AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2024WR037786

Special Collection:
Advancing Interpretable AI/ML Methods for Deeper Insights and Mechanistic Understanding in Earth Sciences: Beyond Predictive Capabilities

Deep Learning Identification of the Governing Equation for Water Flow in Heterogeneous Soils From Data

Wenxiang Song[1], Liangsheng Shi[1], Leilei He[1], Yuanyuan Zha[1], Xiaolong Hu[1], Mehdi Rahmati[2], and Harry Vereecken[2]

[1]State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan, China, [2]Institute of Bio- and Geosciences: Agrosphere (IBG-3), Forschungszentrum Jülich, Jülich, Germany

Key Points:

- A new coarse-grained group sparse regression theory is proposed to identify the governing equation for water flow in heterogeneous soils
- Group sparse regression is incorporated into a deep learning framework to improve its capabilities and robustness
- Soil moisture dynamics in heterogeneous soils and the governing equation are simultaneously identified

Correspondence to:
L. Shi,
liangshs@whu.edu.cn

Citation:
Song, W., Shi, L., He, L., Zha, Y., Hu, X., Rahmati, M., & Vereecken, H. (2025). Deep learning identification of the governing equation for water flow in heterogeneous soils from data. *Water Resources Research*, *61*, e2024WR037786. https://doi.org/10.1029/2024WR037786

Received 19 APR 2024
Accepted 29 JAN 2025

Abstract Despite the remarkable advances in using deep learning for describing and predicting soil water flow, these models inherently cannot deepen our understanding of its underlying physical mechanisms as they are black-box approaches. To address this issue, a novel data-driven equation discovery approach has recently been widely used to facilitate scientific discovery in geoscience disciplines, including soil hydrology. However, due to the inherent complexity of soils, current data-driven discovery approaches cannot deal with heterogeneous soil scenarios. In this study, we present a new group sparse regression theory and a deep learning framework to extend previous studies to be able to identify the governing equations for soil water flow in heterogeneous soils from observational data. Specifically, we focus on discovering equations from only time series of volumetric soil water content data, which are easily accessible. To accommodate it, the underlying assumption of the generalized soil-water content-based governing equation is utilized, and a coarse-grained group sparsity theory is developed. Furthermore, we incorporate the proposed group sparse regression into a new deep-learning framework: Extended-DeepGS (Extended Deep-learning-based Group Sparsity). Through deep-learning identification, it realizes simultaneous reconstructions of soil moisture dynamics and governing equations. A series of comprehensive numerical experiments are designed and conducted to test the performance of the theory and framework, and the results show its robustness. We also summarize the potential effects of soil heterogeneity on the discovery of equations. Finally, we discuss the limitations of the approach, which may inform future developments.

Plain Language Summary Soil water flow is critical to understanding processes like plant growth and water availability. However, it is complex and difficult to model accurately, especially in soils that are not uniform. While conventional deep learning models have made great progress in predicting how water moves through the soils, they are generally black-box and cannot deepen our understanding of soil water flow mechanisms. To address this issue, a novel data-driven equation discovery approach has recently been widely used to facilitate scientific discovery in geoscience. Given the remarkable successes, the use of data-driven discovery to advance the physical modeling of soil water flow processes is rational and promising. However, existing methods have problems when soils are heterogeneous (non-uniform). We have developed a new method that combines advanced sparse regression theory and deep learning to tackle these more complicated soil conditions. Our approach focuses on using readily available data, specifically measurements of soil moisture over time, to determine the equations that govern water flow in different soil types. We designed and conducted comprehensive numerical experiments to test the methods and demonstrate their capabilities. We are taking a solid step forward in discovering the soil water flow equation from data in more complex scenarios.

1. Introduction

Modeling and understanding soil water flow processes are fundamental to hydrological, agricultural, and ecological practice (Vereecken et al., 2015, 2022). Based on observations of phenomena, physical derivations, and a simplified parameterization, governing equations for soil moisture flow processes have been established, including the law of mass conservation, the Darcy-Buckingham (DB) law (Buckingham, 1907; Darcy, 1856), and their combination Richardson-Richards equation (RRE) (Richards, 1931; Richardson, 1922). However, current established governing laws cannot accurately accommodate all scientific and engineering scenarios. For instance, DB and RRE may lose their physical foundation when describing large-scale soil water flow (Or et al., 2015) and they cannot accurately describe complex soil water flow in soils with nonequilibrium conditions and preferential flow pathways (Beven & Germann, 2013). This is because the mechanisms of soil water flow are far too complex

to establish governing equations that can handle all possible conditions and scenarios based on physical principles and empiricism (Narasimhan, 2007).

Recently, the tremendous increase in the availability of soil water content data from various sources (Dorigo et al., 2021) has enabled a significant improvement in the modeling soil water flow processes through the use of techniques such as data assimilation, machine learning, and deep learning (Li et al., 2020; Y. Wang et al., 2024). However, despite these remarkable advances in soil water content prediction, these black-box models are associated with higher model complexity, computational cost, and lack of transparency (Rudin, 2019) and they cannot deepen our understanding of the underlying mechanisms in soil water flow.

The advancement of knowledge and understanding of soil water flow processes is predicated on the formulation of fundamental physical relationships derived from substantial datasets of observed phenomena, experimental and observational data. Automating this process of converting data into knowledge is of great significance. Fortunately, the recent data-driven equation discovery has provided us with a promising way to automatically discover the underlying governing laws from data (Camps-Valls et al., 2023; H. Wang et al., 2023). This novel scientific artificial intelligence approach transforms learned patterns from the data into interpretable and concise physics equations, both in terms of mathematical structure and parameters, which lead us to understand the cause-effect relationship among physical variables (Song et al., 2024). The prospective utility in geosciences encompasses refining existing equations and potentially revealing previously missing governing equations, which has been successfully applied in geophysics (Zeng et al., 2023), ocean modeling (Ross et al., 2023; Zanna & Bolton, 2020), and climate science (Grundner et al., 2024; Xu et al., 2023).

The use of data-driven equation discovery to advance the physical modeling of soil water flow is meaningful and promising given its current success and potential. For instance, it may improve the parameterization of soil hydrological processes at multiple scales. However, inherent complexities such as the multi-phase nature of soils, water-content dependencies, hysteresis, nonequilibrium, and high nonlinearities in water flow in soils that are typically strongly heterogenous in nature, pose challenges unaddressed by other disciplines, resulting in a deficit of viable solutions. This could become even more complicated if the memory effects of the soil are considered, which are already accounted for only loosely in the predictive models (Rahmati et al., 2023, 2024). To overcome these challenges, adaptive methodologies designed for the complexity of soil water flow processes are essential to extract real physical insights from data. Pioneering studies have proposed solutions. For instance, Ghorbani et al. (2021) employed sparse regression techniques to recover the linearized soil moisture flow equation from data. In a subsequent study by Song et al. (2022), the dependence on water content and nonlinearity of soil water flow were considered, using group sparsity techniques to identify nonlinear soil water flow equations from data. In addition, a deep-learning framework proposed by Song et al. (2023) aims to resolve the effects of data sparsity and measurement noise. These studies demonstrate a significant improvement in our ability to distill physical laws for soil moisture flow from data.

Despite recent advancements, current data-driven discovery approaches treat soils as ideal homogeneous systems. However, real-world soils exhibit inherent heterogeneity in their soil properties (Elkateb et al., 2003), a characteristic not captured by existing methods. This heterogeneity leads to spatiotemporal variations in states, parameters, and fluxes impacting soil water flow. A significant challenge in discovering the governing equation for water flow in heterogeneous soils using data-driven techniques lies in disentangling the coupled effects of water content dependency and soil hydraulic properties on observed soil water dynamics. Separating these effects from the data itself remains a significant hurdle. This limitation poses a substantial challenge to the practical application of data-driven equation discovery in the field of soil hydrology.

Here, we advance our previous studies (Song et al., 2022, 2023) to be capable of deriving the governing equations for unsaturated flow in heterogeneous soils from data. Specifically, we consider soil heterogeneity to vary continuously, which is related to weathering, deposition conditions and stress history (Elkateb et al., 2003). The novelty of this study resides in the fact that we focus on discovering equations from only volumetric water content time-series data instead of matric potential data, as the former is much easier to measure with low costs. To achieve this objective, we developed a novel coarse-grained group sparse regression to identify the underlying governing equation structure from data. To cooperate with the proposed sparse regression, we also consider assuming the underlying generalized water content-based governing equation form for describing soil moisture flow in continuously heterogeneous soils (Zha, Yang, et al., 2013). Furthermore, we also incorporate the proposed group sparse regression into a deep-learning framework, extending our previous framework (Song et al., 2023)

into Extended-DeepGS (Extended Deep-learning-based Group Sparsity) to deal with heterogeneous soil scenarios. The resulting framework is capable of accurately reconstructing soil water dynamics and identifying complex coefficient fields simultaneously despite data sparsity and noise. A series of numerical experiments were designed and conducted to test the performance of the theory and framework. Throughout this study, we aimed to put a solid step forward to discover the unsaturated flow governing equation from data in more complex scenarios.

In the remainder of the manuscript, Section 2 presents a brief background of soil moisture flow equations in heterogeneous soils and data-driven equation discovery. Section 3 presents our newly proposed approaches. Section 4 presents and discusses the results in detail. Finally, future perspectives and conclusions are drawn in Section 5.

2. Background

2.1. Data-Driven Equation Discovery

Efforts to discover governing equations from data have historical roots dating back to the era of Kepler. This approach primarily relied on trial and error driven by scientists. In recent years, significant advancements in computational power, scientific artificial intelligence, and data availability have sparked considerable interest and efforts in the automated data-driven discovery of physical laws and governing equations. It is important to note that data-driven discovery differs from traditional inverse modeling. The latter aims to estimate the parameters from data (Nakamura & Potthast, 2015), where the equation structure is usually at least partially given, such as data assimilation (Chang & Zhang, 2019), Gaussian processes regression (Maziar Raissi et al., 2017), and physics-informed neural networks (M. Raissi et al., 2019). Data-driven discovery aims to simultaneously discover underlying equation structure and parameters (i.e., the coefficients of the governing equations). Several approaches can achieve this objective, such as genetic programming and deep reinforcement learning. We ask interested readers to refer to Camps-Valls et al., 2023; Song et al., 2024 for a comprehensive review.

Among various methodologies, sparse regression stands out as particularly well-suited for high-dimensional partial differential equation (PDE) systems due to its low computational cost and minimal hyperparameters (Kutz & Brunton, 2022). Sparse regression is a regression method when sparse feature selection is required (Bertsimas et al., 2020). The term sparse denotes that only a small subset of features is considered essential, while the remaining features are disregarded and effectively set to zero. The core assumption of sparse regression for discovering governing equations from data is that most physical systems are governed by only a few terms, generally represented by derivative terms (Brunton et al., 2016). The sparsity-promoting methods, such as LASSO (Tibshirani, 1996), are designed to solve equation discovery tasks without a brute-force search over all possible combinations. For example, if the goal is to discover the Burgers equations (i.e., $u_t = -uu_x + u_{xx}$), one can assume that the equation is $u_t = \xi_1 u + \xi_2 u_x + \xi_3 u u_x + \xi_4 u_{xx} + \xi_5 u u_{xx} + \xi_6 \sin(u) u_{xx} + \cdots$, where ξ_1 to ξ_∞ are coefficients. Sparse regression ensures that every correct nonzero ξ_i for $i = 1, 2, \cdots$ can be detected. In this study, we developed a novel coarse-grained group sparse regression approach specially for our soil moisture flow equation discovery task.

2.2. Soil Moisture Flow Equation in Heterogeneous Soils

Soil moisture flow in heterogeneous soils is usually described by the RRE with the soil water pressure head ψ as the primary variable or a hybrid form of RRE (Farthing & Ogden, 2017; Zha et al., 2019). However, one can also describe it using water content-based RRE incorporating additional correction terms (Zha, Yang, et al., 2013). Here, we assume that the soil hydraulic properties vary continuously in space. In this case, the mixed-form governing equation for describing unsaturated soil moisture movement is:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(\theta, \boldsymbol{P_1}) \frac{\partial \psi(\theta, \boldsymbol{P_2})}{\partial z}\right] - \frac{\partial}{\partial z}[K(\theta, \boldsymbol{P_1})] \tag{1}$$

where θ is the volumetric water content $\left[\mathrm{L}^3\ \mathrm{L}^{-3}\right]$, ψ is the soil water pressure head [L], t is the time [T], z is the vertical coordinate taken positively downward [L], $z = 0$ is at the top of the profile, K denotes the unsaturated hydraulic conductivity $\left[\mathrm{L}\ \mathrm{T}^{-1}\right]$, $\boldsymbol{P_1}$ denotes the parameters of the unsaturated hydraulic conductivity and $\boldsymbol{P_2}$ represents the parameters of the soil water retention curves (WRC). For simplicity, we assume that these

parameters change continuously in space, which implies the existence of derivatives. However, for further simplification, we assume that these parameters do not change in time, which allows us to expand the derivatives in the above equation using the chain rules. Applying the chain rule to the first term in the right-hand side of the above equation therefore leads to:

$$\frac{\partial\theta}{\partial t}=\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial z}\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial z}+K(\theta,\boldsymbol{P_1})\frac{\partial}{\partial z}\left(\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial z}\right)-\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial z} \tag{2}$$

which can be reorganized as (the detailed steps can be found in Appendix A)

$$\begin{aligned}\frac{\partial\theta}{\partial t}=&\left[\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\theta}\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\boldsymbol{P_2}}\frac{d\boldsymbol{P_2}}{dz}+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\boldsymbol{P_1}}\frac{d\boldsymbol{P_1}}{dz}+2K(\theta,\boldsymbol{P_1})\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\theta\,\partial\boldsymbol{P_2}}\frac{d\boldsymbol{P_2}}{dz}-\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\theta}\right]\frac{\partial\theta}{\partial z}\\&+\left[K(\theta,\boldsymbol{P_1})\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\theta^2}+\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\theta}\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\right]\left(\frac{\partial\theta}{\partial z}\right)^2+\left[K(\theta,\boldsymbol{P_1})\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\right]\frac{\partial^2\theta}{\partial z^2}\\&+\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\boldsymbol{P_1}}\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\boldsymbol{P_2}}\frac{d\boldsymbol{P_1}}{dz}\frac{d\boldsymbol{P_2}}{dz}+K(\theta,\boldsymbol{P_1})\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\boldsymbol{P_2}^2}\left(\frac{d\boldsymbol{P_2}}{dz}\right)^2+K(\theta,\boldsymbol{P_1})\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\boldsymbol{P_2}}\frac{d^2\boldsymbol{P_2}}{dz^2}\\&-\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\boldsymbol{P_1}}\frac{d\boldsymbol{P_1}}{dz}\end{aligned} \tag{3}$$

As K and ψ can be described as functions of θ, $\boldsymbol{P_1}$, and $\boldsymbol{P_2}$, moreover, $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ are functions of spatial coordinates, Equation 3 can be rewritten to a generalized water content-based form:

$$\frac{\partial\theta}{\partial t}=f_0(\theta,z)+f_1(\theta,z)\frac{\partial\theta}{\partial z}+f_2(\theta,z)\frac{\partial^2\theta}{\partial z^2}+f_3(\theta,z)\left(\frac{\partial\theta}{\partial z}\right)^2 \tag{4}$$

where

$$\begin{aligned}f_0(\theta,z)=&\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\boldsymbol{P_1}}\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\boldsymbol{P_2}}\frac{d\boldsymbol{P_1}}{dz}\frac{d\boldsymbol{P_2}}{dz}+K(\theta,\boldsymbol{P_1})\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\boldsymbol{P_2}^2}\left(\frac{d\boldsymbol{P_2}}{dz}\right)^2+K(\theta,\boldsymbol{P_1})\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\boldsymbol{P_2}}\frac{d^2\boldsymbol{P_2}}{dz^2}\\&-\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\boldsymbol{P_1}}\frac{d\boldsymbol{P_1}}{dz}\end{aligned} \tag{5}$$

$$f_1(\theta,z)=\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\theta}\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\boldsymbol{P_2}}\frac{d\boldsymbol{P_2}}{dz}+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\boldsymbol{P_1}}\frac{d\boldsymbol{P_1}}{dz}+2K(\theta,\boldsymbol{P_1})\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\theta\partial\boldsymbol{P_2}}\frac{d\boldsymbol{P_2}}{dz}-\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\theta} \tag{6}$$

$$f_2(\theta,z)=K(\theta,\boldsymbol{P_1})\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta} \tag{7}$$

$$f_3(\theta,z)=K(\theta,\boldsymbol{P_1})\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\theta^2}+\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\theta}\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta} \tag{8}$$

This generalized water content-based equation form for water movement in heterogeneous soils is helpful as a reference equation for our data-driven discovery problem, especially when we can only access volumetric water content time-series data.

