Check for updates

AGU ADVANCING EARTH AND SPACE SCIENCES

JGR Machine Learning and Computation

REVIEW ARTICLE
10.1029/2025JH000703

Special Collection:
Advanced machine learning in solid earth geoscience

Incorporating Deep Learning Into Hydrogeological Modeling: Advancements, Challenges, and Future Directions

Zhenxue Dai[1,2], Chuanjun Zhan[2], Huichao Yin[3], Junjun Chen[4,5], Lulu Xu[1], Yuzhou Xia[1], Songlin Yang[1], Wei Chen[1], Mingxu Cao[1], Zhengyang Du[1], Xiaoying Zhang[1], Bicheng Yan[4], Yue Ma[2], Hao Wang[6], Farzad Moeini[7], Mohamad Reza Soltanian[7,8], Hung Vo Thanh[9], and Kenneth C. Carroll[3]

[1]College of Construction Engineering, Jilin University, Changchun, China, [2]School of Environmental and Municipal Engineering, Qingdao University of Technology, Qingdao, China, [3]Department of Plant & Environmental Sciences, New Mexico State University, Las Cruces, NM, USA, [4]Energy Resources and Petroleum Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [5]National and Local Joint Engineering Laboratory of Internet Application Technology on Mine, China University of Mining and Technology, Xuzhou, China, [6]Water Resources Research Institute of Shandong Province, Jinan, China, [7]Departments of Geosciences, University of Cincinnati, Cincinnati, OH, USA, [8]Department Environmental Engineering, University of Cincinnati, Cincinnati, OH, USA, [9]Waseda Research Institute for Science and Engineering, Waseda University, Tokyo, Japan

Key Points:

- Deep Learning offers advanced capabilities for accurate hydrogeological modeling
- Challenges and limitations of deep learning-based hydrogeological models are discussed
- Future directions for deep learning-based hydrogeological model development are presented

Correspondence to:
C. Zhan, H. Yin and X. Zhang,
zhanchuanjun@qut.edu.cn;
hcyin@nmsu.edu;
xiaoyingzh@jlu.edu.cn

Citation:
Dai, Z., Zhan, C., Yin, H., Chen, J., Xu, L., Xia, Y., et al. (2025). Incorporating deep learning into hydrogeological modeling: Advancements, challenges, and future directions. *Journal of Geophysical Research: Machine Learning and Computation*, *2*, e2025JH000703. https://doi.org/10.1029/2025JH000703

Received 25 MAR 2025
Accepted 11 JUN 2025

Author Contributions:
Conceptualization: Zhenxue Dai, Chuanjun Zhan, Huichao Yin, Xiaoying Zhang, Bicheng Yan, Mohamad Reza Soltanian, Hung Vo Thanh, Kenneth C. Carroll
Formal analysis: Huichao Yin, Junjun Chen, Lulu Xu, Yuzhou Xia, Songlin Yang, Wei Chen, Mingxu Cao, Zhengyang Du, Mohamad Reza Soltanian, Kenneth C. Carroll
Funding acquisition: Zhenxue Dai
Investigation: Zhenxue Dai, Chuanjun Zhan

© 2025 The Author(s). *Journal of Geophysical Research: Machine Learning and Computation* published by Wiley Periodicals LLC on behalf of American Geophysical Union.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract

Hydrogeological modeling is essential for managing groundwater systems, especially in the transport and remediation of contaminants. Traditional modeling methods face challenges due to the increasing complexity and volume of data. Deep learning (DL) has emerged as a promising tool, offering significant improvements in accuracy and efficiency for tasks such as time series prediction, spatial data analysis, and inverse modeling. Although recent applications of DL in hydrogeology have shown potential, many models are still in the testing phase due to limited hydrogeological data, the “black box” nature of DL models hindering interpretability, and the substantial computational resources needed for training. Furthermore, the lack of standardized evaluation benchmarks makes it difficult to compare the performance of different DL models in hydrogeological contexts. To advance DL-based hydrogeological modeling, future research should focus on enhancing data availability through data fusion and public databases, improving model interpretability using physics-informed and explainable DL techniques, and developing more efficient algorithms for training large-scale models. Additionally, exploring new computational paradigms, such as quantum computing, could provide revolutionary solutions for handling the computational challenges associated with training complex models. Establishing standardized benchmarks will also be key for assessing the practical utility of DL models and facilitating their generalization in real-world scenarios. By addressing these challenges and leveraging DL alongside emerging computational technologies, hydrogeological modeling can be significantly advanced, improving the management and remediation of groundwater systems impacted by contaminants.

Plain Language Summary

Groundwater provides drinking water for billions of people worldwide, yet managing and cleaning contaminated groundwater is extremely challenging. Scientists use computer models to understand how water and pollutants move underground, but traditional models struggle with today's massive and complex data. Deep learning (DL), an artificial intelligence method, shows great promise for improving these models. It can help predict how groundwater changes over time, map underground water resources, and identify properties of underground rocks and soil. However, using DL for groundwater studies faces several obstacles: (a) We often do not have enough underground data to properly train these computer models, (b) it is hard to understand how DL makes its predictions, (c) training these models requires expensive computing and time, and (d) we lack standard ways to test whether DL models work well for different groundwater problems. Solutions include: (a) combining different types of data and share databases openly, (b) developing DL models that follow known physics laws and can explain their reasoning, (c) creating faster and more efficient algorithms, and (d) establishing standard tests to compare different DL models fairly. By combining DL with new computing technologies and groundwater expertise, we can better protect and clean our vital underground water supplies.

Methodology: Zhenxue Dai, Chuanjun Zhan, Xiaoying Zhang
Visualization: Junjun Chen, Lulu Xu, Yuzhou Xia, Songlin Yang, Wei Chen, Mingxu Cao, Zhengyang Du, Xiaoying Zhang, Yue Ma, Hao Wang
Writing – original draft: Zhenxue Dai, Chuanjun Zhan, Huichao Yin, Junjun Chen, Lulu Xu, Yuzhou Xia, Songlin Yang, Wei Chen, Mingxu Cao, Zhengyang Du, Xiaoying Zhang, Bicheng Yan, Yue Ma, Hao Wang, Farzad Moeini, Mohamad Reza Soltanian, Hung Vo Thanh, Kenneth C. Carroll
Writing – review & editing: Zhenxue Dai, Chuanjun Zhan, Huichao Yin, Xiaoying Zhang, Bicheng Yan, Yue Ma, Hao Wang, Farzad Moeini, Mohamad Reza Soltanian, Hung Vo Thanh, Kenneth C. Carroll

1. Introduction

Groundwater, a vital and indispensable resource, plays a crucial role in sustaining drinking water, ecosystems, supporting agriculture, and providing industrial water supplies (Hilton & Jasechko, 2023; Scanlon et al., 2023; Soltanian et al., 2024). Effective management and understanding of groundwater systems are essential to address pressing challenges such as water scarcity, contamination, and ecosystem preservation (Cuthbert et al., 2019). Hydrogeological modeling has served as a cornerstone for achieving these objectives, allowing researchers and practitioners to simulate and predict the behavior of groundwater systems under various conditions and into the future (Hammond et al., 2014; Prommer et al., 2003; Pruess et al., 1999; Ren et al., 2022; Wallace et al., 2021).

Since the introduction of Darcy's Law in the mid-19th century, scientists have employed mathematical equations to describe the movement of groundwater, solutes, and other fluids in subsurface environments (Gómez-Hernández & Wen, 1998; Kitanidis, 1997; Neuman, 1972; Rubin & Journel, 1991). The development of hydrogeological models has progressed profoundly over time, with the advent of computer technology in the 1960s marking a turning point (Gómez-Hernández et al., 2017; Šimůnek et al., 2016). Process-based hydrogeological simulation models emerged during this period and have since undergone continuous evolution, refinement, and diversification. Traditional numerical-modeling simulation tools, such as MODFLOW, TOUGH, FEFLOW, and HYDRUS, have become essential in hydrogeology, environmental science, and engineering, offering critical insights for better understanding and managing groundwater systems (Prommer et al., 2003; Xu et al., 2006; Zheng & Wang, 1999).

As the data acquired in the hydrogeology field continues to grow and the problems faced become more complex and multi-scale, traditional hydrogeological modeling methods struggle to cope with the vast amounts of heterogeneous data generated by monitoring networks and remote sensing technologies (Divine et al., 2024; Tsai et al., 2021). Typically, traditional hydrogeological models are limited to handling a single type of data, whereas other data types must be transformed into specific formats or parameters required by the model, often relying on prior knowledge. In some cases, valuable data cannot be effectively transformed into the format needed by the model. For instance, when using stochastic generation models for aquifer structure, geophysical data, borehole data, and hydrological observations must be converted into a format that represents the spatial distribution of lithofacies, similar to borehole data (Carle, 1999; Yeh et al., 2015). Geophysical signals must be transformed into equivalent lithofacies distribution data to be input into the model, whereas hydrological observation data often cannot be directly converted into lithofacies distributions (Jamil et al., 2024). This limits the capability to effectively integrate valuable information for accurately predicting complex groundwater behavior across multiple scales (Sun & Scanlon, 2019). Traditional models typically rely on simplifying assumptions and may not adequately represent the intricate and nonlinear interactions among various subsurface processes, leading to less accurate predictions and limited scalability (Fiori et al., 2010; Rubin & Journel, 1991). Furthermore, traditional models often require significant computational resources and time, which can be a major obstacle for real-time analysis, scenario evaluation, and decision-making (Asher et al., 2015; B. Chen et al., 2018; Luo et al., 2023). The growing complexity of hydrogeological problems and the need for timely and accurate predictions necessitates the development of more advanced modeling approaches.

In the past decade, the rapid advancement of data fusion and artificial intelligence (AI) technologies has led to their integration with hydrogeological numerical simulations, giving rise to a new generation of hydrogeological modeling methods capable of more accurately predicting the behavior of groundwater systems across a wide range of modeling tasks (Shen, 2018; Tahmasebi et al., 2020). Deep learning (DL) algorithms, a subfield of AI, have demonstrated unparalleled capabilities in various applications such as image recognition, natural language processing, and predictive analytics. Their capability to handle large, complex data sets and automatically capture intricate nonlinear relationships (Vereecken et al., 2022) makes them well-suited for tackling the challenges of hydrogeological modeling.

Leveraging the power of DL models, researchers can efficiently process large amounts of data, enabling the assimilation of diverse sources of information to improve the accuracy and reliability of groundwater predictions (X. Li et al., 2023; Q. Zhang et al., 2023; X. Zheng et al., 2023). Unlike traditional methods, DL models can directly extract effective information from original measurement, without the need for prior simplifying assumptions, allowing for more accurate and realistic simulations of complex hydrogeological problems (Bao et al., 2020; Han et al., 2022). For instance, in aquifer structure identification, conventional two-point or multipoint statistical models often rely on specific assumptions, such as stationarity, to represent the spatial variability

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of the aquifer (Dai et al., 2019). However, generative deep learning models, such as generative adversarial networks (GANs) and variational autoencoders (VAEs), eliminate the need for such assumptions and are capable of effectively identifying heterogeneous characteristics of various types of aquifers. This enables the models to capture more complex spatial patterns and nonstationary behaviors inherent in real-world subsurface environments, thus providing a more nuanced understanding of aquifer systems (Zhan et al., 2023). This data-driven approach allows DL models to capture the intricate heterogeneity and nonlinear dynamics of subsurface systems, which are often challenging to represent using traditional process-based models. Additionally, DL models fully utilizes the computational power of modern hardware accelerators, such as graphic processing units (GPUs) and tensor processing units (TPUs) (Haensch et al., 2019), offering computational efficiency far beyond that of traditional numerical models. This is crucial for real-time analysis and informed decision-making in groundwater management.

However, as DL models have been more frequently applied to hydrogeological simulation tasks over the past few years, some issues have become increasingly apparent. For example, in many tasks of application, it is challenging to obtain the large amount of high-quality data required for effectively training DL models (C. Zhan et al., 2024). Hydrogeological data are often sparse, heterogeneous, and subject to measurement errors (Nilsson et al., 2007), posing difficulties in assembling the comprehensive data sets needed for robust model training. Furthermore, the interpretability of DL models remains a concern, as these models often operate as "black boxes," making it difficult to understand the underlying reasoning and decision-making processes. This lack of transparency can hinder the adoption of DL models in critical applications where interpretability and explainability are essential (Tian et al., 2024). There is also a lack of unified evaluation standards and benchmarks for assessing the performance and reliability of DL models in hydrogeological modeling applications, further impeding their widespread adoption (T. Xu et al., 2024). Lastly, issues with the generalizability and lengthy training times of DL models have hindered their practical application (Adombi et al., 2021, 2022). As a result, despite 5–6 years of research on the hydrogeological application of DL algorithms, the vast majority of DL-based hydrogeological models are still difficult to apply in practice.

Given these challenges and the current state of DL applications in hydrogeological modeling, it is necessary to systematically review and assess the progress made in this field. Existing reviews primarily provide broad overviews of DL applications across the entire hydrological domain, rather than focusing specifically on recent advancements in hydrogeological modeling (Ayzel, 2021; Shen, 2018; Sit et al., 2020; Xu & Liang, 2021). Moreover, most reviews aim to describe in detail how various DL models should be applied for the broader scientific community, with less emphasis on the practical challenges and limitations encountered in recent years regarding the application of DL models in this specific field. Additionally, there is a lack of systematic summary and planning for the future development of DL models for related studies. Therefore, this study will not delve into the technical details of various DL models but will instead describe the advantages of applying DL models in hydrogeological simulation, as well as the challenges and limitations faced. Lastly, this study aims to propose potential solutions to address these issues and outlines directions for future research.

2. Advancements of DL in Hydrogeological Modeling

In recent years, the widespread application of DL models has marked a transformative era in hydrogeological modeling. The unique capabilities of DL models have enabled significant advancements in addressing various hydrogeological modeling challenges. This chapter categorizes the application of DL in hydrogeological modeling into three main fields: time series analysis, spatial data analysis, and inverse modeling. We review the applications of DL models in hydrogeological modeling from these three perspectives, highlighting the unique advantages and potential of DL models in relevant studies.

2.1. Time Sequence Prediction

Time sequence prediction is integral to hydrogeological modeling, encompassing tasks such as groundwater level forecasting and mine water inflow prediction. These tasks are pivotal for effective water resource management, environmental planning, and disaster prevention. However, the inherent complexity of temporal relationships in hydrogeological systems, influenced by diverse factors such as climatic variations, geological characteristics, and human activities, poses major challenges for traditional numerical models (Boo et al., 2024). These models, which are typically based on physical processes, often struggle to fully account for the intricate interactions and

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

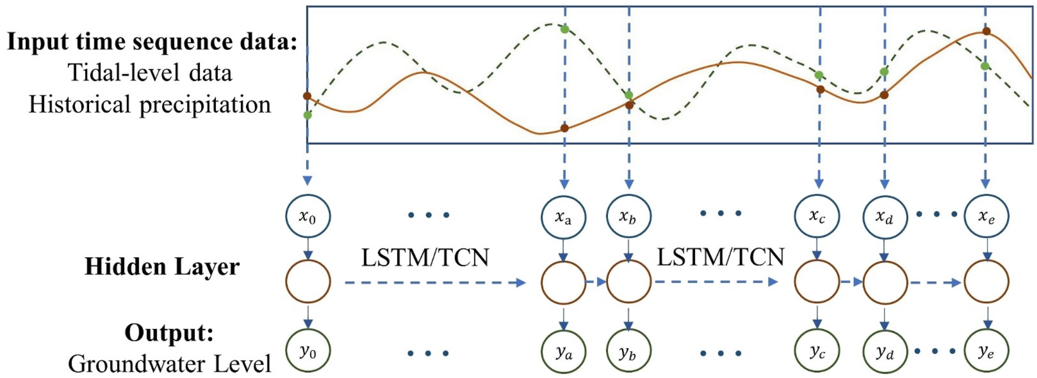

Figure 1. Schematic diagram for predicting coastal groundwater levels with TCN and LSTM (modified from X. Zhang et al. (2023)).

dependencies present in time sequence data, particularly when the data sets are large, diverse, or incomplete (Chen et al., 2020).

DL models, particularly recurrent neural networks (RNNs) and their advanced variants such as long short-term memory (LSTM) networks, have emerged as powerful tools for addressing these challenges (Wunsch et al., 2021). Unlike traditional models, DL approaches can process vast amounts of sequential data, learning directly from the relationships within the data without the need for explicit feature engineering (Nosouhian et al., 2021). This capability enables DL models to capture both short-term fluctuations and long-term dependencies, leading to high-precision predictions in hydrogeological applications.

In groundwater level (i.e., groundwater elevation) forecasting, LSTM networks have been widely applied due to their ability to handle the large volumes of sequential data provided by long-term monitoring networks (Boo et al., 2024; Rajaee et al., 2019). Groundwater levels are influenced by various dynamic factors, such as precipitation, evapotranspiration, and land-use changes, which exhibit both periodic and nonlinear behaviors. Traditional models, although effective for localized and relatively simple systems, often fall short in accurately predicting groundwater levels when faced with complex or large-scale data sets. LSTM models, by contrast, can efficiently process historical observations and related time sequence to uncover and learn the intricate temporal patterns embedded in the data (Tao et al., 2022) (See Figure 1). Studies have demonstrated that LSTM models outperform traditional numerical approaches in both accuracy and adaptability, making groundwater level forecasting one of the most successful applications of DL in hydrogeology (Ali et al., 2024; F. Cui et al., 2022).

Similarly, mine water inflow prediction has greatly benefited from the application of DL models (Chen et al., 2020; Mahmoodzadeh et al., 2021). Accurate forecasting of mine water inflow is essential for ensuring the safety and sustainability of mining operations, as it helps to prevent water-related hazards and optimize resource management. Traditional prediction methods, such as empirical equations and simplified flow models, are often limited by their inability to capture the nonlinear and dynamic nature of water inflow processes, which are influenced by factors such as geological structures, mining activities, and groundwater recharge. DL models, particularly LSTM and temporal convolutional networks (TCNs), have demonstrated superior performance in this domain by effectively learning the complex relationships among these factors as seen in Figure 2 (S. Yang et al., 2023).

One of the key strengths of DL models in time sequence prediction is their capacity to process diverse and large-scale data sets (Sun et al., 2019). For instance, in both groundwater level forecasting and mine water inflow prediction, DL models can integrate data from various sources, such as monitoring stations, satellite observations, and environmental records, to build a comprehensive understanding of the system dynamics (Mo et al., 2022; Moudgil & Rao, 2023; Nourani et al., 2024). Traditional process-based models often require extensive preprocessing and calibration to incorporate such diverse data sources, whereas DL models can seamlessly learn from these inputs in an end-to-end manner. This flexibility significantly enhances the applicability and efficiency of DL models in real-world hydrogeological scenarios.

10.1029/2025JH000703

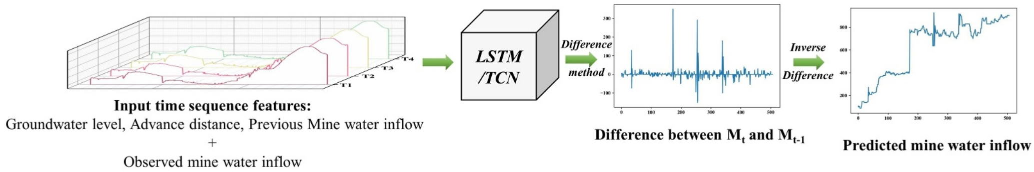

Figure 2. Schematic diagram of predicting mine water inflow process using DIFF-LSTM/DIFF-TCN models (modified from S. Yang, Song, et al. (2023)). M_t and M_{t-1} denote the mine water inflow at time t and time t-1, respectively.

DL models excel in capturing the temporal patterns and dependencies that are crucial for understanding hydrogeological processes (Wunsch et al., 2022). Time sequence prediction tasks typically involve both long-term trends and short-term fluctuations, which are challenging for traditional models to simulate simultaneously. LSTMs, with their specialized architecture, are particularly adept at retaining long-term dependencies while simultaneously focusing on short-term variations. This capability stems from their core components: A memory cell (cell state) and three regulatory gates—input, forget, and output. The cell state acts as a conveyor for information through time. The forget gate selectively decides which information from previous time steps to discard or keep within the cell state, thereby preserving long-term dependencies. Simultaneously, the input gate controls which new information from the current input is added to the cell state, allowing the network to react to short-term variations. The output gate then determines what information from the cell state is passed on as the output for the current time step, effectively balancing these long-term and short-term influences. This dual capability makes them highly effective for tasks including groundwater level forecasting, where the interactions between seasonal cycles, meteorological events, and anthropogenic influences need to be considered and accounted for comprehensively (Fang et al., 2017). Additionally, DL models demonstrate robustness in handling missing or noisy data (Vu et al., 2021), which is a common issue in hydrogeological monitoring networks. This DL capability will be discussed in detail in Section 4.

In summary, DL models, particularly RNNs and LSTM networks, offer substantial advantages in time sequence prediction for hydrogeological modeling (Bai & Tahmasebi, 2023). Their ability to process large and diverse data sets, learn intricate temporal patterns, and accurately model both short-term and long-term dependencies has made them indispensable tools for tasks such as groundwater level forecasting and mine water inflow prediction (Yin et al., 2023).

2.2. Spatial Analysis

Spatial analysis plays a critical role in hydrogeological modeling, serving as the foundation for understanding the location-specific characteristics and variabilities of subsurface systems. These tasks range from analyzing fine-scale pore structures in digital rock physics to understanding large-scale, complex aquifer heterogeneities (L. Chen et al., 2022; Fernàndez-Garcia et al., 2005; Zhang et al., 2024). Accurate spatial analysis is essential for creating reliable groundwater flow and solute transport models, which are necessary for effective water resource management, environmental protection, and pollution remediation (Carroll et al., 2024; Irving & Singha, 2010; Song et al., 2019). However, traditional spatial analysis methods often face significant challenges when dealing with the complexity and heterogeneity inherent in hydrogeological systems. In this context, DL has emerged as a powerful tool to overcome these limitations, offering substantial improvements in handling spatial data in hydrogeology (Cui et al., 2024; Tang et al., 2022).

