Physics-informed neural networks for Richardson-Richards equation: Estimation of constitutive relationships and soil water flux density from volumetric water content measurements

Toshiyuki Bandai[1,1] and Teamrat Ghezzehei[2,2]

[1]University of California, Merced
[2]University of California Merced

November 30, 2022

Abstract

Water retention curves (WRCs) and hydraulic conductivity functions (HCFs) are critical soil-specific characteristics necessary for modeling the movement of water in soils using the Richardson-Richards equation (RRE). Well-established laboratory measurement methods of WRCs and HCFs are not usually unsuitable for simulating field-scale soil moisture dynamics because of the scale mismatch. Hence, the inverse solution of the RRE is used to estimate WRCs and HCFs from field measured data. Here, we propose a physics-informed neural networks (PINNs) framework for the inverse solution of the RRE and the estimation of WRCs and HCFs from only volumetric water content (VWC) measurements. Unlike conventional inverse methods, the proposed framework does not need initial and boundary conditions. The PINNs consists of three linked feedforward neural networks, two of which were constrained to be monotonic functions to reflect the monotonicity of WRCs and HCFs. Alternatively, we also tested PINNs without monotonicity constraints. We trained the PINNs using synthetic VWC data with artificial noise, derived by a numerical solution of the RRE for three soil textures. The PINNs were able to reconstruct the true VWC dynamics. The monotonicity constraints prevented the PINNs from overfitting the training data. We demonstrated that the PINNs could recover the underlying WRCs and HCFs in non-parametric form, without a need for initial guess. However, the reconstructed WRCs at near-saturation–which was not fully represented in the training data–was unsatisfactory. We additionally showed that the trained PINNs could estimate soil water flux density with a broader range of estimation than the currently available methods.

Physics-informed neural networks with monotonicity constraints for Richardson-Richards equation: Estimation of constitutive relationships and soil water flux density from volumetric water content measurements

Toshiyuki Bandai [1], Teamrat A. Ghezzehei [1]

[1]Life and Environmental Science Department, University of California, Merced, 5200, Lake Rd, Merced, CA, 95343, USA

Key Points:

- Hydraulic conductivity functions were precisely estimated from only volumetric water content data using the proposed framework.
- Soil water flux density was accurately derived from the trained physics-informed neural networks.
- Incorporating monotonic neural networks to represent constitutive relationships in physics-informed neural networks prevents overfitting.

Corresponding author: Toshiyuki Bandai, `tbandai@ucmerced.edu`

Abstract

Water retention curves (WRCs) and hydraulic conductivity functions (HCFs) are critical soil-specific characteristics necessary for modeling the movement of water in soils using the Richardson-Richards equation (RRE). Well-established laboratory measurement methods of WRCs and HCFs are not usually unsuitable for simulating field-scale soil moisture dynamics because of the scale mismatch. Hence, the inverse solution of the RRE is used to estimate WRCs and HCFs from field measured data. Here, we propose a physics-informed neural networks (PINNs) framework for the inverse solution of the RRE and the estimation of WRCs and HCFs from only volumetric water content (VWC) measurements. Unlike conventional inverse methods, the proposed framework does not need initial and boundary conditions. The PINNs consists of three linked feedforward neural networks, two of which were constrained to be monotonic functions to reflect the monotonicity of WRCs and HCFs. Alternatively, we also tested PINNs without monotonicity constraints. We trained the PINNs using synthetic VWC data with artificial noise, derived by a numerical solution of the RRE for three soil textures. The PINNs were able to reconstruct the true VWC dynamics. The monotonicity constraints prevented the PINNs from overfitting the training data. We demonstrated that the PINNs could recover the underlying WRCs and HCFs in non-parametric form, without a need for initial guess. However, the reconstructed WRCs at near-saturation–which was not fully represented in the training data–was unsatisfactory. We additionally showed that the trained PINNs could estimate soil water flux density with a broader range of estimation than the currently available methods.

1 Introduction

Accurate prediction of soil moisture dynamics is vital for many applications, including weather forecasts, agricultural water management, and prediction of natural disasters, such as landslides and floods, and drought (Robinson et al., 2008; Babaeian et al., 2019). Notably, detailed information about near-surface soil moisture dynamics is essential for land surface modeling and remote sensing applications.

Mathematically, soil moisture dynamics is described by a non-linear partial differential equation (PDE), commonly referred to as the Richardson-Richards equation (RRE) (Richardson, 1922; Richards, 1931). The RRE is composed of the continuity equation and the Buckingham-Darcy law (Buckingham, 1907) and consists of three primary variables: matric potential ψ, volumetric water content θ, and hydraulic conductivity K. The latter two variables are commonly expressed as functions of matric potential using water retention curves (WRCs) and hydraulic conductivity functions (HCFs), respectively. Furthermore, the two soil hydraulic functions (also referred to as constitutive relationships) are often treated as interdependent by employing conceptual models of unsaturated flow, such as the bundle of capillaries (Mualem, 1976; Burdine, 1953) or angular-pores and slits model (Tuller & Or, 2001). These assumptions simplify soil water dynamics models by allowing WRCs and HCFs to be expressed using a shared set of parameters. Several parametric models have been proposed to describe soil hydraulic functions (Brooks & Corey, 1964; van Genuchten, 1980; Durner, 1994; Kosugi, 1996; Tuller & Or, 2001; Assouline, 2006).

The constitutive relationships embody the characteristic features of soil pore network and are the manifestation of the interactions between soil texture and structure. Hence, the reliability of simulated soil water dynamics largely depends on the accuracy of these soil hydraulic functions (Farthing & Ogden, 2017; Zha et al., 2019). Although well-established laboratory methods for characterizing WRCs and HCFs are available, their direct application for field-scale simulations is typically unsatisfactory because of the scale mismatch as well as sampling and measurement artifacts (Hopmans et al., 2002).

Therefore, it is indispensable to estimate WRCs and HCFs using time-series data from field experiments and the inverse solution of the RRE. Commonly, the inverse problem requires finding the parameters of the constitutive relationships that best describe observed time-series data. In principle, it is possible to fit WRCs and HCFs independently, albeit at the expense of significant increase of the tunable parameters. Several studies also employed free-form functions to estimate WRCs and HCFs (Bitterlich et al., 2004; Iden & Durner, 2007). Inverse methods for characterizing soil hydraulic properties often involve the repeated solution of the forward problem, which requires knowledge of the relevant initial and boundary conditions of the RRE. Global optimization algorithm (Durner et al., 2008) and Gaussian processes (Rai & Tripathi, 2019) are other approaches used to find the best-fitted constitutive relationships.

Here, we propose a deep-learning framework for the inverse solution of the time-dependent RRE and the estimation of both WRCs and HCFs, with fewer assumptions and constraints than approaches described above. The method is based on physics-informed neural networks (PINNs) developed by Raissi et al. (2019). PINNs employs the universal approximation capability of neural networks (Cybenko, 1989) to approximate the solution of PDEs. The neural networks' parameters are trained by minimizing the sum of data-fitting error and the residual of the PDEs simultaneously. This simultaneous fitting enables PINNs to learn the dynamics of the system from measurement data and known physics. This novel PINNs approach has shown promising successes in computational physics (Raissi & Karniadakis, 2018; Raissi et al., 2019; Tartakovsky et al., 2020; He et al., 2020). Notably, Tartakovsky et al. (2020) employed PINNs to determine the hydraulic conductivity function of an unsaturated homogeneous soil from synthetic matric potential data based on the two-dimensional time-independent RRE. In this study, we coupled the PINNs framework with two additional monotonic neural networks (Daniels & Velikova, 2010) to describe the known monotonicity of WRCs and HCFs.

Although matric potential is the variable of choice for training purposes, the range and accuracy of matric potential sensors are still limited (Degré et al., 2017). Therefore, the proposed approach uses only volumetric water content time-series data. There are numerous fully developed methods to measure volumetric water content in fields, including the TDR-array probe (Sheng et al., 2017) and the heat-pulse method (Kamai et al., 2008, 2010).

Unlike conventional inverse methods, this proposed approach does not require the repeated solution of the forward problem. Instead, it simultaneously learns (1) the physics of soil water dynamics as defined by the RRE and the monotonicity of the constitutive relationships and (2) the volumetric water content time-series data. The simultaneous learning eliminates the critical shortcomings of conventional inverse approaches, including (1) the need for initial and boundary conditions to solve the forward problems; (2) the dependence of the optimization algorithms on good prior approximations of WRCs and HCFs; and (3) the need to define the shapes of WRCs and HCFs and their interdependence a priori.