3. Methods

3.1. Problem Statement

In this study, we aim to identify the governing equation describing soil water flow in unsaturated heterogeneous soils from only volumetric water content time-series data measured at fixed depths. We consider one-dimensional vertical water movement in this study. Let us suppose water content time-series data are measured by a sensors at different fixed depths with b timesteps. We can obtain a dataset as

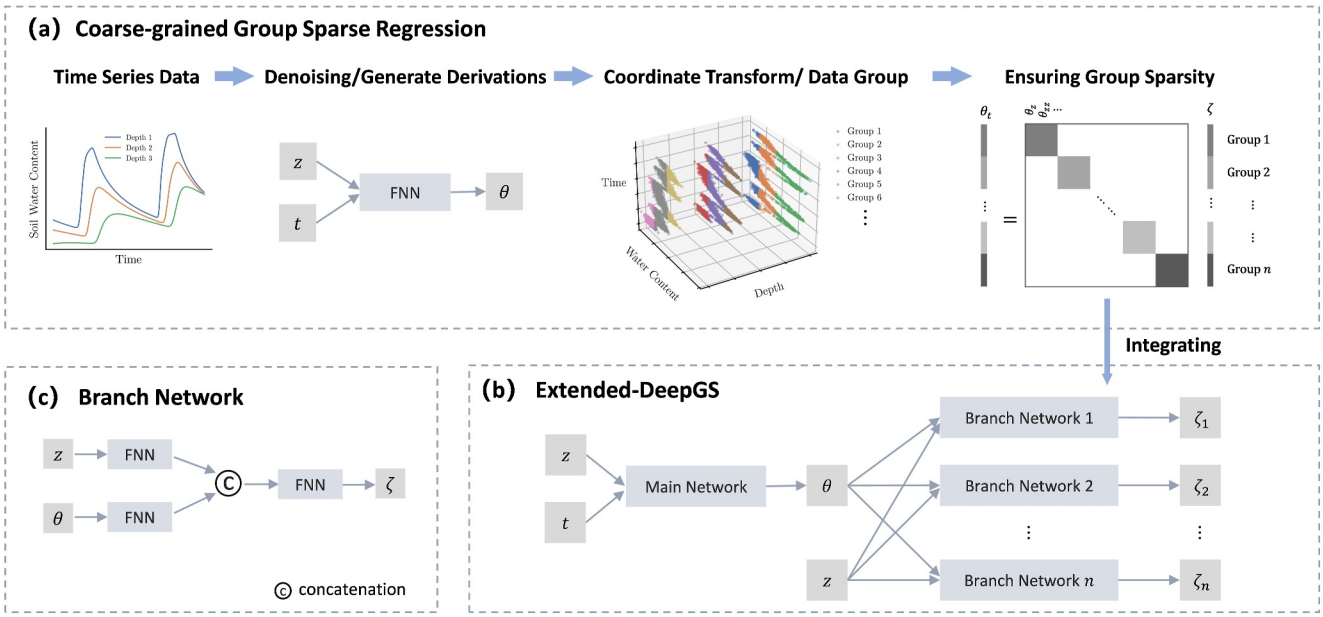

Figure 1. Schematic flowchart of the discovery of the governing equation for soil water flow in heterogeneous soils. (a) Represents the workflow of the proposed coarse-gained group sparse regression. Time series data is used to fit the fully connected neural networks (FNN) for generating denoised high-resolution soil moisture dynamics and derivations, and then the data is split into different groups according to transformed coordinates (theta, depth). Then, grouped data are used as input into a block diagonal structure matrix and solved using a group sparse regression optimization algorithm. (b) Depicts the structure of the proposed Extended-DeepGS. Input to the Extended-DeepGS are the coordinates (z,t) into the neural network, and its outputs are the corresponding θ values to represent the soil moisture dynamics and, together with additional information about depth z, they are connected by the branch networks to represent nonlinear coefficients ξ_i for $i = 1\cdots n$, where n is the number of branch networks determined by the number of non-zero terms in the identified equation structure by group sparse regression. (c) Represents the detailed structure of branch networks. The state variable θ and depth z into the different FNN and then concatenate to map nonlinear coefficient ξ.

$$\begin{bmatrix} \hat{\theta}(z_1,t_1) & \hat{\theta}(z_1,t_2) & \hat{\theta}(z_1,t_3) & \dots & \hat{\theta}(z_1,t_b) \\ \hat{\theta}(z_2,t_1) & \hat{\theta}(z_2,t_2) & \hat{\theta}(z_2,t_3) & \dots & \hat{\theta}(z_2,t_b) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\theta}(z_a,t_1) & \hat{\theta}(z_a,t_2) & \hat{\theta}(z_a,t_3) & \dots & \hat{\theta}(z_a,t_b) \end{bmatrix} \tag{9}$$

where $\hat{\theta}(z_i,t_j)$ denotes the soil volumetric water content data measured by the i^{th} sensor at j^{th} timestep for $i = 1,2,\cdots,a$ and $j = 1,2,\cdots,b$. We aim to discover the soil water flow equation from only this dataset as accurately as possible, including equation structure and parameters (i.e., coefficients). The structure of the governing equation, constitutive relationships, and boundary conditions are not given as prior.

3.2. Coarse-Grained Group Sparse Regression

In this subsection, we introduce our newly proposed coarse-grained group sparse regression for discovering the governing equation from the collected dataset (i.e., Equation 4). The schematic workflow is shown in Figure 1a. We first consider establishing an extensive overdetermined candidate library to include the form of an underlying governing equation based on the sparse regression theory. It is an essential step for avoiding brute-force searching within an undetermined physics space, which is a non-deterministic polynomial-time hard (NP-hard) problem (Virgolin & Pissis, 2022).

We leverage the fact that the potential forms of PDEs generally follow specific patterns. Due to the nonlinearity of unsaturated flow, we need to consider the dependence between state variables and coefficients, that is, coefficients are functions of volumetric water content θ. Nevertheless, in heterogeneous soils, soil hydraulic properties vary in space. Therefore, we assume that the influence of soil hydraulic properties can be directly mapped from spatial coordinates to the final governing equation coefficients, along with the state variable, and then

govern the nonlinear behavior of the system. Under the limited prior information, the equation is postulated to be of the following form:

$$\theta_t = \boldsymbol{N}\left(1, \theta, \theta_z, \theta_{zz}, \dots\right) \cdot \boldsymbol{\xi}(\theta, z) = \sum_{i=1}^{d} N_i \xi_i(\theta, z) \tag{10}$$

where θ_t is the first derivative of θ with respect to time; $\boldsymbol{N}(\cdot)$ is vector and represents the candidate library, which consists of the derivatives of θ to space coordinate z (e.g., θ_z and θ_{zz} are the first and second-order derivatives of θ with respect to z) and their nonlinear combinations; ξ_i is the nonlinear coefficient; N_i is the i^{th} candidate term, and d is the scale of $\boldsymbol{N}(\cdot)$, that is, the number of candidate terms.

To calculate candidates accurately (e.g., θ_z), a fully connected deep neural network (FNN) is used to approximate soil water content measurements $\hat{\theta}(z_i, t_j)$ for $i = 1, 2, \cdots, a$ and $j = 1, 2, \cdots, b$, that is, $\tilde{\theta}\left(z, t; \boldsymbol{\Lambda}^{(0)}\right) = \mathcal{N}_0\left(z, t; \boldsymbol{\Lambda}^{(0)}\right) \rightarrow \hat{\theta}(z, t)$, where $\tilde{\theta}$ represents soil water content derived from the neural network, $\mathcal{N}_0$ represents the FNN (i.e., the main network), and $\boldsymbol{\Lambda}^{(0)} = \left\{\mathrm{W}^{(0)}, \boldsymbol{b}^{(0)}\right\}$ denotes its trainable parameters. It takes the domain coordinates $\{z, t\}$ as input and is followed by multiple hidden layers. This step not only infers high-resolution denoised data but also gets accurate candidates from only sparse, noisy data. After training, candidate terms and θ_t can be easily derived from the trained network by automatic differentiation. In the following, the used soil water content is all derived from trained neural networks, and we use θ to represent $\tilde{\theta}$ for shortness. Then, Equation 10 can be written as

$$\begin{bmatrix} \theta_t(z_1,t_1) \\ \theta_t(z_1,t_2) \\ \vdots \\ \theta_t(z_1,t_b) \\ \theta_t(z_2,t_1) \\ \theta_t(z_2,t_2) \\ \vdots \\ \theta_t(z_a,t_b) \end{bmatrix}_{(a\times b)\times 1} = \left(\begin{bmatrix} 1 & \theta_z(z_1,t_1) & \theta_{zz}(z_1,t_1) & \dots & \dots \\ 1 & \theta_z(z_1,t_2) & \theta_{zz}(z_1,t_2) & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \theta_z(z_1,t_b) & \theta_{zz}(z_1,t_b) & \dots & \dots \\ 1 & \theta_z(z_2,t_1) & \theta_{zz}(z_2,t_1) & \dots & \dots \\ 1 & \theta_z(z_2,t_2) & \theta_{zz}(z_2,t_2) & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots & \dots \\ 1 & \theta_z(z_a,t_b) & \theta_{zz}(z_a,t_b) & \dots & \dots \end{bmatrix}_{(a\times b)\times n} \odot \boldsymbol{\Xi} \right) \mathbf{1}_n \tag{11}$$

where a and b represent the number of input spatial and time coordinates for deriving data from the trained main network, respectively, which are named as grided meta-data points $\mathcal{D}_{meta}$ here, notation $\odot$ denotes the Hadamard product, $\mathbf{1}_n$ represents a column vector of ones with length n, and

$$\boldsymbol{\Xi} = \begin{bmatrix} \xi_1(\theta(z_1,t_1),z_1) & \xi_2(\theta(z_1,t_1),z_1) & \xi_3(\theta(z_1,t_1),z_1) & \dots & \dots \\ \xi_1(\theta(z_1,t_2),z_1) & \xi_2(\theta(z_1,t_2),z_1) & \xi_3(\theta(z_1,t_2),z_1) & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \xi_1(\theta(z_1,t_b),z_1) & \xi_2(\theta(z_1,t_b),z_1) & \xi_3(\theta(z_1,t_b),z_1) & \dots & \dots \\ \xi_1(\theta(z_2,t_1),z_2) & \xi_2(\theta(z_2,t_1),z_2) & \xi_3(\theta(z_2,t_1),z_2) & \dots & \dots \\ \xi_1(\theta(z_2,t_2),z_2) & \xi_2(\theta(z_2,t_2),z_2) & \xi_3(\theta(z_2,t_2),z_2) & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots & \dots \\ \xi_1(\theta(z_a,t_b),z_a) & \xi_2(\theta(z_a,t_b),z_a) & \xi_3(\theta(z_a,t_b),z_a) & \dots & \dots \end{bmatrix}_{(a\times b)\times n} \tag{12}$$

The inherent spatiotemporal variability of the coefficient $\boldsymbol{\Xi}$, characterized by its state variable dependence and spatial dependence, renders it analytically intractable to solve. To address this challenge and to accurately determine the coefficient matrix, a coarse-grained linear approximation is employed. This approach leverages the fact that the coefficients $\xi_i(\theta, z)$ is a function of both volumetric water content and spatial coordinates. Consequently, the coefficients are expected to remain relatively constant for observations with similar volumetric water content values and measured at the same spatial location. Therefore, the data grouping is based on transforming the perspective of spatiotemporal coordinates (z, t) to (z, θ), and then coarse-grained linear regression can be applied. To do this, we define $\mathcal{D}_{(i,j)}$ to be the data point at i^{th} depth at timestep j in the dataset, which includes state variable $\theta(z, t)$ and its derivatives (i.e., candidates) derived by the main network:

$$\mathcal{D}_{(i,j)} = \left\{\theta(z_i, t_j); \theta_t(z_i, t_j); 1; \theta_z(z_i, t_j); \theta_{zz}(z_i, t_j); \cdots\right\} \tag{13}$$

where (z_i, t_j) is the coordinate of $\mathcal{D}_{(i,j)}$. The data points are divided into different sets according to the grouping rules. For data points of different depths, that is, $z = z_i$, the set $\boldsymbol{S}_{(i,k)}$ for $i = 1, 2, 3, \cdots, a$ and $k = 1, 2, 3, \cdots, c$ are defined as

$$\boldsymbol{S}_{(i,k)} = \begin{cases} \left\{\mathcal{D}_{(i,j)} \middle| \theta_{min} \le \theta_{(i,j)} < \theta_{min} + k * win\right\}, & k = 1 \\ \left\{\mathcal{D}_{(i,j)} \middle| \ \theta_{min} + (k-1) * win \le \theta_{(i,j)} < \theta_{min} + k * win\right\}, & 1 < k < c \\ \left\{\mathcal{D}_{(i,j)} \middle| \theta_{min} + (k-1) * win \le \theta_{(i,j)} \le \theta_{max}\right\}, & k = c \end{cases} \tag{14}$$

where $\theta_{(i,j)}$ denotes the corresponding volumetric water content value of the data point $\mathcal{D}_{(i,j)}$, θ_{min}, and θ_{max} denote the lowest and largest volumetric water content value in collected data points at the depth $z = z_i$, win denotes the range of variation of the volumetric water content in the set $\boldsymbol{S}_{(i,j)}$, $c = \sqcup((\theta_{max} - \theta_{min})/win)$, $\sqcup$ is ceiling operation, and c equals the number of sets for each depth.

Now, we can get the regression for a single set $\boldsymbol{S}_{(i,k)}$:

$$\theta_t^{(i,k)} = \boldsymbol{\Theta}(\theta)^{(i,k)} \boldsymbol{\xi}\left(\overline{\theta}^{(i,k)}, z_i\right) \tag{15}$$

where $\overline{\theta}^{(i,k)} = 1/n \sum_1^n \theta$ represents the arithmetic mean of the volumetric water content of the data points in $\boldsymbol{S}_{(i,k)}$. Considering that the variation of θ is small in a single set, we can treat approximate equality in Equation 15 as exact equality. Finally, regressions for all sets are integrated into a single linear system with the following block diagonal structure:

$$\begin{bmatrix} \theta_t^{(1,1)} \\ \vdots \\ \theta_t^{(i,k)} \\ \vdots \\ \theta_t^{(a,c)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Theta}(\theta)^{(1,1)} & & & & \\ & \ddots & & & \\ & & \boldsymbol{\Theta}(\theta)^{(i,k)} & & \\ & & & \ddots & \\ & & & & \boldsymbol{\Theta}(\theta)^{(a,c)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}\left(\overline{\theta}^{(1,1)}, z_1\right) \\ \vdots \\ \boldsymbol{\xi}\left(\overline{\theta}^{(i,k)}, z_i\right) \\ \vdots \\ \boldsymbol{\xi}\left(\overline{\theta}^{(a,c)}, z_a\right) \end{bmatrix} \tag{16}$$

Now, the coarse-grained group sparse regression optimization can be written as

$$\hat{\xi} = argmin_{\xi} \left\| \boldsymbol{\theta_t} - \sum_{g \in \mathcal{G}} \boldsymbol{\Theta}(\theta)_{(g)} \boldsymbol{\xi}_{(g)} \right\|_2^2 + \lambda \sum_{g \varepsilon \mathcal{G}} \left\| \xi_{(g)} \right\|_0 \tag{17}$$

where $\mathcal{G}$ is a collection of groups, each of which contains a subset of the indices enumerating the columns of $\boldsymbol{\Theta}(\boldsymbol{\theta})$ and coefficients in $\boldsymbol{\xi}$. It is an NP-hard optimization problem, and it requires mathematically efficient algorithms to approximate ℓ_0 regularization. We adopt a sequential grouped threshold ridge regression (SGTR) (Rudy et al., 2019) approach to solve our regression optimization, which is mathematically proven to be local minimizers of an unconstrained ℓ_0-penalized least-squares problem. The narrative description of SGTR can be found in **Algorithm** 1.

Leveraging coarse-grained group sparse regression, we can derive the correct governing equation structure along with coarse-grained coefficients, which are discretely estimated based on spatial and water content-based partitioning. In the following section, we will explore how integrating deep learning can lead to more accurate fine-grained coefficients, demonstrating robustness against data sparsity and noise.