Traditional spatial analysis methods, such as two-point and multi-point statistical models, have long been used to interpret subsurface data. These methods typically rely on the statistical relationships between discrete spatial points and use interpolation or stochastic simulations to predict unknown properties (Carle & Fogg, 1997; Sanchez-Vila et al., 2006; Strebelle, 2021; Tahmasebi, 2018; Yeh et al., 2015). Although these methods have been foundational, they are limited by several key factors. First, they often rely on assumptions of stationarity or isotropy, which do not hold in many heterogeneous geological systems. This makes them less effective at accurately capturing the complex and variable nature of subsurface environments (Bonazzi et al., 2022). Second, traditional methods tend to be computationally intensive, particularly when applied to large data sets or high-dimensional spatial data. This computational burden limits their applicability in real-time or large-scale simulations (Harp et al., 2008). Third, many traditional methods depend heavily on domain expertise for parameter

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

selection, model calibration, and interpretation. This dependence introduces subjectivity into the analysis, leading to inconsistencies and reducing the reproducibility of results (Geng et al., 2020).

In contrast, deep learning models, particularly convolutional neural networks (CNNs), have revolutionized spatial analysis in hydrogeology. These models are particularly effective at processing complex and high-dimensional spatial data, such as images or spatial grids, and, owing to their inherent architecture designed to capture spatial structures (Z. Li et al., 2022). CNNs employ convolutional layers with learnable filters that automatically scan the input to detect local patterns. This focus on local regions enables CNNs to thoroughly capture local features. This is critical for hydrogeological modeling, as local heterogeneities, such as variations in permeability, can significantly influence both groundwater flow and solute transport (Dai et al., 2020). Crucially, through weight sharing—applying the same filter across the entire input—CNNs gain two key advantages: they significantly reduce model parameters compared to traditional networks, making them efficient for high-dimensional data, and they achieve translation invariance, enabling them to recognize features regardless of their location (Goodfellow et al., 2016). This structure allows CNNs to learn hierarchical features directly from raw data (Ma et al., 2023; Xiong et al., 2024), initial layers capture basic, low-level features (e.g., edges or gradients), whereas deeper layers progressively combine these to identify increasingly complex large-scale structures and relationships, effectively reflecting the multi-scale heterogeneity inherent in subsurface systems. Unlike traditional statistical methods, which often require assumptions about spatial correlation or explicit human intervention for feature engineering, deep learning, through CNNs, automatically learns representations of spatial relationships from large data sets, making it particularly useful in handling the complexity and heterogeneity of subsurface systems (Y. D. Wang et al., 2021). Deep learning models have been applied to a variety of spatial analysis tasks in hydrogeology, including digital rock image segmentation, mineral facies distribution identification, heterogeneous aquifer structure identification, and hydrogeophysics modeling (Jamil et al., 2024; Kang et al., 2021; S. Song et al., 2023; Sun & Scanlon, 2019; Zhan et al., 2023). These applications demonstrate the power of DL to handle spatial data more efficiently and accurately than traditional methods.

One prominent application of DL in hydrogeology is digital rock image segmentation, which involves analyzing rock samples to understand their pore structures for fluid flow simulation (Karimpouli & Tahmasebi, 2019). Traditional segmentation methods often rely on manual annotation or simple image-processing algorithms, which are time-consuming and prone to errors, especially when dealing with noisy or low-quality images (Saxena et al., 2021; Varfolomeev et al., 2019). In contrast, DL models, particularly CNNs, excel at automatically segmenting rock images into different components, such as pores and minerals, with high accuracy and speed (Wang et al., 2022). These models can process large volumes of rock samples without human intervention, reducing subjectivity and improving consistency. Furthermore, advanced architectures such as U-Net and its variants have been used to segment multi-mineral rock samples, adapting to different mineral compositions and improving the accuracy of petrophysical parameter estimation (See Figure 3). The ability of DL to handle noisy or complex images also allows it to outperform traditional methods in terms of both accuracy and efficiency (Ma et al., 2023).

Similarly, the identification of mineral facies distribution is another crucial spatial task in hydrogeology. Accurate mineral facies mapping is essential for understanding chemical interactions between groundwater and surrounding rock, which affect groundwater quality and solute transport (Chen et al., 2023; Singh & Rao, 2005). Traditional methods for mineral identification, such as optical microscopy or spectroscopy, are labor-intensive and can only identify minerals with clear, distinct characteristics (Mlynarczuk & Skiba, 2017). Recent advancements in DL, particularly in image recognition, have led to significant improvements in mineral facies identification (Hong et al., 2017). CNNs can automatically extract and classify features from large data sets of mineral images, identifying subtle variations in texture and color that would be difficult for human experts to detect (See Figure 4). By using DL models to process mineral images, researchers can more accurately and efficiently map mineral facies distributions, reducing both labor costs and the potential for human mistake (Ramil et al., 2018).

The identification of heterogeneous aquifer structures is another area where DL has shown significant promise. Aquifer heterogeneity is a key factor influencing groundwater flow and solute transport, and accurately identifying aquifer structures is critical for reliable groundwater modeling (Bonazzi et al., 2022; Cirpka et al., 2022). Traditional methods often rely on sparse data from boreholes or geological outcrops, and these data are typically interpolated to generate facies maps. However, this approach, particularly when using two-point statistics such as kriging, can fail to capture fine-scale heterogeneity due to its reliance on pairwise correlations, which may not

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

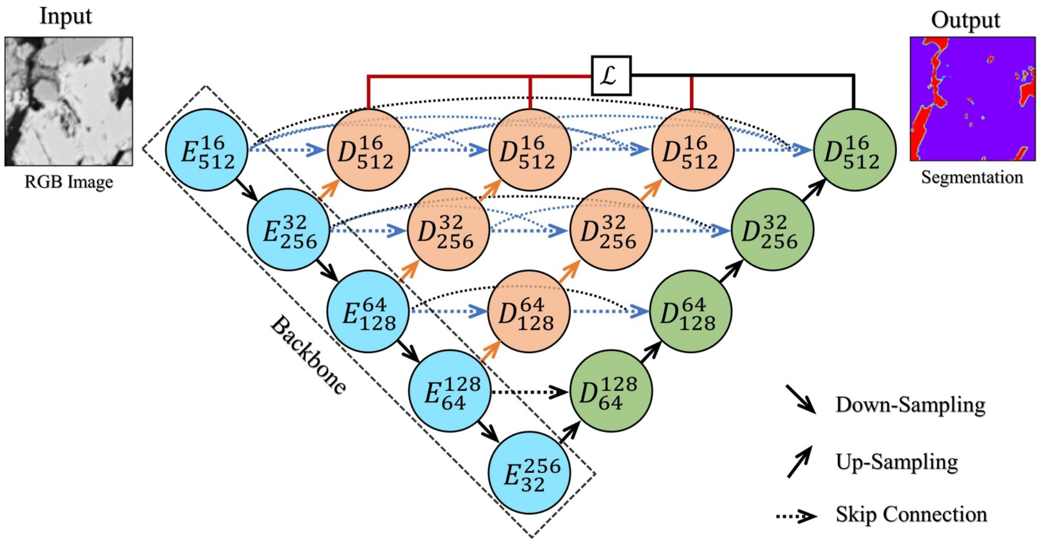

Figure 3. Schematic diagram of the encoder-decoder architecture of a typical segmented CNN network: U-Net++, a U-Net variant with nested and dense skip connections that is widely used in rock image segmentation.

adequately represent complex spatial patterns (Azevedo et al., 2020; El Mezouary et al., 2024). Although multi-point statistics (MPS) methods can utilize image data to capture higher-order spatial structures, they are computationally demanding and less flexible in incorporating multiple constraints. DL models, particularly generative adversarial networks (GANs) and variational autoencoders (VAEs), offer a more robust solution by treating spatial data as images and learning the underlying facies distribution patterns directly from data (Ershadnia et al., 2024; Pan et al., 2023; Zhan, Dai, Soltanian, & Zhang, 2022). These models can generate realistic facies maps that honor observed data while capturing the complex spatial relationships between different facies (See Figure 5). Furthermore, DL models can easily integrate various constraints, such as borehole data and aquifer properties, by modifying the loss function, and they can be efficiently coupled with data assimilation techniques to incorporate indirect measurements, enhancing the accuracy of aquifer structure identification (Moeini et al., 2024; Zhan, Dai, Soltanian, & de Barros, 2022). Additionally, DL models do not depend on specific statistical assumptions, making them suitable for both stationary and non-stationary geological settings.

In hydrogeophysics modeling, DL has proven to be effective in overcoming the limitations of traditional geophysical inversion methods. Traditional hydrogeophysical methods, such as geophysical inversion techniques, often rely on regularization to constrain the inversion process, which can lead to overly smoothed spatial distributions of geophysical parameters, failing to accurately capture the heterogeneity of hydrogeological

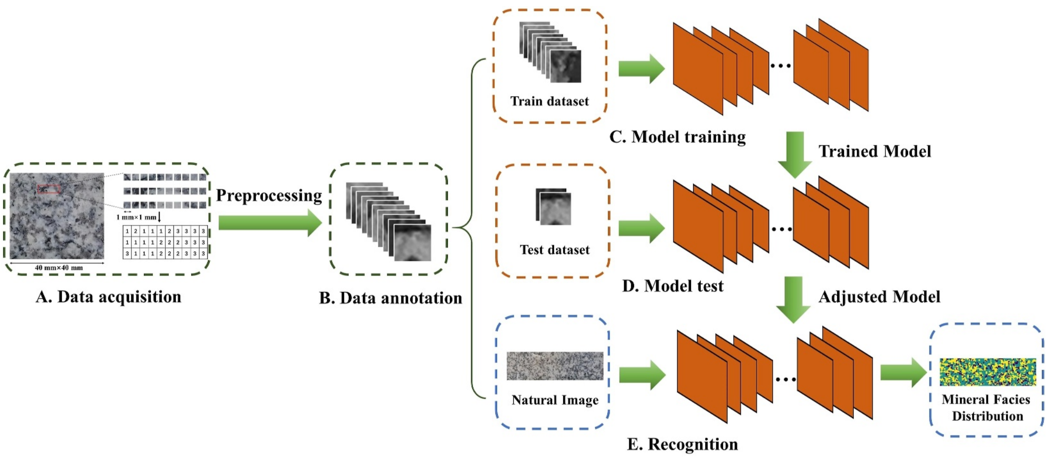

Figure 4. Flowchart of intelligent mineral identification model example.

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

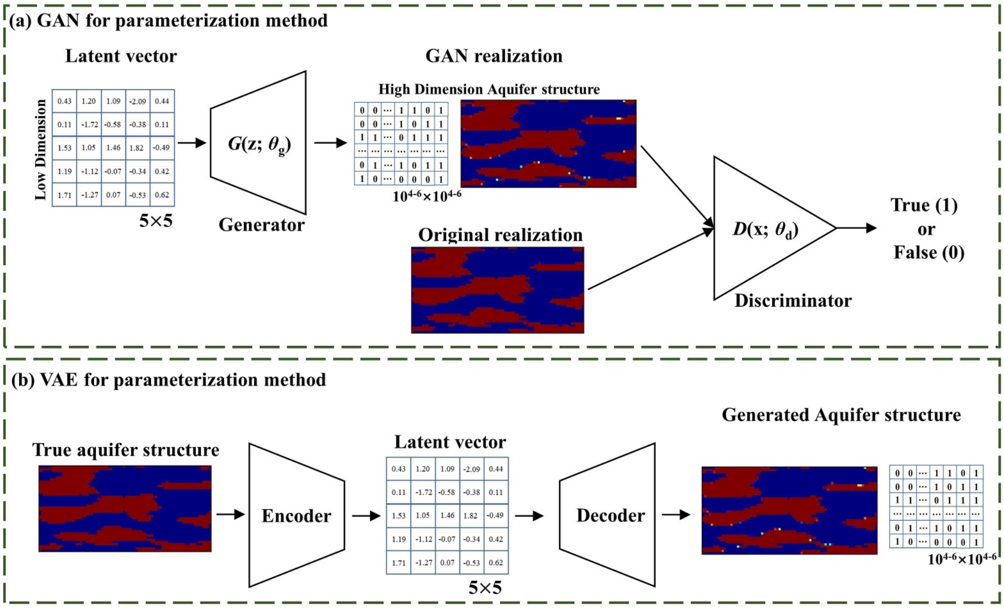

Figure 5. The schematic diagram of parameterization method: (a) GAN; (b) VAE.

systems (Constable et al., 1987; Hyndman et al., 1994; Jamil et al., 2024). These methods also require significant expertise, as practitioners must manually process large volumes of raw monitoring data, resulting in time-intensive workflows that are less adaptable to complex field conditions (Xiong et al., 2024). In contrast, DL models excel by directly mapping raw geophysical monitoring data (e.g., Electrical Resistivity Tomography and Ground Penetrating Radar) to hydrogeological parameters without relying on intermediate petrophysical transformations (See Figure 6). This approach avoids smoothing artifacts and enhances computational efficiency, allowing DL models to capture fine-scale heterogeneity more effectively (Cao et al., 2024, 2025). Additionally, DL-based frameworks, such as encoder-decoder architectures and generative models, streamline the processing of high-dimensional geophysical data, enabling more accurate and efficient integration of these data sets into hydrogeological analyses (S. Li et al., 2020; Tan et al., 2019).

Just as RNNs such as LSTM have advanced the application of DL models in time series analysis in hydrogeological modeling, various CNN-based DL models have similarly propelled their widespread use in spatial data analysis in hydrogeological modeling (Z. Cui et al., 2022; Panahi et al., 2020). CNNs can capture spatial features at different scales through convolutional kernels of varying sizes and pooling layers, enabling DL models to better understand the multi-scale structures in spatial data (Cui et al., 2024; Elmorsy et al., 2022; Liu & Mukerji, 2022). Additionally, since the scale of spatial data structures is often large, traditional models have lower efficiency in processing them. The parallel computing architecture of CNNs allows for efficient computational performance when handling large-scale spatial data (Li et al., 2016; Maggiori et al., 2017).

In hydrogeological modeling, when it comes to spatial data analysis, tasks can be treated as image data can be approached using various CNN-based DL models, such as ResNet, U-Net, and others (Lauzon, 2024; Y. Xu et al., 2024). However, compared to time series data, spatial distribution data are more challenging to obtain and often only reflects partial characteristics of the target, such as rock images obtained by scanning electron microscopy, which only represents part of the rock's features (Cnudde & Boone, 2013), or aquifer outcrop photos obtained through remote sensing technology possibly differing significantly from the actual underground aquifer structure (Bellian et al., 2005; Howell et al., 2014). The difficulty in obtaining training data means that most DL model research in spatial sequence analysis is currently tested using synthetic or laboratory data (C. Zhan et al., 2024). However, as these challenges are addressed and resolved through advancements in data acquisition techniques, such as improved subsurface imaging methods or integration of multiple data sources, DL models are expected to be widely applied in hydrogeological modeling for spatial data analysis in the future, leading to more accurate and reliable predictions of subsurface properties and processes.

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

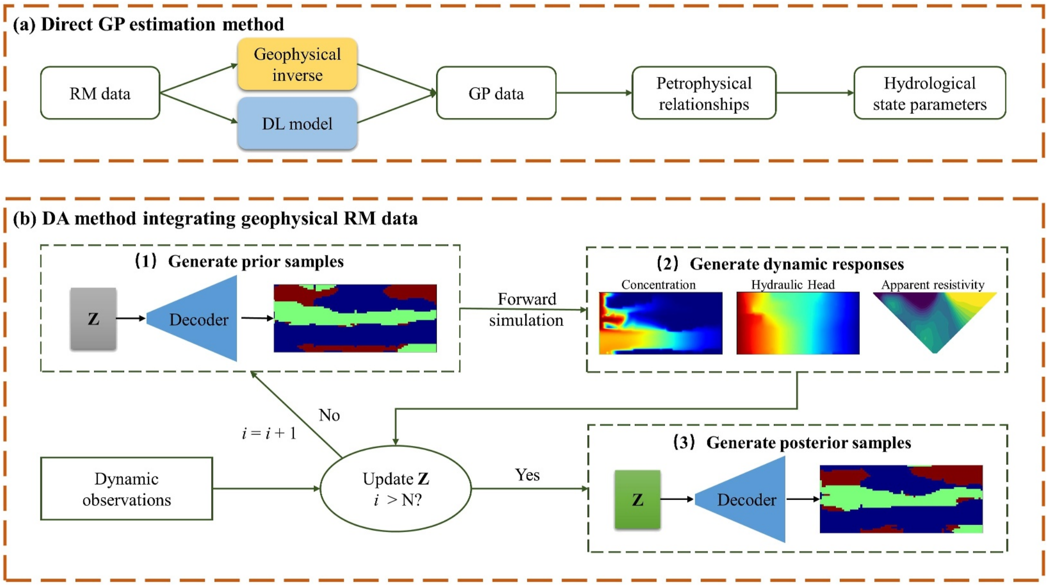

Figure 6. Schematic diagram of integrating DL techniques into hydrogeophysical methods: (a) Direct geophysical parameters (GP) estimation method, (b) DA-based raw monitoring (RM) method.

2.3. Inverse Modeling

In hydrogeological modeling, inverse problems are ubiquitous and crucial for estimating model parameters and understanding subsurface systems. These problems typically involve determining aquifer properties and model parameters that cannot be directly measured, such as aquifer structure, hydraulic conductivity, and transport parameters. Traditional approaches to inverse modeling include local optimization techniques, global optimization algorithms, Bayesian Markov chain Monte Carlo methods, and ensemble Kalman filter variants (Bahrami et al., 2016; Bjarkason et al., 2018; Rajabi et al., 2018; Tsai et al., 2022; Vrugt, 2016). However, these methods face significant challenges that they often require numerous forward model realizations, leading to prohibitive computational costs for complex models, and they struggle with high-dimensional uncertainty in model conditions (Cao et al., 2025). For CPU-intensive high-fidelity numerical models, the computational burden of traditional inversion frameworks becomes particularly problematic, especially when dealing with heterogeneous aquifer systems and complex boundary conditions. Moreover, when dealing with high-dimensional uncertainty in model conditions, these traditional algorithms may struggle to provide reliable parameter estimations due to their inherent limitations in handling complex parameter spaces (Razavi et al., 2012; Yan & Zhou, 2019).

DL approaches have demonstrated remarkable capabilities in addressing these limitations through three primary mechanisms. First, DL models are skillful in accelerating forward simulations through surrogate modeling. Both data-driven and physics-informed neural networks (PINNs) can efficiently approximate complex hydrogeological systems (Tang et al., 2021a; Y. Xu et al., 2024) (See Figure 7). Traditional surrogate methods such as radial basis functions, Gaussian process regression, and polynomial chaos expansion suffer from the curse of dimensionality, where computational costs increase exponentially with input dimensionality (Laloy et al., 2013; Pan et al., 2021). They also face convergence difficulties when dealing with strongly nonlinear relationships (Chen et al., 2021). In contrast, DL model, particularly CNNs, can effectively handle high-dimensional parameter spaces and capture complex nonlinear relationships between model parameters and responses (L. Feng et al., 2024; N. Wang et al., 2023). DL models, with their deep architectures and nonlinear activation functions, possess the universal approximation capability, enabling them to model highly complex and nonlinear relationships inherent in hydrogeological systems. This is particularly advantageous for surrogate modeling, where traditional methods struggle with high-dimensional inputs and strong nonlinearities (Luo et al., 2023). For instance, CNNs are well-suited for processing spatial data, such as parameter fields or concentration distributions, by leveraging local connectivity and shared weights to capture spatial patterns efficiently. Moreover, physics-informed neural networks (PINNs) incorporate physical constraints directly into the learning process, ensuring that the surrogate models adhere to the governing equations of the system (Xu & Zhang, 2024). This not only improves the accuracy

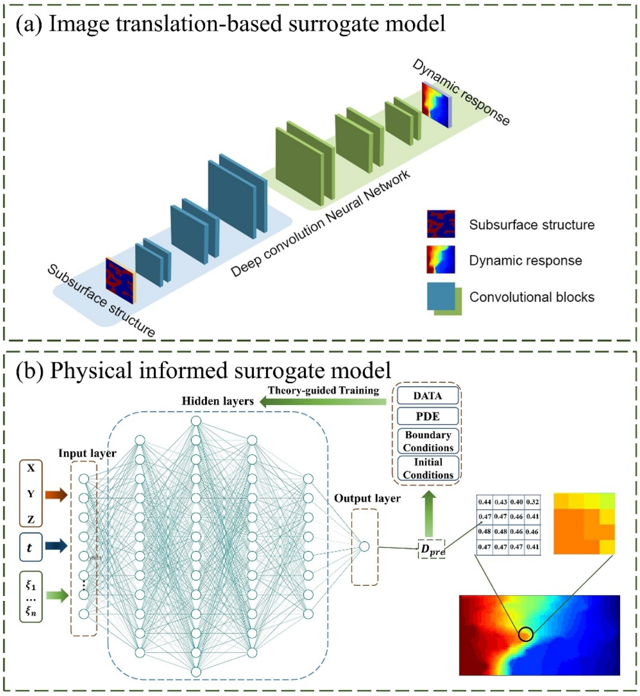

Figure 7. The schematic diagram of: (a) Image-translation-based surrogate model and (b) Physical informed surrogate model (modified from Zhan et al. (2023)).

of the approximations but also allows for better generalization from limited training data, which is often a challenge in hydrogeological applications where data can be scarce or expensive to obtain. Recent studies have demonstrated successful applications of various DL architectures, including U-Net and physics-informed neural networks, in constructing accurate and computationally efficient surrogate models for complex hydrogeological systems (Lauzon, 2024; N. Wang et al., 2024). These DL-based surrogate models have shown remarkable ability in capturing the dominant features of complex systems while maintaining high computational efficiency.

Second, DL offers superior parameter dimension reduction capabilities through generative learning approaches, primarily using GANs and VAEs (Canchumuni et al., 2021; Zhan, Dai, Samper, et al., 2022). These methods effectively map low-dimensional latent vectors to high-dimensional parameter distributions, offering a significant advantage over conventional linear methods such as principal component analysis (PCA). GANs and VAEs leverage the power of deep neural networks (DNNs). The key to their superiority lies in their multi-layered architectures combined with nonlinear activation functions. Without nonlinear activations, a deep stack of layers would mathematically collapse into an equivalent single linear layer. However, by introducing nonlinearities at each layer, DNNs gain the ability to approximate highly complex, arbitrary nonlinear functions (Hornik et al., 1989) and, crucially, to learn hierarchical feature representations—extracting simple patterns at early layers and progressively building more complex and abstract features at deeper layers (LeCun et al., 2015). VAEs (Kingma & Welling, 2014) employ this capability through an encoder (a DNN) that learns a sequence of nonlinear transformations to map high-dimensional parameters onto a lower-dimensional, probabilistic latent space, explicitly designed to represent the data's underlying nonlinear manifold. The training objective, balancing reconstruction loss with a regularization term (like KL divergence), encourages this latent space to be continuous and well-structured, capturing the intrinsic geometry and topology of the data. A decoder (another DNN) then learns the inverse nonlinear mapping from this latent manifold back to the high-dimensional parameter space. GANs (Goodfellow et al., 2014) utilize a different but equally powerful approach. A generator network (DNN) learns to transform simple latent noise vectors into complex, high-dimensional parameter distributions. It is trained through an adversarial process against a discriminator network (DNN) that learns to distinguish real from generated parameters. This dynamic competition forces the generator, with its high capacity for nonlinear mapping, to learn an increasingly sophisticated representation of the true data distribution, implicitly capturing its complex nonlinear manifold structure, including subtle features and dependencies. Therefore, the ability of GANs and VAEs to model these intricate, nonlinear relationships through deep, hierarchical, nonlinear transformations makes them fundamentally more effective than conventional methods such as PCA (Canchumuni et al., 2019) (See Figure 5) for efficient parameter estimation. This is particularly valuable when dealing with non-Gaussian random fields, where traditional dimensionality reduction techniques often perform poorly (Mo et al., 2020). Recent advances in GANs and VAEs, driven by developments in large-scale models and integration with techniques such as self-attention mechanisms and diffusion models, have further improved their ability to characterize aquifer heterogeneity and reduce parameter dimensionality (Cui et al., 2024; Zhang et al., 2024). These improvements have led to more accurate representations of complex geological structures and better handling of spatial correlations in parameter fields.