In this study, we generated synthetic training data by forward modeling of the RRE using HYDRUS-1D (Šimůnek et al., 2013). Using synthetic data has distinct advantages for testing this novel inverse-solution framework. First, it eliminates the uncertainties of field conditions that equally affect other inverse methods. Second, the synthetic data provide information that is not typically available in routine field measurements, including matric potential and soil water flux density at every location and time.

The robustness of using monotonic neural networks to represent WRCs and HCFs in the PINNs is demonstrated by comparing the results with those from the PINNs that lacks the monotonicity constraints. The performance of the framework was further tested by introducing varying degrees of noise to the synthetic volumetric water content data, altering the spacing between the locations at which volumetric water content data were sampled, using different initial weight parameters of the neural networks. The generalization capability of the framework was investigated by training the PINNs with volumetric water content data for three soils (sandy loam, loam, and silt loam soil) and for two different scenarios of the upper boundary condition. Finally, we show the potential application of the PINNs to estimate soil water flux density using only an array of soil moisture sensors.

2 Background

2.1 Richardson-Richards Equation

We consider one-dimensional liquid water flow in a homogeneous rigid soil and ignore water vapor, sink term, and hysteresis. The mass balance of water in the soil leads to the continuity equation:

$$\frac{\partial \theta}{\partial t} = -\frac{\partial q}{\partial z}, \tag{1}$$

where θ is volumetric water content [L^3 L^{-3}]; t is time [T]; z is vertical coordinate (positive upward) [L]; q is soil water flux density [L T^{-1}]. Soil water flux density q is related to matric potential of water in the soil ψ [L] through the Buckingham-Darcy law (Buckingham, 1907):

$$q = -K\left(\frac{\partial \psi}{\partial z} + 1\right), \tag{2}$$

where K is hydraulic conductivity [L T^{-1}]. The two equations (Equation (1) and (2)) are combined to derive the Richardson-Richards equation (RRE) (Richardson, 1922; Richards, 1931):

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K\left(\frac{\partial \psi}{\partial z} + 1\right)\right]. \tag{3}$$

To solve the RRE, matric potential ψ is commonly treated as the primary variable that is dependent on t and z, and volumetric water content θ and hydraulic conductivity K are parameterized through matric potential ψ, as in

$$\frac{\partial \theta(\psi(t,z))}{\partial t} = \frac{\partial}{\partial z}\left[K(\psi(t,z))\left(\frac{\partial \psi(t,z)}{\partial z} + 1\right)\right]. \tag{4}$$

The functions $\theta(\psi)$ and $K(\psi)$ are called constitutive relationships of the RRE and referred to as water retention curves (WRCs) and hydraulic conductivity functions (HCFs), respectively. WRCs and HCFs are commonly expressed by parametric models (e.g., Brooks and Corey (1964); van Genuchten (1980); Durner (1994); Kosugi (1996); Tuller and Or (2001); Assouline (2006)). The WRCs and HCFs for three types of soil (sandy loam, loam, and silt loam soil) using the Mualem-van Genuchen model (van Genuchten, 1980) are shown in Figure 1. As shown in the figure, both WRCs and HCFs are monotonically increasing functions with respect to matric potential ψ, which is an accepted physical principle of water movement in soils. The monotonicity of WRCs and HCFs will be employed to design the architecture of the neural networks in this study later on.

2.2 Feedforward Neural Networks

A standard fully-connected feedforward neural network with three layers (one hidden layer) is introduced here for readers who are not well versed in the topic. The readers should refer to textbooks (e.g., Goodfellow et al. (2016)) for more general explanations.

Given a training dataset $\{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}$, where superscript (i) denotes the ith training data; $\mathbf{x}^{(i)} \in \mathbb{R}^{n_x}$ is input vector for the size of the input n_x, $\mathbf{y}^{(i)} \in \mathbb{R}^{n_y}$ is output vector for the size of the output n_y, a neural network $\hat{f}$ is a mathematical function mapping the input vector $\mathbf{x}^{(i)}$ to predicted output vector $\hat{\mathbf{y}}^{(i)} \in \mathbb{R}^{n_y}$:

$$\hat{\mathbf{y}}^{(i)} = \hat{f}(\mathbf{x}^{(i)}). \tag{5}$$

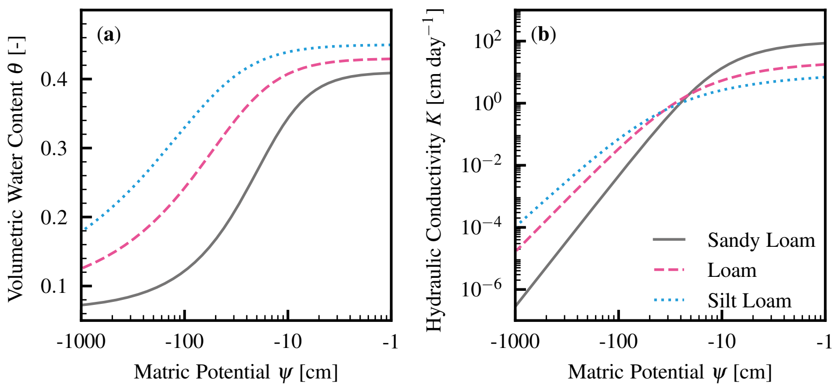

Figure 1. Constitutive relationships for three types of soil (sandy loam, loam, and silt loam soil) generated by using the Mualem-van Genuchten model (van Genuchten, 1980). (**a**) Water retention curves (WRCs). (**b**) Hydraulic conductivity functions (HCFs).

The hat operator represents prediction throughout the paper. The inside of the neural network $\hat{f}$ is commonly represented by layers of units (or neurons), as shown in Figure 2. Herein, $\mathbf{a}^{[L]} \in \mathbb{R}^{n^{[L]}}$ denotes the vector value for the Lth layer of the neural network, where the Lth layer is composed of $n^{[L]}$ units. To calculate the predicted output vector $\hat{\mathbf{y}}^{(i)}$, the input vector $\mathbf{x}^{(i)}$ is entered in the first layer:

$$\mathbf{a}^{[1]} = \mathbf{x}^{(i)}, \tag{6}$$

where the number of units in the first layer $n^{[1]}$ is equal to n_x. Then, the value for the jth unit of the second layer $\mathbf{a}^{[2]}$ is calculated from all the units in the previous layer (i.e., the first layer) with the weight matrix $\mathbf{W}^{[1]}$ and bias vector $\mathbf{b}^{[1]}$ for the first layer in the following way:

$$a_j^{[2]} = g^{[1]}\left(\sum_{k=1}^{n^{[1]}} W_{j,k}^{[1]} a_k^{[1]} + b_j^{[1]}\right), \tag{7}$$

where $g^{[1]}$ is a non-linear activation function for the first layer, such as the hyperbolic tangent function (tanh) shown in Figure 2 (**b**). The jth unit of the third layer is computed from all the units of the second layer (hidden layer):

$$a_j^{[3]} = \sum_{k=1}^{n^{[2]}} W_{j,k}^{[2]} a_k^{[2]} + b_j^{[2]}. \tag{8}$$

Finally, the predicted output vector $\hat{\mathbf{y}}^{(i)}$ is derived from the last layer with an output function h:

$$\hat{y}_j^{(i)} = h(a_j^{[3]}), \tag{9}$$

where the number of the units in the last layer $n^{[3]}$ is equal to n_y. In this study, the sigmoid function (Figure 2 (**c**)) and the exponential function (Figure 2 (**d**)) are used as output functions.

