig. 8. Model performance and train time of different batch combinations. (A) different batch generation methods with the same batch size and number; (B) batch number is 10, the batch size is from 0 to 1200; (C) large batch: $M_B = 200$, 600, 1200 ($N_B = 10$) vs small batch: $M_B = 100$, 300, 600 ($N_B = 20$).

$$f_{NN^i,TNN}(x,y,t) = S_s \frac{\partial NN^i(x,y,t)}{\partial t} - \nabla \bullet \left[TNN(x,y) \nabla NN^i(x,y,t) \right], i = 1,2\cdots n \quad (16)$$

This updated residual function is used in Eqs. (8) and (9) to evaluate the PDE constraint loss for pumping and non-pumping grids. The total loss of the inverse problem ($Loss_{IP}$) is given by:

$$Loss_{IP} = \sum_{i=1,2,\cdots n} Loss_{FP}^i + \lambda_T Loss_T \quad (17)$$

where λ_T is the weight of data matching loss on direct measurements. $Loss_{FP}^i$ is the forward problem loss corresponding to the pumping test i, where a specific PINN surrogate forward model $NN^i(x,y,t)$ is trained to approximate its water heads. All time steps and pumping tests in hydraulic tomography analysis provide equivalent inferences for the inversion, leading to the same weight in the loss function. Although Eq. (17) is labeled as the inverse problem loss, it is the combined loss of the inverse problem and all forward pumping tests. To minimize this loss, we train all networks simultaneously even though it looks more complicated. However, since model coefficients in $NN^i(x,y,t)$ and $TNN(x,y)$ are randomly initialized, their predictions at the beginning are usually far from the reference field. Training one network with meaningless predictions from other networks probably misguides the optimizing direction, which makes it harder to achieve co-convergence. When the inverse problem loss is minimized, HT-PINN can provide an inverse network for estimating hydraulic transmissivity as well as an ensemble surrogate forward model for approximating hydraulic heads under different pumping tests. Fig. 1 shows the flowchart and structure of HT-PINN, including all the loss functions presented above.

2.4. Batch training of collocation points

We design batch (or mini-batch) training for the developed HT-PINN inverse model to solve large-scale inverse problems for estimating high-resolution hydraulic conductivity fields, up to millions of unknowns. Batch training has been commonly used to prevent DNN from overfitting and shorten the training time for big data problems (Jacob et al., 2022). However, in groundwater inverse problems, measurements including both direct and indirect measurements are not large. The primary problem is the high-resolution parameter field to be estimated, which corresponds to computationally expensive forward model simulations. For PINN applications to estimate a high-resolution hydraulic conductivity or permeability field with up to millions of unknowns, it is impossible to include the PDE constraints at all grids. Even a small fraction such as 10 % is computationally expensive. Thus, instead of collocating all grids in one batch, we use a series of mini-batches containing sampled non-pumping grids. Three hyperparameters need to be determined: batch size, number of batches, and sampling method. A large batch size may lead to global optima at the cost of slow convergence (Ioffe and Szegedy, 2015; Li et al., 2021; Masters and Luschi, 2018; Wilson and Martinez, 2003). On the other hand, using mini-batches may have fast convergence, but may not be guaranteed to converge to the global optima. Here, we use three methods for generating batches:

(1) Random and non-recurrent: generate a new batch at each iteration through random sampling from non-pumping grids;
(2) Random and recurrent: generate a certain number of batches by randomly sampling before training and use these batches repeatedly in subsequent training iterations;
(3) Uniform + recurrent: uniformly sample non-pumping grids to generate a certain number of batches and use these batches repeatedly in subsequent training (McCandlish et al., 2018).

3. Numerical experiments

3.1. High-dimensional HT base experiment

A two-dimensional base experiment of HT is implemented in a large-scale high-resolution (1024 × 1024) transmissivity field. The experimental lnT field is generated with a Gaussian spatial covariance function. The generation method implements PCA decomposition on the covariance matrix. 50 top-ranked principal components are retained as the projection matrix, and normally distributed random variables are generated as projections, which retains over 90 % of the total variance (Fig. 2B). The field is scaled to a unit square domain so that the spatial coordinates are dimensionless ($x \in [0,1], y \in [0,1]$). The top and bottom boundaries are impermeable (Neumann boundary) and left and right boundaries are constant-head (Dirichlet boundary). Initially, hydraulic head on the field is 0 m. The geostatistical and hydrogeological parameters are listed in Table 1. The geostatistical parameters for the Gaussian covariance model are similar to previous PINN applications (Tartakovsky et al., 2020). The well network for the HT survey consists of 25 wells, numbered from p1 to p25, evenly distributed in the central area of the domain (Fig. 2A). Pumping events are performed sequentially in 5 pumping wells located at the center (p13) and corners (p1, p5, p21 and p25). For each pumping test event, water is withdrawn at a constant rate from one of the 5 pumping wells, and hydraulic heads are observed from the other 24 monitoring wells as indirect measurements. For transient data, hydraulic heads are evenly observed 10 times within 1 h. The time interval between two consecutive observations is 0.1 h (time input can also be seen as dimensionless given that the scaler is 1 h). For steady-state data, hydraulic heads are only observed once at the steady-state phase. Random noises with a variation of 5 % from the true value are added to corrupt both these indirect measurements and direct measurements of transmissivities. Values of direct and indirect measurements are normalized through dividing by L2 norm before being used by HT-PINN as reference data.

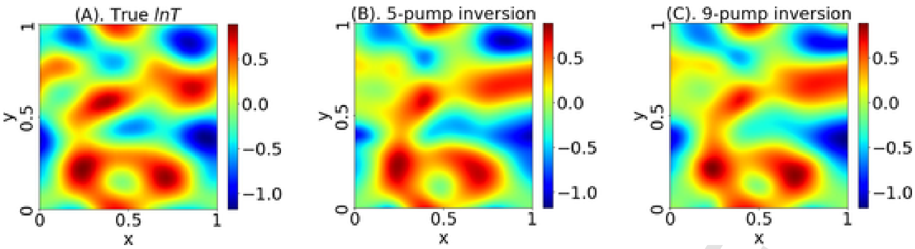

Fig. 9. Inverse estimation results of HT-PINN with a different number of pumping tests (indirect measurements). (A) True *lnT* field; (B) inverse results by HT-PINN with 5 pumping tests; (B) inverse results by HT-PINN with 9 pumping tests.

3.2. HT-PINN implementation

For the steady-state HT experiment described above, the developed HT-PINN contains 5 forward neural networks, $NN^i(x,y,t)$, $i = 1, 5, 13, 21, 25$, corresponding to different pumping well, and 1 inverse network, $TNN(x,y)$. All networks have 6 fully connected layers with tanh activation functions, each containing 50 hidden units. The sample grids for evaluating $Loss_N^i$ and $Loss_D^i$ are uniformly distributed on Neumann (top–bottom) and Dirichlet (left–right) boundaries with N_D and $N_N = 128$. According to the pumping schedule, the well grids are separated as samples for $Loss_p^i$ and $Loss_m^i$ with $N_p = 1$ and $N_m = 24$. Direct measurements of transmissivity are uniformly distributed in the domain with $N_T = 61$ (Fig. 2 B).

