Anonymous authors
Paper under double-blind review

REFERENCES

Zoran Babovi and Branislav Bajat. Research in computing-intensive simulations for nature-oriented civil-engineering and related scientific fields, using machine learning and big data: an overview of open problems. *Journal of Big Data*, 10(1):73, 2023.

Philip Brunner, Craig T Simmons, Peter G Cook, and René Therrien. Modeling surface water-groundwater interaction with modflow: Some considerations. *Groundwater*, 48(2):174–180, 2010.

Silvina Caíno-Lores, Andrei Lapin, Jesús Carretero, and Petter Kropf. Applying big data paradigms to a large scale scientific workflow: Lessons learned and future directions. *Future Generation Computer Systems*, 110:440–452, 2020.

Kyeungwoo Cho and Yeonjoo Kim. Improving streamflow prediction in the wrf-hydro model with lstm networks. *Journal of Hydrology*, 605:127297, 2022.

Anshul Choudhary, John F Lindner, Elliott G Holliday, Scott T Miller, Sudeshna Sinha, and William L Ditto. Physics-enhanced neural networks learn order and chaos. *Physical Review E*, 101(6):062207, 2020.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

Li Daolun, Shen Luhang, Zha Wenshu, Liu Xuliang, and Tan Jieqing. Physics-constrained deep learning for solving seepage equation. *Journal of Petroleum Science and Engineering*, 206: 109046, 2021.

AD Galeev, EV Starovoytova, and SI Ponikarov. Numerical simulation of the consequences of liquefied ammonia instantaneous release using fluent software. *Process safety and environmental protection*, 91(3):191–201, 2013.

Ehsan Haghighat, Maziar Raissi, Adrian Moure, Hector Gomez, and Ruben Juanes. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Computer Methods in Applied Mechanics and Engineering*, 379:113741, 2021.

Arlen W Harbaugh. *MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process*, volume 6. US Department of the Interior, US Geological Survey Reston, VA, USA, 2005.

V Hariharan and M Uma Shankar. A review of visual modflow applications in groundwater modelling. In *IOP Conference Series: Materials Science and Engineering*, volume 263, pp. 032025. IOP Publishing, 2017.

Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

Xin Li, Min Feng, Youhua Ran, Yang Su, Feng Liu, Chunlin Huang, Huanfeng Shen, Qing Xiao, Jianbin Su, Shiwei Yuan, et al. Big data in earth system science and progress towards a digital twin. *Nature Reviews Earth & Environment*, 4(5):319–332, 2023.

Reed M Maxwell, Laura E Condon, and Peter Melchior. A physics-informed, machine learning emulator of a 2d surface water model: what temporal networks and simulation-based inference can help us learn about hydrologic processes. *Water*, 13(24):3633, 2021.

Sarvin Moradi, Burak Duran, Saeed Eftekhar Azam, and Massood Mofid. Novel physics-informed artificial neural network architectures for system and input identification of structural dynamics pdes. *Buildings*, 13(3):650, 2023.

COMSOL Multiphysics. Introduction to comsol multiphysics®. *COMSOL Multiphysics, Burlington, MA, accessed Feb*, 9(2018):32, 1998.

CC Niu, SB Li, JJ Hu, YB Dad, Z Cao, and X Li. Application of machine learning in material informatics: A survey. *Materials Reports*, 34(23):23100–23108, 2020.

Amin Pashaei Kalajahi, Isaac Perez-Raya, and Roshan M D'Souza. Physics informed deep neural net inverse modeling for estimating model parameters in permeable porous media flows. *Journal of Fluids Engineering*, 144(6):061102, 2022.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Chao Song, Tariq Alkhalifah, and Umair Bin Waheed. Solving the frequency-domain acoustic vti wave equation using physics-informed neural networks. *Geophysical Journal International*, 225 (2):846–859, 2021.

Zhong Yi Wan, Pantelis Vlachas, Petros Koumoutsakos, and Themistoklis Sapsis. Data-assisted reduced-order modeling of extreme events in complex dynamical systems. *PloS one*, 13(5): e0197704, 2018.

Jiaji Wang, YL Mo, Bassam Izzuddin, and Chul-Woo Kim. Exact dirichlet boundary physics-informed neural network epinn for solid mechanics. *Computer Methods in Applied Mechanics and Engineering*, 414:116184, 2023.

Nanzhe Wang, Dongxiao Zhang, Haibin Chang, and Heng Li. Deep learning of subsurface flow via theory-guided neural network. *Journal of Hydrology*, 584:124700, 2020.