The collection of the weight matrices $\mathbf{W} = \{\mathbf{W}^{[1]}, \mathbf{W}^{[2]}\}$ and bias vectors $\mathbf{b} = \{\mathbf{b}^{[1]}, \mathbf{b}^{[2]}\}$ are the parameters of the neural network, which are estimated by minimizing a loss function comprising of the output vector $\mathbf{y}^{(i)}$ (training data) and the predicted

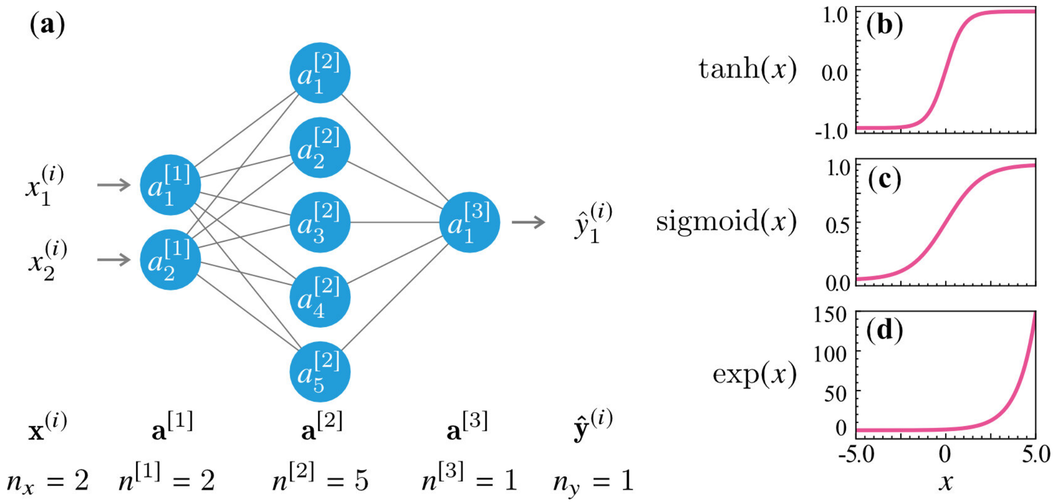

Figure 2. A fully-connected feedforward neural network consisting of three layers (one hidden layer) with activation and output functions. (**a**) A fully-connected feedforward neural network consisting of the input layer with two units, the hidden layer with five units, and the output layer with one unit. (**b**) Hyperbolic tangent function. (**c**) Sigmoid function. (**d**) Exponential function.

output vector $\hat{\mathbf{y}}^{(i)}$. The definition of the loss function varies depending on the purpose of the training, and the loss function used in this study is defined in Equation (14).

It is well known that a feedforward neural network with more hidden layers has a better capability of function approximation (Goodfellow et al., 2016), and such a neural network with more than two hidden layers is called a deep neural network. In such a case, a unit of a hidden layer is computed from all the units of the previous hidden layer in the same way explained above (Equation (7)).

In the next section, three fully-connected feedforward neural networks are combined to construct physics-informed neural networks (PINNs) for the RRE, and the loss function for the PINNs framework is defined to estimate WRCs and HCFs from volumetric water content measurements.

3 Methods

3.1 Physics-Informed Neural Networks with Monotonicity Constraints for RRE

Physics-informed neural networks (PINNs) has been proposed as a deep learning framework to derive the forward and inverse solution of PDEs (Raissi et al., 2019). In this study, PINNs was used to derive the inverse solution of the RRE and the constitutive relationships (i.e., WRCs and HCFs) from a set of volumetric water content time-series data measured at different depths in soils $\{t^{(i)}, z^{(i)}, \theta^{(i)}\}_{i=1}^{i=N}$, where N is the number of measurement data.

PINNs for the RRE was constructed using three fully-connected feedforward neural networks, as shown in Figure 3. The neural network $\hat{f}_\psi$ (Figure 3 (**a**)) is a function mapping from time t and vertical coordinate z into predicted matric potential $\hat{\psi}$:

$$\hat{\psi}^{(i)} = \hat{f}_\psi(t^{(i)}, z^{(i)}; \mathbf{W}_\psi, \mathbf{b}_\psi), \tag{10}$$

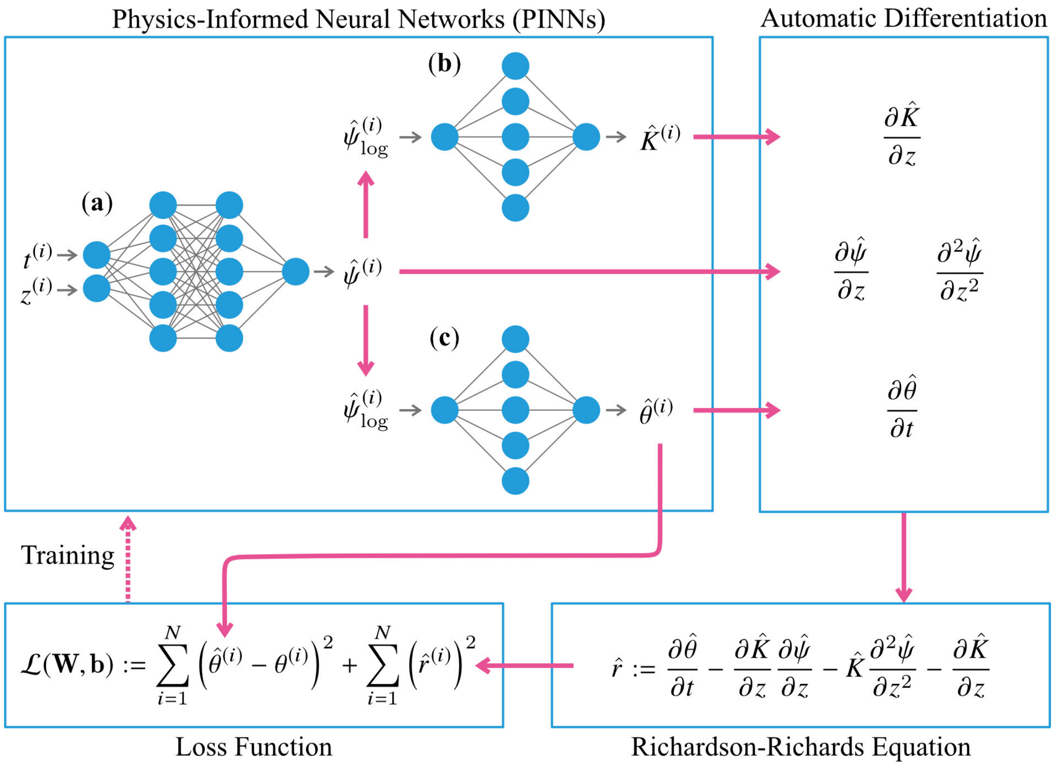

Figure 3. Physics-informed neural networks (PINNs) for the Richardson-Richards equation consisting of three fully-connected feedforward neural networks to predict (**a**) matric potential $\hat{\psi}$, (**b**) hydraulic conductivity $\hat{K}$, and (**c**) volumetric water content $\hat{\theta}$. The number of layers and units in the figure is not actual.

where $\mathbf{W}_\psi$ and $\mathbf{b}_\psi$ are the collection of weight and bias parameters in the neural network. The hyperbolic tangent function (Figure 2 (**b**)) is used for the activation function, as recommended by Raissi et al. (2019). As for the output function, the negative exponential function (i.e., $-\exp(x)$, see Figure 2 (**d**)) is used to force the predicted matric potential to be negative.

The predicted matric potential $\hat{\psi}^{(i)}$ is used to estimate volumetric water content $\hat{\theta}^{(i)}$ and hydraulic conductivity $\hat{K}^{(i)}$ through two distinct neural networks $\hat{f}_\theta$, $\hat{f}_K$ (Figure 3 (**c**) and (**b**), respectively). In other words, the two neural networks are used to represent the WRC and HCF for a given soil. Since WRCs and HCFs become simpler if matric potential is plotted in logarithmic scale, as in Figure 1, the predicted matric potential is converted into logarithmic scale by the following transformation:

$$\hat{\psi}_{\log}^{(i)} = -\log_e(-\hat{\psi}^{(i)}). \tag{11}$$

Then, the predicted matric potential in logarithmic scale $\hat{\psi}_{\log}^{(i)}$ is used as the input value for the two neural networks to represent WRCs and HCFs:

$$\hat{\theta}^{(i)} = \hat{f}_\theta(\hat{\psi}_{\log}^{(i)}; \mathbf{W}_\theta, \mathbf{b}_\theta), \tag{12}$$

$$\hat{K}^{(i)} = \hat{f}_K(\hat{\psi}_{\log}^{(i)}; \mathbf{W}_K, \mathbf{b}_K). \tag{13}$$

The tanh function is used as the activation function for both neural networks. The output functions for $\hat{f}_\theta$ and $\hat{f}_K$ are the sigmoid function and the exponential function, respectively to ensure predicted volumetric water content between 0 and 1 and positive predicted hydraulic conductivity (see Figure 2 (**c**) and (**d**)).

To embrace the monotonicity of WRCs and HCFs, the weight parameters $\mathbf{W}_\psi$ and $\mathbf{W}_K$ are constrained to be non-negative so that $\hat{f}_\theta$ and $\hat{f}_K$ are monotonically increasing functions with respect to the predicted matric potential $\hat{\psi}$ (Daniels & Velikova, 2010). This type of neural networks is called (totally) monotonic neural networks, where the output values depend monotonically on all the variables in the input vector. It is known that a three-layer fully-connected feedforward neural network with non-negative weights can arbitrarily approximate any monotonic scalar functions (Daniels & Velikova, 2010). Readers interested in monotonic neural networks should refer to Daniels and Velikova (2010), where various types of monotonic neural networks are explained.