In the developed HT-PINN, a relax region is defined as a square area (30 m × 30 m) centered on the pumping well. Xu et al. (2021a) discussed that the solution in a pumping test always has control volumes near the pumping point (source/sink) with significantly larger PDE residuals than the outer region. Thus, the loss function may be dominated by the collocation points in this region. Setting the relax region is a tradeoff between the PDE regularization and the solution smoothness. Outside the relax region is the PDE regularization region, where the grid is strictly governed by Eq. (2), i.e., the non-pumping grid. We implement the random and recurrent batching method, i.e., the method (2) discussed in the previous section, to train the HT-PINN. We generate 10 batches, each containing 900 randomly sampled non-pumping collocation points with PDE constraints for each pumping test. These collocation point batches are recurrently used to evaluate $Loss_e$ with $N_e = 900$ at each iteration. In total, we include 9,000 PDE-constrained collocation points for each pumping test and 45,000 collocation points for all the five pumping tests.

For the transient HT experiment, we have 10 determined time steps $t = 0.1, 0.2, \ldots, 1.0$. At each time step, the data composition mainly refers to the steady-state experiment except for the PDE batch size which is reduced to 300 ($N_e = 300$). Aggregately, the number of collocation points in one transient pumping test with 10 time steps will be 30,000 and in the whole experiment will be 150,000. With consideration of unique IC constraints, we add monitoring wells as sample grids for evaluating $Loss_{init}^i$ causing $N_{init} = 24$.

In both types of experiments, neural networks are trained with Adam optimizer on the Google Colab Platform. Training is counted by epoch, each containing 100 iterations for the transient experiment. In each iteration, a specific PDE batch, measurements of hydraulic heads and transmissivities, Neumann and Dirichlet boundary grids, IC grids, and the pumping grid are fed to backpropagation. So one epoch can cover data batches at every time step. HT-PINN is trained for 3000 epochs, and the learning rate decays from 1 × 10^{-3} to 1 × 10^{-4} after 1000 epochs. The weights of each loss term are: $\lambda_m = 10^4$, $\lambda_f = 50$, $\lambda_p = 1$, $\lambda_N = 10^4$, $\lambda_D = 2\times10^4$, $\lambda_T = 10^3$, $\lambda_{init} = 10^4$. These weights are determined to keep each loss term at a similar magnitude to balance their contributions to the total loss (Kingma and Ba, 2017). In a steady-state HT experiment, each epoch only contains 10 iterations to cover the data batches. And only 2000 epochs are needed to achieve comparative convergence with the learning rate decaying from 1 × 10^{-3} to 1 × 10^{-4} after 1000 epochs. The weight of λ_{init} is not considered due to the removal of $Loss_{init}$ and weights of other loss terms are unchanged.

3.3. Quantitative measures

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

$$\epsilon_{NN^i} = \frac{\|\boldsymbol{NN^i}(x,y,t) - \boldsymbol{h^i}(x,y,t)\|_2^2}{\|\boldsymbol{h^i}(x,y,t)\|_2^2}, (x,y) \in \Omega, t \in (0, \mathrm{T}] \tag{18}$$

where $\boldsymbol{h^i}(x,y,t)$ and $\boldsymbol{NN^i}(x,y,t)$ represent the true and approximated hydraulic heads in vector form, respectively. Similarly, the inverse relative residual ϵ_T is evaluated by:

$$\epsilon_T = \frac{\|\boldsymbol{TNN}(x,y) - \boldsymbol{T}(x,y)\|_2^2}{\|\boldsymbol{T}(x,y)\|_2^2}, (x,y) \in \Omega \tag{19}$$

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

$$\varepsilon(x,y) = \frac{|TNN(x,y) - T(x,y)|}{T^{\max} - T^{\min}}, (x,y) \in \Omega \tag{20}$$

4. Results of numerical experiments

4.1. Transient base experiment

4.1.1. Forward model of transient problem evaluation

Fig. 3 shows the performance of the forward model in the transient experiment. Fig. 3A shows the true (first row) and approximated (second row) hydraulic heads at time steps from 0.1 to 1 h in a pumping test with the pumping well located at the left bottom corner (p1) of the well network. The approximated heads are from $NN^1(x,y,t)$ prediction and the true hydraulic heads are solved from a numerical solver. The color map and contour lines show that the NN^1 model prediction agrees with the numerical simulation result at every time step. Specifically, at × = 0 and 1, the NN^1 model reproduces the hydraulic heads of 0 defined by Dirichlet BC, and at y = 0, the contour lines are perpendicular to the boundaries defined by the impermeable Neumann BC. Fig. 3B plots the indirect measurements collected from monitoring wells and model prediction at monitoring wells on the time axis. In the plot, each color indicates a specific monitoring well, stars note indirect measurements, and solid lines note prediction from NN^1. It shows that NN^1 well learns the temporal trends implied by those indirect measurements and initial conditions. At $t = 0$, predicted hydraulic heads are 0 as a result of initial condition constraints. Then, they start dropping down un-

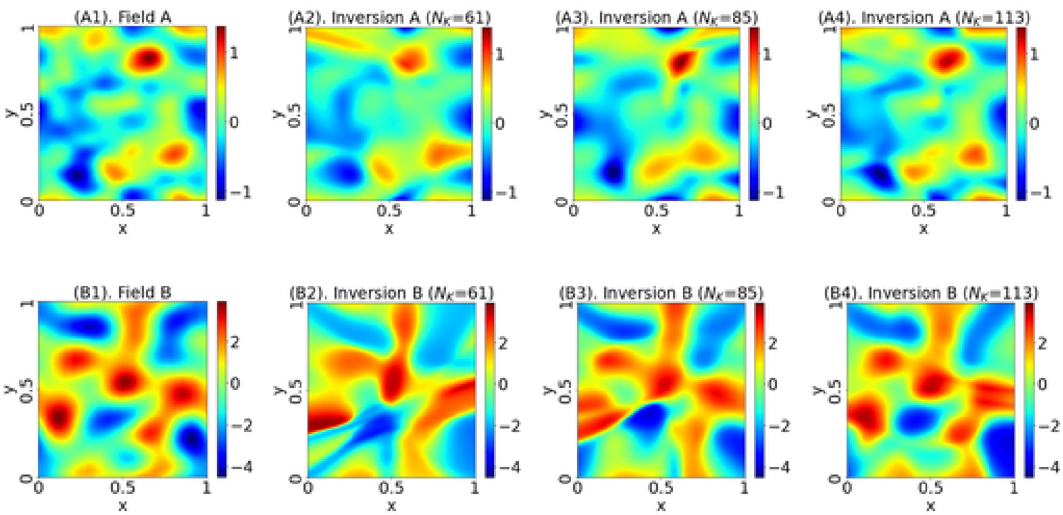

Fig. 10. Inverse results of fields with different structural parameters. (A1) true *lnT* of field A; (A2) – (A4) inverse estimates of field A with 61, 85, and 113 direct measurements; (B1) true *lnT* of field B; (B2) – (B4) inverse estimates of field B with 61, 85, and 113 direct measurements.

til$t = 0.7$. After that, hydraulic heads become unchanged and stay at a steady state. Fig. 3C shows a scattered plot of the whole indirect measurements in all five pumping tests and relative predicted hydraulic heads. The R^2 coefficient of determination is greater than 0.99. The relative residuals of $\epsilon_{NN_t^i}$ for $i = 1, 5, 13, 21, 25$ are 6.14 %, 6.26 %, 6.23 %, 6.58 %, 6.53 %. All five NN models produce a satisfactory prediction of hydraulic heads and are collectively supported as surrogate models for inversion.