Third, DL enables the development of reverse networks that directly map observation data to model parameters. This can be achieved through two main approaches: data-driven methods that learn inverse mapping from extensive training samples, and tandem neural network architectures that combine forward and reverse networks (Sun et al., 2023) (See Figure 8). In the context of inverse modeling, DL facilitates the creation of reverse networks that directly infer model parameters from observation data. This direct mapping is achieved through supervised learning, where the network is trained on a data set comprising pairs of parameters and corresponding observations generated from forward simulations. The power of DL lies in its ability to learn highly nonlinear and

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025JH000703

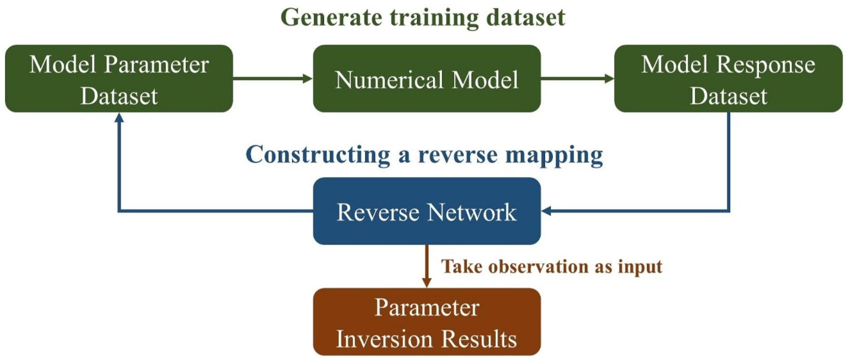

Figure 8. The flowchart for constructing a reverse mapping based on data-driven approaches.

complex mappings, which are prevalent in inverse problems where the relationship between parameters and observations is intricate (Arridge et al., 2019). Unlike traditional optimization-based inversion methods that necessitate iterative forward simulations, DL-based inverse models can provide parameter estimates in a single forward pass once trained, offering substantial computational savings. Additionally, architectures such as tandem neural networks couple forward and reverse networks, allowing for end-to-end training that optimizes both the surrogate model and the inverse mapping simultaneously (Zhan et al., 2025). This integrated approach can lead to improved accuracy and efficiency in parameter estimation (Chen et al., 2021). These approaches eliminate the need for repeated forward simulations during the inversion process, significantly improving computational efficiency.

Furthermore, DL models can be integrated with existing data assimilation algorithms to enhance their performance, either by learning the updating mechanism directly, developing error covariance matrices, or combining results from multiple algorithms (Zhang et al., 2020). Moreover, DL can be seamlessly integrated into existing data assimilation frameworks to enhance their performance. For example, in ensemble-based methods such as the ensemble Kalman filter (EnKF), DL can be employed to learn the error covariance matrices from data, thereby capturing non-Gaussian and nonlinear dependencies that are often approximated in traditional approaches (Cheng & Qiu, 2022). By leveraging historical data or simulations, DL models can provide more accurate representations of the error statistics, leading to better assimilation results. Furthermore, DL can be used to learn the updating mechanism directly, predicting parameter adjustments based on observation discrepancies (Arcucci et al., 2021). This data-driven approach can result in more accurate and efficient assimilation processes, particularly in systems characterized by complex dynamics and high-dimensional state spaces. Additionally, DL facilitates the combination of multiple data assimilation algorithms, allowing for the integration of strengths from different methods to achieve superior parameter estimation (Lang & Qiu, 2021; Tsuyuki & Tamura, 2022). This integration has shown promise in improving the accuracy and efficiency of ensemble-based methods, leading to more reliable parameter estimations in complex hydrogeological systems (He et al., 2025).

The advantages of DL in inverse modeling extend to specific applications such as monitoring network design and contaminant source identification. In monitoring network design, DL approaches serve as efficient surrogate models for uncertainty quantification and can handle non-Gaussian parameter fields and complex uncertainty scenarios (J. Chen et al., 2022; X. Song et al., 2023). Their ability to process both spatial and temporal data makes them particularly suitable for optimizing monitoring networks in heterogeneous aquifer systems. In contaminant source identification, DL models can simultaneously consider multiple parameters, including source characteristics and aquifer properties, providing more comprehensive solutions than traditional methods (Bai & Tahmasebi, 2022; Mo et al., 2019). The capability to establish nonlinear relationships between source parameters and contaminant distributions enables rapid and accurate source identification, which is crucial for environmental remediation efforts (Zheng et al., 2023).

In inverse modeling, DL models fully utilize their advantages in time series prediction and spatial data analysis, and even combine the advantages of these two aspects (Ma et al., 2022). For example, in surrogate model construction, concentration and water level at different times not only have temporal correlations, but the concentration and hydraulic head distribution fields at different times can also be treated as image data (Tang

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

et al., 2021). Thus, DL models can consider both temporal and spatial correlations simultaneously, achieving more accurate predictions. Additionally, the capability of DL models to identify nonlinear relationships also enables their widespread application in identifying and characterizing various unknown relationships, such as the relationship between observation errors and parameter update amounts in data assimilation algorithms (Zhang et al., 2020).

Overall, the application of DL models in inverse modeling mainly brings improvements in efficiency and accuracy. The former benefits from the application of DL models to advanced hardware, whereas the latter mainly relies on a large amount of data to achieve (C. Zhan et al., 2024). The advantages of DL models in the application of inverse modeling are undoubtedly significant, and they have even given rise to entirely new inversion methods, such as inverse networks (N. Wang et al., 2024), which were difficult to achieve with non-DL models. However, because of the difficulty in obtaining training data and the long training time of DL models in practical applications, most of the studies in inverse modeling are still in the testing phase, and there is a lack of research on field applications.

3. Challenges and Limitations of DL-Based Hydrogeological Modeling

DL has proven to be a powerful tool in various scientific domains including hydrogeological modeling due to its capacity to handle large and complex data sets, potentially offering more accurate predictions than traditional methods (Zhi et al., 2024). Although DL has achieved significant success in fields such as autonomous driving, facial recognition, and natural language processing, its application in hydrogeological modeling has not been as straightforward. Most research in this area, particularly over the last 5 years, has been limited to theoretical case studies or laboratory tests. The specific challenges and limitations unique to the deployment of DL in hydrogeological simulation have hindered its practical application in solving real-world hydrogeological modeling problems (Xu & Zhang, 2024). This section aims to provide an objective analysis of various DL applications in hydrogeological simulation. By examining the strengths and weaknesses of these models, we will elucidate the main challenges and limitations currently hindering the application of DL in hydrogeological modeling.

3.1. Data Availability and Quality

Hydrogeological models are complex and need to capture various physical processes that are influenced by many factors such as soil type, rock structure, climate conditions, and human activities, among others (Tao et al., 2022). For DL models to accurately learn and predict these intricate processes, a substantial volume of high-quality data is imperative (Goswami et al., 2022; Lever et al., 2016; H. Wang et al., 2023). The rapid development of DL models has, to some extent, been facilitated by the advent of "big data." Yet, the inherent nature of groundwater systems, being concealed deep beneath the earth's surface, poses a substantial challenge for direct measurement and monitoring (Condon et al., 2021; Kumar, 2014). Typically, data collection for groundwater relies on methods such as drilling and observation wells. The high cost of drilling dictates that groundwater monitoring networks are inherently sparse (Ohmer et al., 2022). In some remote or resource-limited areas, there might be an absence of usable data. Additionally, groundwater data, especially those related to water wells, might be privately owned or subject to privacy constraints (Fitch et al., 2016; Gewin, 2016), further their limiting availability for modeling purposes.

Besides, certain types of data required for hydrogeological models are challenging or impossible to obtain through observational means. The data acquired through existing technologies may not be representative, potentially capturing only a subset of the characteristics of the subject of prediction. For instance, as mentioned in Section 2.3.3, the identification of aquifer structures requires many samples. However, acquiring these samples through current observational techniques is challenging, and the features of aquifer structures obtained through geophysical methods or outcrop observations may not fully represent the true characteristics of the aquifers (Dai et al., 2019; Ikard et al., 2023; Scheibe et al., 2015).

The unique challenge of data scarcity in groundwater systems means that for most hydrogeological modeling tasks, the "big data" desired, and likely required, by DL models is not available. This limits the features that DL models can learn, thereby hindering their full potential (Chilton & Foster, 2024). In scenarios of limited data, the performance of DL models may not be as effective as some traditional process-based models or certain machine learning algorithms that require fewer data samples (Wunsch et al., 2021).

Moreover, even with an abundance of groundwater data, the long-term monitoring data, typically reliant on sensors buried underground, is susceptible to various unforeseeable factors due to the complexity of the subsurface environment. This may lead to measurement errors and potential missing values in groundwater data sets, rendering them incomplete. The potential high error rates in groundwater data, compounded by data scarcity, inevitably introduce bias and uncertainty into the models (Rojas et al., 2008). Therefore, when employing DL for hydrogeological modeling, not only is a cautious approach needed in data collection, cleaning, and preprocessing, but also a focus on techniques for handling missing data and learning from unbalanced or sparse data sets is essential (Fu et al., 2022; Zhou et al., 2020).

3.2. Model Interpretability and Explainability

In applying DL models to hydrogeological modeling, a key issue is their interpretability and explainability (Shen et al., 2023). Interpretability refers to whether a model can clearly show why and how it came to its conclusions based on the data it was given. Explainability involves understanding how the model's design and settings influence its conclusions (Tsang & Benoit, 2023).

In hydrogeological modeling, the goal is not just to predict what will happen, but to understand the physical processes behind these predictions. Traditional hydrogeological models, such as MODFLOW, allow users to see intermediate steps in the modeling process, helping to explain how different factors affect the results. Another useful application is using modeling for hypothesis testing of alternatives through simulation (much lower cost than experimental approaches) which can evaluate the effect of multiple individual processes that may be coupled in natural systems. This can also support investigation of conditions in the past or future. In contrast, DL models often do not provide this level of clarity, making it hard for researchers to understand why the model makes certain predictions (Das & Rad, 2020). This lack of clear explanation can prevent new and important insights into how groundwater systems work.

Moreover, for a tool to be widely accepted, especially in a vital area such as water resource management, those who using it—including policymakers and practitioners—need to trust it. If a DL model works such as a "black box" with its internal workings hidden, it can make people skeptical, slowing down or even preventing its application in real-world groundwater management, despite its potential for highly accurate predictions (Lee & Kam, 2023; Novakovsky et al., 2023; Singha et al., 2020).

When predictions are inaccurate, the challenges to the interpretability of DL models become more evident. If we cannot easily understand the structure of the model or how the settings influence the input, then adjusting the model becomes a guessing game (Saeed & Omlin, 2023). This approach is not only inefficient but also makes it difficult to select the optimal model settings (Lever et al., 2016). The inability to systematically identify and correct model errors will undoubtedly hinder the application of DL-based hydrogeological modeling in practice, as it makes it difficult for users to determine whether the problem lies with the input data, model parameters, or model structure when the model makes incorrect predictions.

Therefore, although DL models bring great hope for more accurate and efficient hydrogeological modeling, their lack of clear interpretability and explicability also poses significant challenges. These issues affect their usefulness for scientific research, their credibility to decision-makers, and their overall effectiveness in practical applications. Overcoming these challenges is crucial for the effective use of DL models in hydrogeological modeling (Jiang et al., 2024).

3.3. Model Performance and Evaluation

Although we have demonstrated the advantages of DL in hydrogeological applications through typical applications in Section 2, there are still certain challenges in achieving the reported performance in practical applications. One of the prominent issues is that DL models are prone to overfitting, especially when there is a lack of enough representative data. Overfitting means that the model performs well on the training data but poorly on new, unseen data (Rice et al., 2020; Ying, 2019).

The poor generalizability of the model is usually caused by both the training data and the model structure. First, for groundwater systems, observation data are mostly limited, and DL models might capture specific patterns of the training data too closely. The reliance of DL models on large data sets and computational power does not necessarily translate into an understanding of hydrogeological principles, thus failing to learn broader,

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

generalizable trends (Nearing et al., 2021). Additionally, DL models might further exacerbate this problem when dealing with small data sets or data sets with many outliers. Most existing studies adopt "shortcut learning" to circumvent this problem, that is, by carefully designing training data and prediction targets to ensure the representativeness of the training data, thus avoiding the problem of poor generalizability. This leads to models that appear effective in controlled test scenarios but perform poorly in real-world applications where conditions are more complex and less predictable (Geirhos et al., 2020). This gap in performance is especially problematic in hydrogeological modeling, where environmental and hydrological processes are constantly changing.

On the other hand, the problem may also come from unreasonable model structure choices and hyperparameter settings. For example, without using the residual network, increase in the depth of DL models could cause an exacerbation of overfitting (Bejani & Ghatee, 2021; Srivastava et al., 2015). The learning rate in hyperparameters, which can be simply regarded as the magnitude of each iteration's update to the network parameters by the DL model, also needs to be matched with the number of training samples each time (Wu et al., 2019).

In the field of hydrogeological modeling, the lack of unified standards for evaluating DL models makes the selection of DL model structures and hyperparameters even more complex. At the same time, the absence of common benchmarks makes it difficult to compare and replicate study results, and researchers often struggle to achieve the performance levels reported in existing literature when applying similar models to their tasks (Cheng et al., 2020; L. Liu et al., 2018). This situation poses challenges for researchers and developers in choosing the most appropriate model or determining effective strategies for model improvement. Much effort is diverted to replicating various algorithms rather than focusing on their application and enhancement.

Therefore, despite the great potential of DL models in improving hydrogeological modeling, their application is hindered by issues such as overfitting, generalization, and the lack of standardized evaluation metrics (T. Xu et al., 2024). Addressing these challenges is vital for ensuring that DL are both practical and reliable for real-world hydrogeological applications.

3.4. Computational Complexity and Resource Requirements

Although DL models excel in complex network structures supported by advanced computational hardware, the scale of data in hydrogeological simulations is vastly larger than typical applications such as image processing. For instance, in solute transport models, grid cells can number in the tens or hundreds of thousands, or even millions—orders of magnitude greater than the pixel resolution commonly used in image-based DL tasks (e.g., 32×32 or 64×64) (Prommer et al., 2003). Image resolutions often result from downscaling higher-resolution natural images. In computer vision, this process is justified because it offers computational gains. It also frequently preserves the salient visual features necessary for tasks such as object recognition (Kim et al., 2018). However, applying a similar reduction in scale to hydrogeological models is profoundly problematic. Upscaling hydrogeological models, analogous to coarsening the simulation grid, can lead to the loss of critical physical information (Durlofsky, 2003). This includes the misrepresentation of fine-scale aquifer heterogeneity, which governs groundwater flow paths and solute dispersion, and the inadequate characterization of localized boundary conditions or small-scale features such as preferential pathways that exert significant control on the groundwater system's behavior (Zehe et al., 2021). The loss of these features does not merely reduce detail but can fundamentally alter the simulated physical processes, rendering model predictions unreliable. Consequently, the effective informational scale required for a single hydrogeological model to adequately capture these essential process-governing features can be several hundred to a thousand times larger than that of a typical downscaled image employed in DL applications (de Graaf et al., 2015). As a result, most existing DL applications in hydrogeology focus on limited-scale problems due to resource constraints, whereas large-scale regional simulations remain rare. The primary limitation arises from insufficient computational resources, especially in terms of accessible GPU capabilities, which are often inadequate for training such large models (Gao et al., 2020). Consequently, as model complexity increases, there is a growing demand for supercomputing resources to train and deploy DL models effectively at scale, which remains a barrier to their widespread application in hydrogeological modeling (Bergen et al., 2019; K. Li et al., 2023).

DL models, especially those with numerous layers (deep networks), tend to be computationally intensive. In hydrogeological modeling, this computational intensity is further compounded by the complexity of the data sets. These data sets are often characterized by high-dimensional grids (e.g., 2D or 3D with a large number of cells), the need to capture highly heterogeneous geological formations, and the representation of nonlinear physical

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

processes such as fluid flow and solute transport. Moreover, the sheer volume of data, which may include multiple time steps and various hydrogeological variables, exacerbates the computational burden. They involve a vast number of parameters that need optimization, a process often entailing complex mathematical computations. To effectively handle this complexity, DL architectures must be carefully designed. For instance, CNNs are frequently employed due to their ability to efficiently process spatial data, whereas RNNs or LSTMs may be utilized to model temporal dependencies. However, these sophisticated architectures typically require a large number of layers and parameters, which in turn demand significant computational resources.

When dealing with large data sets or sophisticated models, computational complexity can become a limiting factor (Justus et al., 2018). This complexity can slow down the training process significantly: depending on the complexity of the model and the size of the data set, training a DL model can take from several hours to days, or even weeks. For example, in a study by Zhan, Dai, Samper, et al. (2022), a DL model named DOCRN was developed for a 100 × 80 grid in a 2D conservative solute transport simulation. Training this model on an NVIDIA Tesla V100s GPU required 5 hr. Once trained, DOCRN could generate concentration and head distribution fields in less than 0.2 s, whereas the traditional numerical model, TOUGHREACT, took approximately 10 s per simulation. However, in the context of a data fusion algorithm that required 40,000 forward simulations, the total time for the traditional approach would be about 111.1 hr. In contrast, the DL approach necessitated generating 10,000 training samples (taking 27.8 hr), training the model (5 hr), and performing 40,000 predictions (2.2 hr), totaling approximately 35 hr. Thus, the overall efficiency gain was around 3 times, which, although beneficial, is less striking than the individual prediction speedup might suggest. This illustrates that for applications requiring a smaller number of simulations, the upfront cost of training a DL model may not be justified. Particularly for large-scale models, the training time alone can be longer than the simulation time using traditional algorithms, potentially offsetting the computational efficiency advantages brought by DL models.

In summary, although DL models promise significant advancements in hydrogeological modeling, their practical application is often hampered by the high computational complexity and the substantial resources they require. Balancing the need for advanced modeling capabilities with the available computational resources remains a crucial challenge in integrating DL into hydrogeological modeling. This challenge is particularly pertinent for researchers and practitioners who may not always have access to the necessary high-end computational infrastructure.

4. Potential Solutions and Future Directions

Having comprehensively delineated the multifaceted challenges and inherent limitations currently impeding the widespread and effective application of DL in hydrogeological modeling in the preceding section, this chapter pivots toward a forward-looking exploration of promising solutions and strategic future research trajectories. The subsequent sections will systematically address key domains requiring concerted advancement. These include: Innovative strategies to surmount data-related impediments, encompassing availability and quality; robust approaches to bolster model interpretability and explainability, thereby moving beyond opaque “black box” paradigms; dedicated pathways to elevate model performance, with an emphasis on generalizability, reliability, and the establishment of standardized benchmarks; and emergent innovations in computational capabilities designed to address the intensive resource demands inherent to sophisticated DL models.

4.1. Data-Related Directions

In the application of DL to hydrogeological modeling, the limitations arising from data handling and management are significant. A thorough examination of these challenges highlights a fundamental fact: The success of DL models in this field is deeply intertwined with the quality and processing of data (X. Li et al., 2023). Issues such as limited data availability, low data quality, and inadequate data processing techniques have hindered the effective implementation of DL in hydrogeological modeling. The lack of comprehensive and diverse data sets often cause models that do not perform well in real-world scenarios, suffering from issues such as overfitting and limited capability to apply learned knowledge to new situations (Alzubaidi et al., 2021).

Addressing these data-related challenges is crucial for the advancement of DL in hydrogeological modeling. This can be approached from several key directions: data collection and sharing, data pre-processing, and data synthesis (See Figure 9). Each of these studies offer a pathway to potentially mitigate existing limitations and enhance the capability of DL models.

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

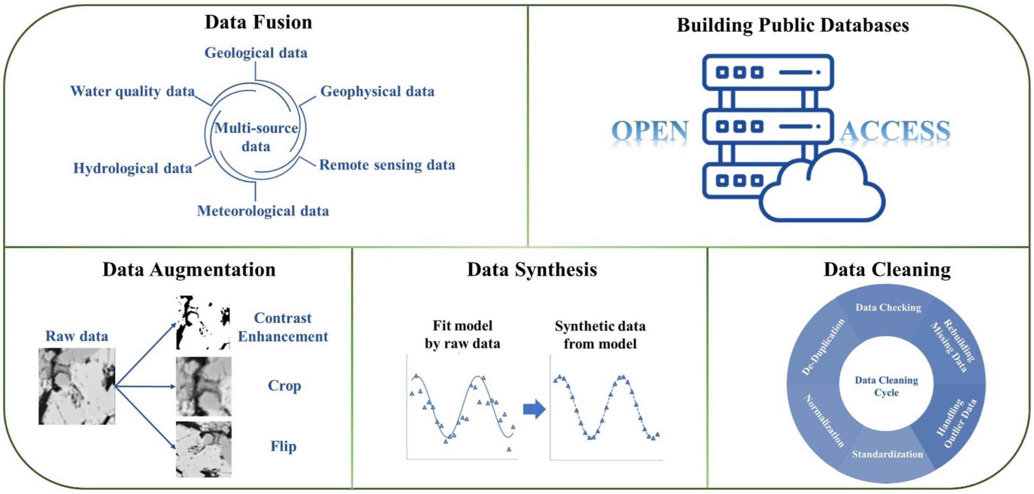

Figure 9. The data-related future Directions in DL-based hydrogeological modeling: Data fusion, building public databases, data augmentation, data synthesis and data cleaning.