Incorporating monotonicity constraints in the neural networks representing WRCs and HCFs honors the physical nature of the movement of water in all soils. This approach is similar to the free-form approach (Bitterlich et al., 2004; Iden & Durner, 2007), where cubic Hermite interpolation was used to approximate WRCs and HCFs. Unlike their studies, our monotonic neural network approach does not assume predetermined saturated water content and saturated hydraulic conductivity because they are not easily available in field applications.

The collection of the parameters in the three neural networks $\mathbf{W} = \{\mathbf{W}_\psi, \mathbf{W}_\theta, \mathbf{W}_K\}$ and $\mathbf{b} = \{\mathbf{b}_\psi, \mathbf{b}_\theta, \mathbf{b}_K\}$ are identified by minimizing a loss function defined as

$$\mathcal{L}(\mathbf{W}, \mathbf{b}) := \sum_{i=1}^{N} (\hat{\theta}^{(i)} - \theta^{(i)})^2 + \sum_{i=1}^{N} (\hat{r}^{(i)})^2, \tag{14}$$

where $\hat{r}$ is the residual of the RRE defined as

$$\hat{r} := \frac{\partial \hat{\theta}}{\partial t} - \frac{\partial}{\partial z}\left[\hat{K}\left(\frac{\partial \hat{\psi}}{\partial z} + 1\right)\right] = \frac{\partial \hat{\theta}}{\partial t} - \frac{\partial \hat{K}}{\partial z}\frac{\partial \hat{\psi}}{\partial z} - \hat{K}\frac{\partial^2 \hat{\psi}}{\partial z^2} - \frac{\partial \hat{K}}{\partial z}. \tag{15}$$

The first term of the loss function (Equation (14)) represents the fitting error of volumetric water content, and the second term represents the contraint by the RRE. This simultaneous learning enables the PINNs to learn the dynamics of water in soils from both volumetric water content data and knowledge in soil physics (i.e., the RRE). In the other studies on PINNs (e.g. Raissi et al. (2019); Tartakovsky et al. (2020); He et al. (2020)), the boundary and initial conditions of PDEs are also included in the loss function. However, we omitted these terms because they are difficult to obtain in real applications.

To calculate the residual of the RRE $\hat{r}$ at all the data points, the derivatives (i.e., $\frac{\partial \hat{\theta}}{\partial t}$, $\frac{\partial \hat{\psi}}{\partial z}$, $\frac{\partial^2 \hat{\psi}}{\partial z^2}$, $\frac{\partial \hat{K}}{\partial z}$) are evaluated at the data points by using automatic differentiation (Nocedal & Wright, 2006). It should be noted that the residual of the RRE $\hat{r}$ can be evaluated at any point in the domain (called collocation points). However, we forced the collocation points to be the same as the measurement locations.

Before training the PINNs, the weight parameters $\mathbf{W}$ are initialized through Xavier initialization (Glorot & Bengio, 2010), and the bias parameters $\mathbf{b}$ are all set to zero. Then, these parameters $\mathbf{W}$ and $\mathbf{b}$ are trained by minimizing the loss function:

$$\min_{\mathbf{W}, \mathbf{b}} \mathcal{L}(\mathbf{W}, \mathbf{b}). \tag{16}$$

The optimization problem was solved by the Adam algorithm (Kingma & Ba, 2014) followed by the L-BFGS-B algorithm (Byrd et al., 1995). This two-step training procedure has been reported to be effective to train PINNs (Raissi et al., 2019; He et al., 2020). In our implementation, the default settings of the Adam optimizer in TensorFlow (Abadi et al., 2015) was used until 300,000 iterations finished. Then, the L-BFGS-B optimizer from Scipy (Virtanen et al., 2020) with $maxcor = 50$, $maxls = 50$, $maxiter = 50,000$,

Table 1. Two scenarios of surface water flux density [cm day^{-1}] (positive upward) were applied to generate synthetic data using HYDRUS-1D (Šimůnek et al., 2013).

Time (day)	Scenario 1	Scenario 2
0.25	-10	-10
0.50	0	0
1.0	0.3	0.3
1.5	0	-5
2.0	0.3	0.3
2.25	-10	-5
2.5	0	-5
3.0	0.3	0.3

$maxfun = 50,000$, $ftol = 2.220446049250313 \times 10^{-16}$, and the default values for the other parameters was applied to achieve the convergence of the loss function. The investigation on the hyperparameters of those optimization algorithms is beyond the scope of the paper. This PINNs framework for the RRE was implemented through TensorFlow 1.14 (Abadi et al., 2015), and the source code is available on https://github.com/ToshiyukiBandai/PINNs_RRE.

3.2 Synthetic Data Generated by HYDRUS-1D

To develop and assess the PINNs framework for the RRE, synthetic soil moisture data were generated by using HYDRUS-1D (Šimůnek et al., 2013). The synthetic data was used for two purposes: (1) to determine the architecture of the neural networks (i.e., the number of hidden layers and units; Section 3.3) (Section 3.3); (2) to investigate the the generalization capability of the PINNs (Section 3.4).

In the HYDRUS-1D simulation, soil moisture dynamics for three days in the 100 cm of homogeneous three soils with different textures (sandy loam, loam, and silt loam soil) were simulated. The soil column was uniformly discretized at a 0.1 cm interval. The initial matric potential was set at -1000 cm for all the depths. The bottom boundary condition was the Neumann boundary condition:

$$\frac{\partial \psi}{\partial z} = 0. \tag{17}$$

The upper boundary was set to the atmospheric upper boundary condition, where two different scenarios of time-dependent surface flux density were applied (see Table 1).

The Mualem-van Genuchen model was used to parameterize WRCs and HCFs in the HYDRUS-1D simulation (van Genuchten, 1980):

$$\theta(\psi) = \theta_r + \frac{\theta_s - \theta_r}{(1 + (-\alpha\psi)^n)^m}, \tag{18}$$

$$K(\theta(\psi)) = K_s S_e^l (1 - (1 - S_e^{1/m})^m)^2, \tag{19}$$

where θ_r, θ_s, α, n, K_s, and l are the Mualem-van Genuchen fitting parameters; $m = 1 - 1/n$; and the effective saturation S_e is defined as

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r}. \tag{20}$$

The Mualem-van Genuchen fitting parameters for the three soils used in this study are summarized in Table 2.

Table 2. The Mualem-van Genuchen fitting parameters for three types of soils (van Genuchten, 1980).

Parameters	Sandy Loam	Loam	Silt Loam
θ_r [cm^3 cm^{-3}]	0.065	0.078	0.067
θ_s [cm^3 cm^{-3}]	0.41	0.43	0.45
α [cm^{-1}]	0.075	0.036	0.02
n [-]	1.89	1.56	1.41
K_s [cm day^{-1}]	106.1	24.96	10.8
l [-]	0.5	0.5	0.5

3.3 Determination of Architecture of Neural Networks

It is known that the architecture of feedforward neural networks (i.e., the number of hidden layers and units) influences their performance. Therefore, the number of hidden layers and units for the three neural networks in the PINNs was determined empirically in two steps.

First, we set the number of hidden layers and units of the two neural networks, $\hat{f}_\theta$ for volumetric water content (Figure 3 (**c**)) and $\hat{f}_K$ for hydraulic conductivity (Figure 3 (**b**)), to 1 hidden layer with 20 units and varied the number of hidden layers and units of the neural network for the predicted matric potential $\hat{f}_\psi$ (Figure 3 (**a**)). Seven different numbers of hidden layers (2, 4, 6, 8, 9, 10, 11) and three different numbers of units (10, 20, 40) were tested.

Second, the number of hidden layers and units of the other two neural networks, $\hat{f}_\theta$ and $\hat{f}_K$, was varied. Three different numbers of layers (1, 2, 3) and units (10, 20, 40) were tested for each neural network.

To determine the architecture of the neural networks in the PINNs, the synthetic data for sandy loam soil for Scenario 1 were used (see Section 3.2). As training data, volumetric water content was sampled every 0.012 day (i.e., 251 data points for a depth) at 10 equally spaced different depths within the top of the 20 cm of the soil column (z = -1, -3, -5, -7, -9, -11, -13, -15, -17, -19 cm) because our study is focused on soil moisture dynamics in near-surface soils.