4.1.2. Inverse model under transient problem evaluation

Fig. 4 shows the inverse modeling results of *TNN* in the transient experiment. By comparing Fig. 4A and B, the estimated transmissivity field from *TNN* well describes the main distribution pattern of the true field. Fig. 4C shows the scatterplot of true and estimated transmissivities for the entire field with over a million values. R^2 coefficient is greater than 0.92. The relative residual ϵ_{TNN} is 10.32 %, and the map accuracy is 94.93 %. In this study, the number of known hydraulic transmissivity data is only about 0.006 % of the number of parameters to be estimated, which is far less than the number in previous studies (Tartakovsky et al., 2020), but the estimation is sufficiently accurate. The total running time of this experiment is about 34,260 s.

4.2. Steady-state base experiment

4.2.1. Forward model of steady-state problem evaluation

Fig. 5 shows the performance of the steady-state forward model. Fig. 5A shows the numerically solved true steady-state hydraulic heads in the pumping test p21 whose location is at the right bottom corner of the well network. Fig. 5B shows the approximated steady-state hydraulic heads from the corresponding forward model $NN^{21}(x, y)$. The map can be regarded as a stagnant moment in the later phase of a pumping test and a comparison of them shows that the NN^{21} model is in good agreement with the numerical simulation result and follows the defined Dirichlet BC and impermeable Neumann BC explicitly as the model in the transient experiment. One more detail is the perpendicularity of the contour line at $y = 1$, which is not shown in Fig. 3A. Fig. 5C shows the scattered plot of the true and predicted steady-state hydraulic heads in all five pumping tests. The R^2 is greater than 0.99. The relative residuals ϵ_{NN^i} for $i = 1, 5, 13, 21, 25$ are 6.00 %, 9.37 %, 6.57 %, 7.13 %, 8.40 %. The steady-state forward model performs equivalently well as the transient model.

4.2.2. Inverse model under steady-state problem evaluation

Fig. 6 shows the inverse modeling results of steady-state *TNN*. Fig. 6A shows the true *lnT* field. Fig. 6B shows *TNN* estimation which presents the main pattern of the true field. Fig. 6C shows the scatterplot of true and estimated transmissivities on the whole field. R^2 coefficient is nearly 0.95. The relative residual ϵ_{TNN_t} is 9.09 %, and the map accuracy is 96.85 %. This performance is even better than that in the transient experiment. More attractively, the total running time of this experiment is only 2769 s, which is much faster than the transient experiment and computationally efficient for such a large-scale inverse problem. For simplicity, in the following sections, HT-PINN in steady-state experiments is used for further discussion.

4.3. Uncertainty quantification

There are mainly-two uncertainty sources for implementing the HT-PINN: initial guesses of model coefficients and randomly selected batches. The implementation is repeated by 50 times with different initial guesses and batch generation to analyze the model uncertainty. Fig. 7A shows the uncertainty of the evaluation metrics: inverse map accuracy, inverse residual, and forward residual. The mean and standard deviation of each metric are presented as dot and error bars. For map accuracy, the mean and standard deviation are 95.82 % and 1.73 %, respectively. For inverse relative residual, the mean and standard deviation are 9.84 % and 0.78 %, respectively. Overall, with a relatively small standard deviation, HT-PINN performs consistently well in all repeats. Thus, it can be concluded that the initial guess of the network coefficients and the random generation of batches have a limited effect on the model performance. This implies that the developed HT-PINN has a robust structure and contains enough coefficients, which make it very likely to converge to a global minimum. In addition, HT-PINN is also robust to error noise added to the direct (transmissivity) and indirect (hydraulic heads) measurements. The PDE constraints enhance the smoothness of model estimation and robustness, which can overcome the influences of added data noises. Fig. 7B shows the variance map of the best estimate. The variance is quite uniformly distributed, with peak zones slightly larger than the rest of the map. The uniform variance is mainly due to the constraints from the direct and indirect measurements distributed over the domain.

Table 3
Inversion performance for random fields with different structural parameters.

Field	σ^2_{lnT}	λ_x (m)	λ_y (m)	N_T	Accuracy (%)	ϵ_T(%)	$E[\epsilon_h]$ (%)
Field A	1	32	24	61	83.32	18.78	10.31
	1	32	24	85	91.66	14.79	10.71
	1	32	24	113	96.66	9.97	10.68
Field B	5	64	48	61	74.19	57.78	37.23
	5	64	48	85	88.32	48.27	31.38
	5	64	48	113	94.85	26.24	28.10

5. Effect of batching training strategy

A specific batch training method is used for the steady-state model presented above. That is, among all the collocation grids with PDE constraints (greater than1 million), 10 batches are randomly generated with 900 points per batch for each pumping test and recurrently use them for training the HT-PINN. To investigate the impact of batch size, the number of batches and batch generation method, the same transmissivity field and HT data of the base experiment are used and the number of epochs and learning rate are kept the same for the model training. Table 2 summarizes all the experiments and results.

5.1. Effect of batch sampling method

To compare the three batch generation methods listed in Table 2, we choose the same batch number, $N_B = 10$, and batch size, $M_B = 100, 300, 600$. The second method of random sampling and non-recurrent batches can be regarded as an extreme case, in which a new batch is generated for each iteration and the batch number is equal to the number of training iterations. Thus, the total collocation points used in the non-recurrent method are much larger than the recurrent methods. Considering that 15,000 iterations and 5 pumping tests are used in this test, the total size of the collocation points used is actually many times the total grid number. Thus, many collocation points are resampled in the iterative process. The method of uniform sampling is designed to cover the entire domain except for the relax region. The map accuracy of the inversion results of different batch generation methods is very close, all around 95 %, which implies that the impact of different batch generation methods on the inverse estimation accuracy is insignificant. For all batch sizes, the training time of the non-recurrent method is consistently slightly longer than the other two recurrent methods. This is because of the resampling at each iteration and the convergence time for new samples. In addition, all methods exhibit an approximately linear increase in training time with increasing batch size, as shown in Fig. 8A.