Nanzhe Wang, Haibin Chang, and Dongxiao Zhang. Deep-learning-based inverse modeling approaches: A subsurface flow example. *Journal of Geophysical Research: Solid Earth*, 126(2): e2020JB020549, 2021a.

Nanzhe Wang, Haibin Chang, and Dongxiao Zhang. Theory-guided auto-encoder for surrogate construction and inverse modeling. *Computer Methods in Applied Mechanics and Engineering*, 385:114037, 2021b.

Kai Yu, Lei Jia, Yuqiang Chen, Wei Xu, et al. Deep learning: yesterday, today, and tomorrow. *Journal of computer Research and Development*, 50(9):1799–1804, 2013.

A Related works

With the evolution of scientific paradigms (Caíno-Lores et al., 2020), researches on natural science has mitigated from empirical and theoretical approaches to computational and data-intensive scientific paradigms driven by computer simulations and big data (Babovi & Bajat, 2023; Li et al., 2023). In the field of computational fluid dynamics, numerical models including FLUENT (Galeev et al., 2013), MODFLOW (Brunner et al., 2010; Harbaugh, 2005; Hariharan & Shankar, 2017), and COMSOL Multiphysics (Multiphysics, 1998), have achieved accurate simulations of physical processes like complex multiphase flow, solute transport, and interactions between surface and soil water within groundwater systems. Important issues in those spatial discretization methods for solving the partial differential equations (PDEs) of groundwater are the computationally intensive. Moreover, the implementation of these models is typically challenging for personnel at water resource management agencies, as highly specialized knowledge and extensive experience in numerical modeling are required. Continuous advancements in data-intensive scientific paradigms have empowered deep neural networks to achieve significant breakthroughs in image recognition (Krizhevsky et al., 2017), meteorological modeling (Wan et al., 2018), material informatics (Niu et al., 2020), hydrology (Maxwell et al., 2021), and mathematics(Choudhary et al., 2020), due to the powerful feature extraction (Yu et al., 2013; Hinton & Salakhutdinov, 2006) and complex function approximation abilities (Cybenko, 1989). Nonetheless, as "black box" models, neural networks exhibit a lack of transparency in their decision-making processes and depend significantly on extensive training data, limiting their use in groundwater research. Recently, by combining physical laws with neural networks, the physics-informed neural network (PINNs) method has improved the interpretability of models and diminished reliance on extensive datasets, providing an efficient and accurate approach for groundwater system modeling, and thus promoting the scientific and reliability of groundwater prediction and management decisions.

The concept of PINNs was first proposed by Raissi et al. (2019). In their approach, the neural network was trained by constructing physical constraints, and optimized by using the residuals based on initial conditions, boundary conditions, and governing equations as loss functions. Neural network derivatives were computed through automatic differentiation (AD). This method effectively solves both continuous and discrete PDEs, marking a new era of integration of physical knowledge and deep learning. The introduction of PINNs has significantly expanded the scope of neural networks in scientific research and engineering applications, while markedly enhancing the prediction accuracy, generalization ability, and interpretability of models. Since then, researchers have continued to explore and innovate within the PINNs framework, applying it to various fields such as fluid mechanics (Haghighat et al., 2021), structural dynamics (Moradi et al., 2023), acoustics (Song et al., 2021), and solid mechanics (Wang et al., 2023). In hydrological research, PINNs offer a new perspective for more convenient and efficient prediction of groundwater flow with their unique physical guidance characteristics. Wang et al. (2020) proposed a theoretically guided neural network model (TgNN) based on the loss function of groundwater seepage differential equations, which improved the generalization performance of the groundwater seepage model with limited observational data. Additionally, the team demonstrated that the physics-informed model could effectively use physical information to predict groundwater seepage responses beyond the training set through groundwater seepage transfer learning tasks. Cho & Kim (2022) used an LSTM network to regress the residual differences between a neural network model and the Weather Research and Forecasting model-Hydro (WRF-Hydro). They developed a hybrid model (WRF-Hydro-LSTM) to predict groundwater levels, and experimental results indicated that this hybrid model achieved higher prediction accuracy, less sensitivity to training datasets, and better generalization performance than the LSTM network or WRF-Hydro model alone. Wang et al. (2021a) proposed a neural network constrained by geostatistical information and a physics-guided autoencoder based on convolutional neural networks (CNNs) (Wang et al., 2021b), applying the physical guidance method to the parameter inversion of groundwater hydraulic conductivity fields. The model's accuracy and practicality were further enhanced. The study by (Pashaei Kalajahi et al., 2022) demonstrated that, even with limited data, physics-guided neural networks could accurately estimate key parameters such as hydraulic conductivity and porosity of the medium, showcasing their strong ability to combine data-driven approaches with physical guidance. Daolun et al. (2021) established a new specialized neuron model incorporating pressure gradient information and proposed an algorithm called signpost neural network (SNN), which significantly improves the accuracy of solving unsteady seepage partial differential equations. Although significant progress has been made in improving groundwater seepage models using PINNs, current models still heavily depend on the adequacy and quality of observed data and remain sensitive to outliers. If the training datasets are insufficient or contain corrupted data, the simulation results may exhibit considerable bias. Moreover, after the model has been trained, its applicability is generally limited to specific hydrogeological parameter settings, making it difficult to generalize to broader or unforeseen hydrogeological scenarios, thus restricting the model's versatility and practicality.