4.1.1. Data Collection and Data Sharing

In DL-based hydrogeological modeling, acquiring a large volume of data is a key first step to developing reliable and robust models. As previously discussed, obtaining comprehensive groundwater data presents numerous challenges. This section explores potential strategies to mitigate these difficulties.

The importance of gathering data from multiple sources cannot be overstated. Often, data from a single source is limited and might not provide the complete information required for accurate prediction models. By integrating various types of data, we can achieve a more holistic view (Schilling et al., 2019). For instance, in hydrological forecasting, besides historical water level data, incorporating climate data and information about rare events can enhance the model (Rajaee et al., 2019). In identifying aquifer structures, limited borehole data can be complemented with hydrological observations (e.g., water levels and concentrations) and geophysical data to form a more accurate picture (Ikard et al., 2023; Jardani et al., 2022; Pearson et al., 2022; Rucker et al., 2021; Zhan, Dai, Soltanian, & de Barros, 2022).

To collect such diverse data sets, advanced data collection technologies must be fully utilized and developed further. Technologies such as satellite remote sensing, drones, and the Internet of Things (IoT) can significantly broaden the amount and range of data collected. Remote sensing, for instance, can provide spatial data over vast areas (Adams et al., 2022; Ma et al., 2015), whereas IoT sensors can offer continuous monitoring of groundwater levels and concentrations, enriching time sequence data (Ighalo et al., 2021). The advancement of these technologies marks a step toward a "big data" era in hydrogeological modeling.

Data sharing plays a crucial role in improving the accessibility, interoperability, and reusability of research data, which are essential for discovery and accelerating scientific progress (Gewin, 2016). DL models depend on large, diverse data sets for training and optimization. Sharing data can provide a wider range of resources, enhancing the model's capability to generalize and improve accuracy. Furthermore, data sharing can help establish standardized problems in hydrogeological modeling, reducing the occurrence of shortcut learning and promoting research transparency and replicability (Geirhos et al., 2020). Governments and academic institutions worldwide are recognizing the importance of data sharing (Lattu, 2023; Sicilia et al., 2017). In the hydrogeology field, initiatives including the Global Groundwater Information System (GGIS) (https://www.un-igrac.org/global-groundwater-information-system-ggis) demonstrate the importance of creating open, standardized databases and promoting data sharing for democratizing data access.

The effective application of DL in hydrogeological modeling is heavily dependent on the strategic collection and sharing of data from multiple sources. By integrating different types of data and leveraging advancements in data collection technologies, alongside promoting data sharing, we can significantly improve the performance and applicability of DL models in hydrogeological studies. These efforts go beyond merely collecting more data; they

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

aim to make data more meaningful and accessible, paving the way for more informed and impactful hydrogeological modeling.

4.1.2. Data Pre-Processing

In hydrogeological modeling, data preprocessing plays a crucial role in the development of DL models (Bernhardt et al., 2022). Unlike other fields, such as surface water monitoring, groundwater systems present more complex and variable environments due to their underground nature. This complexity affects both the accuracy and stability of the data collected by sensors, leading to higher levels of errors, noise, and missing data compared to surface water data sets (Jeong et al., 2020; Z. Wang et al., 2024). Proper preprocessing is essential to address these challenges and improve model performance.

The preprocessing workflow typically starts with data cleaning and anomaly detection, which involves removing duplicates, correcting errors, and identifying outliers (Ariyaluran Habeeb et al., 2019). Groundwater data, often containing various errors due to challenging sensor environments, can significantly skew DL models if left unaddressed (Oladeji et al., 2024). In addition, missing data are common in groundwater data sets (Gao, 2017). Filling these gaps is essential for reducing bias, with methods such as estimation techniques or DL-based approaches such as GANs (Vu et al., 2021). Moreover, hydrogeological time series often exhibit vastly different data collection frequencies due to varied sensor deployment, operational factors, or historical practices. Such temporal heterogeneity can bias models and impair the learning of true temporal dependencies by DL architectures, especially recurrent networks designed for sequential data processing, thereby diminishing predictive reliability (Retike et al., 2022). Resampling techniques are often essential to standardize the data sets to a consistent time interval, ensuring compatibility for integrated analysis (Pazola et al., 2024). This temporal regularization is vital for standardizing input sequences, mitigating biases induced by irregular sampling intervals, and ultimately improving the robustness and accuracy of DL-based hydrogeological predictions.

Another critical step is feature engineering, particularly for handling the heterogeneous nature of groundwater data. Groundwater data sets often come from multiple sensors, capturing diverse information, including chemical, physical, and hydrological properties (Manzione & Castrignanò, 2019; Shao et al., 2022). This results in a wide variety of data formats and structures, making it a challenge to integrate these multi-source, heterogeneous data sets into a form suitable for DL models. Addressing this challenge is key to successful hydrogeological modeling. One approach involves using DL techniques to develop automated methods for extracting and processing data from these diverse sources (Lawley et al., 2023). Such techniques can streamline the conversion of raw, heterogeneous data into a unified format that DL models can interpret.

At the same time, feature engineering plays a vital role in extracting meaningful information from multi-source data sets. By applying feature engineering techniques, important attributes such as scaling, normalization, and the creation of new interaction terms can be introduced to ensure that relevant patterns and relationships within the data are properly captured (Zheng & Casari, 2018). This process not only refines the data but also makes it more suitable for model training (Verdonck et al., 2024), improving the accuracy and reliability of DL models in hydrogeological applications.

In summary, among the potential future developments in DL-based hydrogeological modeling, the establishment of standardized preprocessing workflows appears to be the most immediately feasible option. Given the current volume of data and available technologies, such standardization could significantly advance the application of DL models in this field. For many researchers, the challenge of handling raw, heterogeneous data are often the first and most time-consuming step. Optimizing this process could accelerate progress in applying DL to groundwater modeling.

4.1.3. Data Synthesis

In the realm of hydrogeological modeling, where data acquisition is often expensive and slow, the role of data synthesis becomes crucial. Data synthesis involves creating artificial data through computational simulation based on a small set of real-world data, accurately reflecting the information characteristics of the original data set (Mannino & Abouzied, 2019). This method significantly reduces the need to obtain large volumes of data from the real world, allowing for rapid acquisition of necessary information (Bourou et al., 2021; Rees et al., 2020).

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Notably, in tasks requiring labeled data, data synthesis can automatically generate labels, providing a streamlined approach to data preparation (Sixt et al., 2018).

The heterogeneity of groundwater systems in both time and space, such as the temporal variability of groundwater levels and the heterogeneity of aquifer structure, presents another advantage of data synthesis (Pulla et al., 2024). Data obtained from limited observation points may not encompass all the informational characteristics of the system. Especially for extreme scenarios, synthesizing training data for such conditions is vital for hydrogeological simulation, a task closely linked to human life and safety (e.g., flooding, mine collapse, or land slide). Diverse training data sets generated through synthesis can lead to more precise DL models, capable of handling a wide range of scenarios, including those that are rare or extreme (Nik, 2016).

Regarding methods of data synthesis, we can consider traditional process-based or stochastic models, as well as generative AI models (Fonseca & Bacao, 2023). Traditional models, which are grounded in specific physical equations, can simulate hydrogeological scenarios to produce synthetic data. For instance, groundwater numerical models such as MODFLOW coupled with MT3D can simulate real-world contaminant transport, providing data for solute transport simulation and groundwater contamination risk prediction in DL models. To ensure the synthetic data are reasonable, these process-based models should be calibrated using available observational data. This calibration allows the models to better represent the system's characteristics, thereby generating synthetic data that is more accurate and applicable. Conversely, some data types that are not easily modeled through traditional processes can be estimated using statistical characteristics of real data. For example, Zhan, Dai, Soltanian and Zhang (2022) used borehole data to calculate transition probabilities between different lithologies and generated numerous potential aquifer structures for training a variational autoencoder-based aquifer structure generation model.

In the context of deep learning, generative models such as GANs and VAEs can produce highly realistic and complex data (Park et al., 2018; Wan et al., 2017). However, a critical consideration when employing generative AI is the risk of "hallucinations"—instances where models generate plausible-appearing but physically unrealistic or entirely fictitious patterns that do not correspond to actual hydrogeological phenomena. This is particularly concerning in hydrogeological applications where synthetic data must adhere to fundamental physical laws. However, to ensure that the generated data adheres to physical principles, especially in data-sparse scenarios, it is beneficial to incorporate physical constraints into these models. To mitigate hallucination risks, several strategies should be implemented: (a) Systematic cross-validation against fundamental physical laws, (b) incorporation of domain expertise in the model design and evaluation process, and (c) rigorous statistical comparison between synthetic and real data distributions to identify anomalous patterns. PINNs, which embed physical laws into the learning process (Raissi et al., 2019), have been successfully applied to hydrogeological problems, such as modeling groundwater flow (Secci et al., 2024), and can be adapted for generative tasks to enhance the physical consistency of synthetic data.

Regardless of the method used, rigorous validation of synthetic data sets against available real data and established hydrogeological knowledge is essential to confirm their utility and reliability. This validation must specifically address whether generative AI outputs represent authentic hydrogeological conditions rather than artifactual patterns, requiring both quantitative metrics and expert geological interpretation. This validation process helps ensure that the synthetic data accurately reflects the system's behavior and can be confidently used to train DL models for hydrogeological applications.

To conclude, data synthesis offers a powerful tool for hydrogeological modeling, enabling the generation of diverse and comprehensive data sets necessary for training robust DL models. Through a combination of traditional modeling techniques and advanced generative AI, researchers can create detailed and varied data sets that enhance the precision and applicability of DL models in this critical field. As the field evolves, these synthetic approaches are becoming increasingly important in overcoming data limitations and driving forward the capabilities of DL in hydrogeological modeling.

4.2. Interpretability-Related Directions

In the application of DL methods for hydrogeological modeling, a significant challenge lies in the areas of interpretability and explainability (Razavi, 2021). These models, often perceived as "black boxes," do not provide enough information for understanding how and why certain predictions are made. This lack of clarity in the

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

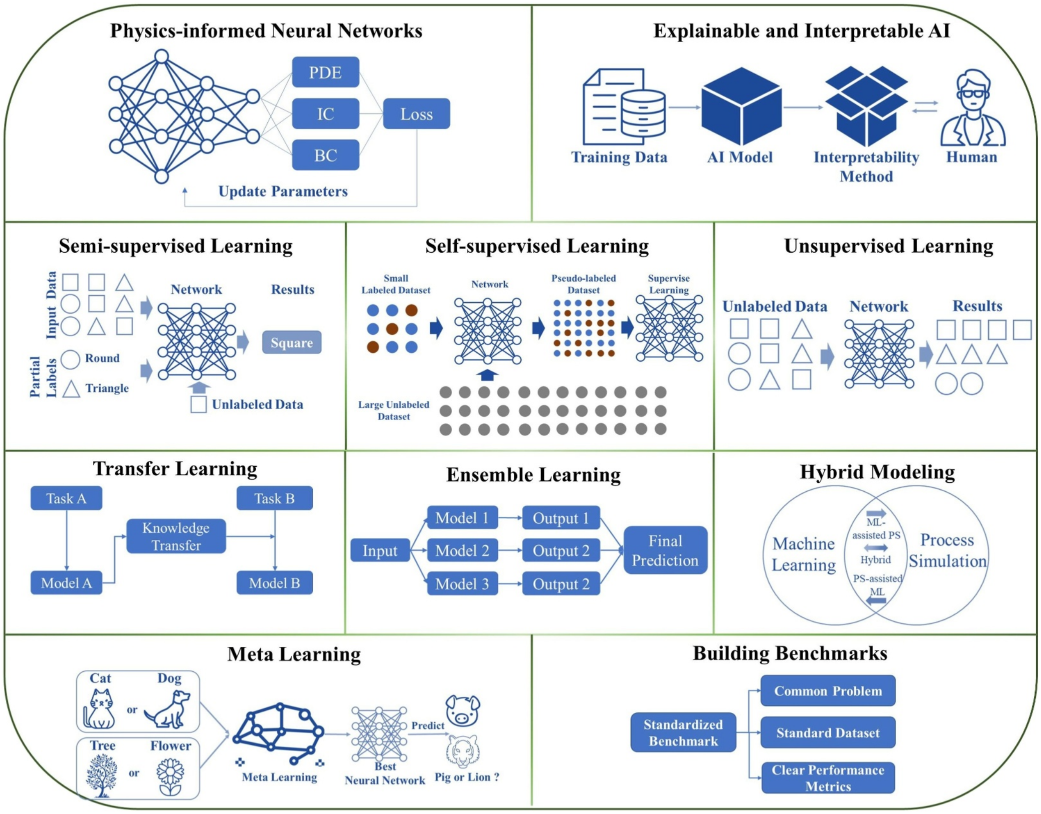

Figure 10. The model-related future directions in DL-based hydrogeological modeling include those related to interpretability (Physics-informed neural networks, Explainable and Interpretable AI), model performance (Semi-supervised learning, Self-supervised learning, Unsupervised learning, Transfer learning, Ensemble learning, Hybrid modeling, Meta learning and Building benchmarks).

decision-making process can hinder the acceptance and trust in DL models, especially in the field of hydrogeology where understanding the underlying processes is as crucial as the predictions themselves.

Improving the interpretability and explainability of DL models is essential for their broader application and acceptance in hydrogeological modeling. Enhanced interpretability not only bolsters confidence in the models among scientists and decision-makers but also aids in identifying and rectifying errors, thereby improving the overall accuracy and reliability of the models (Carvalho et al., 2019; Yoon et al., 2022). To address these challenges, three promising directions emerge: physics-informed DL, explainable DL, and interpretable DL (Figure 10).

Physics-informed DL integrates known physical laws and principles into the learning process, providing a foundational understanding of the model's outputs (Xu & Zhang, 2024). Explainable DL, meanwhile, focuses on developing models with a more transparent decision-making process. This transparency allows users to understand how the model's design and settings influence its conclusions, providing insight into the broader impact of the model's structure on its outputs (Barredo Arrieta et al., 2020). Interpretable DL goes further, aiming to create models where not only are the internal mechanisms transparent, but they also allow for an intuitive understanding of the causal relationships between input features and output predictions (X. Li et al., 2022). This enables stakeholders, including hydrogeologists and non-experts, to grasp the logic behind the model's decisions.

These approaches represent a spectrum, from easier to more complex, in enhancing the interpretability and explainability of DL models. Each offers a potential pathway to overcome the interpretability and explainability challenges in DL applications for hydrogeological modeling, progressively building toward models that are not only powerful in prediction but also clear and understandable in their reasoning.

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025JH000703

4.2.1. Physics-Informed Deep Learning

Physics-informed deep learning (PIDL) has emerged as a transformative approach in scientific computing, particularly in subsurface modeling where integrating physical laws with data-driven models is essential. By embedding physical knowledge into machine learning frameworks, these methods enhance interpretability, generalizability, and scientific consistency. This integration can be achieved through various strategies, such as incorporating physical equations into loss functions, designing physics-inspired architectures, or combining traditional numerical methods with machine learning. As subsurface systems often exhibit complexity that challenges purely data-driven or physics-based approaches, physics-informed machine learning provides a promising path forward, with multiple methodologies contributing to its advancement.

One prominent method is Physics-Informed Neural Networks (PINN), which embed physical equations directly into the training process by augmenting the loss function with terms that enforce adherence to known physical laws (Lagaris et al., 1998). This ensures predictions are both data-consistent and physically plausible. PINN has been successfully applied to subsurface problems, such as groundwater flow modeling (Raissi et al., 2019; Tartakovsky et al., 2020) and solute transport (N. Wang et al., 2021). For instance, in modeling solute transport, the convection-dispersion equation is integrated into the loss function to ensure credible predictions. However, PINN faces challenges, including difficulties in balancing multiple loss terms and high computational costs for large-scale problems (Cuomo et al., 2022).

Neural operators offer another powerful framework designed to learn mappings between function spaces, making them well-suited for solving partial differential equations (PDEs) that govern subsurface processes. Unlike traditional neural networks, which handle fixed-dimensional inputs, neural operators such as Fourier Neural Operators (FNO) (Z. Li et al., 2020) and Deep Operator Networks (DeepONet) (Lu et al., 2021) can process infinite-dimensional data, effectively approximating the solution operators of PDEs. In subsurface applications, they can model complex phenomena such as multiphase flow or heat transfer in porous media. For example, Sun et al. (2024) demonstrated the application of DeepONet in hydrologic modeling, improving streamflow prediction accuracy by bridging ensemble simulations and machine learning. Once trained, neural operators can rapidly evaluate solutions for new parameters or conditions, offering a computationally efficient alternative to traditional numerical solvers. This efficiency is particularly valuable for real-time prediction or uncertainty quantification in studies such as reservoir simulation and groundwater flow.

Neural Differential Equations (NDEs) provide a flexible approach for modeling dynamic systems by parameterizing derivatives with neural networks (R. T. Chen et al., 2018). This property is particularly useful in subsurface modeling, where physical equations might be only partially known or analytically intractable. Their ability to capture complex nonlinear behavior while maintaining physical constraints makes them suitable for hydrogeological applications where data are often scarce. By merging data-driven insights with physical constraints, NDEs enhance predictive accuracy and adaptability in subsurface environments.

Beyond PINNs, physics-embedded machine learning encompasses a broader range of techniques that respect physical symmetries or conservation laws through feature engineering or architectural design. Hybrid models, which integrate traditional numerical solvers with machine learning, also fall under this category. Song et al. (2025) proposed a multi-scale differentiable PIDL method for high-resolution, national-scale hydrologic modeling, significantly improving streamflow prediction accuracy. In subsurface modeling, such approaches can leverage the precision of physical simulators for well-understood processes while using machine learning to handle complex or uncertain components, striking a balance between fidelity and flexibility.

Differentiable programming further advances the field by enabling end-to-end differentiable computational pipelines, allowing for the seamless integration of physical models and neural networks. This paradigm supports the joint optimization of model parameters and physical constants, enhancing predictive accuracy and consistency (D. Feng et al., 2024; Sawadekar et al., 2025). For instance, Shen et al. (2023) proposed differentiable modeling, which combines process models with DL. This approach connects a variable amount of prior physical knowledge to DL models, enhancing interpretability, generalizability, and extrapolation capabilities, and achieving accurate predictions even with less data.

In the past 3 years, the application of Physics-informed DL in hydrogeological modeling has made significant progress, emerging as a highly active research area (Cai et al., 2022; Cai et al., 2024; Jiang et al., 2022; Y. Zhan et al., 2024). Studies have shown that under physical constraints, DL-based hydrogeological models can achieve

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

superior performance with limit data (N. Wang et al., 2024). This is particularly valuable for hydrogeological systems where obtaining extensive data sets can be challenging. However, despite these advancements, challenges remain. Although Physics-Informed DL ensures that model outputs are constrained by physical laws, the exact manner and extent to which these equations influence the results are still not fully understood. This gap highlights the need for further research in Explainable and Interpretable DL, which focuses on making the inner workings of these models more transparent. Understanding how physical equations shape the model's predictions, and identifying the key factors that drive these outputs, is crucial for improving model interpretability. These efforts are essential for advancing Physics-Informed DL beyond its current state, ensuring that it not only produces physically consistent results but also offers clearer insights into the underlying mechanisms that govern those outcomes (Jiang et al., 2024).

4.2.2. Explainable and Interpretable Deep Learning

DL models excel at capturing complex relationships in hydrogeological data, such as predicting groundwater flow or contaminant transport based on aquifer properties and environmental conditions. However, their complexity often obscures how predictions are made, which is critical for building trust and informing management decisions in hydrogeology (Gorelick & Zheng, 2015). Two approaches—explainable DL and interpretable DL—address this by making models more transparent and their outputs easier to understand.

The application of Explainable Deep Learning (Explainable DL) in hydrogeological modeling is driven by the need to understand which inputs have the most significant impact on model decisions, thereby improving users' trust and usability of existing DL models. It involves explaining how models extract and utilize information from geological and hydrological data and how this information influences predictions about groundwater flow, aquifer recharge, or contaminant transport (Kwak & Lee, 2024).

Although not yet widely applied in hydrogeological simulation, various common methods of Explainable DL could be considered for future integration into DL-based hydrogeological models (Dahal et al., 2023). Prominent among these are feature attribution methods, which aim to quantify the influence of each input feature on the model's output. Methods such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) focus on feature contribution explanations (Lundberg & Lee, 2017). SHAP, based on game theory, assigns a “contribution value” to each feature, explaining their impact on model predictions (Niu et al., 2023). LIME approximates the behavior of a DL model at specific input points using a local linear model, providing more accessible explanations (Başağaoğlu et al., 2022). For instance, S. Yang, Song, et al. (2023) used the SHAP method to analyze key factors influencing predictions of mine water inflow, whereas Liu et al. (2022) applied SHAP to interpret the main drivers behind regional groundwater level changes. However, their use in complex hydrogeological tasks, such as spatial data analysis or inverse modeling, remains limited due to their focus on simpler classification and regression problems.

Interpretable DL, by contrast, clarifies how a model processes inputs internally to produce predictions (Linardatos et al., 2021). For example, in a model forecasting groundwater levels, it could reveal how the model balances recent rainfall against long-term climate trends, shedding light on an aquifer's response to weather patterns. Current methods often simplify models, using shallower neural networks or distilling complex models into simpler ones (ElShawi et al., 2021). Another approach is model distillation, where knowledge from a complex model is transferred to a simpler, more transparent model. This method seeks to maintain high performance while increasing interpretability (Gou et al., 2021; X. Liu et al., 2018).

Both approaches have limitations. Explainable DL struggles with tasks such as image-to-image mapping or inverse modeling, which are key in groundwater studies, and most studies rely on a single method, potentially missing broader insights (Alshehri & Rahman, 2023). Interpretable DL sacrifices performance for clarity, limiting its ability to handle intricate data. Future research should develop methods tailored to hydrogeological needs, such as analyzing spatial patterns or incorporating physical laws. Balancing complexity and clarity will be essential to advancing DL in hydrogeology.

4.3. Model Performance-Related Directions

This section explores strategies to enhance the performance of DL in hydrogeological modeling, which encompasses a holistic approach, focusing on improving model generalizability, accuracy, and the establishment of

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

benchmarks, as well as automating the optimization of model parameters and structures (Figure 10). Each aspect contributes uniquely to advancing the effectiveness of DL models in this domain, addressing the need for models that are not only theoretically robust but also practically reliable and efficient in diverse hydrogeological scenarios.

4.3.1. Enhancing Generalizability

The importance of enhancing the generalizability of DL models in hydrogeological modeling cannot be overstated. Generalizability ensures that a model performs consistently across various scenarios and data sets, a crucial aspect for models intended for diverse applications (Tian et al., 2024).