To evaluate the performance of the PINNs, we compared the predicted and true volumetric water content, matric potential, hydraulic conductivity, and soil water flux density. The predicted soil water flux density $\hat{q}$ was derived using the Buckingham-Darcy law (Equation (2)) with the estimated hydraulic conductivity $\hat{K}$ and the gradient of the predicted matric potential $\partial\hat{\psi}/\partial z$. We quantified the prediction error over the time $t \in [0,3]$ day with an interval of 0.012 days and the spatial domain $z \in (-20,0]$ cm with an interval of 0.1 cm for all the four variables in terms of the relative L_2 errors ϵ^γ for $\gamma = \theta, \psi, K, q.$, defined as

$$\epsilon^\gamma := \frac{\sum_{t\in[0,3]}\sum_{z\in(-20,0]}(\hat{\gamma}(t,z)-\gamma(t,z))^2}{\sum_{t\in[0,3]}\sum_{z\in(-20,0]}\gamma(t,z)^2} \tag{21}$$

To demonstrate the effectiveness of including monotonic neural networks in the PINNs, we also trained the PINNs without monotonicity constraints (i.e., standard feedforward neural networks are used to represent WRCs and HCFs) with the same training data.

The architecture of the three neural networks in the PINNs without monotonicity was also determined in the same way as above.

Because the results of training PINNs were affected by the initial values of the weight parameters of the neural networks determined by Xavier initialization (Glorot & Bengio, 2010), three different random seeds were used in the code, and three replicates were obtained for each of those combinations of the number of hidden layers and units. As a result, 63 trainings for $\hat{f}_\psi$ (Figure 3 (**a**)) and 243 ones for $\hat{f}_\theta$ (Figure 3 (**c**)) and $\hat{f}_K$ (Figure 3 (**b**)) were conducted to determined their architecture for the PINNs both with and without monotonicity.

3.4 Application of PINNs to Various Datasets

Different types of data were prepared by using HYDRUS-1D to assess the performance of the PINNs with and without monotonicity constraints. First, we investigated the effect of noise in the training data. To this end, Gaussian noise with the mean of zero and four different values of standard deviation (0, 0.005, 0.01, 0.02) was added to the sampled volumetric water content for sandy loam soil for Scenario 1.

Next, the effect of the sparsity of the training data was studied by using volumetric water content data for sandy loam soil for Scenario 1 without adding noise. We considered three cases for the number of depths at which volumetric water content were sampled: 10 ($z = -1, -3, -5, -7, -9, -11, -13, -15, -17, -19$ cm), 5 ($z = -1, -5, -9, -13, -17$ cm) and 3 ($z = -1, -9, -17$ cm).

Lastly, volumetric water content data for three different types of soils (sandy loam, loam, and silt loam soil) with the two different scenarios of upper boundary condition (see Table 1) were generated. Gaussian noise with the mean of zero and the standard deviation of 0.005 was added to the synthetic data to reflect measurement noise encountered in field applications.

Those training data were applied to the PINNs with and without monotonicity constraints, and the results were evaluated in terms of relative errors defined in Equation (21). For all the cases above, five different random seeds were set in the code to investigate the effects of neural network initialization on the results.

4 Results and Discussions

4.1 Architecture of Neural Networks in PINNs

To determine the number of hidden layers and units of the three neural networks in the PINNs with and without monotonicity, various combinations of layers and units were tested. Figure 4 shows relative error ϵ defined in Equation (21) for volumetric water content θ, matric potential ψ, hydraulic conductivity K, and soil water flux density q for different numbers of hidden layers and units for the neural network $\hat{f}_\psi$ (Figure 3 (**a**)) of the PINNs with and without monotonicity while the architecture of the other two neural networks are fixed (1 hidden layer with 20 units). For the PINNs with monotonic neural networks (left column), relative error for volumetric water content ϵ^θ, hydraulic conductivity ϵ^K, and soil water flux density ϵ^q decreased with the increase in number of units, with 40 units resulting in the lowest error (the Pearson correlation coefficient is provided in Table S1 in the supplementary information.).

The lowest arithmetic mean of relative error was observed when the number of hidden layers is 4 for volumetric water content θ, 6 for hydraulic conductivity K, and 8 for soil water flux density q when the number of units is 40. Clear trends were not obtained for relative error for matric potential ϵ_ψ. Because relative error for soil water flux density q reflects the predictive accuracy of the PINNs for both matric potential ψ and hydraulic conductivity K fields, we set the neural network for the predicted matric potential $\hat{f}_\psi$ to 8 layers with 40 units.

For the PINNs without monotonicity constraints (right column), the architecture of the neural network for the predicted matric potential $\hat{f}_\psi$ was set to 6 hidden layers with 40 units, which coincides with the lowest relative error of soil water flux density ϵ_q. We observed a non-linear correlation between the number of hidden layers and relative error for volumetric water content θ, hydraulic conductivity K, and soil water flux density q; relative error reached the lowest when the number of hidden layers was 6 and increased again. This is clear evidence that the PINNs without monotonicity was overfitting the training data. On the other hand, such a non-linear behavior was minimized for the PINNs with monotonicity constraints, which means imposing monotonicity can prevent the PINNs from overfitting the training data. In addition, the variability of relative errors between different initializations of the neural networks was lower for the PINNs with monotonic neural networks than the PINNs with non-monotonic neural networks. This further demonstrates the benefit of the monotonicity constraints in improving the stability and reliability of the training.

After determining the architecture of the neural network for the predicted matric potential $\hat{f}_\psi$, the number of hidden layers and units for the other two neural networks, $\hat{f}_\theta$ and $\hat{f}_K$, was varied. We did not observe clear trends of relative error for different neural network architectures for the PINNs with and without monotonicity (see Table S1 and Figure S1 in the supplementary information). However, the performance of the PINNs without monotonicity constraints was much more sensitive to the neural network architecture. This implied that incorporating monotonicity constraints stabilized the training, which enabled us to determine the neural network structure easier than the PINNs without monotonicity constraints. As a result, the architecture of the two neural networks was set as follows: 1 hidden layer with 40 units for the PINNs with monotonicity and 3 hidden layers with 40 units for without monotonicity for the neural network for the predicted volumetric water content $\hat{f}_\theta$; 3 hidden layer with 40 units for PINNs with monotonicity and 2 hidden layers with 20 units for without monotonicity for the neural network for the predicted hydraulic conductivity $\hat{f}_K$.

4.2 Effect of Noise and Sparsity of Training Data

To investigate the effect of measurement noise on the performance of the PINNs, Gaussian noise with mean of zero and different values of standard deviation (0, 0.005, 0.01, 0.02) was added to the synthetic volumetric water content data (see Section 3.4), which was used to train the PINNs with and without monotonicity constraints. Figure 5 (**a**) shows relative error for soil water flux density ϵ^q for different values of noise added to the true volumetric water content data. For the PINNs with and without monotonicity constraints, relative error increased with the standard deviation of noise, although the effect of the noise was substantially lower for the PINNs with monotonicity. On the other hand, the PINNs without monotonicity constraints exhibited consistently large relative error for all levels of noise. These observations indicate that monotonicity constraints are critical for ensuring stability and reliability when fitting noisy data. Therefore, PINNs without monotonic neural networks is not practically feasible for field applications.

The number of measurement locations at which simulated volumetric water content was sampled data were varied from 10 to 5 and 3 to investigate the effect of the sparsity of the training data. Figure 5 (**b**) illustrates that smaller relative error for soil water flux density ϵ^q was observed for denser training data. Although PINNs have been shown to be effective for sparse training data (Raissi et al., 2019; Tartakovsky et al., 2020), the PINNs for this application needs dense volumetric water content measurements (e.g., 2 cm interval).