5.2. Effect of batch size

We then choose to use the method with recurrent random samples to study the effect of batch number and size. With constant batch number, $N_B = 10$, the batch size M_B is set from 0 to 1200, used in HT-PINN to evaluate non-pumping PDE loss ($Loss_e^i$). We repeat such tests ten times, each time generating new batches and reinitializing the network for uncertainty quantification. Fig. 8B shows the mean and standard deviation for accuracy and training time. Results are also summarized in Table 2 along with the inverse relative residuals. We should notice that when $M_B = 0$, the training data does not contain any non-pumping grids and $Loss_e^i$ is not evaluated. That is, the PDE constraint has no role in the total loss. Therefore, the HT-PINN is relegated to a DNN that only learns inversion from data fitting. The mean accuracy and relative residual of the DNN are hardly as satisfactory as the developed HT-PINN. Additionally, the associated inversion uncertainty is much larger for the DNN, suggesting that the DNN is significantly affected by the initialization of network coefficients. The training time of HT-PINN is longer than DNN because it uses AD to evaluate the partial derivatives in PDE loss terms and takes longer to converge. The results show that the DNN is not as good as the HT-PINN in providing robust predictions given the small number of direct measurements in our experiment, or that the DNN may require much larger data to achieve the accuracy of HT-PINN with physical constraints (LeCun et al., 2015).

Staring from $M_B = 100$, the HT-PINN gains high accuracy and small uncertainty. It illustrates that HT-PINN behaves more robustly to sparse data. At the batch size $M_B = 100$, a total of 5000 PDE grids are collocated, which is about 0.48 % of the total grids in the domain. Since random sampling and PDE regions do not prohibit repetition, there may be duplicate grids sampled and the actual grids with PDE coverage may be lower. However, such a low coverage makes it possible to achieve a high level of mean accuracy and low uncertainty already, and scaling up the batch size does not improve accuracy and reduce uncertainty (Fig. 8B). Similar discussions can be found in other groundwater applications, where only a single batch with different sizes is considered. For example, Tartakovsky et al. (2020) used one batch of 300 PDE grids to invert a groundwater flow field with a resolution of 32 × 32 without pumping. Its coverage is about 30 %. In another groundwater flow and transport application, He et al. (2020) used a batch of only 200 PDE grids to invert a field with a resolution of 256 × 128. The coverage is about 0.61 %. Xu et al. (2021a) used 5000 PDE grids to train a pumping test forward model surrogate in a field with a resolution of 51 × 51. We should note that neither absolute numbers nor coverage can provide us with a constant collocation size threshold, as the results depend on the specific forward PDE problem, field resolutions, and underlying parameter field smoothness. The uncertainty in all cases (standard deviation around 1.5 %) is small, indicating that the network initialization and random sampling have little effect on the inverse estimation. In contrast, training time increases with the number of total PDE collocation grids (Fig. 8B). Thus, we can choose fewer PDE grids since no manifest improvement is seen after $M_B = 100$ (Hoffer et al., 2018; Nitish et al., 2017).

5.3. Effect of batch number

To examine the effect of batch number, we control for the total number of collocated PDE grids and investigate whether training with a small number of large batches or a large number of small batches is superior. The total number of collocated grids for a forward problem is set to 2000, 6000, and 12,000 for $N_B = 20$ batches with $M_B = 100$, 300, and 600, respectively. Considering the HT-PINN performance is quite stable with a small uncertainty, we do not repeat the test and only compare with the average performance of $N_B = 10$ and $M_B = 200$, 600, 900. Fig. 8C shows the comparison and Table 2 lists the specific numbers. There is no significant difference in inversion accuracy within each pair, with the largest difference being around 1 %. However, training the model with 20 small batches takes longer for the same number of epochs. It may be because training with more batches needs to execute more switching batches. In extreme cases, when the batch size is too large, the model performance suffers severely. We generate one batch with 3000 and 9000 grids ($N_B = 1$ and $M_B = 3000$, 9000). The model is trained for 6000 epochs using a single batch with the learning rate decreased from 1×10^{-3} to 1×10^{-4}. The training time is orders of magnitude longer than $N_B = 10$ and $M_B = 300$, 900, but the inverse accuracy is unmatched due to the high-dimensional inverse problem caused by large batches that is harder to converge. In the supplementary material (Fig. S1), we also present a small-scale (64 × 64) field and use all the grids in a single batch to train the model. The training time is also orders of magnitude longer than the high-resolution base experiment presented above.

Table 4
Model performance of RGA and HT-PINN on inversing fields with different resolutions.

Inverse Method	Field Resolution	Accuracy (%)	ϵ_T (%)	Runtime (s)	Iterations /Epochs	Time per iteration /100 epochs (s)
RGA	64 × 64	90.55	34.03	40	7	5.69
	128 × 128	92.93	27.91	202	7	28.87
	256 × 256	90.42	30.92	975	8	121.83
	512 × 512	90.79	42.63	5868	12	489.00
	1024 × 1024	95.90	18.66	30,236	14	2159.71
HT-PINN	64 × 64	94.90	25.91	1902	2000	95.10
	128 × 128	92.96	27.54	1949	2000	97.47
	256 × 256	90.57	17.68	1968	2000	98.40
	512 × 512	90.14	27.4	1832	2000	91.54
	1024 × 1024	90.35	11.6	2001	2000	100.05

6. Data demands for HT-PINN

In this section, we investigate the performance of HT-PINN on different data volumes of indirect and direct measurements. We choose the random and recurrent method for batch generation, with a batch number of 10 and a batch size of 100. The number and distribution of direct and indirect measurements are provided in different scenarios. All data is corrupted with 5 % noise.

6.1. Effect of indirect measurements

For the indirect measurement of hydraulic heads, two pumping scenarios are considered: one scenario is the same as the base steady-state experiment, with pumping tests performed at 5 wells, i.e., p1, p5, p13, p21 and p25, and the other includes pumping tests at 9 uniformly located wells, i.e., p1, p3, p5, p11, p13, p15, p21, p23 and p25. In the second scenario, the number of forward networks and indirect measurements are 9 and 216, respectively. The transmissivity field and direct measurements remain the same as the base experiment, shown in Fig. 2B.

Fig. 9 shows the true transmissivity field and inverse estimates for both scenarios. For the scenario with 5 pumping tests, the map accuracy is 93.58 %, the inverse relative residual is 10.33 %, and the mean of five forward model relative residuals is 10.17 %. In contrast, the inversion performance of the 9-pumping test scenario is slightly better, with a map accuracy of 96.40 %, an inverse relative residual of 9.73 %, and an average of the nine forward model relative residuals of 8.83 %. This is because more indirect measurements and PDE constraints help improve inversion accuracy. Correspondingly, the computation cost increases with the number of pumping tests. In this experiment, the training time is about 80 and 154 s per 100 epochs for 5 and 9 pumping tests, respectively. The training time increases approximately linearly with the number of pumping tests. Thus, using more pumping tests may not always be the optimal plan for training HT-PINN, given the increased computational and experimental cost and slight enhancement in performance (Fig. 10).