Algorithm 1. Coarse-Grained Group Sparse Regression

Input: Measurements $\mathcal{D}_m$, grided meta-data points $\mathcal{D}_{meta}$, range of variation of the volumetric water content in a set win, threshold sequences number n_{tol}, the maximum iterations of training n and the maximum iterations of sparse regression n_{sr}

Initialize the trainable parameters for the main network $\boldsymbol{\Lambda}^{(0)} = \{\mathbb{W}^{(0)}, \boldsymbol{b}^{(0)}\}$

for $iter = 1, 2, \cdots, n$ **do**

$$\left\{\hat{\boldsymbol{\Lambda}}^{(0)}\right\} = \arg\min_{\{\boldsymbol{\Lambda}^{(0)}\}}\left\{\mathcal{L}_d\left(\boldsymbol{\Lambda}^{(0)}; \mathcal{D}_m\right)\right\}.$$

Predict the soil water content θ and candidates over grided meta-data points $\mathcal{D}_{meta}$, and ensemble them:

$$\mathcal{D}_{(i,\,j)} = \{\theta(z_i, t_j); \theta_t(z_i, t_j); 1; \theta_z(z_i, t_j); \theta_{zz}(z_i, t_j); \cdots\}$$

Divide data points into different sets according to the grouping rules. Calculate θ_{min} and θ_{max} as the lowest and largest volumetric water content value in $\mathcal{D}_{pre}$. For data points of different depths, that is, $z = z_i$, the set $\boldsymbol{S}_{(i,\,k)}$ for $i = 1, 2, 3, \cdots, a$ and $k = 1, 2, 3, \cdots, c$ are defined as

$$\boldsymbol{S}_{(i,\,1)} = \left\{\mathcal{D}_{(i,\,j)} \middle| \theta_{min} \le \theta_{(i,\,j)} < \theta_{min} + 1 * win\right\}$$

$$\boldsymbol{S}_{(i,\,k)} = \left\{\mathcal{D}_{(i,\,j)} \middle| \ \theta_{min} + (k-1) * win \le \theta_{(i,\,j)} < \theta_{min} + k * win\right\}$$

$$\boldsymbol{S}_{(i,\,c)} = \left\{\mathcal{D}_{(i,\,j)} \middle| \theta_{min} + (c-1) * win \le \theta_{(i,\,j)} \le \theta_{max}\right\}$$

Construct the regression equation in each set, for example, $\boldsymbol{S}_{(i,\,k)}$:

$$\theta_t^{(i,\,k)} = \boldsymbol{\Theta}(\theta)^{(i,\,k)}\boldsymbol{\xi}\left(\bar{\theta}^{(i,\,k)}, z_i\right).$$

Integrate all sets into a single linear system with a block diagonal structure

$$\begin{bmatrix} \theta_t^{(1,1)} \\ \vdots \\ \theta_t^{(i,k)} \\ \vdots \\ \theta_t^{(a,c)} \end{bmatrix} = \begin{bmatrix} \Theta(\theta)^{(1,1)} & & & & \\ & \ddots & & & \\ & & \Theta(\theta)^{(i,k)} & & \\ & & & \ddots & \\ & & & & \Theta(\theta)^{(a,c)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}\left(\bar{\theta}^{(1,1)}, z_1\right) \\ \vdots \\ \boldsymbol{\xi}\left(\bar{\theta}^{(i,k)}, z_i\right) \\ \vdots \\ \boldsymbol{\xi}\left(\bar{\theta}^{(a,c)}, z_a\right) \end{bmatrix}.$$

For k sets and d candidate terms, define groups as

$$\mathcal{G} = \{\{j + d \cdot i : i = 1, \ldots, k\} : j = 1, \ldots, d\}.$$

Initialize coefficients with ridge regression:

$$\hat{\xi} = argmin_{\xi\bar{\theta}} \|\boldsymbol{\theta_t} - \boldsymbol{\Theta}(\boldsymbol{\theta})\boldsymbol{\xi}\|_2^2 + \lambda\|\boldsymbol{\xi}\|_2^2.$$

Determine the upper and lower bound of specified thresholds

$$\varepsilon : \varepsilon_{max} = \max\left(\|\boldsymbol{\xi}_{(g)}\|_2\right)$$

$$\varepsilon_{min} = \min\left(\|\boldsymbol{\xi}_{(g)}\|_2\right)$$

for $l = 1, 2, \cdots, n_{tol}$ **do**

    Determine the specified threshold ε:

$$\varepsilon = iter*(\varepsilon_{max} - \varepsilon_{min})/n_{tol} + \varepsilon_{min}.$$

for $iter = 1, 2, \cdots, n_{sr}$ **do**

Remove groups with smaller 2-norms than the specified threshold ε:

$$\mathcal{G} = \{g \in \mathcal{G} : \|\boldsymbol{\xi}\|_2 > \varepsilon\}.$$

Refit the remaining groups:

$$\hat{\xi} = \text{argmin}_{\xi\bar{\theta}} \|\boldsymbol{\theta}_t - \sum_{g \in \mathcal{G}} \boldsymbol{\Theta}(\boldsymbol{\theta})_{(g)} \boldsymbol{\xi}_{(g)}\|_2^2 + \lambda \|\boldsymbol{\xi}\|_2^2.$$

Obtain the linearized approximating coefficients with unbiased regression:

$$\hat{\boldsymbol{\xi}}_{(g)} = \text{argmin}_{\xi\bar{\theta}} \|\boldsymbol{\theta}_t - \sum_{g \in \mathcal{G}} \boldsymbol{\Theta}(\boldsymbol{\theta})_{(g)} \boldsymbol{\xi}_{(g)}\|_2^2.$$

Calculate the AIC-inspired loss, p denotes the number of nonzero terms, ε is a penalty term:

$$\mathcal{L}_l = N \ln\left(\|\boldsymbol{\Theta}(\boldsymbol{\theta})\hat{\boldsymbol{\xi}}(\bar{\theta}) - \boldsymbol{\theta}_t\|_2^2/N + \varepsilon\right) + 2p$$

if $l = 1$ **or** $\mathcal{L}_l < \mathcal{L}_{l-1}$:

$$\boldsymbol{\xi}_{best} = \hat{\boldsymbol{\xi}}$$

end if

end for

end for

Output: The best coefficients $\boldsymbol{\xi}_{best}$

3.3. Extended-DeepGS: Deep-Learning Framework Integration

Here, we introduce the Extended-DeepGS (Extended Deep-learning-based Group Sparsity) framework for integrating deep learning with the newly proposed coarse-grained group sparse regression. The schematic architecture of Extended-DeepGS is shown in Figure 1b. It is based on the DeepGS framework (Song et al., 2023) and extended to deal with heterogeneous soil scenarios, predict soil moisture dynamics, and accurately reconstruct nonlinear coefficients.

The FNN utilized for approximating soil moisture dynamics is referred to as the main network here, that is, $\hat{\theta}(z,t) = \mathcal{N}_0(z,t;\boldsymbol{\Lambda}^{(0)})$, where $\boldsymbol{\Lambda}^{(0)} = \{\mathrm{W}^{(0)}, \boldsymbol{b}^{(0)}\}$ denotes the trainable parameters in FNN. It takes spatiotemporal coordinates $\{z,t\}$ as inputs, followed by multiple hidden layers, except for the output layer, which employs linear activation; the other layers utilize the sine function as the activation function to consider the potential periodicity of soil moisture dynamics. Essentially, the main network acts as a nonlinear function to approximate the potential solutions of water content. In addition, it is connected to the governing equations by automatic differentiation, as the output θ is differentiable with respect to spatiotemporal coordinates, and the required information of candidates can be easily computed.

To approximate the nonlinear coefficients $\boldsymbol{\xi}$ with dependence on the state variable θ and spatial coordinate z, several branch networks are employed. The schematic architecture of branch networks is depicted in Figure 1c. Two simple FNNs take the state variable θ and spatial coordinate z as inputs and then concatenate them to one FNN to play the role of nonlinear mapping between θ, z, and $\xi_i(\theta,z)$. It can be expressed as $\widetilde{\xi}_i(\theta,z) = \mathcal{N}_i(\theta,z;\boldsymbol{\Lambda}^{(i)})$. The main network and branch networks collectively furnish the essential information required for reconstructing the governing equations.

The Extended-DeepGS framework is optimized by the total loss function $\mathcal{L}$, consisting of the data loss $\mathcal{L}_d$, the governing equation residual loss $\mathcal{L}_p$, and an extra regularization term:

$$\mathcal{L} = \omega_d \mathcal{L}_d(\boldsymbol{\Lambda}^{(0)}; \mathcal{D}_m) + \omega_p \mathcal{L}_p(\boldsymbol{\Lambda}^{(0)}, \boldsymbol{\Lambda}^{(i)}; \mathcal{D}_r) + \gamma\|\boldsymbol{\xi}\|_0 \tag{18}$$

$$\mathcal{L}_d(\boldsymbol{\Lambda}^{(0)}; \mathcal{D}_m) = \frac{1}{N_d}\left\|\tilde{\theta}^{(0)} - \hat{\theta}\right\|_2^2 \tag{19}$$

$$\mathcal{L}_p\left(\mathbf{\Lambda}^{(0)}, \mathbf{\Lambda}^{(i)}; \mathcal{D}_r\right) = \frac{1}{N_r}\left\| \theta_t^{(0)} - \sum_{i=1}^{d} N_i^{(0)} \xi_i(\theta, z)^{(i)} \right\|_2^2 \tag{20}$$

where ω_d, ω_p, and γ are the relative weightings of loss functions, $\mathcal{D}_m$ represents measured data points, $\mathcal{D}_r$ represents residual points, the superscript (i) denotes value output by i^{th} branch networks, $\hat{\theta}$ denotes the measured volumetric water content, N_m and N_r are the number of measurements and residual points, respectively. The data loss $\mathcal{L}_d$ is evaluated on measurements and, therefore, ensures the approximation of soil moisture dynamics. The residuals of governing equations $\mathcal{L}_p$ are evaluated on a large number of residual points $\mathcal{D}_r$, which is randomly selected within spatiotemporal domains and requires no real measurements. $\|\boldsymbol{\xi}\|_0$ ensures the sparsity of underlying governing equations. Due to the mathematical optimization challenge of $\|\boldsymbol{\xi}\|_0$, we decompose the optimization into easy-to-handle sub-problems. First, we adopt the coarse-grained group sparse regression as a proxy for ℓ_0 regularization to realize sparsity. When the corresponding candidate N_i is discovered by this sparse regression step, the $\xi_i(\theta, z)$ is nonzero. Then, the framework is optimized by minimizing data loss $\mathcal{L}_d$ and residuals of governing equations $\mathcal{L}_p$. The minimum of the total loss function $\mathcal{L}$ assures that accurate soil moisture dynamics and the underlying governing equations can be identified simultaneously. All the training is realized by the Adam algorithm (Kingma & Ba, 2015). The training strategies of Extended-DeepGS are summarized in Algorithm 2.

Algorithm 2. Training Strategies of Extended-DeepGS

Input: Measurements $\mathcal{D}_m$, grided meta-data points $\mathcal{D}_{pre}$, residual points $\mathcal{D}_{pre}$, the relative weighting of loss functions ω_d, ω_p, the maximum iterations of neural network training n_{train}

Based on the measurements $\mathcal{D}_m$ and meta-data points $\mathcal{D}_{meta}$, run coarse-grained group sparse regression as shown in **Algorithm 1** to determine the equation structure $\hat{\boldsymbol{\xi}}$

Initialize the trainable parameters for the main network $\mathbf{\Lambda}^{(0)} = \{\mathbb{W}^{(0)}, \boldsymbol{b}^{(0)}\}$ and branch networks $\mathbf{\Lambda}^{(i)} = \{\mathbb{W}^{(i)}, \boldsymbol{b}^{(i)}\}$ for $i = 1, 2, \cdots$

for $iter = 1, 2, \cdots, n_{train}$ **do**

$$\{\hat{\mathbf{\Lambda}}^{(0)}, \hat{\mathbf{\Lambda}}^{(i)}\} = \arg\min_{\{\mathbf{\Lambda}^{(0)}, \mathbf{\Lambda}^{(i)}\}} \{\omega_d \mathcal{L}_d(\mathbf{\Lambda}^{(0)}; \mathcal{D}_m) + \omega_p \mathcal{L}_p(\mathbf{\Lambda}^{(0)}, \mathbf{\Lambda}^{(i)}; \mathcal{D}_m, \mathcal{D}_r)\}.$$

end

Output: The best solution $\mathbf{\Lambda}_{best}^{(0)} = \hat{\mathbf{\Lambda}}^{(0)}$, $\widehat{\xi_{best}}(\theta, z) = \mathbf{\Lambda}_{best}^{(i)} = \hat{\mathbf{\Lambda}}^{(i)}$

3.4. Performance Validations

To evaluate the performance of the proposed methods, the identified equation from the data is compared with the true reference governing equation. Here, the moisture-form RRE (i.e., Equation 4) is assumed to be the true referenced governing equation. In addition, the Mualem-van Genuchten model (Van Genuchten, 1980) is used to parameterize the constitutive relationship within the reference governing equation:

$$\theta(\psi) = \theta_r + (\theta_s - \theta_r)/((1 + (-\alpha\psi)^n)^m) \tag{21}$$

$$K(\theta(\psi)) = K_s S_e^l \left(1 - \left(1 - S_e^{1/m}\right)^m\right)^2 \tag{22}$$

where θ_s is the saturated water content $\left[\mathrm{L}^3\,\mathrm{L}^{-3}\right]$, θ_r is the residual water content $\left[\mathrm{L}^3\,\mathrm{L}^{-3}\right]$, $S_e = (\theta - \theta_r)/(\theta_s - \theta_r)$ is the effective saturation, K_s is the saturated hydraulic conductivity $\left[\mathrm{L\,T}^{-1}\right]$, and $\alpha\left[\mathrm{L}^{-1}\right]$ and $n\,[-]$ are shape parameters. The shape parameter l is assumed to be 0.5, and $m = 1 - 1/n$. When substituting the Mualem-van Genuchten model (Equations 21 and 22) into RRE (Equation 4), we get the reference true equation in this study. Due to the lengthy expressions, the expanded form is given in Appendix A. It should be noted that selecting a specific constitutive relationship model is solely used to compare the discovered governing equation with the reference actual equation. The proposed data-driven discovery method does not necessitate

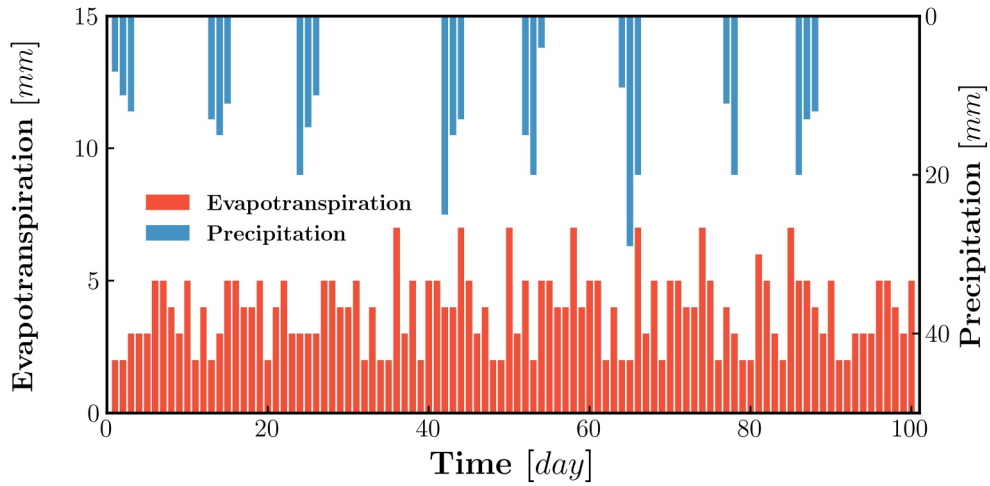

Figure 2. Time series of daily averaged precipitation and evapotranspiration for generating synthetic data sets.

prior knowledge of such constitutive relationship models. Here, fewer differences between the discovered equation and the reference governing equation indicate better performance. The relative error is defined as

$$\varepsilon = \frac{1}{d}\sum_{i=1}^{d}\left(\left(\xi_i - \hat{\xi}_i\right)^2 / \left\|\xi_i\right\|_2\right) \tag{23}$$

where d denotes the number of candidate terms; here, we choose the first-order spatial derivatives to the third-order spatial derivatives of volumetric water content θ and their second-order combinations as the candidate terms. In addition, only when the correct PDE structure (i.e., derivative terms) is identified will the error be evaluated.