To improve model generalizability, it's essential to consider factors related to the training data, the training process, and the model architecture (Yoshida & Miyato, 2017). From the data perspective, increasing the diversity and volume of training data and implementing effective data preprocessing are key steps, as outlined in Section 4.1. During the training process, techniques such as early stopping and cross-validation can help prevent overfitting, thus improving generalizability (Rice et al., 2020). However, it's noteworthy that many current studies in DL for hydrogeological modeling do not set aside a validation set, relying only on training and test sets, which may limit the capability to accurately assess the model's generalizability.

To further improve the generalization capability of the model from the model architecture perspective, it is necessary to utilize methods such as reinforcement learning, transfer learning, unsupervised learning, semi-supervised learning, and self-supervised learning (Arulkumaran et al., 2017; Singh et al., 2016). These methods can increase a model's capability to generalize and reduce dependency on extensive labeled data sets. Transfer learning, for instance, uses knowledge from one task to improve another related task's simulation accuracy, reducing reliance on new task-specific data (Weiss et al., 2016). Semi-supervised learning allows models to learn from both labeled and unlabeled data (X. Yang et al., 2023), and unsupervised learning enables models to find patterns and structure in unlabeled data (Usama et al., 2019). Self-supervised learning involves training models on unlabeled data by generating synthetic labels, offering an innovative approach to model training (Jaiswal et al., 2021).

Enhancing the generalizability of DL models in hydrogeological modeling requires a multifaceted approach. It involves not only diversifying and preprocessing training data but also incorporating advanced learning methodologies that reduce overfitting and dependency on labeled data. By employing these strategies, more robust and adaptable models can be developed, capable of delivering reliable predictions across varying hydrogeological scenarios.

4.3.2. Improving Reliability

Improving the reliability of DL approaches in hydrogeological modeling is a longstanding and central concern. High accuracy in these models ensures reliable predictions, which is vital for effective decision-making and management.

Initially, to improve reliability, researchers increased the layers in ANNs beyond three, giving rise to early DL models. This architectural change significantly enhanced the models' capability to depict nonlinear relationships, thus improving reliability. However, as layers increased, overfitting became more prevalent. The introduction of residual networks (ResNets) allowed for the use of deeper models with tens or even hundreds of layers, further advancing reliability (He et al., 2016). Therefore, developing advanced DL architectures remains a key area of focus. Similarly, developing models tailored to specific data types (e.g., RNNs, CNNs, and graph neural network (GNNs)), is also vital as these models can extract relevant information more effectively for reliable simulations.

Before revolutionary innovations in model architecture, ensemble models, which combine predictions from multiple individual models, also enhance reliability (Sagi & Rokach, 2018). These individual models could be all DL models, all traditional models, or a mix of both. Ensemble models often achieve higher reliability than single models by averaging out individual errors. They are particularly beneficial when individual models have different strengths and weaknesses, as the ensemble can leverage these advantages while minimizing drawbacks (Zounemat-Kermani et al., 2021). Ensemble models also exhibit robustness against overfitting and noise in data since they rely on multiple models, reducing the risk of the final prediction being overly influenced by any specific model or data peculiarity. Creating hybrid and ensemble models requires careful design and validation. It's crucial

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to ensure that the models within the ensemble complement each other, having distinct strengths and weaknesses. Additionally, methods used to combine individual model predictions significantly impacts the performance of the hybrid or ensemble model, necessitating careful selection and optimization.

4.3.3. Building Benchmarks

Establishing benchmarks in DL for hydrogeological modeling is highly important for successful applications. Although most studies focus on enhancing model performance, the varying data sets and scenarios used across different research efforts make it challenging to accurately assess and compare model performance. This lack of standardized benchmarks hinders the evaluation of the effectiveness of various DL models in hydrogeological simulation.

To address this, methods for building benchmarks involve identifying common problems or scenarios in hydrogeological modeling and creating standardized data sets and evaluation metrics. For instance, the 2022 Groundwater Time Series Modeling Challenge (Collenteur et al., 2024) provided a standardized data set of hydraulic-head time series from four monitoring wells across diverse hydrogeological settings, enabling the comparison of data-driven models such as lumped-parameter, machine learning, and DL approaches. Performance was assessed using metrics such as Nash-Sutcliffe efficiency (NSE) and mean absolute error (MAE), applied over a 15-year calibration period and a 5-year validation period. Another approach involves synthetic benchmarking scenarios, as exemplified by (T. Xu et al., 2024), who developed five 2D groundwater flow cases for inverting hydraulic conductivity from pressure data. These scenarios, varying in heterogeneity and observation layouts, were paired with high-fidelity reference solutions generated via Markov chain Monte Carlo (MCMC) methods, using metrics such as normalized MAE and energy distance to quantify accuracy. This process includes collating data from multiple sources, ensuring data quality, and defining clear performance metrics that can be universally applied (Cheng et al., 2024). These benchmarks would then provide a basis for comparing different models on a level playing field, facilitating an objective evaluation of their performance (W. Wang et al., 2024).

The process of building benchmarks is complex and lengthy. It requires careful consideration of the diversity of hydrogeological modeling scenarios and the intricacies of DL models. For example, benchmarks must account for nonlinear responses in transient flow or extreme conditions such as droughts necessitating data sets with sufficient temporal resolution (e.g., daily or weekly) and spatial coverage. The challenge lies in creating benchmarks that are representative of real-world conditions while being fair and applicable to various models (L. Liu et al., 2018). In conclusion, the development of benchmarks in DL for hydrogeological modeling is a critical step toward a more systematic and objective assessment of model performance. It will provide a foundation for comparing different models, fostering a deeper understanding of their strengths and limitations, and guiding future research and development in this field. Moreover, the integration of open-source software for machine learning and data processing is equally vital in advancing open science. Open-source tools not only facilitate the development and application of DL models but also ensure methodological transparency and reproducibility, which are critical for fostering hydrogeological research.

4.3.4. Automatic Optimization of Model Parameters and Structures

The capability of DL models to enhance simulation efficiency is a significant reason for their application in hydrogeological modeling. However, in practice, the time spent on adjusting model parameters and structures often surpasses the efficiency gains of the models. In addition, selecting appropriate parameters and structures is also critical for optimal simulation results, and relying solely on manual trial-and-error methods is often insufficient for finding the best solutions.

To automate this optimization process, methods such as genetic algorithms, grid search, and Bayesian optimization are employed to explore the parameter space systematically and efficiently (Luo, 2016; Yang & Shami, 2020). Neural architecture search (NAS) (Ren et al., 2021; Zoph & Le, 2016) and meta-learning (Vettoruzzo et al., 2024) can automatically design model structures based on different data sets, further streamlining the optimization process.

The application of these automatic optimization techniques in hydrogeological modeling presents unique challenges. Groundwater systems typically feature less data, greater spatio-temporal variability, and larger scales

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Boggs et al., 1992) compared to other fields where DL models are commonly applied. Consequently, DL models for hydrogeological simulations demand more computational power, have more parameters, and require more complex structures, making automatic optimization in this domain particularly challenging.

4.4. Computational Capability-Related Directions

In the evolving landscape of DL applications for hydrogeological modeling, one of the predominant challenges is the significant demand for computational resources. The complexity and depth of DL models necessitate substantial computational power, which often poses a barrier, especially in resource-constrained settings (Mehonic & Kenyon, 2022; Thompson et al., 2020). Addressing this challenge requires not only maximizing the efficiency of existing computational resources but also exploring innovative computational paradigms (Zhu et al., 2023). This section delves into two main directions: the effective utilization of current computational resources and the adoption of cutting-edge computational methods.

The first direction involves optimizing the use of current computational capabilities, such as high-performance computing (HPC) and cloud computing. Each of these technologies offers unique advantages in processing large data sets and complex models (S. Wang et al., 2024), which are integral to DL in hydrogeological modeling. Leveraging these technologies can significantly enhance the efficiency and feasibility of running sophisticated DL models, making them more accessible and practical for widespread use.

The second direction explores the potential of emerging advanced computational approaches, such as quantum computing, photonic computing, and biological computing (Sood & Chauhan, 2024). These novel computing paradigms hold the promise of revolutionizing the processing capabilities for DL models, offering solutions that could surpass the limitations of traditional computing infrastructures. The exploration of these advanced technologies could lead to groundbreaking developments in the field of hydrogeological modeling, enabling more complex, accurate, and faster computations than ever before.

4.4.1. Maximizing Existing Computational Resources

In the short term, without significant advancements in computational hardware and techniques, HPC systems will remain essential for those seeking to leverage DL models (Huerta et al., 2020). This is particularly true for tasks such as spatial data analysis, which demand high GPU capacity. For instance, training a 3D residual network for groundwater flow prediction on a domain of $80 \times 60 \times 20$ grid cells can require over 30 GB of GPU memory, exceeding the capacity of most consumer-grade GPUs (typically 8–24 GB) (Zhan, Dai, Soltanian, & Zhang, 2022). Even the most advanced consumer-grade GPUs often fall short in terms of training speed and the volume of data they can process effectively.

HPC systems offer parallel processing capabilities, allowing multiple processors or cores to perform computations simultaneously. They typically include advanced hardware such as GPUs or TPUs, which are optimized for the matrix operations commonly used in DL, drastically reducing the time required for model training and evaluation (Jena et al., 2022; More et al., 2021). For instance, Commodity Off-The-Shelf HPC systems with GPU clusters can train a neural network with 1 billion parameters, equivalent to approximately 569 conservative solute transport surrogate models (each with a 40×80 grid and 1,757,446 parameters), on just 3 machines in a few days, compared to traditional systems such as DistBelief requiring 1,000 machines with 16,000 CPU cores. Similarly, an 11 billion parameter network, equivalent to about 6,263 such solute transport models, can be trained on 16 machines in approximately 3 days, with a single mini-batch update (96 images) taking less than 0.6 s and a full epoch over 10 million images completed in about 17 hr, achieving a 99.7% reduction in machine usage compared to earlier approaches (Coates et al., 2013). Specifically, training a single solute transport surrogate model on an HPC system takes an estimated 5 min, compared to 2.5 hr on a personal computer, yielding a speedup of approximately 29.6 times.

Complementing HPC, cloud computing platforms provide scalable and adaptable resources, allowing researchers to flexibly adjust their computational capacity to meet the varying demands of DL models (Jauro et al., 2020; Jones et al., 2015). The pay-as-you-go pricing model offers a cost-effective solution. Taking cloud computing costs in mainland China as an example, the hourly fee for a V100s 32G GPU is approximately 5–10 RMB (equivalent to about $0.70–1.40 USD). Compared to purchasing a dedicated server equipped with one V100s GPU, which costs around 100,000 RMB (approximately $13,880 USD), using cloud computing can result in

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

savings of about 80%–90% for projects requiring less than 2,000 GPU-hours annually. This is particularly for researchers new to DL-based hydrogeological modeling who may be unsure of their computational needs. Cloud platforms also simplify data management, model deployment, and performance monitoring, making it easier for researchers to deploy their DL models on the cloud, enabling simulations and predictions even when local computational hardware is limited (Arif, 2022; Y. Wang et al., 2024).

By combining HPC systems with cloud computing, researchers can maximize existing computational resources while benefiting from a flexible and user-friendly computational framework. This integration allows resource-intensive DL models to be deployed anywhere with internet access, significantly improving the efficiency of model development and deployment (Dhanasekaran et al., 2024). The synergy between HPC and cloud computing is transforming the implementation of DL models in hydrogeological research, making these models more accessible and feasible across a wide range of scenarios.

4.4.2. Exploring Advanced Computational Paradigms

The popularity of large-scale models such as ChatGPT has brought attention to the significant leap in learning and predictive capabilities when DL model parameters surpass a certain threshold (Ray, 2023; Wei et al., 2022). Currently, in the broader field of Earth sciences, research utilizing large models based on existing frameworks has already begun to emerge (Hu et al., 2023). For example, Deng et al. (2024) proposed a large language model named K2, which enables question-and-answer functionality in the field of Earth sciences, mainly leveraging large models to extract and utilize vast amounts of information. Similarly, in the field of hydrogeology, large models can be used to effectively extract multi-source information required for hydrogeological modeling. For instance, by training targeted large models, structured borehole data could be extracted from repetitive and unstructured geological descriptions (Ghorbanfekr et al., 2025). Additionally, with the capability of large models to process and output multi-modal data, it may be possible to use a single model to simultaneously achieve tasks such as time sequence analysis, spatial data analysis, and inversion modeling.

However, the computational resources required for developing such large models are beyond the capacity of individual users. Currently, the development of these mainstream large models is confined to institutions such as OpenAI, Google, Facebook, Baidu, etc., due to their immense computational demands, which even existing cloud computing platforms struggle to support. To democratize the development and exploration of large models on personal computers, it's imperative to venture into new computational modes (Végh, 2023).

In the quest for the next generation of computing units, quantum computing and photonic computing emerge as burgeoning frontiers with vast potential (Sahimi & Tahmasebi, 2022; Xu et al., 2021). Quantum computing leverages the principles of quantum mechanics to process information in a radically different way, offering exponential speedups for specific types of problems (Valdez & Melin, 2023). Although these advanced computational paradigms do not universally outperform current silicon-based computing technology, quantum computing excels at optimization problems and linear algebra operations through quantum parallelism, which could accelerate the gradient descent algorithms and matrix multiplications fundamental to neural network training. Photonic computing, on the other hand, uses photons instead of electrons for computation, promising high-speed and energy-efficient processing (Xu & Jin, 2023). Its inherent parallelism and low-latency matrix-vector multiplication capabilities align well with the convolution operations and forward propagation in deep neural networks. Additionally, the analog nature of photonic processors makes them particularly suited for the continuous-valued computations prevalent in DL models.

Recent efforts have been made by scholars to address the challenges associated with applying quantum algorithms in hydrogeological modeling (Golden et al., 2022; Henderson et al., 2023). For instance, Henderson et al. (2024) proposed efficient methods for preparing the b-register and extracting valuable information from quantum computers that solve linear systems representing geologic fracture networks. Their findings indicate that quantum computing is feasible in this domain. With the rapid advancements in quantum hardware, quantum computing is expected to fundamentally transform the methodologies used in groundwater modeling. These new computational paradigms differ significantly from current algorithms, and most research remains in the theoretical and exploratory stage. The practical application of these paradigms to DL models in hydrogeological modeling is still a considerable distance away, requiring further research and development.

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2025JH000703

5. Conclusions

The integration of DL into hydrogeological modeling marks the dawn of a transformative era. Utilizing large data sets and complex algorithms, DL models have demonstrated exceptional capability in capturing intricate relationships and patterns that traditional models may overlook. This has resulted in more accurate and efficient models, particularly in areas such as time series prediction, spatial data analysis, and inverse modeling.

However, integrating DL into hydrogeological modeling is not without challenges. The scarcity and quality of data, coupled with the "black box" nature of DL models, pose significant obstacles. Groundwater systems are inherently complex, and data collection is often limited by practical conditions and cost, leading to sparse and sometimes incomplete data sets. This presents a considerable challenge for DL models, which typically require large amounts of high-quality data for training. The interpretability and explainability of DL models remain a concern, as their "black box" nature may hinder their acceptance and credibility, especially in a field where understanding the underlying processes is crucial for decision-making and policy formulation. Additionally, the computational complexity and resource requirements of DL models can be prohibitive, limiting their accessibility and applicability in real-world scenarios.

Despite these challenges, the future of DL-based hydrogeological modeling looks promising. Efforts to improve data collection and sharing, enhance data preprocessing techniques, and develop data synthesis methods are underway, addressing the critical issue of data scarcity and quality. Advances in model interpretability, such as the development of physics-informed neural networks and explainable AI, are making strides in unraveling the "black box" nature of DL models. Furthermore, innovations in model performance, through approaches such as transfer learning and hybrid modeling, are continually improving the accuracy and efficiency of DL models. Additionally, the exploration of novel computing paradigms holds significant potential to revolutionize DL-based hydrogeological modeling. Quantum computing, with its capability to handle complex computations and large data sets more efficiently than classical computing, could dramatically enhance the speed and scalability of DL models, enabling more sophisticated analyses and simulations of groundwater systems.

In conclusion, DL models have undoubtedly demonstrated their potential to drive the development of hydrogeological modeling. However, to truly advance DL-based hydrogeological modeling for practical applications, it is crucial to fully recognize the current limitations of DL models and to think critically about how to address these limitations in real-world applications. The journey of applying deep learning to hydrogeological modeling continues, but instead of ignoring the inherent limitations of DL models through elaborate what-if scenarios, as many past studies have done, the focus should shift to exploring how DL-based hydrogeological modeling can be used to solve real-world problems.

Data Availability Statement

No data were used for the research described in the article.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (NSFC: U2267217, 42141011), Shandong Key Water Conservancy Science and Technology Project (2024370203001957), and Shandong Provincial Natural Science Foundation (ZR2024QD095).

References

Adams, K. H., Reager, J. T., Rosen, P., Wiese, D. N., Farr, T. G., Rao, S., et al. (2022). Remote sensing of groundwater: Current capabilities and future directions. *Water Resources Research*, *58*(10), e2022WR032219. https://doi.org/10.1029/2022WR032219

Adombi, A. V. D. P., Chesnaux, R., & Boucher, M.-A. (2021). Review: Theory-guided machine learning applied to hydrogeology—State of the art, opportunities and future challenges. *Hydrogeology Journal*, *29*(8), 2671–2683. https://doi.org/10.1007/s10040-021-02403-2

Adombi, A. V. D. P., Chesnaux, R., & Boucher, M.-A. (2022). Comparing numerical modelling, traditional machine learning and theory-guided machine learning in inverse modeling of groundwater dynamics: A first study case application. *Journal of Hydrology*, *615*, 128600. https://doi.org/10.1016/j.jhydrol.2022.128600

Ali, A. S. A., Jazaei, F., Babakhani, P., Ashiq, M. M., Bakhshaee, A., & Waldron, B. (2024). An overview of deep learning applications in groundwater level modeling: Bridging the gap between academic research and industry applications. *Applied Computational Intelligence and Soft Computing*, *2024*(1), 9480522. https://doi.org/10.1155/2024/9480522

Alshehri, F., & Rahman, A. (2023). Coupling machine and deep learning with explainable artificial intelligence for improving prediction of groundwater quality and decision-making in arid region, Saudi Arabia. *Water*, *15*(12), 2298. https://doi.org/10.3390/w15122298

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., et al. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Arcucci, R., Zhu, J., Hu, S., & Guo, Y.-K. (2021). Deep data assimilation: Integrating deep learning with data assimilation. *Applied Sciences*, *11*(3), 1114. https://doi.org/10.3390/app11031114

Arif, T. M. (2022). *Introduction to deep learning for engineers: Using Python and Google cloud platform*. Springer Nature.

Ariyaluran Habeeb, R. A., Nasaruddin, F., Gani, A., Targio Hashem, I. A., Ahmed, E., & Imran, M. (2019). Real-time big data processing for anomaly detection: A survey. *International Journal of Information Management*, *45*, 289–307. https://doi.org/10.1016/j.ijinfomgt.2018.08.006

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Arridge, S., Maass, P., Öktem, O., & Schönlieb, C.-B. (2019). Solving inverse problems using data-driven models. *Acta Numerica*, *28*, 1–174. https://doi.org/10.1017/s0962492919000059

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, *34*(6), 26–38. https://doi.org/10.1109/MSP.2017.2743240

Asher, M. J., Croke, B. F. W., Jakeman, A. J., & Peeters, L. J. M. (2015). A review of surrogate models and their application to groundwater modeling. *Water Resources Research*, *51*(8), 5957–5973. https://doi.org/10.1002/2015WR016967

Ayzel, G. V. (2021). Deep neural networks in hydrology: The new generation of universal and efficient models. *Vestnik of Saint Petersburg University Earth Sciences*, *66*(1). https://doi.org/10.21638/spbu07.2021.101

Azevedo, L., Paneiro, G., Santos, A., & Soares, A. (2020). Generative adversarial network as a stochastic subsurface model reconstruction. *Computational Geosciences*, *24*(4), 1673–1692. https://doi.org/10.1007/s10596-020-09978-x

Bahrami, S., Doulati Ardejani, F., & Baafi, E. (2016). Application of artificial neural network coupled with genetic algorithm and simulated annealing to solve groundwater inflow problem to an advancing open pit mine. *Journal of Hydrology*, *536*, 471–484. https://doi.org/10.1016/j.jhydrol.2016.03.002

Bai, T., & Tahmasebi, P. (2022). Characterization of groundwater contamination: A transformer-based deep learning model. *Advances in Water Resources*, *164*, 104217. https://doi.org/10.1016/j.advwatres.2022.104217

Bai, T., & Tahmasebi, P. (2023). Graph neural network for groundwater level forecasting. *Journal of Hydrology*, *616*, 128792. https://doi.org/10.1016/j.jhydrol.2022.128792

Bao, J., Li, L., & Redoloza, F. (2020). Coupling ensemble smoother and deep learning with generative adversarial networks to deal with non-Gaussianity in flow and transport data assimilation. *Journal of Hydrology*, *590*, 125443. https://doi.org/10.1016/j.jhydrol.2020.125443

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Başağaoğlu, H., Chakraborty, D., Lago, C. D., Gutierrez, L., Şahinli, M. A., Giacomoni, M., et al. (2022). A review on interpretable and explainable artificial intelligence in hydroclimatic applications. *Water*, *14*(8), 1230. https://doi.org/10.3390/w14081230

Bejani, M. M., & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review*, *54*(8), 6391–6438. https://doi.org/10.1007/s10462-021-09975-1

Bellian, J. A., Kerans, C., & Jennette, D. C. (2005). Digital outcrop models: Applications of terrestrial scanning Lidar technology in stratigraphic modeling. *Journal of Sedimentary Research*, *75*(2), 166–176. https://doi.org/10.2110/jsr.2005.013

Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*(6433), eaau0323. https://doi.org/10.1126/science.aau0323

Bernhardt, M., Castro, D. C., Tanno, R., Schwaighofer, A., Tezcan, K. C., Monteiro, M., et al. (2022). Active label cleaning for improved dataset quality under resource constraints. *Nature Communications*, *13*(1), 1161. https://doi.org/10.1038/s41467-022-28818-3

Bjarkason, E. K., Maclaren, O. J., O'Sullivan, J. P., & O'Sullivan, M. J. (2018). Randomized truncated SVD levenberg-marquardt approach to geothermal natural state and history matching. *Water Resources Research*, *54*(3), 2376–2404. https://doi.org/10.1002/2017WR021870

Boggs, J. M., Young, S. C., Beard, L. M., Gelhar, L. W., Rehfeldt, K. R., & Adams, E. E. (1992). Field study of dispersion in a heterogeneous aquifer: 1. Overview and site description. *Water Resources Research*, *28*(12), 3281–3291. https://doi.org/10.1029/92WR01756

Bonazzi, A., Dentz, M., & de Barros, F. P. J. (2022). Mixing in multidimensional porous media: A numerical study of the effects of source configuration and heterogeneity. *Transport in Porous Media*, *146*(1–2), 369–393. https://doi.org/10.1007/s11242-022-01822-3

Boo, K. B. W., El-Shafie, A., Othman, F., Khan, M. M. H., Birima, A. H., & Ahmed, A. N. (2024). Groundwater level forecasting with machine learning models: A review. *Water Research*, *252*, 121249. https://doi.org/10.1016/j.watres.2024.121249

Bourou, S., El Saer, A., Velivassaki, T.-H., Voulkidis, A., & Zahariadis, T. (2021). A review of tabular data synthesis using GANs on an IDS dataset. *Information*, *12*(9), 375. https://doi.org/10.3390/info12090375

Cai, H., Liu, S., Shi, H., Zhou, Z., Jiang, S., & Babovic, V. (2022). Toward improved lumped groundwater level predictions at catchment scale: Mutual integration of water balance mechanism and deep learning method. *Journal of Hydrology*, *613*, 128495. https://doi.org/10.1016/j.jhydrol.2022.128495

Cai, H., Shi, H., Zhou, Z., Liu, S., & Babovic, V. (2024). Explaining the mechanism of multiscale groundwater drought events: A new perspective from interpretable deep learning model. *Water Resources Research*, *60*(7), e2023WR035139. https://doi.org/10.1029/2023WR035139

Canchumuni, S. W. A., Castro, J. D. B., Potratz, J., Emerick, A. A., & Pacheco, M. A. C. (2021). Recent developments combining ensemble smoother and deep generative networks for facies history matching. *Computational Geosciences*, *25*(1), 433–466. https://doi.org/10.1007/s10596-020-10015-0

Canchumuni, S. W. A., Emerick, A. A., & Pacheco, M. A. C. (2019). History matching geological facies models based on ensemble smoother and deep generative models. *Journal of Petroleum Science and Engineering*, *177*, 941–958. https://doi.org/10.1016/j.petrol.2019.02.037

Cao, M., Dai, Z., Chen, J., Yin, H., Zhang, X., Wu, J., et al. (2025). An integrated framework of deep learning and entropy theory for enhanced high-dimensional permeability field identification in heterogeneous aquifers. *Water Research*, *268*, 122706. https://doi.org/10.1016/j.watres.2024.122706

Cao, M., Dai, Z., Jia, S., Samper, J., Ling, H., Du, Z., et al. (2024). Identification of solute transport parameters in fractured granites with heterogeneous apertures. *Journal of Hydrology*, *633*, 130938. https://doi.org/10.1016/j.jhydrol.2024.130938

Carle, S. F. (1999). *T-PROGS: Transition probability geostatistical software, version 2.1*. Department of land, Air and water resources. University of California.