6.2. Effect of direct measurements

As discussed by Tartakovsky et al. (2020), the requirement for direct measurements of local transmissivity is related to the spatial distribution characteristics of the field. In the base experiment, 61 direct measurements of transmissivity are used to constrain the parameter field, which is significantly lower than previous research (Tartakovsky et al. (2020)). To investigate the effect of direct measurements, we change the structural parameters to generate random fields with different variance and correlation length, and invert them by feeding HT-PINN a different number of direct measurements. Table 3 lists the structural parameters used to generate two random fields. Both fields have a resolution of 1024 × 1024. Compared with the base experiment, field A has a shorter correlation length and a smaller scale of peak-to-valley regionalization. In contrast, field B has the same correlation length as the base experiment, but with much greater variance, representing a highly heterogeneous field. We consider three cases with the number of measurements, $N_K =$ 61, 85 and 113, corresponding to 0.006 %, 0.008 % and 0.011 % of the parameters to be estimated, respectively, and the locations are uniformly distributed. HT-PINN uses five pumping tests, the same as the base experiment, and the batch generation uses the random and recurrent sampling method with a batch number of 10 and a batch size of 100. Due to the large value range of field B, we modify the network structure by adding an *exponential* activation function to the output layer. The gradient propagating from the *exponential* activation function decays faster, therefore, we train the networks with more epochs and higher learning rates to minimize the loss function. We set the learning rate to 1×10^{-3} for the first 2000 epochs and 1×10^{-4} for the other 1000 epochs.

In Fig. 8(A2), the HT-PINN trained with 61 measurements can locate large patterns in the peak and valley region but loses many details. In contrast, Fig. 8(A3) and (A4) show that as the number of measurements increases to 85 and 113, HT-PINN gradually solves this problem and provides a better description of the distribution details. This is also illustrated by the map accuracy and inverse estimate relative residuals summarized in Table 3. The relative residuals of forward models are comparable for all three cases, indicating that the fitting to the indirect measurements of hydraulic heads are similar because they are all underdetermined inverse problems. Compared with the base experiment, where 61 direct measurements provide high inversion map accuracy, it is advantageous to have more direct measurements to estimate transmissivity fields with shorter correlation lengths.

Fig. 8(B2) – (B4) show the estimation of the highly heterogeneous field, field B. When 113 direct measurements are used as the training data, HT-PINN can provide a much smoother estimation with a higher inverse accuracy. In addition, Table 3 shows that the relative residuals of both inverse estimates and forward models are much larger than the base experiment and all the experiments presented above. This is because in highly heterogeneous transmissivity fields, where both direct and indirect measurements differ by orders of magnitude, optimizing the loss contribution is more challenging when assimilating these different types of data together. Compared to the effect of indirect measurements, we can see that the addition of direct measurements can improve the inversion, especially in highly heterogeneous fields and fields with short correlation lengths. Furthermore, in the supplementary material (Fig. S3), we also show that our base experiment results are better than Kriging results of the direct measurements based on the corrupted data in the base experiment and true covariance function.

7. Model scalability for field resolution

We compare the performance of the developed HT-PINN with batch training techniques and a recently developed reformulated geostatistical approach (RGA) to inversely estimate fields with resolutions ranging from coarse (64 × 64) to fine (1024 × 1024). Except for resolution, other field characteristics including spatial covariance and domain BCs and the pumping test strategy are the same as the base experiment. For HT-PINN, direct measurements of transmissivity are collected at fixed spatial coordinates in the domain, regardless of the field resolution. The method of random and recurrent sampling is used to generate 10 batches of 100 grids with PDE constraints in each forward model. RGA is a gradient-based optimization approach. As a dimensionality reduction method, RGA projects spatially correlated transmissivity on dominant principal components of its covariance matrix and directly estimates these projections. Hence, the number of forward model runs and normal equations to be solved is reduced to the number of retained