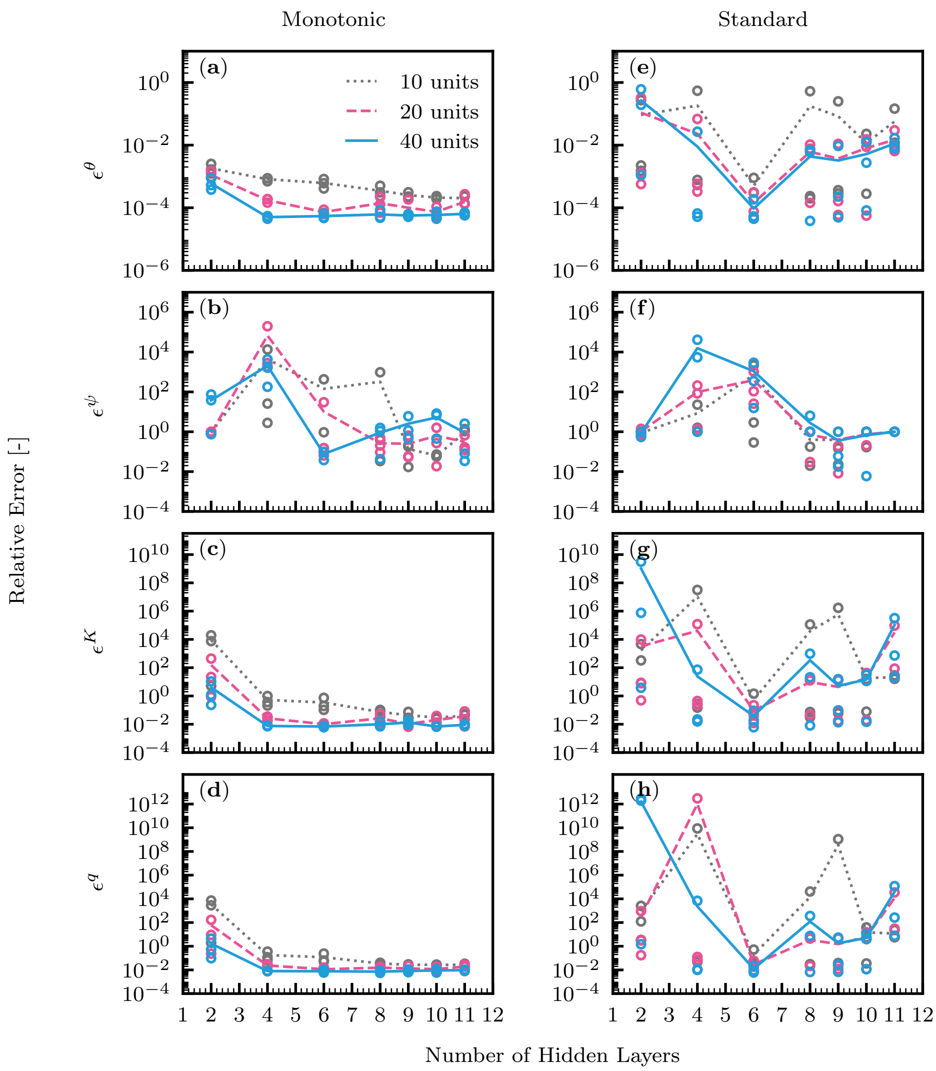

Figure 4. Relative error ϵ for volumetric water content θ, matric potential ψ, hydraulic conductivity K, and soil water flux density q for different numbers of hidden layers and units in the neural network for the predicted matric potential $\hat{f}_\psi$ (Figure 3 (**a**)); with (left column) and without monotonicity (right column). The architecture of the other two neural networks are set to 1 hidden layer with 20 units each. The lines represent the arithmetic mean of five replicates for each neural network architecture.

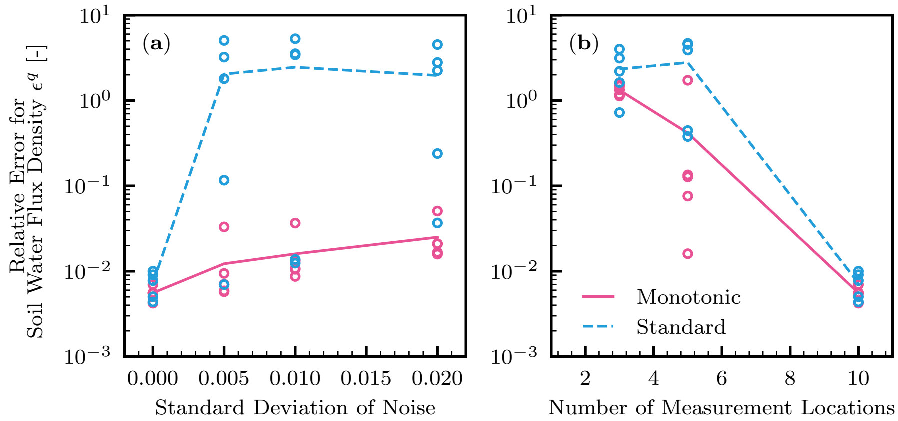

Figure 5. Relative error for soil water flux density ϵ^q for different values of standard deviation of noise (**a**) and measurement locations at which synthetic volumetric water content data were sampled (**b**). The number of measurement locations was varied from 10 (z = -1, -3, -5, -7, -9, -11, -13, -15, -17, -19 cm) to 5 ($z = -1, -5, -9, -13, -17$ cm) and 3 ($z = -1, -9, -17$ cm). The lines represent the arithmetic mean of five replicates.

.

4.3 Generalization Capability of PINNs

The generalization capability of the PINNs with and without monotonicity constraints was assessed with noisy synthetic volumetric water content data generated by HYDRUS-1D for three types of soils (sandy loam, loam, silt loam soil) with two different scenarios of upper boundary conditions (see Table 1). Table 3 shows relative error for volumetric water content ϵ^θ, matric potential ϵ^ψ, hydraulic conductivity ϵ^K, and soil water flux density ϵ^q. The PINNs without monotonicity constraints could not produce satisfactory results, which is shown by the large values of relative error for hydraulic conductivity ϵ^K and soil water flux density ϵ^q for both scenarios. This is mainly caused by the noise in the training data, which was indicated in Figure 5 (**a**). Also, poor generalization capability of the PINNs without monotonicity constraints is implied by the fact that higher relative error was observed for loam and silt loam soil. Therefore, in the following sections, we focus on the results of the PINNs with monotonicity constraints. While the trainings were conducted with five different random seeds initializing the weight parameters of the neural networks, we provide the results that show medium performance in terms of relative error for soil water flux density ϵ^q.

4.3.1 Volumetric water content

Figure 6 shows predicted volumetric water content by the PINNs with monotonicity constraints from noisy training data for sandy loam soil for the two scenarios. The PINNs could precisely capture the true distribution of soil moisture from the training data with the noise (standard deviation of 0.005). The PINNs could capture the distribution well for the other two soils as well (shown in Figure S2 and S3 in the supporting information).

Larger errors were observed when the upper boundary condition changed abruptly (e.g., $t = 1.5$ day for Scenario 2 in Figure 6 (**e**)). This indicated that the neural networks used in the study could not represent such a sharp change in soil moisture dynam-

Table 3. Relative error (arithmtic mean (± standard devation)) for volumetric water content ϵ^{θ}, matric potential ϵ^{ψ}, hydraulic conductivity ϵ^{K}, and soil water flux density ϵ^{q} for the PINNs with and without monotonicity constraints trained by noisy volumetric water content data for three soils (sandy loam, loam, silt loam soil) for two scenarios (Scenario 1 and 2). The arithmetic mean and standard deviation were calculated from five replicates.

Relative Error	Sandy Loam	Loam	Silt Loam
		PINNs with monotonicity constraints	
		Scenario 1	
ϵ^{θ}	$1.05(\pm 0.75) \times 10^{-4}$	$4.02(\pm 0.26) \times 10^{-5}$	$3.60(\pm 0.48) \times 10^{-5}$
ϵ^{ψ}	$4.21(\pm 0.38) \times 10^{-1}$	$6.79(\pm 5.99) \times 10^{2}$	$1.14(\pm 1.17) \times 10^{1}$
ϵ^{K}	$3.01(\pm 4.78) \times 10^{-2}$	$3.34(\pm 0.63) \times 10^{-2}$	$2.87(\pm 0.23) \times 10^{-1}$
ϵ^{q}	$1.22(\pm 1.05) \times 10^{-2}$	$1.55(\pm 0.32) \times 10^{-2}$	$2.27(\pm 0.25) \times 10^{-2}$
		Scenario 2	
ϵ^{θ}	$4.89(\pm 0.34) \times 10^{-5}$	$3.03(\pm 0.30) \times 10^{-5}$	$3.66(\pm 2.51) \times 10^{-5}$
ϵ^{ψ}	$4.19(\pm 0.43) \times 10^{-1}$	$9.42(\pm 18.0) \times 10^{-1}$	$1.17(\pm 1.09)$
ϵ^{K}	$5.33(\pm 0.70) \times 10^{-3}$	$2.47(\pm 0.74) \times 10^{-2}$	$5.18(\pm 3.98) \times 10^{-1}$
ϵ^{q}	$5.48(\pm 0.53) \times 10^{-3}$	$1.01(\pm 0.09) \times 10^{-2}$	$3.49(\pm 2.55) \times 10^{-2}$
		PINNs without monotonicity constraints	
		Scenario 1	
ϵ^{θ}	$2.38(\pm 2.27) \times 10^{-3}$	$8.38(\pm 9.01) \times 10^{-4}$	$7.25(\pm 5.80) \times 10^{-4}$
ϵ^{ψ}	$1.13(\pm 0.58)$	$1.19(\pm 2.14) \times 10^{1}$	$4.46(\pm 7.14)$
ϵ^{K}	$5.98(\pm 5.70)$	$1.08(\pm 1.33) \times 10^{5}$	$1.54(\pm 1.36) \times 10^{5}$
ϵ^{q}	$2.04(\pm 1.92)$	$1.30(\pm 1.61) \times 10^{4}$	$1.15(\pm 1.02) \times 10^{4}$
		Scenario 2	
ϵ^{θ}	$1.50(\pm 1.23) \times 10^{-3}$	$3.13(\pm 3.11) \times 10^{-4}$	$3.19(\pm 2.50) \times 10^{-4}$
ϵ^{ψ}	$2.76(\pm 3.75)$	$3.62(\pm 5.26)$	$2.30(\pm 2.74)$
ϵ^{K}	$2.02(\pm 1.74)$	$1.11(\pm 2.05) \times 10^{4}$	$5.95(\pm 6.51) \times 10^{4}$
ϵ^{q}	$9.69(\pm 8.30) \times 10^{-1}$	$2.32(\pm 4.30) \times 10^{3}$	$7.84(\pm 8.66) \times 10^{3}$