Carle, S. F., & Fogg, G. E. (1997). Modeling spatial variability with one and multidimensional continuous-lag Markov chains. *Mathematical Geology*, *29*(7), 891–918. https://doi.org/10.1023/A:1022303706942

Carroll, K. C., Brusseau, M. L., Tick, G. R., & Soltanian, M. R. (2024). Rethinking pump-and-treat remediation as maximizing contaminated groundwater. *Science of the Total Environment*, *918*, 170600. https://doi.org/10.1016/j.scitotenv.2024.170600

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(8), 832. https://doi.org/10.3390/electronics8080832

Chen, B., Harp, D. R., Lin, Y., Keating, E. H., & Pawar, R. J. (2018). Geologic CO2 sequestration monitoring design: A machine learning and uncertainty quantification based approach. *Applied Energy*, *225*, 332–345. https://doi.org/10.1016/j.apenergy.2018.05.044

Chen, C., He, W., Zhou, H., Xue, Y., & Zhu, M. (2020). A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China. *Scientific Reports*, *10*(1), 3904. https://doi.org/10.1038/s41598-020-60698-9

Chen, J., Dai, Z., Dong, S., Zhang, X., Sun, G., Wu, J., et al. (2022). Integration of deep learning and information theory for designing monitoring networks in heterogeneous aquifer systems. *Water Resources Research*, *58*(10), e2022WR032429. https://doi.org/10.1029/2022WR032429

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chen, J., Dai, Z., Yang, Z., Pan, Y., 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 Resources Research*, *57*(12), e2021WR030595. https://doi.org/10.1029/2021WR030595

Chen, L., He, A., Zhao, J., Kang, Q., Li, Z.-Y., Carmeliet, J., et al. (2022). Pore-scale modeling of complex transport phenomena in porous media. *Progress in Energy and Combustion Science*, *88*, 100968. https://doi.org/10.1016/j.pecs.2021.100968

Chen, R. T., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. (2018). Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, *31*.

Chen, W., Dai, Z., Samper, J., Ling, H., Zhan, C., Ma, F., et al. (2023). Impact of reactive mineral facies distributions on radionuclide sorption properties in multiscale heterogeneous granite rocks. *Hydrogeology Journal*, *31*(6), 1581–1597. https://doi.org/10.1007/s10040-023-02672-z

Cheng, G., Xie, X., Han, J., Guo, L., & Xia, G. S. (2020). Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 3735–3756. https://doi.org/10.1109/JSTARS.2020.3005403

Cheng, S., & Qiu, M. (2022). Observation error covariance specification in dynamical systems for data assimilation using recurrent neural networks. *Neural Computing & Applications*, *34*(16), 13149–13167. https://doi.org/10.1007/s00521-021-06739-4

Cheng, Y., Wang, H., Bao, Y., & Lu, F. (2024). Appearance-based gaze estimation with deep learning: A review and benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.48550/arXiv.2104.12668

Chilton, J., & Foster, S. (2024). Long datasets for improved understanding, management and protection of groundwater. *Hydrogeology Journal*, *32*(2), 347–352. https://doi.org/10.1007/s10040-023-02759-7

Cirpka, O. A., Stettler, M.-M., & Dentz, M. (2022). Spatial Markov model for the prediction of travel-time-based solute dispersion in three-dimensional heterogeneous media. *Water Resources Research*, *58*(6), e2022WR032215. https://doi.org/10.1029/2022WR032215

Cnudde, V., & Boone, M. N. (2013). High-resolution X-ray computed tomography in geosciences: A review of the current technology and applications. *Earth-Science Reviews*, *123*, 1–17. https://doi.org/10.1016/j.earscirev.2013.04.003

Coates, A., Huval, B., Wang, T., Wu, D., Catanzaro, B., & Andrew, N. (2013). *Deep learning with COTS HPC systems* proceedings of the 30th international conference on machine learning. *Proceedings of Machine Learning Research*. https://proceedings.mlr.press/v28/coates13.html

Collenteur, R. A., Haaf, E., Bakker, M., Liesch, T., Wunsch, A., Soonthornrangsan, J., et al. (2024). Data-driven modelling of hydraulic-head time series: Results and lessons learned from the 2022 groundwater time series modelling challenge. *Hydrology and Earth System Sciences*, *28*(23), 5193–5208. https://doi.org/10.5194/hess-28-5193-2024

Condon, L. E., Kollet, S., Bierkens, M. F. P., Fogg, G. E., Maxwell, R. M., Hill, M. C., et al. (2021). Global groundwater modeling and monitoring: Opportunities and challenges. *Water Resources Research*, *57*(12), e2020WR029500. https://doi.org/10.1029/2020WR029500

Constable, S. C., Parker, R. L., & Constable, C. G. (1987). Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, *52*(3), 289–300. https://doi.org/10.1190/1.1442303

Cui, F., Al-Sudani, Z. A., Hassan, G. S., Afan, H. A., Ahammed, S. J., & Yaseen, Z. M. (2022). Boosted artificial intelligence model using improved alpha-guided grey wolf optimizer for groundwater level prediction: Comparative study and insight for federated learning technology. *Journal of Hydrology*, *606*, 127384. https://doi.org/10.1016/j.jhydrol.2021.127384

Cui, Z., Chen, Q., & Liu, G. (2022). Characterization of subsurface hydrogeological structures with convolutional conditional neural processes on limited training data. *Water Resources Research*, *58*(12), e2022WR033161. https://doi.org/10.1029/2022WR033161

Cui, Z., Chen, Q., Liu, G., & Xun, L. (2024). SA-RelayGANs: A novel framework for the characterization of complex hydrological structures based on GANs and self-attention mechanism. *Water Resources Research*, *60*(1), e2023WR035932. https://doi.org/10.1029/2023WR035932

Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, *92*(3), 88. https://doi.org/10.1007/s10915-022-01939-z

Cuthbert, M. O., Gleeson, T., Moosdorf, N., Befus, K. M., Schneider, A., Hartmann, J., & Lehner, B. (2019). Global patterns and dynamics of climate–groundwater interactions. *Nature Climate Change*, *9*(2), 137–141. https://doi.org/10.1038/s41558-018-0386-4

Dahal, K., Sharma, S., Shakya, A., Talchabhadel, R., Adhikari, S., Pokharel, A., et al. (2023). Identification of groundwater potential zones in data-scarce mountainous region using explainable machine learning. *Journal of Hydrology*, *627*, 130417. https://doi.org/10.1016/j.jhydrol.2023.130417

Dai, Z., Zhan, C., Dong, S., Yin, S., Zhang, X., & Soltanian, M. R. (2020). How does resolution of sedimentary architecture data affect plume dispersion in multiscale and hierarchical systems? *Journal of Hydrology*, *582*, 124516. https://doi.org/10.1016/j.jhydrol.2019.124516

Dai, Z., Zhan, C., Soltanian, M. R., Ritzi, R. W., & Zhang, X. (2019). Identifying spatial correlation structure of multimodal permeability in hierarchical media with Markov chain approach. *Journal of Hydrology*, *568*, 703–715. https://doi.org/10.1016/j.jhydrol.2018.11.032

Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371*.

de Graaf, I. E. M., Sutanudjaja, E. H., van Beek, L. P. H., & Bierkens, M. F. P. (2015). A high-resolution global-scale groundwater model. *Hydrology and Earth System Sciences*, *19*(2), 823–837. https://doi.org/10.5194/hess-19-823-2015

Deng, C., Zhang, T., He, Z., Chen, Q., Shi, Y., Xu, Y., et al. (2024). K2: A foundation language model for geoscience knowledge understanding and utilization. In *Proceedings of the 17th ACM international conference on web search and data mining*. Merida, Mexico. https://doi.org/10.1145/3616855.3635772

Dhanasekaran, S., Rajput, K., Yuvaraj, N., Aeri, M., Shukla, R. P., & Singh, S. K. (2024). Utilizing cloud computing for distributed training of deep learning models. In *2024 second International Conference on Data Science and Information System (ICDSIS)*.

Divine, C., Fortner III, E., Barton, C. O., Cisco, C., Hollister, C., Profusek, D., & Spurlin, M. (2024). Leveraging hydrogeologic-based data—Reduce, repurpose, reimagine. *Groundwater Monitoring & Remediation*, *44*(3), 13–21. https://doi.org/10.1111/gwmr.12673

Durlofsky, L. J. (2003). Upscaling of geocellular models for reservoir flow simulation: A review of recent progress. In *7th international forum on reservoir simulation bühl/Baden-Baden, Germany*.

El Mezouary, L., Hadri, A., Kharrou, M. H., Fakır, Y., Elfarchouni, A., Bouchaou, L., & Chehbouni, A. (2024). Contribution to advancing aquifer geometric mapping using machine learning and deep learning techniques: A case study of the AL Haouz-Mejjate aquifer, Marrakech, Morocco. *Applied Water Science*, *14*(5), 102. https://doi.org/10.1007/s13201-024-02162-x

Elmorsy, M., El-Dakhakhni, W., & Zhao, B. (2022). Generalizable permeability prediction of digital porous media via a novel multi-scale 3D convolutional neural network. *Water Resources Research*, *58*(3), e2021WR031454. https://doi.org/10.1029/2021WR031454

ElShawi, R., Sherif, Y., Al-Mallah, M., & Sakr, S. (2021). Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Computational Intelligence*, *37*(4), 1633–1650. https://doi.org/10.1111/coin.12410

Ershadnia, R., Moeini, F., Abolfazl Hosseini, S., Dai, Z., & Reza Soltanian, M. (2024). Predicting multiphase flow behavior of methane in shallow unconfined aquifers using conditional deep convolutional generative adversarial network. *Journal of Hydrology*, *641*, 131674. https://doi.org/10.1016/j.jhydrol.2024.131674

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Fang, K., Shen, C., Kifer, D., & Yang, X. (2017). Prolongation of SMAP to spatiotemporally seamless coverage of continental U.S. Using a deep learning neural network. *Geophysical Research Letters*, *44*(21), 11–39. https://doi.org/10.1002/2017GL075619
Feng, D., Beck, H., de Bruijn, J., Sahu, R. K., Satoh, Y., Wada, Y., et al. (2024). Deep dive into hydrologic simulations at global scale: Harnessing the power of deep learning and physics-informed differentiable models (δHBV-globe1.0-hydroDL). *Geoscientific Model Development*, *17*(18), 7181–7198. https://doi.org/10.5194/gmd-17-7181-2024
Feng, L., Mo, S., Sun, A. Y., Wu, J., & Shi, X. (2024). Uncertainty quantification of CO_2 plume migration in highly channelized aquifers using probabilistic convolutional neural networks. *Advances in Water Resources*, *183*, 104607. https://doi.org/10.1016/j.advwatres.2023.104607
Fernàndez-Garcia, D., Illangasekare, T. H., & Rajaram, H. (2005). Differences in the scale-dependence of dispersivity estimated from temporal and spatial moments in chemically and physically heterogeneous porous media. *Advances in Water Resources*, *28*(7), 745–759. https://doi.org/10.1016/j.advwatres.2004.12.011
Fiori, A., Boso, F., de Barros, F. P. J., De Bartolo, S., Frampton, A., Severino, G., et al. (2010). An indirect assessment on the impact of connectivity of conductivity classes upon longitudinal asymptotic macrodispersivity. *Water Resources Research*, *46*(8). https://doi.org/10.1029/2009WR008590
Fitch, P., Brodaric, B., Stenson, M., & Booth, N. (2016). Integrated groundwater data management. *Integrated Groundwater Management: Concepts, Approaches and Challenges*, 667–692. https://doi.org/10.1007/978-3-319-23576-9_26
Fonseca, J., & Bacao, F. (2023). Tabular and latent space synthetic data generation: A literature review. *Journal of Big Data*, *10*(1), 115. https://doi.org/10.1186/s40537-023-00792-7
Fu, G., Jin, Y., Sun, S., Yuan, Z., & Butler, D. (2022). The role of deep learning in urban water management: A critical review. *Water Research*, *223*, 118973. https://doi.org/10.1016/j.watres.2022.118973
Gao, Y. (2017). Dealing with missing data in hydrology: Data analysis of discharge and groundwater time-series in Northeast Germany.
Gao, Y., Liu, Y., Zhang, H., Li, Z., Zhu, Y., Lin, H., & Yang, M. (2020). Estimating GPU memory consumption of deep learning models. In *Proceedings of the 28th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering*. Virtual Event. https://doi.org/10.1145/3368089.3417050
Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence*, *2*(11), 665–673. https://doi.org/10.1038/s42256-020-00257-z
Geng, X., Boufadel, M. C., Lee, K., & An, C. (2020). Characterization of pore water flow in 3-D heterogeneous permeability fields. *Geophysical Research Letters*, *47*(3), e2019GL086879. https://doi.org/10.1029/2019GL086879
Gewin, V. (2016). Data sharing: An open mind on open data. *Nature*, *529*(7584), 117–119. https://doi.org/10.1038/nj7584-117a
Ghorbanfekr, H., Kerstens, P. J., & Dirix, K. (2025). Classification of geological borehole descriptions using a domain adapted large language model. *Applied Computing and Geosciences*, *25*, 100229. https://doi.org/10.1016/j.acags.2025.100229
Golden, J., O'Malley, D., & Viswanathan, H. (2022). Quantum computing and preconditioners for hydrological linear systems. *Scientific Reports*, *12*(1), 22285. https://doi.org/10.1038/s41598-022-25727-9
Gómez-Hernández, J. J., Butler, J. J., Fiori, A., Bolster, D., Cvetkovic, V., Dagan, G., & Hyndman, D. (2017). Introduction to special section on Modeling highly heterogeneous aquifers: Lessons learned in the last 30 years from the MADE experiments and others. *Water Resources Research*, *53*(4), 2581–2584. https://doi.org/10.1002/2017WR020774
Gómez-Hernández, J. J., & Wen, X.-H. (1998). To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Advances in Water Resources*, *21*(1), 47–61. https://doi.org/10.1016/s0309-1708(96)00031-0
Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.
Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*.
Gorelick, S. M., & Zheng, C. (2015). Global change and the groundwater management challenge. *Water Resources Research*, *51*(5), 3031–3051. https://doi.org/10.1002/2014WR016825
Goswami, S., Kontolati, K., Shields, M. D., & Karniadakis, G. E. (2022). Deep transfer operator learning for partial differential equations under conditional shift. *Nature Machine Intelligence*, *4*(12), 1155–1164. https://doi.org/10.1038/s42256-022-00569-2
Gou, J., Yu, B., Maybank, S. J., & Tao, D. (2021). Knowledge distillation: A survey. *International Journal of Computer Vision*, *129*(6), 1789–1819. https://doi.org/10.1007/s11263-021-01453-z
Haensch, W., Gokmen, T., & Puri, R. (2019). The next generation of deep learning hardware: Analog computing. *Proceedings of the IEEE*, *107*(1), 108–122. https://doi.org/10.1109/JPROC.2018.2871057
Hammond, G. E., Lichtner, P. C., & Mills, R. T. (2014). Evaluating the performance of parallel subsurface simulators: An illustrative example with PFLOTRAN. *Water Resources Research*, *50*(1), 208–228. https://doi.org/10.1002/2012WR013483
Han, Z., Kang, X., Wu, J., & Shi, X. (2022). Characterization of the non-Gaussian hydraulic conductivity field via deep learning-based inversion of hydraulic-head and self-potential data. *Journal of Hydrology*, *610*, 127830. https://doi.org/10.1016/j.jhydrol.2022.127830
Harp, D. R., Dai, Z., Wolfsberg, A. V., Vrugt, J. A., Robinson, B. A., & Vesselinov, V. V. (2008). Aquifer structure identification using stochastic inversion. *Geophysical Research Letters*, *35*(8). https://doi.org/10.1029/2008GL033585
He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), Computer vision – ECCV 2016 *cham*.
He, L., Cheng, H., Nan, Z., Gong, Y., Guo, H., Mao, J., & Zhang, J. (2025). Improving joint identification of groundwater contaminant source and non-Gaussian distributed conductivity field using a deep learning-based ensemble smoother. *Journal of Hydrology*, *658*, 133202. https://doi.org/10.1016/j.jhydrol.2025.133202
Henderson, J. M., Kath, J., Golden, J. K., Percus, A. G., & O'Malley, D. (2024). Addressing quantum's "fine print" with efficient state preparation and information extraction for quantum algorithms and geologic fracture networks. *Scientific Reports*, *14*(1), 3592. https://doi.org/10.1038/s41598-024-52759-0
Henderson, J. M., Podzorova, M., Cerezo, M., Golden, J. K., Gleyzer, L., Viswanathan, H. S., & O'Malley, D. (2023). Quantum algorithms for geologic fracture networks. *Scientific Reports*, *13*(1), 2906. https://doi.org/10.1038/s41598-023-29643-4
Hilton, A., & Jasechko, S. (2023). Widespread aquifer depressurization after a century of intensive groundwater use in USA. *Science Advances*, *9*(37), eadh2992. https://doi.org/10.1126/sciadv.adh2992
Hong, K., Han, E., & Kang, K. (2017). Determination of geological strength index of jointed rock mass based on image processing. *Journal of Rock Mechanics and Geotechnical Engineering*, *9*(4), 702–708. https://doi.org/10.1016/j.jrmge.2017.05.001
Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8
Howell, J. A., Martinius, A. W., & Good, T. R. (2014). The application of outcrop analogues in geological modelling: A review, present status and future outlook. *Geological Society, London, Special Publications*, *387*(1), 1–25. https://doi.org/10.1144/SP387.12

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Hu, Y., Mai, G., Cundy, C., Choi, K., Lao, N., Liu, W., et al. (2023). Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages. *International Journal of Geographical Information Science*, *37*(11), 2289–2318. https://doi.org/10.1080/13658816.2023.2266495

Huerta, E. A., Khan, A., Davis, E., Bushell, C., Gropp, W. D., Katz, D. S., et al. (2020). Convergence of artificial intelligence and high performance computing on NSF-supported cyberinfrastructure. *Journal of Big Data*, *7*(1), 88. https://doi.org/10.1186/s40537-020-00361-2

Hyndman, D. W., Harris, J. M., & Gorelick, S. M. (1994). Coupled seismic and tracer test inversion for aquifer property characterization. *Water Resources Research*, *30*(7), 1965–1977. https://doi.org/10.1029/94WR00950

Ighalo, J. O., Adeniyi, A. G., & Marques, G. (2021). Internet of Things for water quality monitoring and assessment: A comprehensive review. In A. E. Hassanien, R. Bhatnagar, & A. Darwish (Eds.), *Artificial intelligence for sustainable development: Theory, practice and future applications* (pp. 245–259). Springer International Publishing. https://doi.org/10.1007/978-3-030-51920-9_13

Ikard, S. J., Carroll, K. C., Rucker, D. F., Adams, R. F., & Brooks, S. C. (2023). Geoelectric characterization of hyporheic exchange flow in the bedrock-lined streambed of East Fork Poplar Creek, Oak Ridge, Tennessee. *Geophysical Research Letters*, *50*(8), e2022GL102616. https://doi.org/10.1029/2022GL102616

Irving, J., & Singha, K. (2010). Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. *Water Resources Research*, *46*(11). https://doi.org/10.1029/2009WR008340

Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D., & Makedon, F. (2021). A survey on contrastive self-supervised learning. *Technologies*, *9*(1), 2. https://doi.org/10.3390/technologies9010002

Jamil, A., Rucker, D. F., Lu, D., Brooks, S. C., Tartakovsky, A. M., Cao, H., & Carroll, K. C. (2024). Comparison of machine learning and electrical resistivity arrays to inverse modeling for locating and characterizing subsurface targets. *Journal of Applied Geophysics*, *229*, 105493. https://doi.org/10.1016/j.jappgeo.2024.105493

Jardani, A., Vu, T. M., & Fischer, P. (2022). Use of convolutional neural networks with encoder-decoder structure for predicting the inverse operator in hydraulic tomography. *Journal of Hydrology*, *604*, 127233. https://doi.org/10.1016/j.jhydrol.2021.127233

Jauro, F., Chiroma, H., Gital, A. Y., Almutairi, M., Abdulhamid, S. i. M., & Abawajy, J. H. (2020). Deep learning architectures in emerging cloud computing architectures: Recent development, challenges and next research trend. *Applied Soft Computing*, *96*, 106582. https://doi.org/10.1016/j.asoc.2020.106582

Jena, B., Nayak, G. K., & Saxena, S. (2022). High-performance computing and its requirements in deep learning. *High-Performance Medical Image Processing*, 255–288. https://doi.org/10.1201/9781003190011-12

Jeong, J., Park, E., Chen, H., Kim, K.-Y., Shik Han, W., & Suk, H. (2020). Estimation of groundwater level based on the robust training of recurrent neural networks using corrupted data. *Journal of Hydrology*, *582*, 124512. https://doi.org/10.1016/j.jhydrol.2019.124512

Jiang, S., Sweet, L.-b., Blougouras, G., Brenning, A., Li, W., Reichstein, M., et al. (2024). How interpretable machine learning can benefit process understanding in the geosciences. *Earth's Future*, *12*(7), e2024EF004540. https://doi.org/10.1029/2024EF004540

Jiang, S., Zheng, Y., Wang, C., & Babovic, V. (2022). Uncovering flooding mechanisms across the contiguous United States through interpretive deep learning on representative catchments. *Water Resources Research*, *58*(1), e2021WR030185. https://doi.org/10.1029/2021WR030185

Jones, D., Jones, N., Greer, J., & Nelson, J. (2015). A cloud-based MODFLOW service for aquifer management decision support. *Computers and Geosciences*, *78*, 81–87. https://doi.org/10.1016/j.cageo.2015.02.014

Justus, D., Brennan, J., Bonner, S., & McGough, A. S. (2018). Predicting the computational cost of deep learning models. In *2018 IEEE international conference on big data (Big Data)*.