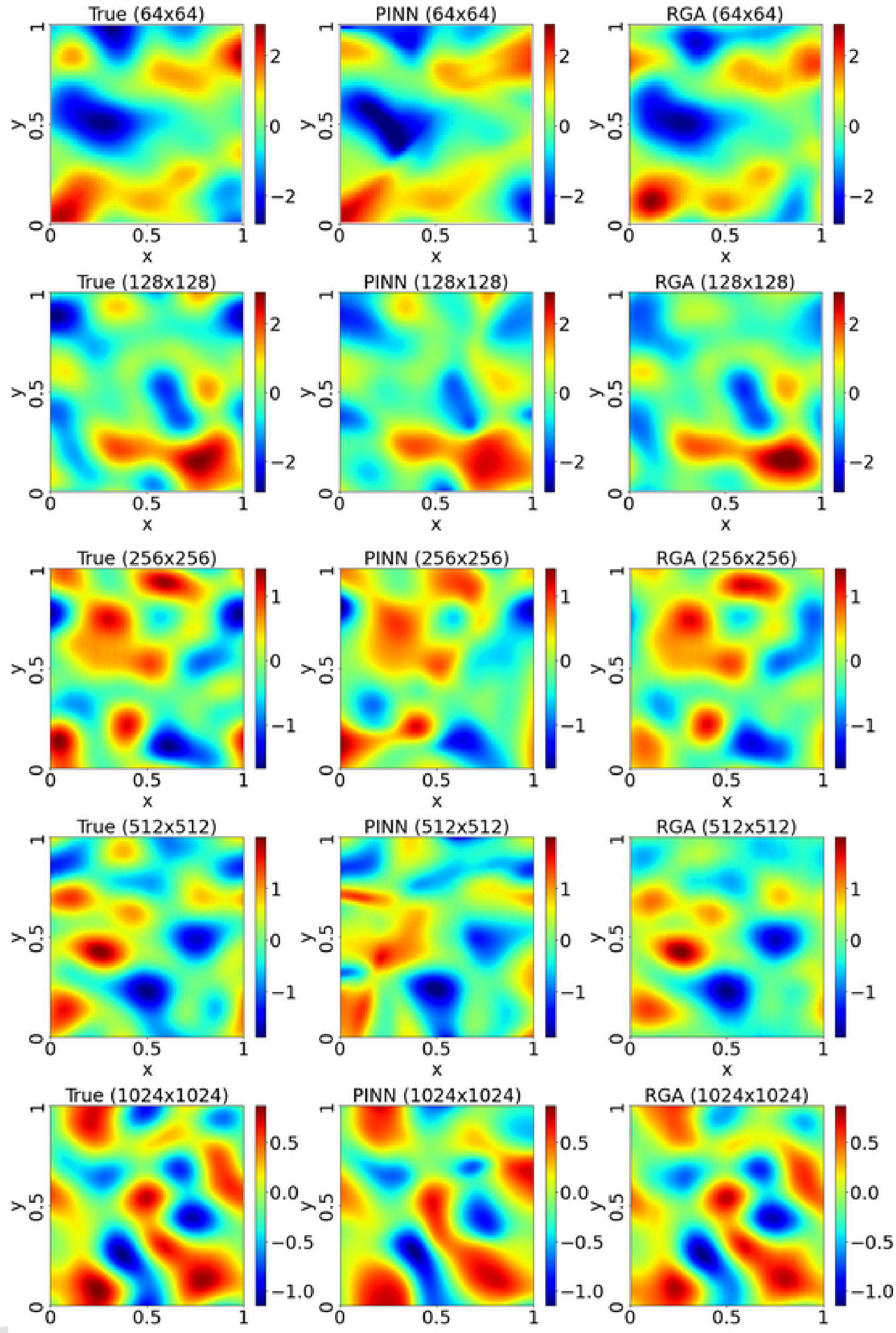

Fig. 11. RGA and HT-PINN inverse results for transmissivity fields with different resolutions. The left column contains the true fields, the center column contains HT-PINN results, and the right column contains RGA results.

principal components, which is usually much smaller than the field dimension. In this experiment, we use the correct covariance model to calculate the principal components. This condition is not necessary because biased spatial covariance can be iteratively corrected (Zhao and Luo, 2021a). The first 100 principal components are retained, which means that there are 100 unknown projections to be estimated. The RGA forward model for simulating hydraulic tomography is a numerical finite element solver that is also used to generate the reference data of hydraulic heads for both HT-PINN and RGA. Computation is implemented on a desktop computer with an Intel® Core™ i7–7700 CPU at 3.60 GHz and 16.0 GB RAM.

Table 4 summarizes the map accuracy, the inverse relative residuals, and running time for each model. Fig. 11 shows the inverse results of HT-PINN and RGA for each resolution. Fig. 12 directly compares the map accuracy and running time of the two methods. Both methods produce high-quality inversions with over 90 % accuracy at all resolutions. For HT-PINN, the map accuracy slightly decreases from the coarse resolution (64 × 64) to the fine resolution (1024 × 1024). This is mainly because we use constant batch numbers and sizes for all resolutions. Therefore, for coarse resolution, the collocation points may cover almost the entire grid, while for fine resolution, they are only sparsely distributed. The accuracy can be further improved with more epochs. However, this also depends on the random field generated, as the same sampling method yields over 95 % accuracy in the high-resolution base experiment (see Table 2). The model scalability is reflected by both the data requirements and trends of the running time.

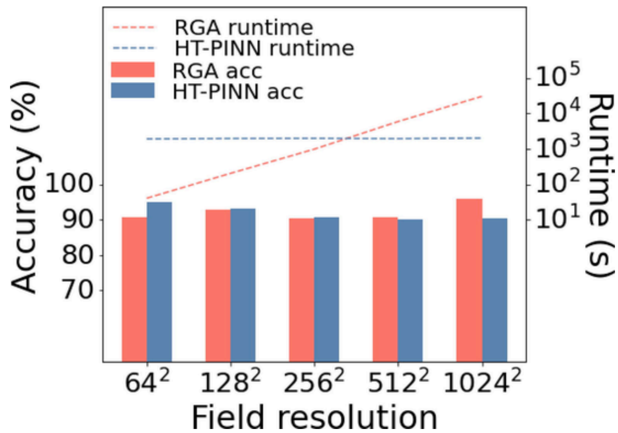

Fig. 12. RGA and HT-PINN scalability on field resolution. Bars show model inverse accuracy, and dashed lines show model running time.

For the iterative method of RGA, the running time per iteration increases with the field resolution, and the number of iterations also increases. This is due to the increased computational cost of forward model simulations required to determine the Jacobian matrix and associated matrix computation. For HT-PINN with batch training techniques, the running time remains almost the same for all resolutions. This is because the mesh-free nature of HT-PINN makes the running time only depend on the batch size and number rather than the field resolution, which is more advantageous for fine-resolution inversion. In terms of total running time, HT-PINN with batch training surpasses RGA after fields finer than 512×512. In fact, according to Zhao et al. (2022), RGA can be implemented with upscaled principal components, i.e., UPCIA, which has already achieved constant runtime for estimating fields with different resolutions. The total runtime of UPCIA depends on upscaling factors. For example, the principal components of dimension 1024×1024 can be upscaled to 16×16 so that a 1024×1024 field can be inverted faster than all the cases above. Even though the developed HT-PINN may not be better than UPCIA on computation efficiency, it still shows broad applicability and robustness to various resolutions considering that the network structure of HT-PINN and the batch generation method remain unchanged at all resolutions. We have also successfully applied the developed HT-PINN to groundwater flow inverse problems without pumping, which have been the main cases discussed in the literature (Fig. S2 in the supplementary material).

8. Conclusion

The developed HT-PINN involves a neural network model for transmissivity and a series of neural network models to describe steady-state and transient sequential pumping tests. It jointly trains all neural network models by minimizing the total loss function including data fitting errors and PDE constraints. New advances and findings include:

(1) To the best of our knowledge, this is the first PINN application to invert the transmissivity field with pumping test data, especially multiple pumping tests. Considering that the pumping test is one of the main field tests for aquifer characterization, the development of HT-PINN is necessary and important, especially for extending PINN to field applications.
(2) We incorporate batch training technique into HT-PINN to accurately and efficiently invert the high-dimensional transmissivity field with over a million unknowns (1024×1024 resolution), a significant advance over the literature. The data requirements are suitable, and the number of direct measurements is only 0.006 % of the estimated parameters in the presented high-dimensional field.
(3) We systematically study the effect of different batch training methods, including batch generation methods, batch number and size. Compared to batch number and size, the batch generation method has a negligible impact on the inversion results and running time. For the Gaussian field used in our experiment, a few batches (10 in our experiments) consisting of hundreds of randomly sampled collocation grids can meet the minimum requirement of PDE constraints and yield satisfactory inversion results. The batch training technique is more efficient than a single batch with the same total collocation grids.
(4) The data requirements for indirect and direct measurements are studied for random fields with different structural parameters. In our experiment, 5 pumping test hydraulic tomography can provide sufficient indirect measurements of the hydraulic head for high accuracy inversion. Performing more pumping tests may only slightly enhance the inversion. In contrast, increasing direct measurements of transmissivity can greatly improve the inversion, as direct measurements are often limited and expensive in the field, especially for highly heterogeneous fields and fields with short correlation lengths.
(5) Compared with the gradient-based RGA, the developed HT-PINN exhibits great scalability in inverting fields with different resolutions due to its mesh-free nature. In specific, the computational cost of HT-PINN remains almost unchanged while maintaining high inversion accuracy for high-resolution fields. Conversely, the computational cost of RGA increases significantly with increasing field resolution due to forward model simulations required to determine the Jacobian matrix. Furthermore, data requirements for HT-PINN do not increase with problem dimensionality. This shows that the developed HT-PINN with batch training technique is particularly effective for large-scale inverse modeling of high-resolution fields.