ics. For the same reason, larger errors were observed just after the initial condition ($t = 0$ day). Also, the PINNs could not reproduce the true volumetric water content at depths that are not covered in the training data (i.e., near the surface and lower than $z = -19$ cm). This means the PINNs could not extrapolate the volumetric water content data while it could interpolate. Similar trends were observed for the other two soils (see Figure S2 and S3 in the information).

4.3.2 Residual of RRE

The PINNs minimizes the data fitting error, as well as the residual of the RRE defined by Equation (15). The absolute value of the residual of the RRE for sandy loam soil at three times for the two scenarios is shown in Figure 7. The values in the spatial domain were small (less than 10^{-3}), which means the RRE was satisfied in the spatial domain of interest (i.e., (-20cm, 0cm]). Larger deviations from zero were observed near the surface and lower than the lowest virtual sensor ($z = -19cm$). This corresponds to the fact that the collocation points at which the residual of the RRE is evaluated were set to the measurement locations. This error may be minimized by distributing more collocation points in the spatial domain, including near the surface. Tartakovsky et al. (2020) reported that the accuracy of the PINNs improved if larger numbers of collocation points were provided. The drawback of increasing the number of collocation points is increased in computational demand. Further investigations are needed for seeking an efficient strategy to distribute the collocation points to achieve a better performance of the PINNs. The results for the other soils are provided in Figure S4 and S5 in the supporting information.

4.3.3 Water Retention Curves

Predicting matric potential from the noisy volumetric water content corresponds to estimating WRCs, which is one of the primary goals of the study. The PINNs with monotonicity constraints could not precisely predict the WRCs for the three soils, as shown in Figure 8. Especially, the prediction was not satisfactory for low and high volumetric water content, where the training data points were not provided. This suggests the difficulty in representing the two characteristics of WRCs by using a monotonic neural network: monotonicity and well defined upper and lower limits (saturation and dryness, respectively). This weakness of the current PINNs needs to be fixed in future research. Nevertheless, the predicted WRCs were surprisingly similar to the true WRCs in the middle range regardless of the fact that any actual value of matric potential was not used to train the PINNs.

How does the PINNs with monotonicity constraints learn WRCs from only volumetric water content data? A possible explanation is that matric potential is estimated from the gradient of matric potential $\partial\hat{\psi}/\partial z$, which is calculated in the residual of the RRE $\hat{r}$. Also, a matric potential of zero at saturation is implied by forcing matric potential to be negative while imposing the monotonically increasing relationship between matric potential and volumetric water content. These two explanations partly support the possibility that the PINNs with monotonicity constraints can predict WRCs from only volumetric water content if sufficient numbers and quality of training data are given.

4.3.4 Hydraulic Conductivity Functions

The estimated HCFs for the three soils for the two scenarios are shown in Figure 9. It should be noted that hydraulic conductivity is plotted against volumetric water content, not matric potential, as in Figure 1, because the estimated values of matric potential do not match the actual values, unlike volumetric water content.

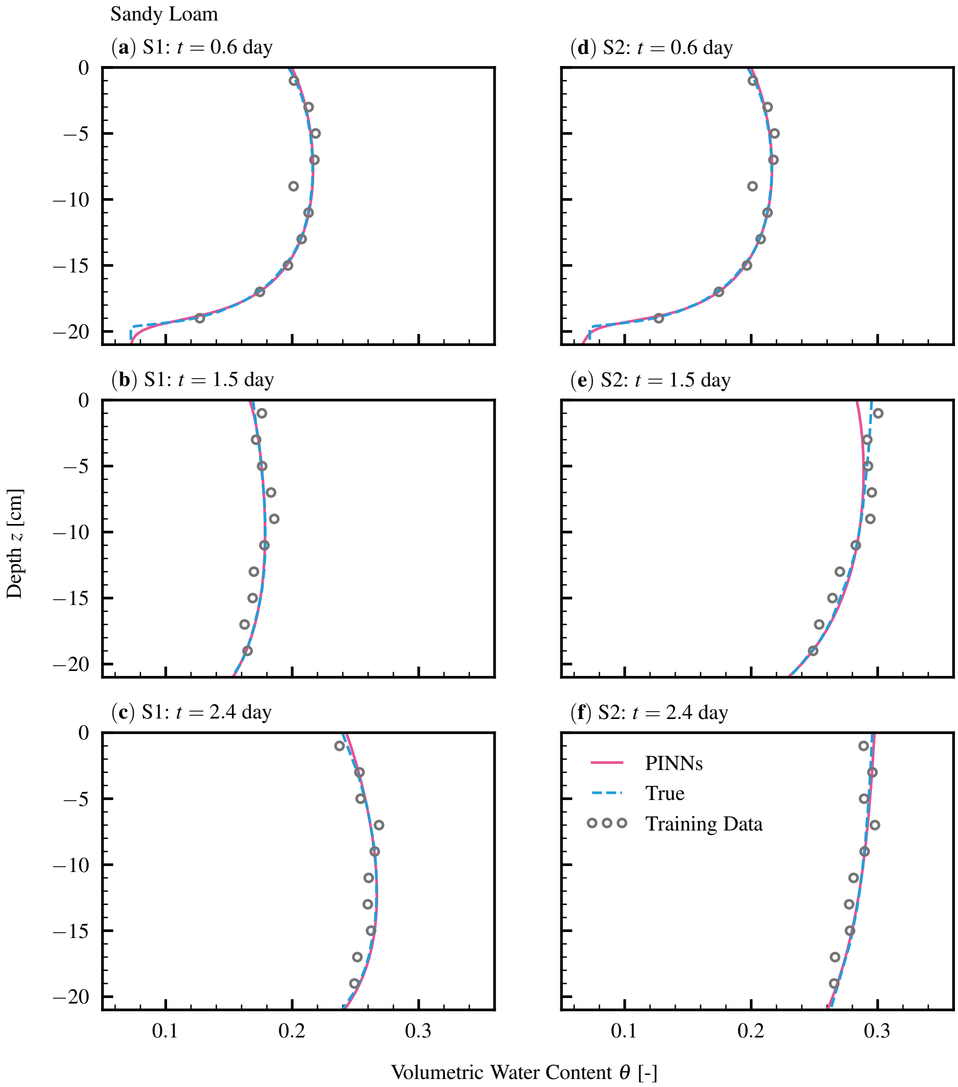

Figure 6. Predicted volumetric water content (PINNs) and noisy synthetic training data (Training Data) for sandy loam soil for the two scenarios at three different times. The dotted lines represent the synthetic data before adding the noise (True). Scenario 1 (S1): (**a**) $t = 0.6$ day, (**b**) $t = 1.5$ day, and (**c**) $t = 2.4$ day. Scenario 2 (S2): (**d**) $t = 0.6$ day, (**e**) $t = 1.5$, and (**f**) $t = 2.4$ day.