3.5. Numerical Experiments

In this study, we design and conduct numerical experiments to comprehensively evaluate the proposed coarse-grained group sparse regression and Extended-DeepGS framework. Frist, as for generating synthetic datasets, the one-dimensional vertical soil moisture flow is simulated. We consider five soil hydraulic parameters that vary continuously in space $\{K_s, \alpha, n, \theta_s, \theta_r\}$ to simulate the natural heterogeneous soils. To ensure positive values of soil hydraulic parameters, we assume that the parameters following the logarithmic normal distribution, and the different variances were also employed to test the effect of the strength of heterogeneous, and the integral scale is set as 0.25 m. The soils are set as a one-layer 1.0 m bare soil column, which is discretized into 100 nodes at 0.01 m intervals. In each node, the hydraulic parameters are generated by random fields using the turning bands method (Mantoglou & Wilson, 1982). The visualizations of generated parameters are given in Appendix C. The temporal step is set as 0.01 days, and the simulated time is 100 days. The synthetic datasets are generated by solving RRE with the generalized Ross method (Zha, Shi, et al., 2013). The upper boundary is set as the atmospheric boundary, and the bottom is set as a free drainage boundary. The soil column is initialized by running a steady-state equation with 2 mm flux to get a non-uniform soil moisture profile to simulate the natural situations. The settings of the upper boundary, including the daily averaged time series of precipitation and evapotranspiration, are given in Figure 2.

Following simulation, data are exclusively sampled at predetermined locations to emulate soil moisture measurements conducted by in situ sensors. Moreover, noise is subsequently incorporated into the sampled data. Leveraging the synthetic datasets, a series of numerical experiments is devised and executed as follows. First, a case is set to illustrate the characteristics of the Extended-DeepGS. In this case, the soil water content measurements were sampled in heterogeneous loam soils by fixed-depth sensors at different locations $z_m \in \{-0.05\ \text{m},\ -0.15\ \text{m},\ -0.25\ \text{m}\}$. The sampling temporal resolution is 0.01 day. The 10% Gaussian noise is added to the sampled data as

$$\theta^*(z,t) = \theta(z,t) + \sigma * std(\theta) * \mathcal{N}(0,1) \tag{24}$$

Table 1
The Mualem-van Genuchten Soil Hydraulic Parameters

	$K_s[\mathrm{m\ day^{-1}}]$		$\alpha[\mathrm{m^{-1}}]$		$n[-]$		$\theta_s[\mathrm{m^3\ m^{-3}}]$		$\theta_r[\mathrm{m^3\ m^{-3}}]$	
Texture	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Sandy loam	1.061	0.4	7.5	1.5	1.89	0.1	0.41	0.02	0.065	0.01
Loam	0.2496	0.2	3.6	1.0	1.56	0.1	0.43	0.02	0.078	0.01
Silt Loam	0.108	0.1	2.0	0.5	1.41	0.1	0.45	0.02	0.067	0.01
Loam (Homogenous)	0.2496	0	3.6	0	1.56	0	0.43	0	0.078	0
Loam (Higher Heterogeneous)	0.2496	0.3	3.6	1.5	1.56	0.2	0.43	0.03	0.078	0.02

where $\theta^*(z,t)$ is noisy measured volumetric water content data, $\sigma = 0.1$, representing the Gaussian noise level. The numerical experiments are repeatedly conducted with five different random seeds to reduce the influence of random initializations of neural networks.

Second, a case is set to test the effects of soil heterogeneities. We consider three scenarios of different soils: homogeneous loam soils, relatively low heterogeneous loam soils, and relatively high heterogeneous loam soils. The soil hydraulic parameters are given in Table 1. The measurements of spatial and temporal sampling strategies are the same as in the first case.

Third, a case is set to test the effects of soil textures. Three typical soil textures are considered to test their effects, including sandy loam, loam, and silt loam soil, and their parameters are given in Table 1. The measurements of spatial and temporal sampling strategies are the same as in the first case.

Finally, a case is set to test measurement uncertainties, including measurement noise, data missing, and spatial sampling intervals. The measurements were sampled by fixed-depth sensors at different locations: $z_m^{(1)} \in \{-0.05\,\mathrm{m},\ -0.1\,\mathrm{m},\ -0.15\,\mathrm{m}\}$, $z_m^{(2)} \in \{-0.05\,\mathrm{m},\ -0.15\,\mathrm{m},\ -0.25\,\mathrm{m}\}$, $z_m^{(3)} \in \{-0.05\,\mathrm{m},\ -0.1\,\mathrm{m},\ -0.2\,\mathrm{m}\}$, and $z_m^{(4)} \in \{-0.05\,\mathrm{m},\ -0.2\,\mathrm{m},\ -0.35\,\mathrm{m}\}$ that is, 0.05 m, 0.1 m, and 0.15 m uniform sampling intervals and 0.05 m, 0.1 m non-uniform sampling intervals. The sampling temporal resolution is still set as 0.01 *day*, and the Gaussian noise of different levels is added to the sampled data. In addition, the effects of random $\{20\%, 40\%, 60\%\}$ data missing during the simulation period is also considered. For a fair comparison, we consider the error of reconstructing the coefficients and moisture dynamics at spatial locations from 0.05 to 0.15 m. All the cases and the codes are implemented by Python using the Pytorch library (Paszke et al., 2019), and the detailed hyper-parameters setting can be found in Appendix B.

4. Results and Discussions

4.1. Identifying Governing Equations and Reconstructing Soil Moisture Dynamics

We first show the case study that illustrates the characteristics of the Extended-DeepGS framework when identifying governing equations for the soil moisture flow in heterogeneous soils from volumetric water content data. Table 2 summarizes the process of identifying governing equations and reconstructing soil moisture dynamics. It shows that the correct PDE structure is identified during the group sparse regression step. At this stage,

Table 2
Identified Governing Equations and Reconstructed Soil Moisture Dynamics

		Relative error	
Stage	Identified PDE	Coefficients	Soil moisture dynamics
Group Sparse Regression Step	$\frac{\partial\theta}{\partial t} = \xi_0(\theta,z) + \xi_1(\theta,z)\frac{\partial\theta}{\partial z} + \xi_2(\theta,z)\frac{\partial^2\theta}{\partial z^2} + \xi_4(\theta,z)\left(\frac{\partial\theta}{\partial z}\right)^2$	2.97×10^{0}	1.52×10^{-3}
Deep-learning Reconstruction Step	$\frac{\partial\theta}{\partial t} = \xi_0(\theta,z) + \xi_1(\theta,z)\frac{\partial\theta}{\partial z} + \xi_2(\theta,z)\frac{\partial^2\theta}{\partial z^2} + \xi_4(\theta,z)\left(\frac{\partial\theta}{\partial z}\right)^2$	1.24×10^{-1}	1.90×10^{-5}
True PDE	$\frac{\partial\theta}{\partial t} = f_0(\theta,z) + f_1(\theta,z)\frac{\partial\theta}{\partial z} + f_2(\theta,z)\frac{\partial^2\theta}{\partial z^2} + f_3(\theta,z)\left(\frac{\partial\theta}{\partial z}\right)^2$		

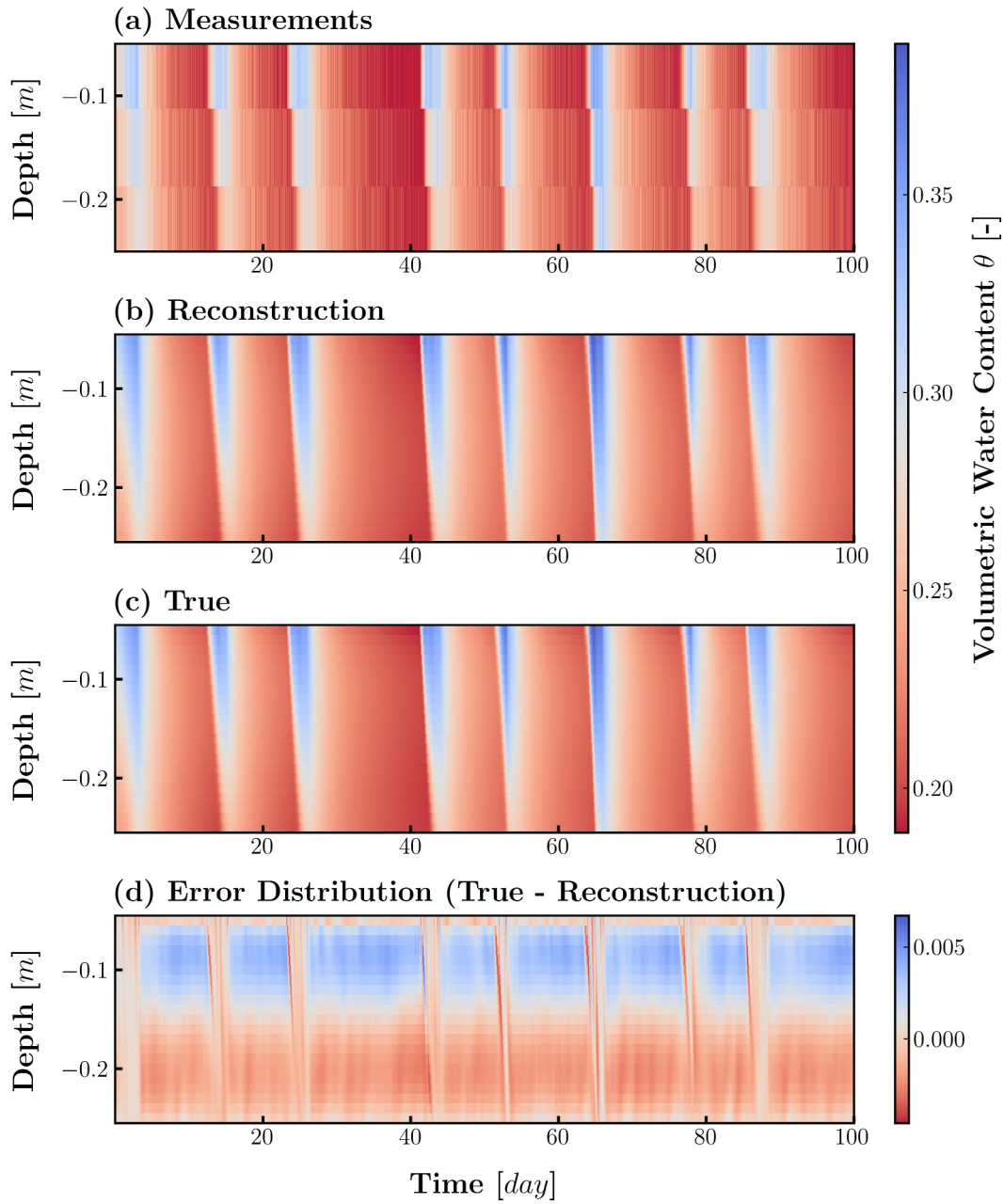

Figure 3. (a) Visualization of sparse and noisy soil volumetric water content time series measurements at locations in loamy soils. (b) and (c) represents the reconstructed and true soil moisture spatiotemporal dynamics, respectively. (d) The error spatiotemporal distribution of the reconstructed soil moisture dynamics, which equals the true value minus the reconstructed value.

the deep neural networks initially fit the soil moisture dynamics through purely data-driven processes. Group sparse regression can identify a governing equation, including equation structure and coefficient field, based on the coarse-grained constructed information of candidates. After a deep-learning reconstruction step, the accuracy of coefficient fields significantly improves, and the relative error of reconstructed soil moisture dynamics also reduces.

For visual inspection, Figure 3 shows the reconstructed and true spatiotemporal soil moisture dynamics along with the error distribution. It can be found that the reconstructed soil moisture dynamics are generally consistent with the true values, with larger errors mainly occurring at the sharp fronts where the soil hydraulic conductivity

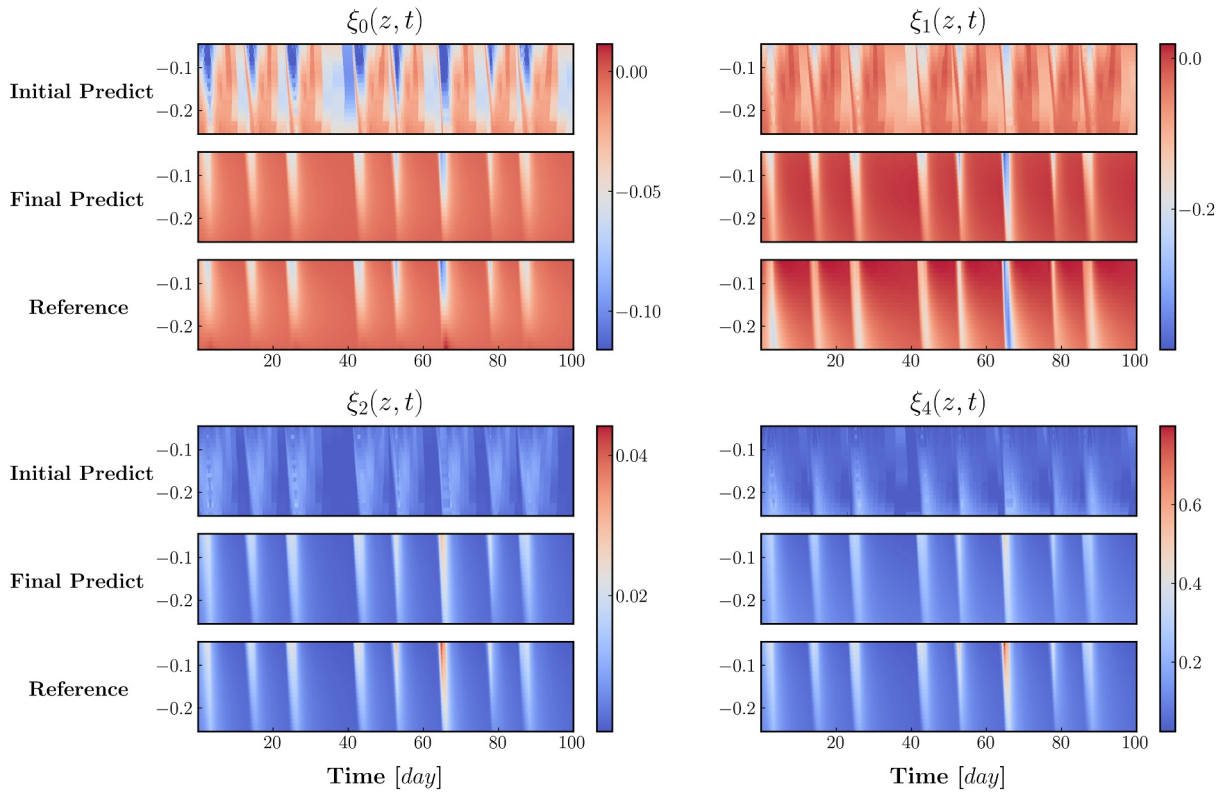

Figure 4. Comparison of identified spatiotemporal coefficients and reference spatiotemporal coefficients for true governing equations. $\xi_0(z,t)$, $\xi_1(z,t)$, $\xi_2(z,t)$, and $\xi_4(z,t)$ are the coefficient of $1, \theta_z, \theta_{zz}$ and ${\theta_z}^2$, respectively. The initial prediction represents the estimated coarse-grained coefficients at the group sparse regression step, which are discretely estimated based on spatial and water content-based partitioning. The final prediction represents the fine-grained coefficients, which are identified by the deep-learning reconstruction step.

changes dramatically in space. Additionally, Figure 4 depicts the identified spatiotemporal coefficient field and reference values. In the initial prediction step, group sparse regression gives an initial guess of coefficient fields, where errors are obviously present. This is due to the fact that the final estimated coefficient field $\xi_i(\theta,z)$ is a coarse-grained average estimate, which is strongly influenced by data error and data representativeness. Subsequently, after the deep-learning reconstruction step, the accuracy of reconstructed coefficients significantly improves, with slight differences in the near-saturated region. It can also be seen that the error distribution of the identified coefficient field is relevant to the error distribution of soil moisture dynamics. This is attributed to the fact that the estimation of the coefficients depends on the reconstructed soil moisture dynamics. This approach captures the growing trend of the coefficients in near-saturation but underestimates it. As a result, error propagation leads to the loss of nonlinear information in the coefficient maps. Additionally, the presence of nonuniform hydraulic parameters can worsen the challenge of accurately estimating. However, in the majority of volumetric water content ranges, the impact of these factors is minimal, and the coefficients can be accurately determined. Therefore, it seems that the observation of soil water content provides sufficient information about the governing equation to identify it effectively.

Then, the relationship between the coefficients ξ_i for $i = 0,1,2,4$ and θ can be derived from the reconstructed coefficients at each depth. Since these values vary spatially, the average of the coefficients along with their variations across space are presented in Figure 5. It is observed that identified soil hydraulic functions are close to the reference curve overall, with slight differences at the near-saturated region. In addition, the captured variations in coefficients induced by spatial changes of soil hydraulic parameters are shown in the identified coefficients ξ_0 and ξ_1. From Equations 4–8, we can see that the reference coefficients ξ_0 and ξ_1 is related to spatial changes in soil hydraulic parameters, while ξ_2 and ξ_4 not. Therefore, the results are reasonable, and the methods can estimate soil hydraulic functions.