Kang, X., Kokkinaki, A., Kitanidis, P. K., Shi, X., Lee, J., Mo, S., & Wu, J. (2021). Hydrogeophysical characterization of nonstationary DNAPL source zones by integrating a convolutional variational autoencoder and ensemble smoother. *Water Resources Research*, *57*(2), Article e2020WR028538. https://doi.org/10.1029/2020wr028538

Karimpouli, S., & Tahmasebi, P. (2019). Segmentation of digital rock images using deep convolutional autoencoder networks. *Computers and Geosciences*, *126*, 142–150. https://doi.org/10.1016/j.cageo.2019.02.003

Kim, H., Choi, M., Lim, B., & Lee, K. M. (2018). Task-aware image downscaling. In *Proceedings of the European conference on computer vision (ECCV)*.

Kingma, D. P., & Welling, M. (2014). In *Auto-encoding variational bayes*. *arXiv.org*.

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

Kumar, C. (2014). *Groundwater data requirement and analysis*. GRIN Publishing.

Kwak, K., & Lee, E. H. (2024). Impact of road transport system on groundwater quality inferred from explainable artificial intelligence (XAI). *Science of the Total Environment*, *917*, 170388. https://doi.org/10.1016/j.scitotenv.2024.170388

Lagaris, I. E., Likas, A., & Fotiadis, D. I. (1998). Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, *9*(5), 987–1000. https://doi.org/10.1109/72.712178

Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., & Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resources Research*, *49*(5), 2664–2682. https://doi.org/10.1002/wrcr.20226

Lang, J., & Qiu, F. (2021). Hybrid data assimilation based on multilayer perceptron. *Journal of Physics: Conference Series*, *1948*(1), 012161. https://doi.org/10.1088/1742-6596/1948/1/012161

Lattu, A. (2023). Open science in China: An open and closed case. https://doi.org/10.17617/2.3553727

Lauzon, D. (2024). A U-Net architecture as a surrogate model combined with a geostatistical spectral algorithm for transient groundwater flow inverse problems. *Advances in Water Resources*, *189*, 104726. https://doi.org/10.1016/j.advwatres.2024.104726

Lawley, C. J. M., Gadd, M. G., Parsa, M., Lederer, G. W., Graham, G. E., & Ford, A. (2023). Applications of natural language processing to geoscience text data and prospectivity modeling. *Natural Resources Research*, *32*(4), 1503–1527. https://doi.org/10.1007/s11053-023-10216-1

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

Lee, E., & Kam, J. (2023). Deciphering the black box of deep learning for multi-purpose dam operation modeling via explainable scenarios. *Journal of Hydrology*, *626*, 130177. https://doi.org/10.1016/j.jhydrol.2023.130177

Lever, J., Krzywinski, M., & Altman, N. (2016). Model selection and overfitting. *Nature Methods*, *13*(9), 703–704. https://doi.org/10.1038/nmeth.3968

Li, K., Persaud, D., Choudhary, K., DeCost, B., Greenwood, M., & Hattrick-Simpers, J. (2023). Exploiting redundancy in large materials datasets for efficient machine learning with less data. *Nature Communications*, *14*(1), 7283. https://doi.org/10.1038/s41467-023-42992-y

Li, S., Liu, B., Ren, Y., Chen, Y., Yang, S., Wang, Y., & Jiang, P. (2020). Deep-learning inversion of seismic data. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(3), 2135–2149. https://doi.org/10.1109/TGRS.2019.2953473

Li, X., Feng, M., Ran, Y., Su, Y., Liu, F., Huang, C., et al. (2023). Big Data in Earth system science and progress towards a digital twin. *Nature Reviews Earth and Environment*, *4*(5), 319–332. https://doi.org/10.1038/s43017-023-00409-w

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., et al. (2022). Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems*, *64*(12), 3197–3234. https://doi.org/10.1007/s10115-022-01756-8

Li, X., Zhang, G., Huang, H. H., Wang, Z., & Zheng, W. (2016). Performance analysis of GPU-based convolutional neural networks. In *2016 45th international conference on parallel processing (ICPP)*.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2022). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(12), 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable AI: A review of machine learning interpretability methods. *Entropy*, *23*(1), 18. https://doi.org/10.3390/e23010018

Liu, L., Wu, Y., Wei, W., Cao, W., Sahin, S., & Zhang, Q. (2018). Benchmarking deep learning frameworks: Design considerations, metrics and beyond. In *2018 IEEE 38th international conference on distributed Computing Systems (ICDCS)*.

Liu, M., & Mukerji, T. (2022). Multiscale fusion of digital rock images based on deep generative adversarial networks. *Geophysical Research Letters*, *49*(9), e2022GL098342. https://doi.org/10.1029/2022GL098342

Liu, Q., Gui, D., Zhang, L., Niu, J., Dai, H., Wei, G., & Hu, B. X. (2022). Simulation of regional groundwater levels in arid regions using interpretable machine learning models. *Science of the Total Environment*, *831*, 154902. https://doi.org/10.1016/j.scitotenv.2022.154902

Liu, X., Wang, X., & Matwin, S. (2018). Improving the interpretability of deep neural networks with knowledge distillation. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*.

Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, *3*(3), 218–229. https://doi.org/10.1038/s42256-021-00302-5

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Luo, G. (2016). A review of automatic selection methods for machine learning algorithms and hyper-parameter values. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *5*(1), 18. https://doi.org/10.1007/s13721-016-0125-6

Luo, J., Ma, X., Ji, Y., Li, X., Song, Z., & Lu, W. (2023). Review of machine learning-based surrogate models of groundwater contaminant modeling. *Environmental Research*, *238*, 117268. https://doi.org/10.1016/j.envres.2023.117268

Ma, X., Zhang, K., Wang, J., Yao, C., Yang, Y., Sun, H., & Yao, J. (2022). An efficient spatial-temporal convolution recurrent neural network surrogate model for history matching. *SPE Journal*, *27*(2), 1160–1175. https://doi.org/10.2118/208604-pa

Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, *51*, 47–60. https://doi.org/10.1016/j.future.2014.10.029

Ma, Z., He, X., Kwak, H., Gao, J., Sun, S., & Yan, B. (2023). Enhancing rock image segmentation in digital rock physics: A fusion of generative ai and state-of-the-art neural networks. *arXiv preprint arXiv:2311.06079*.

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(2), 645–657. https://doi.org/10.1109/TGRS.2016.2612821

Mahmoodzadeh, A., Mohammadi, M., Noori, K. M. G., Khishe, M., Ibrahim, H. H., Ali, H. F. H., & Abdulhamid, S. N. (2021). Presenting the best prediction model of water inflow into drill and blast tunnels among several machine learning techniques. *Automation in Construction*, *127*, 103719. https://doi.org/10.1016/j.autcon.2021.103719

Mannino, M., & Abouzied, A. (2019). Is this real? Generating synthetic data that looks real. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*.

Manzione, R. L., & Castrignanò, A. (2019). A geostatistical approach for multi-source data fusion to predict water table depth. *Science of the Total Environment*, *696*, 133763. https://doi.org/10.1016/j.scitotenv.2019.133763

Mehonic, A., & Kenyon, A. J. (2022). Brain-inspired computing needs a master plan. *Nature*, *604*(7905), 255–260. https://doi.org/10.1038/s41586-021-04362-w

Mlynarczuk, M., & Skiba, M. (2017). The application of artificial intelligence for the identification of the maceral groups and mineral components of coal. *Computers and Geosciences*, *103*, 133–141. https://doi.org/10.1016/j.cageo.2017.03.011

Mo, S., Zabaras, N., Shi, X., & Wu, J. (2019). Deep autoregressive neural networks for high-dimensional inverse problems in groundwater contaminant source identification. *Water Resources Research*, *55*(5), 3856–3881. https://doi.org/10.1029/2018WR024638

Mo, S., Zabaras, N., Shi, X., & Wu, J. (2020). Integration of adversarial autoencoders with residual dense convolutional networks for estimation of non-Gaussian hydraulic conductivities. *Water Resources Research*, *56*(2), e2019WR026082. https://doi.org/10.1029/2019WR026082

Mo, S., Zhong, Y., Forootan, E., Shi, X., Feng, W., Yin, X., & Wu, J. (2022). Hydrological droughts of 2017–2018 explained by the Bayesian reconstruction of GRACE(-FO) fields. *Water Resources Research*, *58*(9), e2022WR031997. https://doi.org/10.1029/2022WR031997

Moeini, F., Ershadnia, R., Rubinstein, R. L., Versteeg, R., Li, P., McGarr, J. T., et al. (2024). Employing generative adversarial neural networks as surrogate model for reactive transport modeling in the hyporheic zone. *Journal of Hydrology*, *639*, 131485. https://doi.org/10.1016/j.jhydrol.2024.131485

More, N., Galphade, M., Nikam, V. B., & Banerjee, B. (2021). High-performance computing: A deep learning perspective. In A. Suresh & S. Paiva (Eds.), *Deep learning and edge computing solutions for high performance computing* (pp. 247–268). Springer International Publishing. https://doi.org/10.1007/978-3-030-60265-9_15

Moudgil, P. S., & Rao, G. S. (2023). Groundwater levels estimation from GRACE/GRACE-FO and hydro-meteorological data using deep learning in Ganga River basin, India. *Environmental Earth Sciences*, *82*(19), 441. https://doi.org/10.1007/s12665-023-11137-1

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., et al. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, *57*(3), e2020WR028091. https://doi.org/10.1029/2020WR028091

Neuman, S. P. (1972). Theory of flow in unconfined aquifers considering delayed response of the water table. *Water Resources Research*, *8*(4), 1031–1045. https://doi.org/10.1029/WR008i004p01031

Nik, V. M. (2016). Making energy simulation easier for future climate – Synthesizing typical and extreme weather data sets out of Regional Climate Models (RCMs). *Applied Energy*, *177*, 204–226. https://doi.org/10.1016/j.apenergy.2016.05.107

Nilsson, B., Højberg, A. L., Refsgaard, J. C., & Troldborg, L. (2007). Uncertainty in geological and hydrogeological data. *Hydrology and Earth System Sciences*, *11*(5), 1551–1561. https://doi.org/10.5194/hess-11-1551-2007

Niu, X., Lu, C., Zhang, Y., Zhang, Y., Wu, C., Saidy, E., et al. (2023). Hysteresis response of groundwater depth on the influencing factors using an explainable learning model framework with Shapley values. *Science of the Total Environment*, *904*, 166662. https://doi.org/10.1016/j.scitotenv.2023.166662

Nosouhian, S., Nosouhian, F., & Khoshouei, A. K. (2021). A review of recurrent neural network architecture for sequence learning: Comparison between LSTM and GRU. https://doi.org/10.20944/preprints202107.0252.v1

Nourani, V., Jabbarian Paknezhad, N., Ng, A., Wen, Z., Dabrowska, D., & Üzelaltınbulat, S. (2024). Application of the machine learning methods for GRACE data based groundwater modeling, a systematic review. *Groundwater for Sustainable Development*, *25*, 101113. https://doi.org/10.1016/j.gsd.2024.101113

Novakovsky, G., Dexter, N., Libbrecht, M. W., Wasserman, W. W., & Mostafavi, S. (2023). Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nature Reviews Genetics*, *24*(2), 125–137. https://doi.org/10.1038/s41576-022-00532-2

Ohmer, M., Liesch, T., & Wunsch, A. (2022). Spatiotemporal optimization of groundwater monitoring networks using data-driven sparse sensing methods. *Hydrology and Earth System Sciences*, *26*(15), 4033–4053. https://doi.org/10.5194/hess-26-4033-2022

Oladeji, E., Parsekian, A., & Grana, D. (2024). Machine learning-based hydrofacies classification: Effects of noise and regularization. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(3), e2024JH000197. https://doi.org/10.1029/2024JH000197

Pan, W., Chen, J., Mohamed, S., Jo, H., Santos, J. E., & Pyrcz, M. J. (2023). Efficient subsurface modeling with sequential patch generative adversarial neural networks. In *SPE annual technical conference and exhibition*.

Pan, Y., Zeng, X., Xu, H., Sun, Y., Wang, D., & Wu, J. (2021). Evaluation of Gaussian process regression kernel functions for improving groundwater prediction. *Journal of Hydrology*, *603*, 126960. https://doi.org/10.1016/j.jhydrol.2021.126960

Panahi, M., Sadhasivam, N., Pourghasemi, H. R., Rezaie, F., & Lee, S. (2020). Spatial prediction of groundwater potential mapping based on Convolutional Neural Network (CNN) and Support Vector Regression (SVR). *Journal of Hydrology*, *588*, 125033. https://doi.org/10.1016/j.jhydrol.2020.125033

Park, N., Mohammadi, M., Gorde, K., Jajodia, S., Park, H., & Kim, Y. (2018). Data synthesis based on generative adversarial networks. *arXiv preprint arXiv:1806.03384*, *11*(10), 1071–1083. https://doi.org/10.14778/3231751.3231757

Pazola, A., Shamsudduha, M., French, J., MacDonald, A. M., Abiye, T., Goni, I. B., & Taylor, R. G. (2024). High-resolution long-term average groundwater recharge in Africa estimated using random forest regression and residual interpolation. *Hydrology and Earth System Sciences*, *28*(13), 2949–2967. https://doi.org/10.5194/hess-28-2949-2024

Pearson, A. J., Rucker, D. F., Tsai, C.-H., Fuchs, E. H., & Carroll, K. C. (2022). Electrical resistivity monitoring of lower Rio Grande River-Groundwater intermittency. *Journal of Hydrology*, *613*, 128325. https://doi.org/10.1016/j.jhydrol.2022.128325

Prommer, H., Barry, D. A., & Zheng, C. (2003). MODFLOW/MT3DMS-Based reactive multicomponent transport modeling. *Groundwater Series*, *41*(2), 247–257. https://doi.org/10.1111/j.1745-6584.2003.tb02588.x

Pruess, K., Oldenburg, C. M., & Moridis, G. (1999). TOUGH2 user's guide version 2.

Pulla, S. T., Yasarer, H., & Yarbrough, L. D. (2024). Synthetic time series data in groundwater analytics: Challenges, insights, and applications. *Water*, *16*(7), 949. https://doi.org/10.3390/w16070949

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). 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. https://doi.org/10.1016/j.jcp.2018.10.045

Rajabi, M. M., Ataie-Ashtiani, B., & Simmons, C. T. (2018). Model-data interaction in groundwater studies: Review of methods, applications and future directions. *Journal of Hydrology*, *567*, 457–477. https://doi.org/10.1016/j.jhydrol.2018.09.053

Rajaee, T., Ebrahimi, H., & Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology*, *572*, 336–351. https://doi.org/10.1016/j.jhydrol.2018.12.037

Ramil, A., López, A. J., Pozo-Antonio, J. S., & Rivas, T. (2018). A computer vision system for identification of granite-forming minerals based on RGB data and artificial neural networks. *Measurement*, *117*, 90–95. https://doi.org/10.1016/j.measurement.2017.12.006

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

Razavi, S. (2021). Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling. *Environmental Modelling and Software*, *144*, 105159. https://doi.org/10.1016/j.envsoft.2021.105159

Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Review of surrogate modeling in water resources. *Water Resources Research*, *48*(7). https://doi.org/10.1029/2011WR011527

Rees, E. M., Nightingale, E. S., Jafari, Y., Waterlow, N. R., Clifford, S., B. Pearson, C. A., et al. (2020). COVID-19 length of hospital stay: A systematic review and data synthesis. *BMC Medicine*, *18*(1), 270. https://doi.org/10.1186/s12916-020-01726-3

Ren, P., Xiao, Y., Chang, X., Huang, P.-y., Li, Z., Chen, X., & Wang, X. (2021). A comprehensive survey of neural architecture search: Challenges and solutions. *ACM Computing Surveys*, *54*(4), 1–34. Article 76. https://doi.org/10.1145/3447582

Ren, W., Ershadnia, R., Wallace, C. D., LaBolle, E. M., Dai, Z., de Barros, F. P. J., & Soltanian, M. R. (2022). Evaluating the effects of multiscale heterogeneous sediments on solute mixing and effective dispersion. *Water Resources Research*, *58*(9), e2021WR031886. https://doi.org/10.1029/2021WR031886

Retike, I., Bikše, J., Kalvāns, A., Dēliņa, A., Avotniece, Z., Zaadnoordijk, W. J., et al. (2022). Rescue of groundwater level time series: How to visually identify and treat errors. *Journal of Hydrology*, *605*, 127294. https://doi.org/10.1016/j.jhydrol.2021.127294

Rice, L., Wong, E., & Kolter, Z. (2020). *Overfitting in adversarially robust deep learning* proceedings of the 37th international conference on machine learning. *Proceedings of Machine Learning Research*. https://proceedings.mlr.press/v119/rice20a.html

Rojas, R., Feyen, L., & Dassargues, A. (2008). Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging. *Water Resources Research*, *44*(12). https://doi.org/10.1029/2008WR006908

Rubin, Y., & Journel, A. G. (1991). Simulation of non-Gaussian space random functions for modeling transport in groundwater. *Water Resources Research*, *27*(7), 1711–1721. https://doi.org/10.1029/91wr00838

Rucker, D. F., Tsai, C.-H., Carroll, K. C., Brooks, S., Pierce, E. M., Ulery, A., & Derolph, C. (2021). Bedrock architecture, soil texture, and hyporheic zone characterization combining electrical resistivity and induced polarization imaging. *Journal of Applied Geophysics*, *188*, 104306. https://doi.org/10.1016/j.jappgeo.2021.104306

Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, *263*, 110273. https://doi.org/10.1016/j.knosys.2023.110273

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery*, *8*(4), e1249. https://doi.org/10.1002/widm.1249

Sahimi, M., & Tahmasebi, P. (2022). The potential of quantum computing for geoscience. *Transport in Porous Media*, *145*(2), 367–387. https://doi.org/10.1007/s11242-022-01855-8

Sanchez-Vila, X., Guadagnini, A., & Carrera, J. (2006). Representative hydraulic conductivities in saturated groundwater flow. *Reviews of Geophysics*, *44*(3). https://doi.org/10.1029/2005RG000169

Sawadekar, K., Song, Y., Pan, M., Beck, H., McCrary, R., Ullrich, P., et al. (2025). Improving differentiable hydrologic modeling with interpretable forcing fusion. *Journal of Hydrology*, *659*, 133320. https://doi.org/10.1016/j.jhydrol.2025.133320

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Saxena, N., Day-Stirrat, R. J., Hows, A., & Hofmann, R. (2021). Application of deep learning for semantic segmentation of sandstone thin sections. *Computers and Geosciences*, *152*, 104778. https://doi.org/10.1016/j.cageo.2021.104778

Scanlon, B. R., Fakhreddine, S., Rateb, A., de Graaf, I., Famiglietti, J., Gleeson, T., et al. (2023). Global water resources and the role of groundwater in a resilient water future. *Nature Reviews Earth and Environment*, *4*(2), 87–101. https://doi.org/10.1038/s43017-022-00378-6

Scheibe, T. D., Murphy, E. M., Chen, X., Rice, A. K., Carroll, K. C., Palmer, B. J., et al. (2015). An analysis platform for multiscale hydrogeologic modeling with emphasis on hybrid multiscale methods. *Groundwater Series*, *53*(1), 38–56. https://doi.org/10.1111/gwat.12179

Schilling, O. S., Cook, P. G., & Brunner, P. (2019). Beyond classical observations in hydrogeology: The advantages of including exchange flux, temperature, tracer concentration, residence time, and soil moisture observations in groundwater model calibration. *Reviews of Geophysics*, *57*(1), 146–182. https://doi.org/10.1029/2018RG000619

Secci, D., Godoy, A., & Gómez-Hernández, J. J. (2024). Physics-Informed Neural Networks for solving transient unconfined groundwater flow. *Computers and Geosciences*, *182*, 105494. https://doi.org/10.1016/j.cageo.2023.105494

Shao, S., Yang, X., & Jia, C. (2022). Combining multi-source data to evaluate the leakage pollution and remediation effects of landfill. *Journal of Hydrology*, *610*, 127889. https://doi.org/10.1016/j.jhydrol.2022.127889

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558–8593. https://doi.org/10.1029/2018wr022643

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., et al. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Reviews Earth and Environment*, *4*(8), 552–567. https://doi.org/10.1038/s43017-023-00450-9

Sicilia, M.-A., García-Barriocanal, E., & Sánchez-Alonso, S. (2017). Community curation in open dataset repositories: Insights from Zenodo. *Procedia Computer Science*, *106*, 54–60. https://doi.org/10.1016/j.procs.2017.03.009

Šimůnek, J., Van Genuchten, M. T., & Šejna, M. (2016). Recent developments and applications of the HYDRUS computer software packages. *Vadose Zone Journal*, *15*(7), 1–25. https://doi.org/10.2136/vzj2016.04.0033

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. In *2016 3rd international conference on computing for Sustainable Global Development (INDIACom)*.