In addition, the performance of HT-PINN shows higher inversion accuracy compared to DNN with no physics information (i.e., PDE constraints) included in the loss function. It demonstrates that PINN is more capable than DNN in solving inverse problems with sparse data, particularly for problems with well-known forward models (Huang et al., 2022; Wang et al., 2021b; Xu et al., 2021b).

Although the effectiveness of the developed HT-PINN with batch training technique has been improved to be efficient for inverting high-dimensional fields presented, there is still a long way to go before we can claim that HT-PINN is superior to other inverse methods. A serious issue is that the developed HT-PINN relies on direct measurements to constrain the smoothness of model predictions, and the dependence becomes stronger when the geostatistical pattern is complex and variable (Bengio et al., 2006), such as highly heterogeneous and short correlation length. GA relies on the prior information of spatial covariance to regularize the underlying parameter field, which also requires direct measurements to estimate. But it is not necessary to have accurate prior structural parameters because biased spatial covariance can be iteratively corrected (Zhao and Luo, 2021a). That is, it may not be necessary to have many direct measurements to obtain an accurate estimation of the spatial covariance. A feasible solution for HT-PINN is to incorporate additional constraints, such as spatial correlation, and adapt it to inverse problems with fewer direct measurements. However, it is challenging to account for the uncertainty of spatial covariance. Another limitation is that most of the parameter fields used in the literature and this study are smooth, either described by a Gaussian covariance model or only by a few principal components (i.e., dominant large-scale distribution patterns) of an exponential covariance model. Since there is no smoothness regularization in the developed HT-PINN, it has the potential to apply to parameter fields with complex distribution patterns, such as non-Gaussian fields. However, it can be data-hungry and requires much more collocation points to capture small-scale and complex patterns of variation (Dong et al., 2019; Klepikova et al., 2020). These all need to be addressed in the application of PINN for inverse modeling of large-scale complex parameter fields.

CRediT authorship contribution statement

Quan Guo : Methodology, Investigation, Software, Validation, Formal analysis, Writing – original draft. **Yue Zhao :** Validation, Formal analysis. **Chunhui Lu :** Methodology, Writing – review & editing. **Jian Luo :** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Code and data of base experiment are available at: https://github.com/QuanGuo/HT-PINN

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2022.128828.

References

Ambikasaran, S., Li, J.Y., Kitanidis, P.K., Darve, E., 2013. Large-scale stochastic linear inversion using hierarchical matrices. Comput. Geosci. 17 (6), 913–927. https://doi.org/10.1007/s10596-013-9364-0.

Bengio, Y., Delalleau, O., Roux, N., 2006. The curse of highly variable functions for local kernel machines. Adv. Neur. In 18, 107–114.

Bottou, L., Bousquet, O., 2008. The tradeoffs of large scale learning. Adv. Neur. In 20, 161–168.

Broyden, C.G., 1965. A class of methods for solving nonlinear simultaneous equations. Math. Comput. 19 (92), 577–593.

Chen, J., Dai, Z., Yang, Z., Pan, Y.u., Zhang, X., Wu, J., Reza Soltanian, M., 2021. An improved tandem neural network architecture for inverse modeling of multicomponent reactive transport in porous media. Water Resour. Res. 57 (12). https://doi.org/10.1029/2021WR030595.

Chen, J., Dai, Z., Dong, S., Zhang, X., Sun, G., Wu, J., Ershadnia, R., Yin, S., Soltanian, M.R., 2022. Integration of deep learning and information theory for designing monitoring networks in heterogeneous aquifer systems. Water Resour. Res. 58 (10). https://doi.org/10.1029/2022WR032429.

Griewank, A., 2003. A mathematical view of automatic differentiation. Acta Numer. 12, 321–398. https://doi.org/10.1017/S0962492902000132.

He, Q., Tartakovsky, A.M., 2021. Physics-Informed Neural Network Method for Forward and Backward Advection-Dispersion Equations. Water Resour. Res. 57 (7), e2020WR029479. DOI:https://doi.org/10.1029/2020WR029479.

He, Q., Barajas-Solano, D., Tartakovsky, G., Tartakovsky, A.M., 2020. Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. Adv. Water Resour. 141, 103610. https://doi.org/10.1016/j.advwatres.2020.103610.

Hoffer, E., Hubara, I., Soudry, D., 2018. Train longer, generalize better: closing the generalization gap in large batch training of neural networks. arXiv. DOI:https://doi.org/10.48550/arXiv.1705.08741.

Huang, Z., Kurotori, T., Pini, R., Benson, S.M., Zahasky, C., 2022. Three-Dimensional Permeability Inversion Using Convolutional Neural Networks and Positron Emission Tomography. Water Resour. Res. 58 (3), e2021WR031554. DOI:https://doi.org/10.1029/2021WR031554.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: Francis, B., David, B. (Eds.), Proceedings of the 32nd International Conference on Machine Learning. PMLR, Proceedings of Machine Learning Research. pp. 448–456.

Jacob, D., Stern, S., Andrea, M., Douglas, 2022. Learning without a global clock: asynchronous learning in a physics-driven learning. Network. arXiv. DOI:https://doi.org/10.48550/arXiv.2201.04626.

Jagtap, A.D., Kharazmi, E., Karniadakis, G.E., 2020. Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. Comput. Methods Appl. Mech. Eng. 365, 113028. https://doi.org/10.1016/j.cma.2020.113028.

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 Resour. Res. 53 (5), 4444–4458.

Karniadakis, G., 2019. VPINNs: Variational physics-informed neural networks for solving partial differential equations. arXiv. DOI:https://doi.org/10.48550/arXiv.1912.00873.

Kharazmi, E., Zhang, Z., Karniadakis, G.E.M., 2021. hp-VPINNs: Variational physics-informed neural networks with domain decomposition. Comput. Methods Appl. Mech. Eng. 374, 113547. https://doi.org/10.1016/j.cma.2020.113547.

Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. arXiv. DOI: https://doi.org/10.48550/arXiv.1412.6980.

Kitanidis, P.K., 1995. Quasi-linear geostatistical theory for inversing. Water Resour. Res. 31 (10), 2411–2419. https://doi.org/10.1029/95WR01945.

Kitanidis, P., Lee, J., 2014. Principal Component Geostatistical Approach for large-dimensional inverse problems. Water Resour. Res. 50, 5428–5443. https://doi.org/10.1002/2013WR014630.

Klein, O., Cirpka, O.A., Bastian, P., Ippisch, O., 2017. Efficient geostatistical inversion of transient groundwater flow using preconditioned nonlinear conjugate gradients. Adv. Water Resour. 102, 161–177. https://doi.org/10.1016/j.advwatres.2016.12.006.