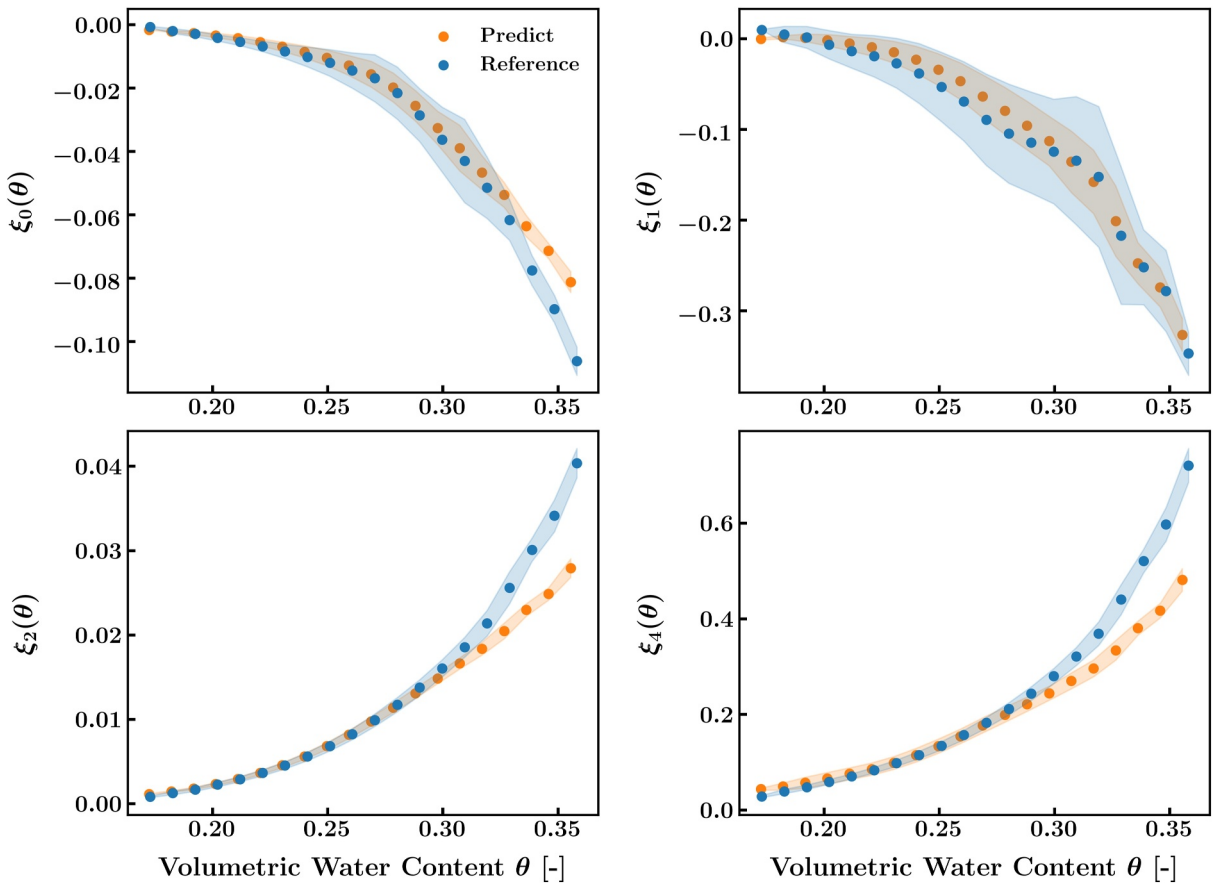

Figure 5. Comparison of reconstructed coefficients (orange points) and reference coefficients (blue points) for four coefficients. The points represent the mean value at different locations in space. The shaded regions represent one standard deviation of coefficients in space. ξ_0, ξ_1, ξ_2 and ξ_4 are the coefficient of constant terms, θ_z, θ_{zz}, and ${\theta_z}^2$, respectively.

The features of the Extended DeepGS framework are evident from the results presented above. The group sparse regression identifies the optimal parsimonious governing equation structure and provides a coarse-grained estimate of the coefficients. The soil moisture dynamics and governing equations are then refined more optimally through deep-learning steps based on the initial information of governing equations. The ability of physics-informed deep learning to infer underlying physics information from data based on imperfect governing equations is consistent with previous studies (Song et al., 2023), even when additional spatial dependence is added in coefficients. Therefore, we initially demonstrate the capabilities of the Extended-DeepGS framework for identifying governing equations for water flow in heterogeneous soils from soil water content data.

4.2. Effects of Soil Heterogeneities on Identifying Governing Equations

Coarse-grained group sparse regression theory and the Extended-DeepGS framework are proposed to extend previous studies to establish the governing equations for water flow in heterogeneous soils based on soil water content observations. Therefore, it is necessary to analyze the effects of soil heterogeneities on identifying governing equations. Therefore, three scenarios of different soils are considered: homogeneous loam soils, relatively low heterogeneous loam soils, and relatively high heterogeneous loam soils.

It is found that when dealing with the datasets of homogeneous soils, the newly developed coarse-gained group sparse regression identifies the equation as

$$\frac{\partial \theta}{\partial t} = \xi_1(\theta, z)\frac{\partial \theta}{\partial z} + \xi_2(\theta, z)\frac{\partial^2 \theta}{\partial z^2} + \xi_4(\theta, z)\left(\frac{\partial \theta}{\partial z}\right)^2 \tag{25}$$

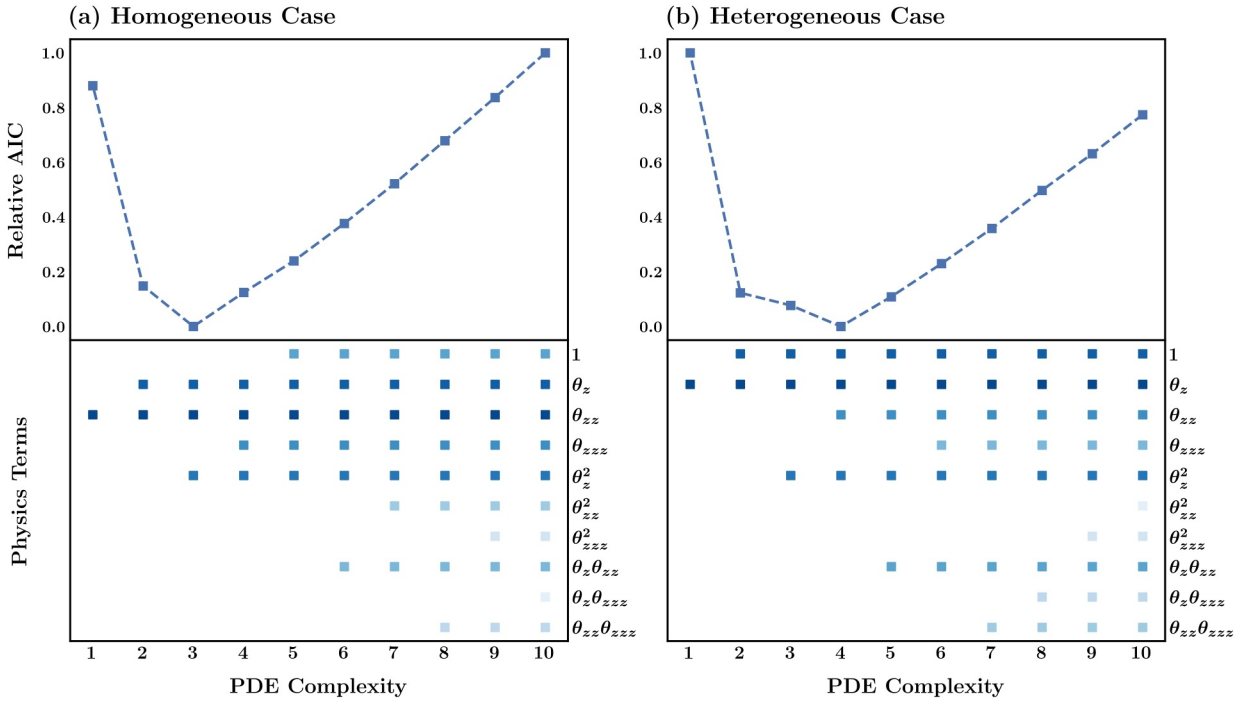

Figure 6. Varying PDE complexities ε (*x*-axis) versus discovered derivative terms (i.e., the PDE structure) (upper *y*-axis) and the corresponding relative AIC loss (lower *y*-axis) for the homogenous case and heterogenous case.

while the method identifies the following equation when dealing with the datasets of heterogenous soils:

$$\frac{\partial\theta}{\partial t} = \xi_0(\theta,z) + \xi_1(\theta,z)\frac{\partial\theta}{\partial z} + \xi_2(\theta,z)\frac{\partial^2\theta}{\partial z^2} + \xi_4(\theta,z)\left(\frac{\partial\theta}{\partial z}\right)^2 \tag{26}$$

For visual inspection, Figure 6 depicts varying complexities of underlying governing equations versus the corresponding relative AIC (Akaike information criterion), which is related to Pareto front analysis. It balances the governing equations between precisions and complexities, finding the optimal parsimonious equations. It can be found that as the PDE complexities increase, the relative AIC decreases sharply and then turns to increases at certain PDE complexities. In the decreasing period of AIC, the physics terms correctly support the soil moisture flow system. In contrast, the equations gradually overfit when PDE complexities further increase. In the homogeneous soil case, extra correcting term $\xi_0(\theta,z)$ is not relative to the physical system as the spatial varieties of soil hydraulic parameters introduce it. Therefore, it is reasonable to see that $\xi_0(\theta,z)$ will disappear in the identified equation for homogeneous soil scenarios. This also demonstrates that the proposed method can account for the correction term resulting from the heterogeneity of the soil hydraulic parameters.

To evaluate the effects on the identified equation precision, the relative error of the reconstructed soil moisture dynamics and identified coefficients is shown in Figure 7. It is clear that the error for both soil moisture dynamics and coefficients increases when soils change from homogeneous to relatively low heterogeneous and relatively high heterogeneous. The results may be attributed to the fact that the additional effects of soil heterogeneities on the underlying equation coefficients will exacerbate the difficulty of the inverse identification processes. The method can only estimate the heterogeneity of the entire profile based on the limited information available (i.e., the sparse, noisy soil water content measurements), so the accuracy naturally decreases. Therefore, the results imply that the current approach can be enhanced by incorporating additional geostatistical data. Nevertheless, the results indicate that current approaches are effective in dealing with heterogeneous soils.

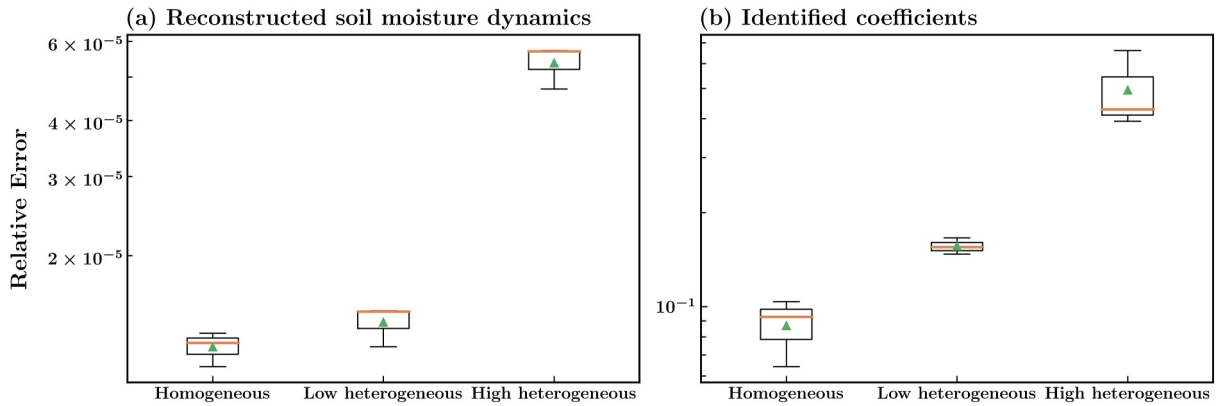

Figure 7. The relative error comparison of the reconstructed soil moisture dynamics and identified coefficients for different datasets of homogenous soils and heterogeneous soils. The box is the interquartile range (IQR), with whiskers that extend 1.5 times the IQR from the box edges. The orange line and the green triangle in the box indicate the median and mean, respectively.

4.3. Effects of Soil Textures

Soil hydraulic properties and their spatial variations exist as complex coefficient maps in discovered PDEs. The range of coefficients is determined by the soil hydraulic properties associated with different soil textures, especially considering the highly nonlinear characteristic of the coefficients. Therefore, we tested the impact of different soil textures on the accuracy of the governing equation identified by Extended-DeepGS. Here, the capabilities of Extended-DeepGS for dealing with different soil hydraulic properties are assessed with three types of heterogeneous soils: sandy loam, loam, and silt loam soils.

Figure 8 presents a case for comparing spatiotemporal coefficients of identified and reference textures of soils for visual inspection. In this case, the correct governing equation structures are identified in scenarios of three soils. The spatiotemporal distributions of the identified coefficients are found to be in general agreement with the true values for the different soil textures. Larger errors are mainly found at the sharp fronts of higher soil moisture contents, consistent with previous cases.

The error comparison results are summarized and depicted in Figure 9. It is observed that the relative error of the reconstructed soil moisture dynamics and equation coefficients grows with increasing noise levels for all types of soils. There were no significant differences between soils in terms of errors in reconstructed soil moisture dynamics. However, when it comes to coefficients, dealing with silty loam soils resulted in slightly better outcomes than loamy soils, while dealing with sandy loam soils resulted in the relatively largest error. Meanwhile, it is clear that when dealing with silty loam soils, Extended-DeepGS is also capable of tolerating higher levels of noise and can identify the correct PDE structure compared to the other two soils. This situation may be attributed to the stronger nonlinearity of the coefficients resulting from the presence of higher conductivity and their spatial heterogeneity, which are more challenging to capture.

Nevertheless, the above results suggest that recovery of the governing equations from the data is not sensitive to varying soil textures. However, soils with higher conductivity may pose challenges in identifying the correct equation structure and recovering accurate coefficients. It demonstrates the broad applicability of Extended-DeepGS among different soil textures.

4.4. Effects of Measurement Uncertainties and Sampling Intervals

Uncertainties of soil water content measurements, such as data noise, data missing or gap, and varying sampling intervals, will affect the accuracy of the identified governing equations. The effects of these factors may be more pronounced, especially when dealing with heterogeneous soils. Therefore, a case is considered for testing measurement uncertainties together, including measurement noise and spatial sampling intervals. In addition, the effects of data missing with data noise are also tested.