Singh, V., & Rao, S. M. (2005). Application of image processing and radial basis neural network techniques for ore sorting and ore classification. *Minerals Engineering*, *18*(15), 1412–1420. https://doi.org/10.1016/j.mineng.2005.03.003

Singha, S., Pasupuleti, S., Singha, S. S., & Kumar, S. (2020). Effectiveness of groundwater heavy metal pollution indices studies by deep-learning. *Journal of Contaminant Hydrology*, *235*, 103718. https://doi.org/10.1016/j.jconhyd.2020.103718

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, *82*(12), 2635–2670. https://doi.org/10.2166/wst.2020.369

Sixt, L., Wild, B., & Landgraf, T. (2018). RenderGAN: Generating realistic labeled data [methods]. *Frontiers in Robotics and AI*, *5*. https://doi.org/10.3389/frobt.2018.00066

Soltanian, M. R., Moeini, F., Dai, Z., Sawyer, A. H., Fleckenstein, J. H., Doherty, J., et al. (2024). Sustainability nexus AID: Groundwater. *Sustainability Nexus Forum*, *32*(1), 21. https://doi.org/10.1007/s00550-024-00557-7

Song, S., Zhang, D., Mukerji, T., & Wang, N. (2023). GANSim-surrogate: An integrated framework for stochastic conditional geomodelling. *Journal of Hydrology*, *620*, 129493. https://doi.org/10.1016/j.jhydrol.2023.129493

Song, X., Chen, X., Ye, M., Dai, Z., Hammond, G., & Zachara, J. M. (2019). Delineating facies spatial distribution by integrating ensemble data assimilationand indicator geostatistics with level-set transformation. *Water Resources Research*, *55*(4), 2652–2671. https://doi.org/10.1029/2018WR023262

Song, X., Ren, H., Hou, Z., Lin, X., Karanovic, M., Tonkin, M., et al. (2023). Predicting future well performance for environmental remediation design using deep learning. *Journal of Hydrology*, *617*, 129110. https://doi.org/10.1016/j.jhydrol.2023.129110

Song, Y., Bindas, T., Shen, C., Ji, H., Knoben, W. J. M., Lonzarich, L., et al. (2025). High-resolution national-scale water modeling is enhanced by multiscale differentiable physics-informed machine learning. *Water Resources Research*, *61*(4), e2024WR038928. https://doi.org/10.1029/2024WR038928

Sood, V., & Chauhan, R. P. (2024). Archives of quantum computing: Research progress and challenges. *Archives of Computational Methods in Engineering*, *31*(1), 73–91. https://doi.org/10.1007/s11831-023-09973-2

Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Training very deep networks. *Advances in Neural Information Processing Systems*, *28*.

Strebelle, S. (2021). Multiple-point statistics simulation models: Pretty pictures or decision-making tools? *Mathematical Geosciences*, *53*(2), 267–278. https://doi.org/10.1007/s11004-020-09908-8

Sun, A. Y., Jiang, P., Shuai, P., & Chen, X. (2024). Bridging hydrological ensemble simulation and learning using deep neural operators. *Water Resources Research*, *60*(10), e2024WR037555. https://doi.org/10.1029/2024wr037555

Sun, A. Y., & Scanlon, B. R. (2019). How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions. *Environmental Research Letters*, *14*(7), Article 073001. https://doi.org/10.1088/1748-9326/ab1b7d

Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., & Zhong, Z. (2019). Combining physically based modeling and deep learning for fusing GRACE satellite data: Can we learn from mismatch? *Water Resources Research*, *55*(2), 1179–1195. https://doi.org/10.1029/2018WR023333

Sun, Y., Wang, L., Gao, K., Ying, S., Lin, W., Humphreys, K. L., et al. (2023). Self-supervised learning with application for infant cerebellum segmentation and analysis. *Nature Communications*, *14*(1), 4717. https://doi.org/10.1038/s41467-023-40446-z

Tahmasebi, P. (2018). Multiple point statistics: A review. In B. S. Daya Sagar, Q. Cheng, & F. Agterberg (Eds.), *Handbook of mathematical geosciences: Fifty years of IAMG* (pp. 613–643). Springer International Publishing. https://doi.org/10.1007/978-3-319-78999-6_30

Tahmasebi, P., Kamrava, S., Bai, T., & Sahimi, M. (2020). Machine learning in geo- and environmental sciences: From small to large scale. *Advances in Water Resources*, *142*, 103619. Article 103619. https://doi.org/10.1016/j.advwatres.2020.103619

Tan, C., Lv, S., Dong, F., & Takei, M. (2019). Image reconstruction based on convolutional neural network for electrical resistance tomography. *IEEE Sensors Journal*, *19*(1), 196–204. https://doi.org/10.1109/JSEN.2018.2876411

Tang, K., Da Wang, Y., McClure, J., Chen, C., Mostaghimi, P., & Armstrong, R. T. (2022). Generalizable framework of unpaired domain transfer and deep learning for the processing of real-time synchrotron-based X-ray microcomputed tomography images of complex structures. *Physical Review Applied*, *17*(3), 034048. https://doi.org/10.1103/PhysRevApplied.17.034048

Tang, M., Liu, Y., & Durlofsky, L. J. (2021a). Deep-learning-based surrogate flow modeling and geological parameterization for data assimilation in 3D subsurface flow. *Computer Methods in Applied Mechanics and Engineering*, *376*, 113636. Article 113636. https://doi.org/10.1016/j.cma.2020.113636

Tang, M., Liu, Y., & Durlofsky, L. J. (2021b). Deep-learning-based surrogate flow modeling and geological parameterization for data assimilation in 3D subsurface flow. *Computer Methods in Applied Mechanics and Engineering*, *376*, 113636. https://doi.org/10.1016/j.cma.2020.113636

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tao, H., Hameed, M. M., Marhoon, H. A., Zounemat-Kermani, M., Heddam, S., Kim, S., et al. (2022). Groundwater level prediction using machine learning models: A comprehensive review. *Neurocomputing*, *489*, 271–308. https://doi.org/10.1016/j.neucom.2022.03.014

Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., & Barajas-Solano, D. (2020). Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems. *Water Resources Research*, *56*(5), e2019WR026731. https://doi.org/10.1029/2019WR026731

Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2020). The computational limits of deep learning. *arXiv preprint arXiv:2007.05558*.

Tian, W., Zhang, Z., Bouffard, D., Wu, H., Xin, K., Gu, X., & Liao, Z. (2024). Enhancing interpretability and generalizability of deep learning-based emulator in three-dimensional lake hydrodynamics using Koopman operator and transfer learning: Demonstrated on the example of lake Zurich. *Water Research*, *249*, 120996. https://doi.org/10.1016/j.watres.2023.120996

Tsai, C.-H., Rucker, D. F., Brooks, S. C., Ginn, T., & Carroll, K. C. (2022). Transient storage model parameter optimization using the simulated annealing method. *Water Resources Research*, *58*(7), e2022WR032018. https://doi.org/10.1029/2022WR032018

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., et al. (2021). From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. *Nature Communications*, *12*(1), 5988. https://doi.org/10.1038/s41467-021-26107-z

Tsang, W. K., & Benoit, D. F. (2023). Interpretability and explainability in machine learning. In Y. Ohsawa (Ed.), *Living beyond data: Toward sustainable value creation* (pp. 89–100). Springer International Publishing. https://doi.org/10.1007/978-3-031-11593-6_5

Tsuyuki, T., & Tamura, R. (2022). Nonlinear data assimilation by deep learning embedded in an ensemble Kalman filter. *Journal of the Meteorological Society of Japan*.

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. l. A., Elkhatib, Y., et al. (2019). Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE Access*, *7*, 65579–65615. https://doi.org/10.1109/ACCESS.2019.2916648

Valdez, F., & Melin, P. (2023). A review on quantum computing and deep learning algorithms and their applications. *Soft Computing*, *27*(18), 13217–13236. https://doi.org/10.1007/s00500-022-07037-4

Varfolomeev, I., Yakimchuk, I., & Safonov, I. (2019). An application of deep neural networks for segmentation of microtomographic images of rock samples. *Computers*, *8*(4), 72. https://doi.org/10.3390/computers8040072

Végh, J. (2023). Why does von Neumann obstruct deep learning? In *2023 IEEE 23rd international symposium on Computational Intelligence and Informatics (CINTI)*.

Verdonck, T., Baesens, B., Óskarsdóttir, M., & vanden Broucke, S. (2024). Special issue on feature engineering editorial. *Machine Learning*, *113*(7), 3917–3928. https://doi.org/10.1007/s10994-021-06042-2

Vereecken, H., Amelung, W., Bauke, S. L., Bogena, H., Brüggemann, N., Montzka, C., et al. (2022). Soil hydrology in the Earth system. *Nature Reviews Earth and Environment*, *3*(9), 573–587. https://doi.org/10.1038/s43017-022-00324-6

Vettoruzzo, A., Bouguelia, M. R., Vanschoren, J., Rognvaldsson, T., & Santosh, K. (2024). Advances and challenges in meta-learning: A technical review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(7), 1–20. https://doi.org/10.1109/TPAMI.2024.3357847

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling and Software*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Vu, M. T., Jardani, A., Massei, N., & Fournier, M. (2021). Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network. *Journal of Hydrology*, *597*, 125776. https://doi.org/10.1016/j.jhydrol.2020.125776

Wallace, C. D., Tonina, D., McGarr, J. T., de Barros, F. P. J., & Soltanian, M. R. (2021). Spatiotemporal dynamics of nitrous oxide emission hotspots in heterogeneous riparian sediments. *Water Resources Research*, *57*(12), e2021WR030496. https://doi.org/10.1029/2021WR030496

Wan, Z., Zhang, Y., & He, H. (2017). Variational autoencoder based synthetic data generation for imbalanced learning. In *2017 IEEE symposium series on computational intelligence (SSCI)*.

Wang, H., Dalton, L., Fan, M., Guo, R., McClure, J., Crandall, D., & Chen, C. (2022). Deep-learning-based workflow for boundary and small target segmentation in digital rock images using UNet++ and Ik-EBM. *Journal of Petroleum Science and Engineering*, *215*, 110596. https://doi.org/10.1016/j.petrol.2022.110596

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., et al. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47–60. https://doi.org/10.1038/s41586-023-06221-2

Wang, N., Chang, H., & Zhang, D. (2021). Deep-learning-based inverse modeling approaches: A subsurface flow example. *Journal of Geophysical Research: Solid Earth*, *126*(2), e2020JB020549. https://doi.org/10.1029/2020JB020549

Wang, N., Chang, H., & Zhang, D. (2023). Inverse modeling for subsurface flow based on deep learning surrogates and active learning strategies. *Water Resources Research*, *59*(7), e2022WR033644. https://doi.org/10.1029/2022WR033644

Wang, N., Kong, X.-Z., & Zhang, D. (2024). Physics-Informed Convolutional Decoder (PICD): A novel approach for direct inversion of heterogeneous subsurface flow. *Geophysical Research Letters*, *51*(13), e2024GL108163. https://doi.org/10.1029/2024GL108163

Wang, S., Zheng, H., Wen, X., & Fu, S. (2024). Distributed high-performance computing methods for accelerating deep learning training. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *3*(3), 108–126. https://doi.org/10.60087/jklst.v3.n3.p108-126

Wang, W., Shuai, Y., Yang, Q., Zhang, F., Zeng, M., & Li, M. (2024). A comprehensive computational benchmark for evaluating deep learning-based protein function prediction approaches. *Briefings in Bioinformatics*, *25*(2). https://doi.org/10.1093/bib/bbae050

Wang, Y., Yang, C., Lan, S., Zhu, L., & Zhang, Y. (2024). End-edge-cloud collaborative computing for deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *26*(4), 1–2683. https://doi.org/10.1109/COMST.2024.3393230

Wang, Y. D., Shabaninejad, M., Armstrong, R. T., & Mostaghimi, P. (2021). Deep neural networks for improving physical accuracy of 2D and 3D multi-mineral segmentation of rock micro-CT images. *Applied Soft Computing*, *104*, 107185. https://doi.org/10.1016/j.asoc.2021.107185

Wang, Z., Lu, W., & Chang, Z. (2024). Application of observed data denoising based on variational mode decomposition in groundwater pollution source recognition. *Science of the Total Environment*, *946*, 174374. https://doi.org/10.1016/j.scitotenv.2024.174374

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., et al. (2022). Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

Wu, Y., Liu, L., Bae, J., Chow, K. H., Iyengar, A., Pu, C., et al. (2019). Demystifying learning rate policies for high accuracy training of deep neural networks. In *2019 IEEE international conference on big data (big data)*.

Wunsch, A., Liesch, T., & Broda, S. (2021). Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), Convolutional Neural Networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). *Hydrology and Earth System Sciences*, *25*(3), 1671–1687. https://doi.org/10.5194/hess-25-1671-2021

Wunsch, A., Liesch, T., & Broda, S. (2022). Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. *Nature Communications*, *13*(1), 1221. https://doi.org/10.1038/s41467-022-28770-2

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Xiong, H., Li, J., Li, Z., & Zhang, Z. (2024). GPR-GAN: A ground-penetrating radar data generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–14. https://doi.org/10.1109/TGRS.2023.3337172

Xu, R., & Zhang, D. (2024). Forward prediction and surrogate modeling for subsurface hydrology: A review of theory-guided machine-learning approaches. *Computers and Geosciences*, *188*, 105611. https://doi.org/10.1016/j.cageo.2024.105611

Xu, T., & Liang, F. (2021). Machine learning for hydrologic sciences: An introductory overview. *Wiley Interdisciplinary Reviews-Water*, *8*(5). Article e1533. https://doi.org/10.1002/wat2.1533

Xu, T., Sonnenthal, E., Spycher, N., & Pruess, K. (2006). TOUGHREACT-A simulation program for non-isothermal multiphase reactive geochemical transport in variably saturated geologic media: Applications to geothermal injectivity and CO2 geological sequestration. *Computers and Geosciences*, *32*(2), 145–165. https://doi.org/10.1016/J.CAGEO.2005.06.014

Xu, T., Xiao, S., Reuschen, S., Wildt, N., Hendricks Franssen, H. J., & Nowak, W. (2024a). Towards a community-wide effort for benchmarking in subsurface hydrological inversion: Benchmarking cases, high-fidelity reference solutions, procedure and a first comparison. *Hydrology and Earth System Sciences Discussions*, *2024*, 1–64. https://doi.org/10.5194/hess-2024-60

Xu, T., Xiao, S., Reuschen, S., Wildt, N., Hendricks Franssen, H. J., & Nowak, W. (2024b). Towards a community-wide effort for benchmarking in subsurface hydrological inversion: Benchmarking cases, high-fidelity reference solutions, procedure, and first comparison. *Hydrology and Earth System Sciences*, *28*(24), 5375–5400. https://doi.org/10.5194/hess-28-5375-2024

Xu, X., Tan, M., Corcoran, B., Wu, J., Boes, A., Nguyen, T. G., et al. (2021). 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature*, *589*(7840), 44–51. https://doi.org/10.1038/s41586-020-03063-0

Xu, X.-Y., & Jin, X.-M. (2023). Integrated Photonic Computing beyond the von Neumann Architecture. *ACS Photonics*, *10*(4), 1027–1036. https://doi.org/10.1021/acsphotonics.2c01543

Xu, Y., Lu, W., Pan, Z., Wang, Z., Luo, C., & Bai, Y. (2024). Intelligent enhanced particle filter with deep residual network surrogate for accurate groundwater pollution source characterization. *Journal of Hydrology*, *642*, 131904. https://doi.org/10.1016/j.jhydrol.2024.131904

Yan, L., & Zhou, T. (2019). Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse problems. *Journal of Computational Physics*, *381*, 110–128. https://doi.org/10.1016/j.jcp.2018.12.025

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061

Yang, S., Lian, H., Xu, B., Thanh, H. V., Chen, W., Yin, H., & Dai, Z. (2023). Application of robust deep learning models to predict mine water inflow: Implication for groundwater environment management. *Science of the Total Environment*, *871*, 162056. https://doi.org/10.1016/j.scitotenv.2023.162056

Yang, X., Song, Z., King, I., & Xu, Z. (2023). A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, *35*(9), 8934–8954. https://doi.org/10.1109/TKDE.2022.3220219

Yeh, T.-C., Khaleel, R., & Carroll, K. C. (2015). *Flow through heterogeneous geologic media*. Cambridge University Press.

Yin, H., Zhang, G., Wu, Q., Yin, S., Soltanian, M. R., Thanh, H. V., & Dai, Z. (2023). A deep learning-based data-driven approach for predicting mining water inrush from coal seam floor using microseismic monitoring data. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–15. https://doi.org/10.1109/TGRS.2023.3300012

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, *1168*(2), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

Yoon, C. H., Torrance, R., & Scheinerman, N. (2022). Machine learning in medicine: Should the pursuit of enhanced interpretability be abandoned? *Journal of Medical Ethics*, *48*(9), 581–585. https://doi.org/10.1136/medethics-2020-107102

Yoshida, Y., & Miyato, T. (2017). Spectral norm regularization for improving the generalizability of deep learning. *arXiv preprint arXiv:1705.10941*.

Zehe, E., Loritz, R., Edery, Y., & Berkowitz, B. (2021). Preferential pathways for fluid and solutes in heterogeneous groundwater systems: Self-organization, entropy, work. *Hydrology and Earth System Sciences*, *25*(10), 5337–5353. https://doi.org/10.5194/hess-25-5337-2021

Zhan, C., Dai, Z., Jiao, J. J., Soltanian, M. R., Yin, H., & Carroll, K. C. (2025). Toward artificial general intelligence in hydrogeological modeling with an integrated latent diffusion framework. *Geophysical Research Letters*, *52*(3), e2024GL114298. https://doi.org/10.1029/2024GL114298

Zhan, C., Dai, Z., Samper, J., Yin, S., Ershadnia, R., Zhang, X., et al. (2022). An integrated inversion framework for heterogeneous aquifer structure identification with single-sample generative adversarial network. *Journal of Hydrology*, *610*, 127844. https://doi.org/10.1016/j.jhydrol.2022.127844

Zhan, C., Dai, Z., Soltanian, M. R., & de Barros, F. P. J. (2022). Data-worth analysis for heterogeneous subsurface structure identification with a stochastic deep learning framework. *Water Resources Research*, *58*(11), e2022WR033241. https://doi.org/10.1029/2022WR033241

Zhan, C., Dai, Z., Soltanian, M. R., & Zhang, X. (2022). Stage-wise stochastic deep learning inversion framework for subsurface sedimentary structure identification. *Geophysical Research Letters*, *49*(1), e2021GL095823. https://doi.org/10.1029/2021GL095823

Zhan, C., Dai, Z., Yang, Z., Zhang, X., Ma, Z., Thanh, H. V., & Soltanian, M. R. (2023). Subsurface sedimentary structure identification using deep learning: A review. *Earth-Science Reviews*, *239*, 104370. https://doi.org/10.1016/j.earscirev.2023.104370

Zhan, C., Dai, Z., Yin, S., Carroll, K. C., & Soltanian, M. R. (2024). Conceptualizing future groundwater models through a ternary framework of multisource data, human expertise, and machine intelligence. *Water Research*, *257*, 121679. https://doi.org/10.1016/j.watres.2024.121679

Zhan, Y., Guo, Z., Yan, B., Chen, K., Chang, Z., Babovic, V., & Zheng, C. (2024). Physics-informed identification of PDEs with LASSO regression, examples of groundwater-related equations. *Journal of Hydrology*, *638*, 131504. https://doi.org/10.1016/j.jhydrol.2024.131504

Zhang, J., Zheng, Q., Wu, L., & Zeng, L. (2020). Using deep learning to improve ensemble smoother: Applications to subsurface characterization. *Water Resources Research*, *56*(12), e2020WR027399. https://doi.org/10.1029/2020WR027399

Zhang, Q., Li, P., Ren, X., Ning, J., Li, J., Liu, C., et al. (2023). A new real-time groundwater level forecasting strategy: Coupling hybrid data-driven models with remote sensing data. *Journal of Hydrology*, *625*, 129962. https://doi.org/10.1016/j.jhydrol.2023.129962

Zhang, X., Dong, F., Chen, G., & Dai, Z. (2023). Advance prediction of coastal groundwater levels with temporal convolutional and long short-term memory networks. *Hydrology and Earth System Sciences*, *27*(1), 83–96. https://doi.org/10.5194/hess-27-83-2023

Zhang, X., Jiang, S., Wei, J., Wu, C., Xia, X., Wang, X., et al. (2024). Non-Gaussian hydraulic conductivity and potential contaminant source identification: A comparison of two advanced DLPM-based inversion framework. *Journal of Hydrology*, *638*, 131540. https://doi.org/10.1016/j.jhydrol.2024.131540

Zheng, A., & Casari, A. (2018). *Feature engineering for machine learning: Principles and techniques for data scientists*. O'Reilly Media, Inc.

Zheng, C., & Wang, P. P. (1999). MT3DMS: A modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems; documentation and user's guide.

Zheng, N., Jiang, S., Xia, X., Kong, W., Li, Z., Gu, S., & Wu, Z. (2023). Efficient estimation of groundwater contaminant source and hydraulic conductivity by an ILUES framework combining GAN and CNN. *Journal of Hydrology*, *621*, 129677. https://doi.org/10.1016/j.jhydrol.2023.129677

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Zhi, W., Appling, A. P., Golden, H. E., Podgorski, J., & Li, L. (2024). Deep learning for water quality. *Nature Water*, *2*(3), 228–241. https://doi.org/10.1038/s44221-024-00202-z

Zhou, F., Yang, S., Fujita, H., Chen, D., & Wen, C. (2020). Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowledge-Based Systems*, *187*, 104837. https://doi.org/10.1016/j.knosys.2019.07.008

Zhu, S., Yu, T., Xu, T., Chen, H., Dustdar, S., Gigan, S., et al. (2023). Intelligent computing: The latest advances, challenges, and future. *Intelligent Computing*, *2*, 0006. https://doi.org/10.34133/icomputing.0006

Zoph, B., & Le, Q. V. (2016). Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*. https://doi.org/10.48550/arXiv.1611.01578

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, *598*, 126266. https://doi.org/10.1016/j.jhydrol.2021.126266

29935210, 2025, 2, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JH000703 by New Mexico State University, Wiley Online Library on [06/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License