B BASIC LAW OF GROUNDWATER SEEPAGE

Without considering changes in water density, the flow of groundwater in a three-dimensional aquifer within porous media can be expressed by the following partial differential equation:

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial H}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}\frac{\partial H}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial H}{\partial z}\right)+W=\mu_s M\frac{\partial H}{\partial t} \tag{1}$$

Where t is time (T); K_{xx}, K_{yy}, and K_{zz} are the hydraulic conductivities (LT^{-1}) along the x, y, and z axes, respectively. H is the hydraulic head (L) at the corresponding space-time point; W is the source/sink term of groundwater (such as precipitation, pumping, etc.); μ_s is the specific storage (L^{-1}) of the aquifer below the free surface. M is the thickness of the aquifer. Combined with fixed solution conditions (boundary conditions and initial conditions), it can reflect the water balance relationship per unit volume and per unit time under Darcy flow conditions.

C Deep Neural Network

The Deep Neural Network (DNN) consists of an input layer, several hidden layers, and an output layer, where each layer contains multiple neurons connected by a weight matrix (which may include bias terms). Consider a deep neural network with L layers, denoted as $Y = f_{net}(X, \theta)$, with n^l neurons in the $L\text{-}th$ layer. The number of neurons in the input layer equals the dimension of the input feature vector X, and the number of neurons in the output layer equals the dimensionality of the output vector Y. The value of the $i\text{-}th$ neuron in $L\text{-}th$ layer, $z_{l,i}$, is computed as the product of the weight $W_{l,i}^T$ and neurons from the previous layer, plus the bias term $b_{l,i}$:

$$z_{l,i} = W_{l,i}^T a_{l-1} + b_{l,i} \tag{2}$$

The output of neurons is nonlinearly transformed by the activation function, which provides nonlinear capability to the network. After processing $z_{l,i}$ with the activation function σ, the output of the $i\text{-}th$ neuron in the $L\text{-}th$ layer, $a_{l,i}$, Lis obtained:

$$a_{l,i} = \sigma(z_{l,i}) \tag{3}$$

The final output of the network is produced by the output layer of the multi-layer neural network, and the network training process relies on a loss function that quantitatively assesses the difference between the network's predictions and the actual target values.

D Performance metrics

To evaluate the performance of the model, three main indicators were used for quantitative analysis: Mean Absolute Error (MAE), Root Mean Square Error ($RMSE$), and Determination Coefficient (R^2). Their definitions are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

Where n represents the total number of samples, y_i represents the reference output from MODFLOW of the $i\text{-}th$ sample, $\hat{y}_i$ represents the predicted value of the $i\text{-}th$ sample by the DNN model, and $\bar{y}_i$ is the average of the reference values of all samples. Ideally, the closer MAE and $RMSE$ are to 0, the smaller the prediction error and the higher the accuracy of the model. The closer the R^2 value is to 1, the better prediction the DNN model achieves.

E Details of PI-RGSM-K model

Given that the PI-RGSM-K model is a specific application of the PI-RGSM model under conditions of a heterogeneous hydraulic conductivity field, and since the structures of both models are similar, most of the experimental settings are identical to PI-RGSM as follows:

(1) As the hydraulic conductivity field has been integrated into the physics-informed loss function, the model input features are spatial-temporal coordinates (x, y, t), canal water levels (H_a, H_b), and source/sink term W. The model still uses a random method to generate input data for self-supervised training (shown in Table 1).

(2) A heterogeneous hydraulic conductivity field is introduced, with K defined as a function that varies linearly with x (i.e., $K = -0.01x + 0.8$). The hydraulic conductivity linearly changes from $0.8m/d$ at the left boundary to $0.4m/d$ at the right boundary.

Table 1: Input features and value ranges of the PI-RGSM-K

Input features	$x(m)$	$y(m)$	$t(d)$	$W(m/d)$	$H_a(m)$	$H_b(m)$
Value ranges	[0,40]	[0,10]	[0,100]	[0,0.007]	[2,3.5]	[2,3.5]