The results are summarized and depicted in Figure 10. It can be seen that at low data noise levels, the noise effect on the accuracy of the identified coefficients and reconstructed soil moisture dynamics is insignificant. Once the

10.1029/2024WR037786

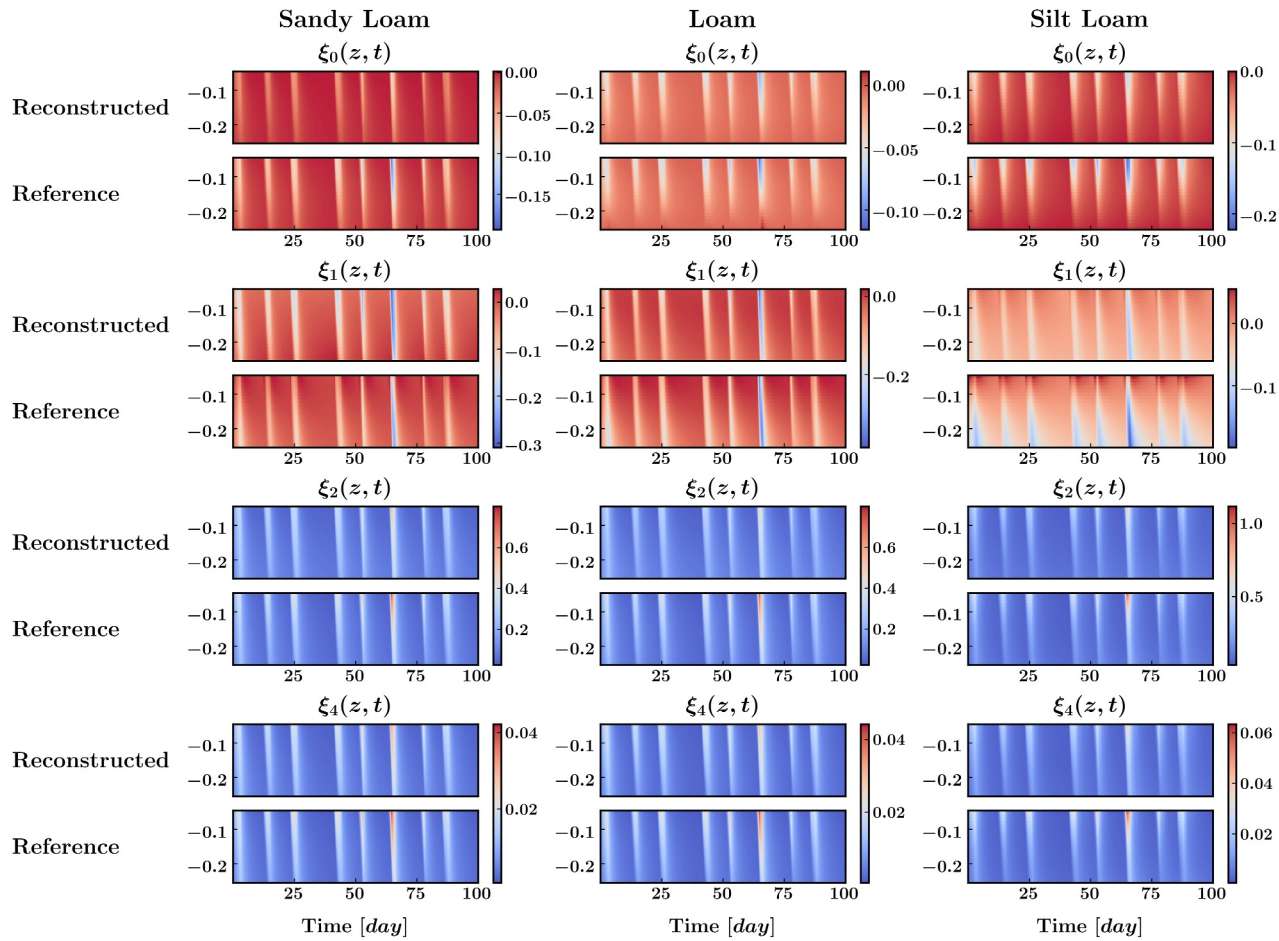

Figure 8. Comparison of identified spatiotemporal coefficients and reference spatiotemporal coefficients of the true governing equation when dealing with datasets of sandy loam, loam, and silt loam soils. The Gaussian noise is set as 10% here. $\xi_0(z,t)$, $\xi_1(z,t)$, $\xi_2(z,t)$, and $\xi_4(z,t)$ are the coefficient of 1, θ_z, θ_{zz} and ${\theta_z}^2$, respectively.

noise reaches a certain level, these errors increase relatively rapidly. This phenomenon is particularly pronounced in terms of the coefficient identification accuracy. The results imply that the precise identification of governing equations for water flow in heterogeneous soils is a highly ill-posed task. It may inherently require precise

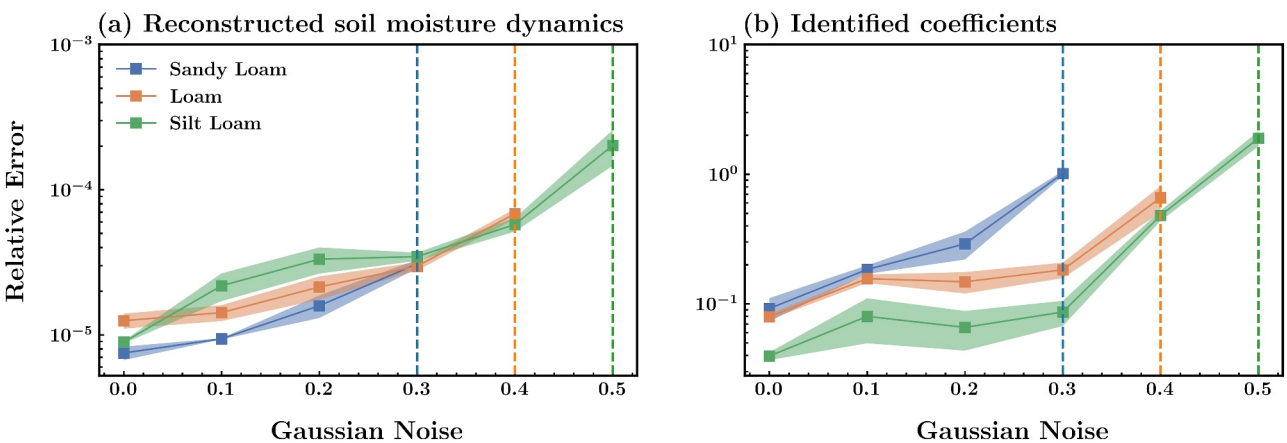

Figure 9. The relative error of the reconstructed soil moisture dynamics and identified coefficients of governing equations at different degrees of Gaussian noise for datasets of sandy loam, loam, and silt loam soils. The shaded regions represent one standard deviation. The vertical lines indicate the highest noise level that Extended-DeepGS can correctly discover the governing PDE structure for water flow in heterogeneous soils.

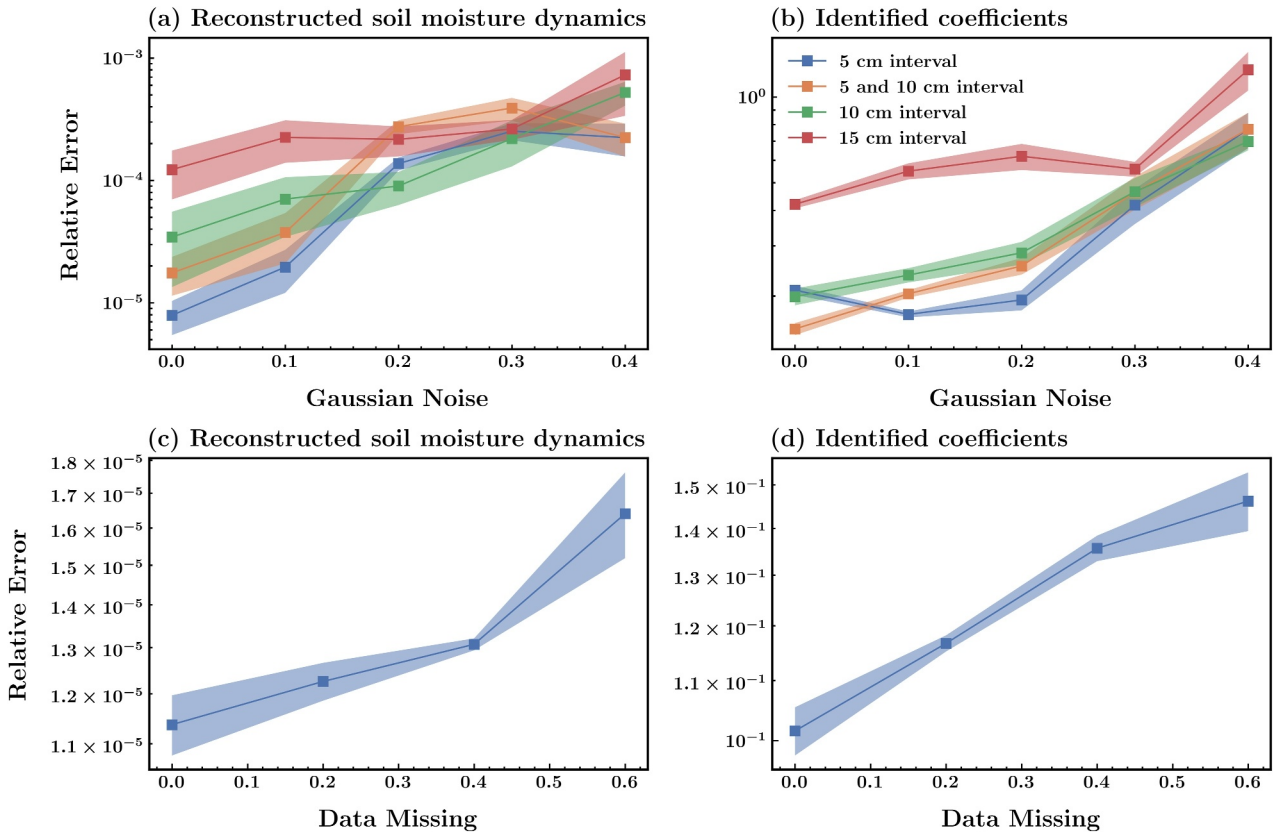

Figure 10. (a) and (b) The relative error of the reconstructed soil moisture dynamics and identified coefficients of governing equations when dealing with different levels of Gaussian noise with uniform 5 cm intervals, nonuniform 5 and 10 cm intervals, uniform 10 cm intervals and uniform 15 cm intervals of soil water measurements. (c) and (d) The relative error of the reconstructed soil moisture dynamics and identified coefficients when dealing with different levels of data missing. The shaded regions represent one standard deviation.

information, which may be lost in highly noisy data. Furthermore, it has been observed that in low-noise scenarios, particularly with regard to coefficients, the identification results are significantly affected by the size of the sampling interval. In addition, nonuniform sampling does not affect the accuracy of reconstructed coefficients and soil moisture dynamics. However, the error is obviously higher when adopting 0.15 m spatial sampling intervals than 0.05 m and 0.1 m, while the difference between the latter two is minimal. It emphasizes the critical role of sufficiently dense spatial sampling intervals in capturing nonlinear soil moisture dynamics and coefficients. Nevertheless, this distinction becomes less pronounced as the noise level increases, suggesting that the information conveyed by dense spatial intervals is negated by data noise.

Regarding the effects of data missing, the error increases linearly with the presence of random missing data; however, its variation remains minimal. This can be attributed to the neural network's ability to operate without a fixed input grid, effectively learning accurate dynamics from randomly distributed spatiotemporal data. Consequently, the method demonstrates robustness against missing data. In summary, the results demonstrate the robustness of the proposed approaches on data noise and data sampling intervals at a certain level. Nevertheless, considering the complexity of heterogeneous soils, it may require a high demand on measurement qualities and sampling strategy.

5. Limitations

Despite the promising progress, we recognize that there are still limitations that require further improvements. First, we assume that soil hydraulic properties vary continuously in space. This assumption is valid in soils primarily formed through the weathering of parent material, where gradual changes in soil hydraulic properties occur over depths. Other soil processes such as erosion, deposition, and stratification may introduce more abrupt changes or layers in soil properties, leading to discontinuities (Elkateb et al., 2003). As for soils with

discontinuous hydraulic properties, we can leverage neural networks with domain decomposition (Bandai & Ghezzehei, 2022) to deal with it. Nevertheless, soil layering information is still required. How to correctly identify the governing equations without the layering information is still an open problem. In addition, how to further interpret the estimated coefficients is also to be resolved. For example, we can validate the Darcy-Buckingham law based on discovered equation coefficients, and we can establish new soil hydraulic functions based on them.

Secondly, the proposed method can be further improved by integrating geostatistical data, which would enhance its ability to accurately model spatial variability, particularly in cases where soil properties exhibit complex heterogeneity. The incorporation of geostatistical information allows for a more detailed representation of soil characteristics, enabling the method to better capture local variations in soil properties, which is critical for increasing prediction accuracy in complex environments. In addition, the use of advances from other deep learning approaches, such as physics-informed generative adversarial networks (GANs) (Yang et al., 2020), has significant potential to improve the performance of the model. Physics-informed GANs incorporate the underlying physical laws directly into the GANs training process, which can help to identify more accurate governing equations, even when dealing with highly complex soil environments. By using these methods, the approach can better account for the uncertainties and complexities inherent in soil moisture dynamics in heterogenous soils, thereby improving the overall reliability and robustness of predictions in different environmental settings.

Finally, the proposed approaches require validation using experimental data under laboratory and field conditions. Spatial variation of soil hydraulic properties is just one element in the complexity of natural soils. It is essential to account for the effects of soil water-heat coupling processes, hysteresis and memory effects, preferential flow processes, and root water uptake processes, which sometimes play a dominant role in natural soil water flow processes. Therefore, future research should address these challenges to facilitate the discovery of physical laws and previously undisclosed governing equations for soil moisture flow.

6. Conclusions

In this study, we presented a new coarse-grained group sparse regression and a novel deep-learning framework, Extended-DeepGS. These approaches extend previous studies by being able to identify the underlying governing equation for soil moisture flow and reconstruct soil moisture dynamics in heterogeneous soils. Based on designed numerical experiments and synthetic data, we analyzed the characteristics of coarse-grained group sparse regression and the Extended-DeepGS framework for identifying governing equations from data. Group sparse regression can initially identify a governing equation, including equation structure and coarse-grained coefficient field. Then, the deep-learning reconstruction step identifies high-resolution precise coefficients and soil moisture dynamics.

We investigated the effects and differences between heterogeneous and homogeneous soils in identifying the governing equation for soil water flow. The proposed group sparse regression could identify the extra correction term in the governing equation for describing water movement in heterogeneous conditions. In addition, the coefficients can be constructed well using the proposed methods. Furthermore, the results indicated that the error of the identified governing equations and soil moisture dynamics increases as heterogeneity increases.

The impact of soil textures on the discovery of the governing equations of water flow was assessed. The results indicated that the recovery of governing equations from data is not sensitive to differences in soil textures. However, soils with higher conductivity may pose challenges in discovering accurate soil moisture dynamics and equations. Finally, we evaluated the effectiveness of Extended-DeepGS in handling various data spatial sampling strategies and noise levels. The results demonstrated the robustness of Extended-DeepGS when dealing with certain levels of sparse and noisy data. It showed that using smaller sampling node spacing may improve identification accuracy in low noise conditions but fails in high noise conditions. It may require higher demand for measurement qualities and a sampling strategy for heterogeneous soil conditions. In summary, this study showed the potential of using data-driven discovery methods to identify the soil water flow governing equation in heterogenous soils from observed soil water content data.

Appendix A

In Appendix A, we show the detailed steps for deriving Equation 3 from Equation 2. The first term on the right side of Equation 2 can be expanded as

$$\left(\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\theta}\frac{\partial\theta}{\partial z}+\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial \boldsymbol{P_1}}\frac{d\boldsymbol{P_1}}{dz}\right)\left(\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\frac{\partial\theta}{\partial z}+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial \boldsymbol{P_2}}\frac{d\boldsymbol{P_2}}{dz}\right)$$
$$=\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\theta}\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\left(\frac{\partial\theta}{\partial z}\right)^2+\left[\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\theta}\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial \boldsymbol{P_2}}\frac{d\boldsymbol{P_2}}{dz}+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial \boldsymbol{P_1}}\frac{d\boldsymbol{P_1}}{dz}\right]\frac{\partial\theta}{\partial z}$$
$$+\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial \boldsymbol{P}}\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial \boldsymbol{P_2}}\frac{d\boldsymbol{P_1}}{dz}\frac{d\boldsymbol{P_2}}{dz} \tag{A1}$$

The second term on the right side of Equation 2 can be expanded as

$$K(\theta,\boldsymbol{P_1})\frac{\partial}{\partial z}\left[\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\frac{\partial\theta}{\partial z}+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial \boldsymbol{P_2}}\frac{d\boldsymbol{P_2}}{dz}\right]=K(\theta,\boldsymbol{P_1})\left[\frac{\partial}{\partial z}\left(\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\right)\frac{\partial\theta}{\partial z}+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\frac{\partial^2\theta}{\partial z^2}\right.$$
$$\left.+\frac{\partial}{\partial z}\left(\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial \boldsymbol{P_2}}\right)\frac{d\boldsymbol{P_2}}{dz}+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial \boldsymbol{P_2}}\frac{d^2\boldsymbol{P_2}}{dz^2}\right]$$
$$=K(\theta,\boldsymbol{P_1})\left[\left(\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\theta^2}\frac{\partial\theta}{\partial z}+\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\theta\,\partial \boldsymbol{P_2}}\frac{d\boldsymbol{P_2}}{dz}\right)\frac{\partial\theta}{\partial z}\right.$$
$$+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\frac{\partial^2\theta}{\partial z^2}+\left(\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial \boldsymbol{P_2}^2}\frac{d\boldsymbol{P_2}}{dz}+\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\theta\,\partial \boldsymbol{P_2}}\frac{\partial\theta}{\partial z}\right)\frac{d\boldsymbol{P_2}}{dz}$$
$$\left.+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial \boldsymbol{P_2}}\frac{d^2\boldsymbol{P_2}}{dz^2}\right]$$
$$=K(\theta,\boldsymbol{P_1})\left[\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\theta^2}\left(\frac{\partial\theta}{\partial z}\right)^2+2\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial\theta\,\partial \boldsymbol{P_2}}\frac{d\boldsymbol{P_2}}{dz}\frac{\partial\theta}{\partial z}\right.$$
$$\left.+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial\theta}\frac{\partial^2\theta}{\partial z^2}+\frac{\partial^2\psi(\theta,\boldsymbol{P_2})}{\partial \boldsymbol{P_2}^2}\left(\frac{d\boldsymbol{P_2}}{dz}\right)^2+\frac{\partial\psi(\theta,\boldsymbol{P_2})}{\partial \boldsymbol{P_2}}\frac{d^2\boldsymbol{P_2}}{dz^2}\right] \tag{A2}$$