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

Laloy, E., Hérault, R., Jacques, D., Linde, N., 2018. Training-image based geostatistical inversion using a spatial generative adversarial neural network. Water Resour. Res. 54 (1), 381–406. https://doi.org/10.1002/2017WR022148.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

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 Resour. Res. 50, 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 Resour. Res. 52 (7), 5213–5231. https://doi.org/10.1002/2015WR018483.

Li, M., Zhang, T., Chen, Y., Smola, A.J., 2021. Efficient mini-batch training for stochastic optimization. ACM. doi 10 (1145/2623330), 2623612.

Liu, X., Kitanidis, P., 2011. Large-scale inverse modeling with an application in hydraulic tomography. Water Resour. Res. 47 (2). 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 Resour. Res. 49 (10), 6587–6600. https://doi.org/10.1002/wrcr.20489.

Masters, D., Luschi, C., 2018. Revisiting Small Batch Training for Deep Neural Networks. arXiv. DOI:https://doi.org/10.48550/arXiv.1804.07612.

McCandlish, S., Kaplan, J., Amodei, D., OpenAi, 2018. An Empirical Model of Large-Batch Training. arXiv. DOI:https://doi.org/10.48550/arXiv.1812.06162.

Nitish, Mudigere, D., Nocedal, J., Smelyanskiy, M., Ping, 2017. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. arXiv. DOI:https://doi.org/10.48550/arXiv.1609.04836.

Nowak, W., Cirpka, O.A., 2004. A modified Levenberg–Marquardt algorithm for quasi-linear geostatistical inversing. Adv. Water Resour. 27 (7), 737–750. https://doi.org/10.1016/j.advwatres.2004.03.004.

Nowak, W., Tenkleve, S., Cirpka, O.A., 2003. Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities. Math. Geol. 35 (1), 53–66. https://doi.org/10.1023/A:1022365112368.

Obiefuna, G.I., Eslamian, S., 2019. An evaluation of groundwater storage potentials in a semiarid climate. Earth sciences in the 21st century series. Nova Science Publishers Inc.

Pang, G., D'Elia, M., Parks, M., Karniadakis, G., 2020. nPINNs: Nonlocal physics-informed neural networks for a parametrized nonlocal universal Laplacian operator. Algorithms and applications. J. Comput. Phys. 422, 109760. https://doi.org/10.1016/j.jcp.2020.109760.

Raissi, M., Perdikaris, P., Karniadakis, G., 2017a. Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations. arXiv abs/1711.10561. DOI:https://doi.org/10.48550/arXiv.1711.10561.

Raissi, M., Perdikaris, P., Karniadakis, G., 2017b. Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations. arXiv abs/1711.10566. DOI:https://doi.org/10.48550/arXiv.1711.10566.

Raissi, M., Karniadakis, G.E., 2018. Hidden physics models: Machine learning of nonlinear partial differential equations. Journal of Computational Physics 357, 125–141.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2017c. Inferring solutions of differential equations using noisy multi-fidelity data. J. Comput. Phys. 335, 736–746. https://doi.org/10.1016/j.jcp.2017.01.060.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2017d. Machine learning of linear differential equations using Gaussian processes. J. Comput. Phys. 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. Gas. Sci. Technol 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 Resour. Res., pp. e2019WR026731. DOI:https://doi.org/10.1029/2019WR026731.

Wang, N., Chang, H., Zhang, D., 2021b. Theory-guided Auto-Encoder for surrogate construction and inverse modeling. Comput. Methods Appl. Mech. Eng. 385, 114037. https://doi.org/10.1016/j.cma.2021.114037.

Wang, N., Chang, H., Zhang, D., 2021a. Deep-Learning-Based Inverse Modeling Approaches: A Subsurface Flow Example. J. Geophys. Res.: Solid Earth 126 (2), e2020JB020549. DOI:https://doi.org/10.1029/2020JB020549.

Wang, N., Zhang, D., Chang, H., Li, H., 2020. Deep learning of subsurface flow via theory-guided neural network. J. Hydrol. 584, 124700. https://doi.org/10.1016/j.jhydrol.2020.124700.

Wilson, D.R., Martinez, T.R., 2003. The general inefficiency of batch training for gradient descent learning. Neural Networks 16 (10), 1429–1451. https://doi.org/10.1016/S0893-6080(03)00138-2.

Xu, R., Wang, N., Zhang, D., 2021a. Solution of diffusivity equations with local sources/sinks and surrogate modeling using weak form Theory-guided Neural Network. Adv. Water Resour. 153, 103941. https://doi.org/10.1016/j.advwatres.2021.103941.

Xu, R., Zhang, D., Rong, M., Wang, N., 2021b. Weak form theory-guided neural network (TgNN-wf) for deep learning of subsurface single- and two-phase flow. J. Comput. Phys. 436, 110318. https://doi.org/10.1016/j.jcp.2021.110318.

Yang, L., Zhang, D., Karniadakis, G.E., 2020. Physics-Informed Generative Adversarial Networks for Stochastic Differential Equations. SIAM J. Sci. Comput. 42 (1), A292–A317. https://doi.org/10.1137/18M1225409.

Yang, L., Meng, X., Karniadakis, G.E., 2021. B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. J. Comput. Phys. 425, 109913. https://doi.org/10.1016/j.jcp.2020.109913.

Yang, Y., Perdikaris, P., 2019. Adversarial Uncertainty Quantification in Physics-Informed Neural Networks. J. Comput. Phys. 394, 136–152. https://doi.org/10.1016/j.jcp.2019.05.027.

Yeh, T., Liu, S., 2000. Hydraulic tomography: Development of a new aquifer test method. Water Resour. Res. 36, 2095–2105. https://doi.org/10.1029/2000WR900114.

Zhao, Y., Luo, J., 2021b. A Quasi-Newton Reformulated Geostatistical Approach on Reduced Dimensions for Large-Dimensional Inverse Problems. Water Resour. Res. 57 (1), e2020WR028399. DOI:https://doi.org/10.1029/2020WR028399.

Zhao, Y., Guo, Q., Lu, C., Luo, J., 2022. High-Dimensional Groundwater Flow Inverse Modeling by Upscaled Effective Model on Principal Components. Water Resour. Res. 58 (7), e2022WR032610. DOI:https://doi.org/10.1029/2022WR032610.

Zhao, Y., Luo, J., 2020. Reformulation of Bayesian Geostatistical Approach on Principal Components. Water Resour. Res. 56. https://doi.org/10.1029/2019WR026732.

Zhao, Y., Luo, J., 2021a. Bayesian inverse modeling of large-scale spatial fields on iteratively corrected principal components. Adv. Water Resour. 151, 103913. https://doi.org/10.1016/j.advwatres.2021.103913.

Zhu, Y., Zabaras, N., Koutsourelakis, P.-S., Perdikaris, P., 2019. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. J. Comput. Phys. 394, 56–81. https://doi.org/10.1016/j.jcp.2019.05.024.