The third term on the right side of Equation 2 can be expanded as

$$-\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial\theta}\frac{\partial\theta}{\partial z}-\frac{\partial K(\theta,\boldsymbol{P_1})}{\partial \boldsymbol{P_1}}\frac{d\boldsymbol{P_1}}{dz} \tag{A3}$$

Then the equations can be merged into Equation 3. Second, we show the expanded form of the generalized water content-based RRE in heterogeneous soils when combining the Mualem-van Genuchten model (Van Genuchten, 1980). The parameters $\{K_s,\alpha,n,\theta_s,\theta_r\}$ exist spatially varying continuously in space. The expanded form is as follows:

$$\frac{\partial\theta}{\partial t}=f_0(\theta,z)+f_1(\theta,z)\frac{\partial\theta}{\partial z}+f_2(\theta,z)\frac{\partial^2\theta}{\partial z^2}+f_3(\theta,z)\left(\frac{\partial\theta}{\partial z}\right)^2 \tag{A4}$$

where

$$f_0(\theta,z)=-\frac{\partial K}{\partial K_s}\frac{dK_s}{dz}+\left(\frac{\partial K}{\partial n}\frac{\partial\psi}{\partial n}+K\frac{\partial^2\psi}{\partial n^2}\right)\left(\frac{dn}{dz}\right)^2+K\frac{\partial^2\psi}{\partial\boldsymbol{\alpha^2}}\left(\frac{d\alpha}{dz}\right)^2+K\frac{\partial\psi}{\partial\alpha}\frac{d^2\alpha}{dz^2}+K\frac{\partial\psi}{\partial n}\frac{d^2n}{dz^2}-\frac{\partial K}{\partial n}\frac{dn}{dz}$$
$$+\left(\frac{\partial K}{\partial\theta_r}\frac{\partial\psi}{\partial\theta_r}+K\frac{\partial^2\psi}{\partial\theta_r^2}\right)\left(\frac{d\theta_r}{dz}\right)^2+K\frac{\partial\psi}{\partial\theta_r}\frac{d^2\theta_r}{dz^2}-\frac{\partial K}{\partial\theta_r}\frac{d\theta_r}{dz}+\left(\frac{\partial K}{\partial\theta_s}\frac{\partial\psi}{\partial\theta_s}+K\frac{\partial^2\psi}{\partial\theta_s^2}\right)\left(\frac{d\theta_s}{dz}\right)^2$$
$$+K\frac{\partial\psi}{\partial\theta_s}\frac{d^2\theta_s}{dz^2}-\frac{\partial K}{\partial\theta_s}\frac{d\theta_s}{dz} \tag{A5}$$

$$f_1(\theta,z)=-\frac{\partial K}{\partial\theta}+\frac{\partial\psi}{\partial\theta}\frac{\partial K}{\partial K_s}\frac{dK_s}{dz}+\frac{\partial K}{\partial\theta}\frac{\partial\psi}{\partial\alpha}\frac{d\alpha}{dz}+2K\frac{\partial^2\psi}{\partial\theta\partial\alpha}\frac{d\alpha}{dz}+\left(\frac{\partial K}{\partial\theta}\frac{\partial\psi}{\partial n}+\frac{\partial K}{\partial n}\frac{\partial\psi}{\partial\theta}+2K\frac{\partial^2\psi}{\partial\theta\partial n}\right)\frac{dn}{dz}$$
$$+\left(\frac{\partial K}{\partial\theta}\frac{\partial\psi}{\partial\theta_r}+\frac{\partial K}{\partial\theta_r}\frac{\partial\psi}{\partial\theta}+2K\frac{\partial^2\psi}{\partial\theta\,\partial\theta_r}\right)\frac{d\theta_r}{dz}++\left(\frac{\partial K}{\partial\theta}\frac{\partial\psi}{\partial\theta_s}+\frac{\partial K}{\partial\theta_s}\frac{\partial\psi}{\partial\theta}+2K\frac{\partial^2\psi}{\partial\theta\,\partial\theta_s}\right)\frac{d\theta_s}{dz} \tag{A6}$$

$$f_2(\theta, z) = K\frac{\partial \psi}{\partial \theta} \tag{A7}$$

$$f_3(\theta, z) = K\frac{\partial^2 \psi}{\partial \theta^2} + \frac{\partial K}{\partial \theta}\frac{\partial \psi}{\partial \theta} \tag{A8}$$

The derivative of parameters with respect to space, such as dn/dz, is calculated by finite difference methods. Other expanded terms are given as follows:

$$\psi = -\alpha^{-1}\left(S_e^{-1/m} - 1\right)^{1/n} \tag{A9}$$

$$\frac{\partial \psi}{\partial \theta} = -\left(mn\alpha(\theta_s - \theta_r)S_e\left(S_e^{1/m} - 1\right)\right)^{-1}\left(S_e^{-1/m} - 1\right)^{1/n} \tag{A10}$$

$$\frac{\partial^2 \psi}{\partial \theta^2} = \alpha^{-1}\left(mn(\theta_s - \theta_r)S_e\left(S_e^{1/m} - 1\right)\right)^{-2}\left(S_e^{-1/m} - 1\right)^{1/n}\left(n(m+1)\,S_e^{1/m} - mn - 1\right) \tag{A11}$$

$$\frac{\partial K}{\partial \theta} = K_s\left(2S_e^{1/2}(\theta_s - \theta_r)\right)^{-1}\left(1 - \left(1 - S_e^{1/m}\right)^m\right)\left(\left(5S_e^{1/m} - 1\right)\left(1 - S_e^{1/m}\right)^{m-1} + 1\right) \tag{A12}$$

$$\frac{\partial K}{\partial K_s} = S_e^{1/2}\left(1 - \left(1 - S_e^{1/m}\right)^m\right)^2 \tag{A13}$$

$$\frac{\partial \psi}{\partial \alpha} = \alpha^{-2}\left(S_e^{-1/m} - 1\right)^{1/n} \tag{A14}$$

$$\frac{\partial^2 \psi}{\partial \alpha^2} = -2\alpha^{-3}\left(S_e^{-1/m} - 1\right)^{1/n} \tag{A15}$$

$$\frac{\partial^2 \psi}{\partial \theta \partial \alpha} = \alpha^{-2}\left(mn(\theta_s - \theta_r)S_e\left(S_e^{1/m} - 1\right)\right)^{-1}\left(S_e^{-1/m} - 1\right)^{1/n} \tag{A16}$$

$$\frac{\partial K}{\partial n} = -2K_s S_e^{1/2} n^{-2}\left(1 - S_e^{1/m}\right)^m\left(1 - \left(1 - S_e^{1/m}\right)^m\right)\left(m^{-1}\left(1 - S_e^{1/m}\right)^{-1} S_e^{1/m}\ln S_e + \ln\left(1 - S_e^{1/m}\right)\right) \tag{A17}$$

$$\frac{\partial \psi}{\partial n} = -\alpha^{-1}m^{-2}n^{-2}S_e^{-1/m}\left(S_e^{-1/m} - 1\right)^{-m}\left(m^2\left(S_e^{1/m} - 1\right)\ln\left(S_e^{-1/m} - 1\right) - (m-1)\ln(S_e)\right) \tag{A18}$$

$$\begin{aligned}\frac{\partial^2 \psi}{\partial n^2} = {} & 2\alpha^{-1}m^{-2}n^{-3}S_e^{-1/m}\left(S_e^{-1/m} - 1\right)^{-m}\left(m^2\left(S_e^{1/m} - 1\right)\ln\left(S_e^{-1/m} - 1\right) - (m-1)\ln(S_e)\right) \\ & - \alpha^{-1}n^{-4}m^{-4}\left(S_e^{1/m} - 1\right)^{-1}S_e^{-1/m}\left(S_e^{-1/m} - 1\right)^{-m}\left(-\left(S_e^{1/m} - 1\right)^2 m^4\left(\ln\left(S_e^{-1/m} - 1\right)\right)^2\right. \\ & \left. + 2m\left(S_e^{1/m} - 1\right)\ln(S_e)\left((m-1)\,m\ln\left(S_e^{-1/m} - 1\right) - 1\right) - (m-1)\left(S_e^{1/m} + m - 1\right)(\ln(S_e))^2\right)\end{aligned} \tag{A19}$$

$$\begin{aligned}\frac{\partial^2 \psi}{\partial \theta\,\partial n} = {} & -\alpha^{-1}(\theta_s - \theta_r)^{-1}n^{-2}m^{-3}\left(S_e^{1/m} - 1\right)^{-1}S_e^{-(m+1)/m}\left(S_e^{-1/m} - 1\right)^{-m}\left((m-1)\left(m - 1 + S_e^{1/m}\right)\ln(S_e)\right. \\ & \left. - m\left(S_e^{1/m} - 1\right)\left(m(m-1)\ln\left(S_e^{-1/m} - 1\right) - 1\right)\right)\end{aligned} \tag{A20}$$

$$\frac{\partial \psi}{\partial \theta_r} = -\left(mn\alpha(\theta_s - \theta_r)S_e\left(S_e^{1/m} - 1\right)\right)^{-1}\left(S_e^{-1/m} - 1\right)^{1/n}(1 - S_e) \tag{A21}$$

$$\frac{\partial K}{\partial \theta_r} = K_s\left(2S_e^{1/2}(\theta_s - \theta_r)\right)^{-1}\left(1 - \left(1 - S_e^{1/m}\right)^m\right)\left(\left(5S_e^{1/m} - 1\right)\left(1 - S_e^{1/m}\right)^{m-1} + 1\right)(1 - S_e) \tag{A22}$$

$$\frac{\partial^2 \psi}{\partial \theta\,\partial \theta_r} = \alpha^{-1}\left(mn(\theta_s - \theta_r)S_e\left(S_e^{1/m} - 1\right)\right)^{-2}\left(S_e^{-1/m} - 1\right)^{1/n}\left(n(m+1)\,S_e^{1/m} - mn - 1\right)(1 - S_e) \tag{A23}$$

$$\frac{\partial^2 \psi}{\partial \theta_r{}^2} = \alpha^{-1}\left(mn(\theta_s - \theta_r)S_e\left(S_e{}^{1/m} - 1\right)\right)^{-2}\left(S_e{}^{-1/m} - 1\right)^{1/n}\left(n(m+1)\,S_e{}^{1/m} - mn - 1\right)(1 - S_e)^2 \tag{A24}$$

$$\frac{\partial \psi}{\partial \theta_r} = -S_e\left(mn\alpha(\theta_s - \theta_r)S_e\left(S_e{}^{1/m} - 1\right)\right)^{-1}\left(S_e{}^{-1/m} - 1\right)^{1/n} \tag{A25}$$

$$\frac{\partial K}{\partial \theta_s} = K_s S_e\left(2S_e{}^{1/2}(\theta_s - \theta_r)\right)^{-1}\left(1 - \left(1 - S_e{}^{1/m}\right)^m\right)\left(\left(5S_e{}^{1/m} - 1\right)\left(1 - S_e{}^{1/m}\right)^{m-1} + 1\right) \tag{A26}$$

$$\frac{\partial^2 \psi}{\partial \theta\, \partial \theta_s} = S_e\alpha^{-1}\left(mn(\theta_s - \theta_r)S_e\left(S_e{}^{1/m} - 1\right)\right)^{-2}\left(S_e{}^{-1/m} - 1\right)^{1/n}\left(n(m+1)\,S_e{}^{1/m} - mn - 1\right) \tag{A27}$$

$$\frac{\partial^2 \psi}{\partial \theta_s{}^2} = S_e{}^2\alpha^{-1}\left(mn(\theta_s - \theta_r)S_e\left(S_e{}^{1/m} - 1\right)\right)^{-2}\left(S_e{}^{-1/m} - 1\right)^{1/n}\left(n(m+1)\,S_e{}^{1/m} - mn - 1\right) \tag{A28}$$

Appendix B

Here, we introduce the hyper-parameter settings within the Extended-DeepGS. First, the determinations of neural networks are based on an empirical study, and the principle is ensuring the neural networks are deep enough to approximate soil moisture dynamics and complex coefficients. The number of hidden layers of the main networks is set as $\{5, 6, 7, 8\}$ with 50 nodes a layer. The branch networks of input θ are set as $\{3, 4, 5\}$ hidden layers with 20 nodes a layer and input z are set as 2 hidden layers with 20 nodes. Except for the output layer, which employs linear activation; the other layers utilize the sine function as the activation function to consider the potential periodicity of soil moisture dynamics. In addition, adaptive activation functions (Jagtap et al., 2020) are used to accelerate training. Training maximum iterations n_{train} are set up to 1×10^5 steps. All the training is realized by the Adam optimizer (Kingma & Ba, 2015). The first case study setting is used to determine the optimal neural networks.

The results of the relative error of the reconstructed coefficients are summarized and depicted in Figure B1. It seems that the results produced by the various network configurations are highly close. Therefore, combining

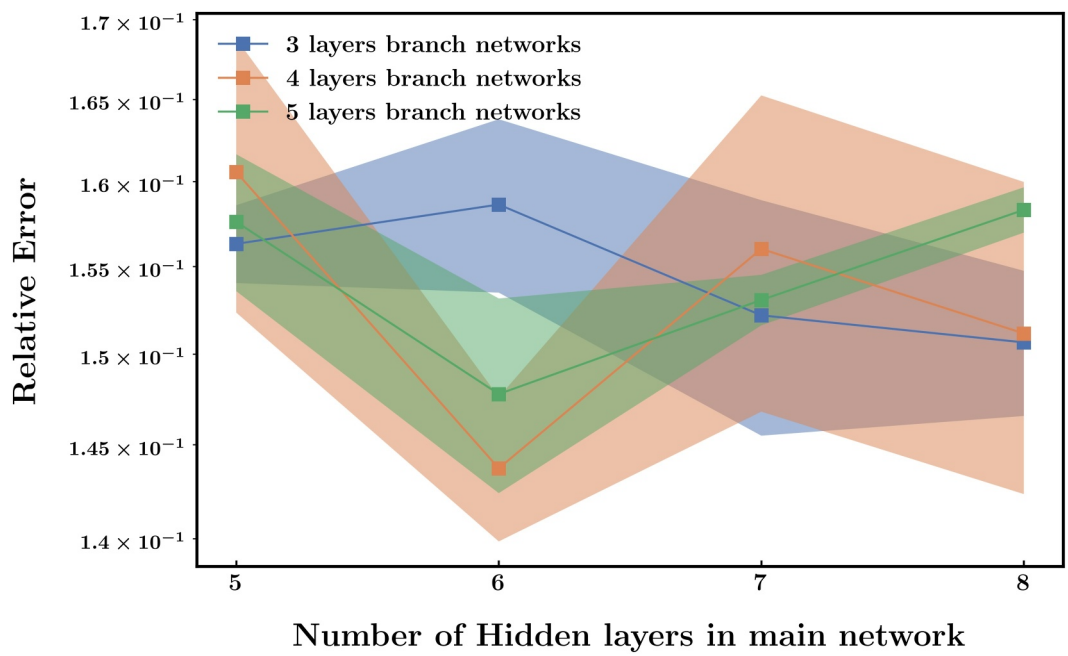

Figure B1. The relative error of the reconstructed coefficients when Extended-DeepGS adopts different numbers of hidden layers.

computational cost and accuracy, the main networks are set as six hidden layers with 50 nodes, and the branch networks are set as four hidden layers (input z), and two hidden layers (input θ) with 20 nodes. The selection of the relative weight parameters of the loss function is chosen empirically, as all adaptive selection methods do not work well, as reported previously (Bandai & Ghezzehei, 2022; Song et al., 2023). The relative weight parameter ω_p is set as 10 and ω_d is set as 1 in this study. The regularization term is realized by group sparse regression, and therefore does not need to be set. The number of residual points is set as 2×10^5, and the Halton sequence is set as the residual points sampling strategy. As for the hyper-parameters in coarse-grained group sparse regression, the maximum number of iterations is set as 10 and the threshold sequences number is set as 500, to allow the algorithm to heuristically locate an optimal threshold. The *win* is a hyper-parameter that should be set before running the algorithms. The selection of *win* mainly follow some observational experiences and generally need not to be changed across scenarios, within the orders of magnitude of 0.01 to 0.001."

Appendix C

This appendix presents the generated soil hydraulic parameters utilized as references in the study. Figure C1 displays the five Mualem-van Genuchten Soil Hydraulic Parameters based on the experimental setup.

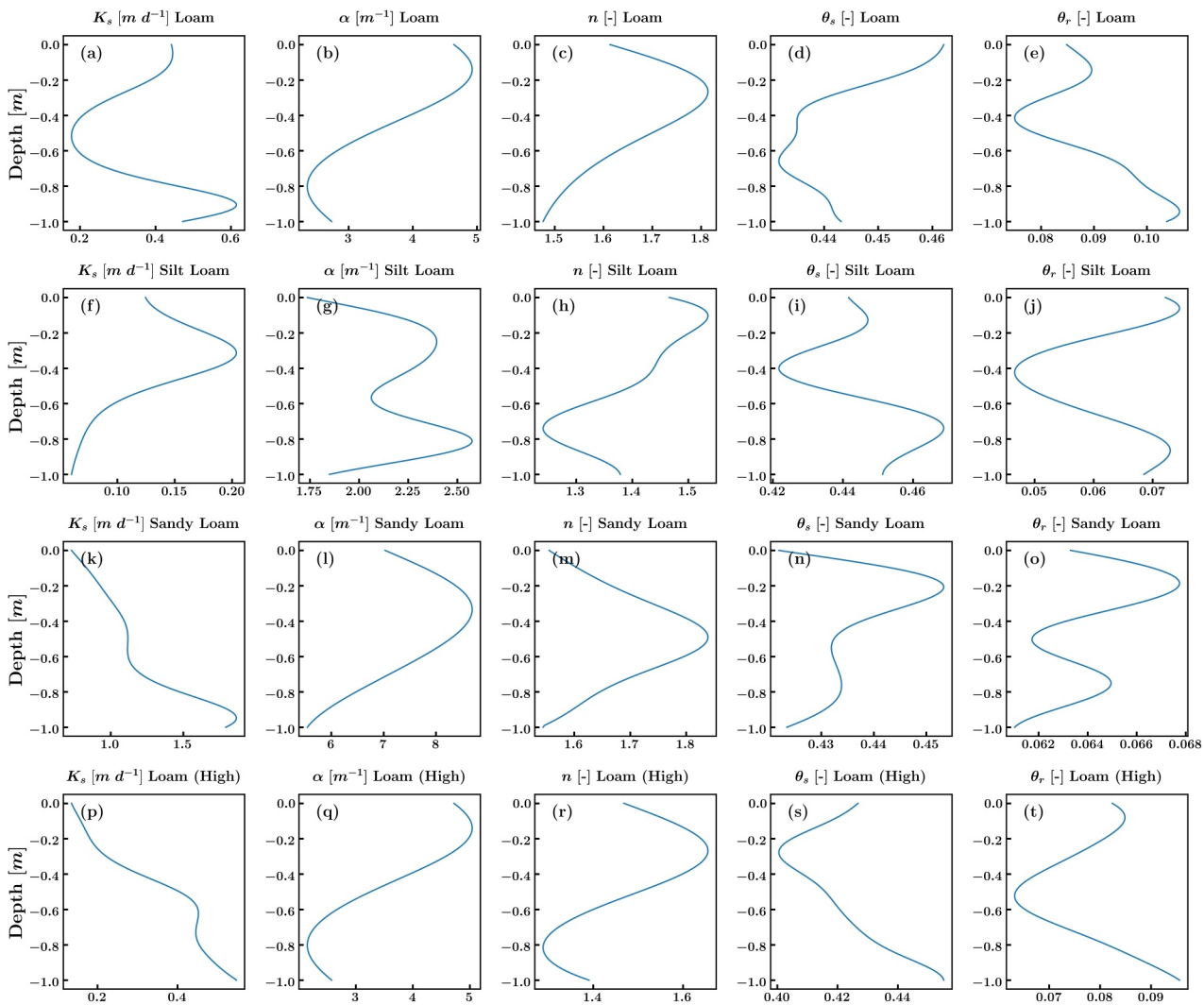

Figure C1. (a)–(t) Visualizations of changes of five Mualem-van Genuchten soil hydraulic parameters in space.

Notation

$\wedge$	Hat indicating predicted values or functions
(i)	Superscript (i)
$\|\cdot\|_2$	2-Norm of a matrix or a vector
a	Number of sensors
b	Timesteps of measured data
$\boldsymbol{b}$	Bias vector in neural networks
c	The number of sets for each depth
$\mathcal{D}_{(i,j)}$	Data point measured by the sensor at i th sensor at j timestep
$\mathcal{D}_m$	Measured data points
$\mathcal{D}_{meta}$	Grided metadata points
$\mathcal{D}_r$	Residual points
$f_i(\theta, z)$	i th coefficient function
g	A group that contains a subset of the indices enumerating the columns of the candidate library matrix and coefficients
$\mathcal{G}$	A collection of groups
K	Unsaturated hydraulic conductivity $\left[\mathrm{L\ T^{-1}}\right]$
K_s	Saturated hydraulic conductivity $\left[\mathrm{L\ T^{-1}}\right]$
l	Mualem-van Genuchen model parameter $[-]$
$\mathcal{L}$	Total Loss
$\mathcal{L}_d$	Data loss
$\mathcal{L}_p$	PDE residual loss
m	Mualem-van Genuchen model parameter $[-]$
n	Mualem-van Genuchen model parameter $[-]$
$N(\cdot)$	Candidate library
N_i	i th candidate term
N_m	Number of measurements
N_r	Number of residual points
$\mathcal{N}_i$	i th neural network
$\boldsymbol{P_1}$	Parameter vector of the unsaturated hydraulic conductivity mode
$\boldsymbol{P}_2$	Parameter vector of the soil water retention curves (WRC)
S_e	Effective saturation $[-]$
$\boldsymbol{S}_{(i,k)}$	Set in group sparse regression
t	Time [T]
U	Dataset matrix
$\mathbf{W}$	Weight matrix in neural networks

win	The range of variation of the volumetric water content in a set $\left[L^3\ L^{-3}\right]$
z	Vertical coordinate or depth (positive upward) [L]
z_m	Measurement location (positive upward) [L]
α	Mualem-van Genuchen model parameter $\left[L^{-1}\right]$
θ	Volumetric water content $\left[L^3\ L^{-3}\right]$
$\tilde{\theta}$	Estimated volumetric water content by neural networks $\left[L^3\ L^{-3}\right]$
$\hat{\theta}$	Measured volumetric water content $\left[L^3\ L^{-3}\right]$
θ_s	Saturated water content $\left[L^3\ L^{-3}\right]$
θ_r	Residual water content $\left[L^3\ L^{-3}\right]$
θ_{min}	Lowest volumetric water content value in a set $\left[L^3\ L^{-3}\right]$
θ_{max}	Largest volumetric water content value in a set $\left[L^3\ L^{-3}\right]$
Θ	Candidate library matrix
$\xi_i(\theta, z)$	i th coefficient scalar in the underlying equation
$\boldsymbol{\xi}$	Coefficient vector
$\boldsymbol{\Lambda}^{(i)}$	Parameter vector in i th neural network
ω_d	Relative weightings of data loss
ω_p	Relative weightings of PDE residual loss
γ	Relative weightings of the regularization term

Data Availability Statement

The data and codes used in this paper are available on the Zenodo website (Song, 2024).

Acknowledgments
This study was supported by the National Natural Science Foundation of China Grant 52179038.

References

Bandai, T., & Ghezzehei, T. A. (2022). Forward and inverse modeling of water flow in unsaturated soils with discontinuous hydraulic conductivities using physics-informed neural networks with domain decomposition. *Hydrology and Earth System Sciences*, *26*(16), 4469–4495. https://doi.org/10.5194/hess-26-4469-2022

Bertsimas, D., Pauphilet, J., & Van Parys, B. (2020). Sparse regression: Scalable algorithms and empirical performance. *Statistical Science*, *35*(4), 555–578. https://doi.org/10.1214/19-STS701

Beven, K., & Germann, P. (2013). Macropores and water flow in soils revisited: Review. *Water Resources Research*, *49*(6), 3071–3092. https://doi.org/10.1002/wrcr.20156

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *113*(15), 3932–3937. https://doi.org/10.1073/pnas.1517384113

Buckingham, E. (1907). *Studies on the movement of soil moisture*. Govt. Print. Off.

Camps-Valls, G., Gerhardus, A., Ninad, U., Varando, G., Martius, G., Balaguer-Ballester, E., et al. (2023). Discovering causal relations and equations from data. *Physics Reports*, *1044*, 1–68. https://doi.org/10.1016/j.physrep.2023.10.005

Chang, H., & Zhang, D. (2019). Identification of physical processes via combined data-driven and data-assimilation methods. *Journal of Computational Physics*, *393*, 337–350. https://doi.org/10.1016/j.jcp.2019.05.008

Darcy, H. (1856). The public fountains of the city of dijon.

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., et al. (2021). The international soil moisture network: Serving earth system science for over a decade. *Hydrology and Earth System Sciences*, *25*(11), 5749–5804. https://doi.org/10.5194/hess-25-5749-2021

Elkateb, T., Chalaturnyk, R., & Robertson, P. K. (2003). An overview of soil heterogeneity: Quantification and implications on geotechnical field problems. *Canadian Geotechnical Journal*, *40*(1), 1–15. https://doi.org/10.1139/t02-090

Farthing, M. W., & Ogden, F. L. (2017). Numerical solution of richards' equation: A review of advances and challenges. *Soil Science Society of America Journal*, *81*(6), 1257–1269. https://doi.org/10.2136/sssaj2017.02.0058

Ghorbani, A., Sadeghi, M., & Jones, S. B. (2021). Towards new soil water flow equations using physics-constrained machine learning. *Vadose Zone Journal*, *20*(May). https://doi.org/10.1002/vzj2.20136

Grundner, A., Beucler, T., Gentine, P., & Eyring, V. (2024). Data-driven equation discovery of a cloud cover parameterization. *Journal of Advances in Modeling Earth Systems*, *16*(3), e2023MS003763. https://doi.org/10.1029/2023MS003763

Jagtap, A. D., Kawaguchi, K., & Karniadakis, G. E. (2020). Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. *Journal of Computational Physics*, *404*, 109136. https://doi.org/10.1016/j.jcp.2019.109136

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *Iclr*, 1–15.
Kutz, J. N., & Brunton, S. L. (2022). Parsimony as the ultimate regularizer for physics-informed machine learning. *Nonlinear Dynamics*, *107*(3), 1801–1817. https://doi.org/10.1007/s11071-021-07118-3
Li, P., Zha, Y., Shi, L., Tso, C. H. M., Zhang, Y., & Zeng, W. (2020). Comparison of the use of a physical-based model with data assimilation and machine learning methods for simulating soil water dynamics. *Journal of Hydrology*, *584*(October 2019), 124692. https://doi.org/10.1016/j.jhydrol.2020.124692
Mantoglou, A., & Wilson, J. L. (1982). The Turning Bands Method for simulation of random fields using line generation by a spectral method. *Water Resources Research*, *18*(5), 1379–1394. https://doi.org/10.1029/WR018i005p01379
Nakamura, G., & Potthast, R. (2015). *Inverse modeling*. IOP Publishing.
Narasimhan, T. N. (2007). Central ideas of buckingham (1907): A century later. *Vadose Zone Journal*, *6*(4), 687–693. https://doi.org/10.2136/vzj2007.0080
Or, D., Lehmann, P., & Assouline, S. (2015). Natural length scales define the range of applicability of the Richards equation for capillary flows: Natural length scales for application of Richards equation. *Water Resources Research*, *51*(9), 7130–7144. https://doi.org/10.1002/2015WR017034
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library.
Rahmati, M., Amelung, W., Brogi, C., Dari, J., Flammini, A., Bogena, H., et al. (2024). Soil moisture memory: State-of-the-art and the way forward. *Reviews of Geophysics*, *62*(2), e2023RG000828. https://doi.org/10.1029/2023RG000828
Rahmati, M., Or, D., Amelung, W., Bauke, S. L., Bol, R., Hendricks Franssen, H.-J., et al. (2023). Soil is a living archive of the Earth system. *Nature Reviews Earth & Environment*, *4*(7), 421–423. https://doi.org/10.1038/s43017-023-00454-5
Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). Machine learning of linear differential equations using Gaussian processes. *Journal of Computational Physics*, *348*(1), 683–693. https://doi.org/10.1016/j.jcp.2017.07.050
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
Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, *1*(5), 318–333. https://doi.org/10.1063/1.1745010
Richardson, L. F. (1922). *Weather prediction by numerical process*. Cambridge University Press.
Ross, A., Li, Z., Perezhogin, P., Fernandez-Granda, C., & Zanna, L. (2023). Benchmarking of machine learning ocean subgrid parameterizations in an idealized model. *Journal of Advances in Modeling Earth Systems*, *15*(1), e2022MS003258. https://doi.org/10.1029/2022MS003258
Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x
Rudy, S., Alla, A., Brunton, S. L., & Kutz, J. N. (2019). Data-driven identification of parametric partial differential equations. *SIAM Journal on Applied Dynamical Systems*, *18*(2), 643–660. https://doi.org/10.1137/18M1191944
Song, W. (2024). Extended-deepgs [dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.14551650
Song, W., Jiang, S., Camps-Valls, G., Williams, M., Zhang, L., Reichstein, M., et al. (2024). Towards data-driven discovery of governing equations in geosciences. *Communications Earth & Environment*, *5*(1), 589. https://doi.org/10.1038/s43247-024-01760-6
Song, W., Shi, L., Hu, X., Wang, Y., & Wang, L. (2023). Reconstructing the unsaturated flow equation from sparse and noisy data: Leveraging the synergy of group sparsity and physics-informed deep learning. *Water Resources Research*, *59*(5), e2022WR034122. https://doi.org/10.1029/2022WR034122
Song, W., Shi, L., Wang, L., Wang, Y., & Hu, X. (2022). Data-driven discovery of soil moisture flow governing equation: A sparse regression framework. *Water Resources Research*, *58*(8), e2022WR031926. https://doi.org/10.1029/2022wr031926
Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society - Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, *44*(5), 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x
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 & Environment*, *3*(9), 573–587. https://doi.org/10.1038/s43017-022-00324-6
Vereecken, H., Huisman, J. A., & Franssen, H. J. H. (2015). Soil hydrology: Recent methodological advances, challenges, and perspectives. *Water Resources Research*, *5*(3), 2. https://doi.org/10.1002/2014wr016852
Virgolin, M., & Pissis, S. P. (2022). Symbolic regression is NP-hard. *TMLR*, *1*(2016), 1–11.
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, Y., Shi, L., Hu, Y., Hu, X., Song, W., & Wang, L. (2024). A comprehensive study of deep learning for soil moisture prediction. *Hydrology and Earth System Sciences*, *28*(4), 917–943. https://doi.org/10.5194/hess-28-917-2024
Xu, H., Chen, Y., Zeng, Z., Li, N., Li, J., & Zhang, D. (2023). Interpretable AI-driven discovery of terrain-precipitation relationships for enhanced climate insights. *arXiv*. https://doi.org/10.48550/arXiv.2309.15400
Yang, L., Zhang, D., & Karniadakis, G. E. (2020). Physics-informed generative adversarial networks for stochastic differential equations. *SIAM Journal on Scientific Computing*, *42*(1), A292–A317. https://doi.org/10.1137/18M1225409
Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, *47*(17), e2020GL088376. https://doi.org/10.1029/2020GL088376
Zeng, J., Xu, H., Chen, Y., & Zhang, D. (2023). Deep learning discovery of macroscopic governing equations for viscous gravity currents from microscopic simulation data. *Computational Geosciences*, *27*(123456789), 987–1000. https://doi.org/10.1007/s10596-023-10244-z
Zha, Y., Shi, L., Ye, M., & Yang, J. (2013). A generalized ross method for two- and three-dimensional variably saturated flow. *Advances in Water Resources*, *54*, 67–77. https://doi.org/10.1016/j.advwatres.2013.01.002
Zha, Y., Yang, J., Shi, L., & Song, X. (2013). Simulating one-dimensional unsaturated flow in heterogeneous soils with water content-based richards equation. *Vadose Zone Journal*, *12*(2), 1–13. https://doi.org/10.2136/vzj2012.0109
Zha, Y., Yang, J., Zeng, J., Tso, C. M., Zeng, W., & Shi, L. (2019). Review of numerical solution of Richardson–Richards equation for variably saturated flow in soils. *WIREs Water*, *6*(5), e1364. https://doi.org/10.1002/wat2.1364