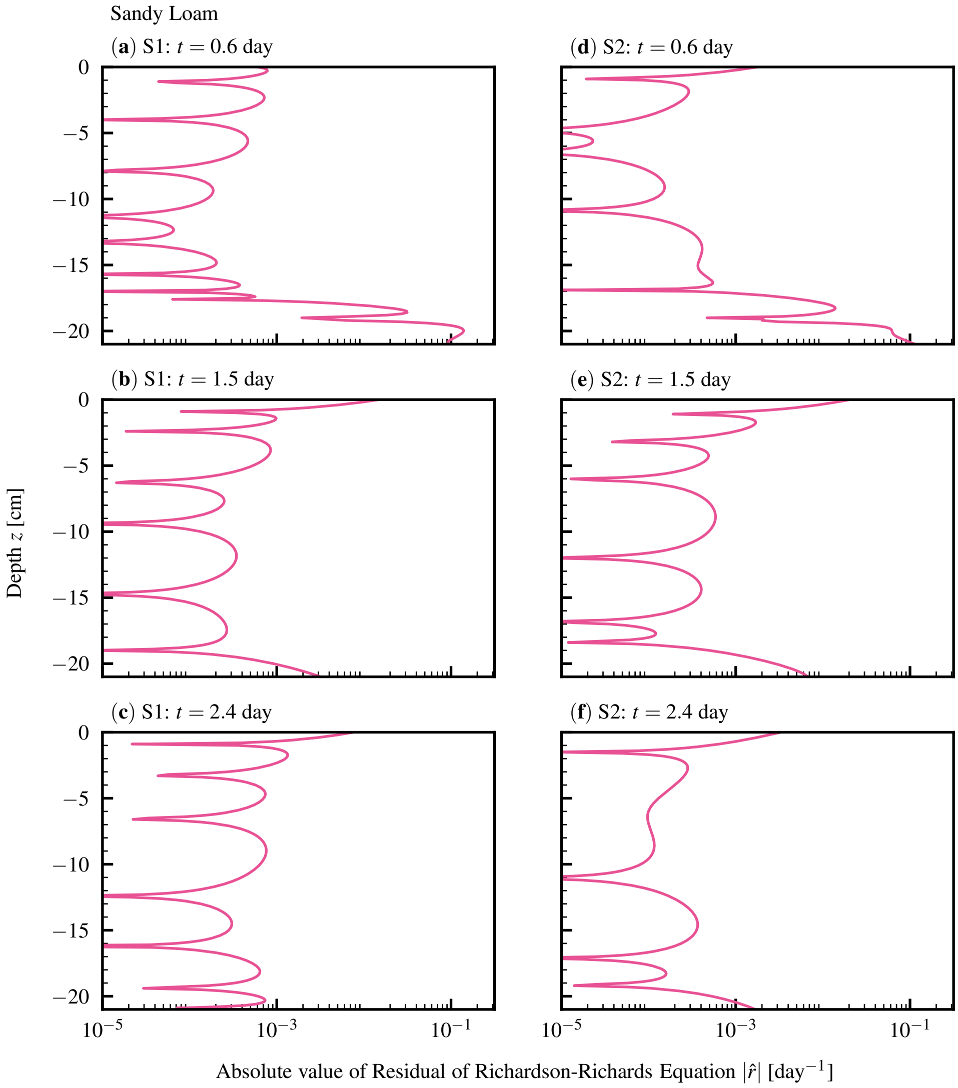

Figure 7. The absolute value of the residual of the Richardson-Richards equation at three different times for sandy loam soil for the two scenarios. Scenario 1 (S1): (**a**) $t = 0.6$ day, (**b**) $t = 1.5$ day, and (**c**) $t = 2.4$ day. Scenario 2 (S2): (**d**) $t = 0.6$ day, (**e**) $t = 1.5$, and (**f**) $t = 2.4$ day.

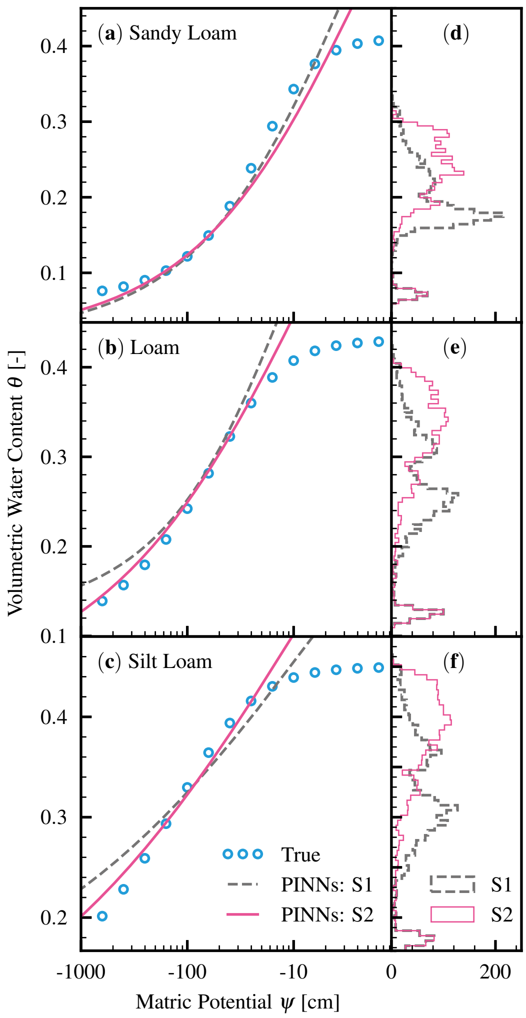

Figure 8. Comparison of true water retention curves (True) to the ones predicted by the PINNs with monotonicity constraints for the three soils for the two scenarios (S1: Scenario 1, S2: Scenario 2) with the histogram of the noisy training data. Water retention curve for (**a**) sandy loam, (**b**) loam, and (**c**) silt loam. Histogram of the training data for (**d**) sandy loam, (**e**) loam, and (**f**) silt loam.

The PINNs with monotonicity constraints could estimate the HCFs, especially for the range of the volumetric water content that is covered in the training data. On the other hand, the PINNs could not precisely extrapolate the HCFs; dryness and near saturation. As for a drier range of HCFs, although some of the training data are distributed in the range, they did not contribute to the learning of the HCFs. This is caused by the fact that these data correspond to the initial volumetric content, which increased rapidly due to the prescribed upper boundary conditions, and the PINNs could not capture the abrupt change well.

Hydraulic conductivity was estimated through minimizing the residual of the RRE, which contains hydraulic conductivity (see Equation (15)). Tartakovsky et al. (2020) reported that HCFs could be estimated from matric potential measurements using PINNs with the time-independent RRE. Considering our result and their findings, we conclude that hydraulic conductivity can be estimated from only either volumetric water content or matric potential.

The advantage of the PINNs approach over the other studies to estimate HCFs was that we did not assume any information about HCFs a priori, such as saturated water content and saturated hydraulic conductivity. Also, the neural network for HCFs is separated from WRCs, which prevents the error in WRCs from propagating into HCFs. Considering these advantages, we conclude that the current framework of PINNs for the RRE is a powerful way to estimate HCFs from only volumetric water content data, which has never been attained to the best of our knowledge.

4.3.5 Soil Water Flux Density

In this section, we will show that the current PINNs framework with monotonicity constraints can be used to estimate soil water flux density from noisy volumetric water content data.

The comparison of the estimated soil water flux density to the true one calculated by HYDRUS-1D at three different depths ($z = -1, -9, -17$ cm) for sandy loam soil for the two scenarios is shown in Figure 10. It was found that the PINNs with monotonicity constraints could estimate soil water flux density from noisy volumetric water content measurements. Larger errors were observed at wetting fronts and near the surface, where soil water flux density changed abruptly. Although larger relative error was observed for loam and silt loam (see Table 3), especially for Scenario 1, the PINNs with monotonicity constraints could reasonably capture the trend of soil water flux density, which is shown in Figure S6 and S7 in the supporting information.

The advantage of this approach over the available heat pulse method (Kamai et al., 2008, 2010) is that this method can estimate soil water flux density lower than 1 cm day^{-1} (see Figure S8, S9, and S10 in the supporting information). Because continuous measurement of volumetric water content at different depths is becoming popular with an advanced TDR array (Sheng et al., 2017), this PINNs approach can be used to estimate soil water flux density in fields. This finding has a significant implication in the application of land surface modeling, where soil water flux density near the surface is critical.

5 Summary and Conclusions

A framework of estimating soil hydraulic functions or constitutive relationships of the Richardson-Richards equation (RRE) (i.e., water retention curves (WRCs) and hydraulic conductivity functions (HCFs)) from noisy volumetric water content measurements was proposed using physics-informed neural networks (PINNs). The PINNs for the RRE was designed by endowing the neural networks with the monotonicity of WRCs

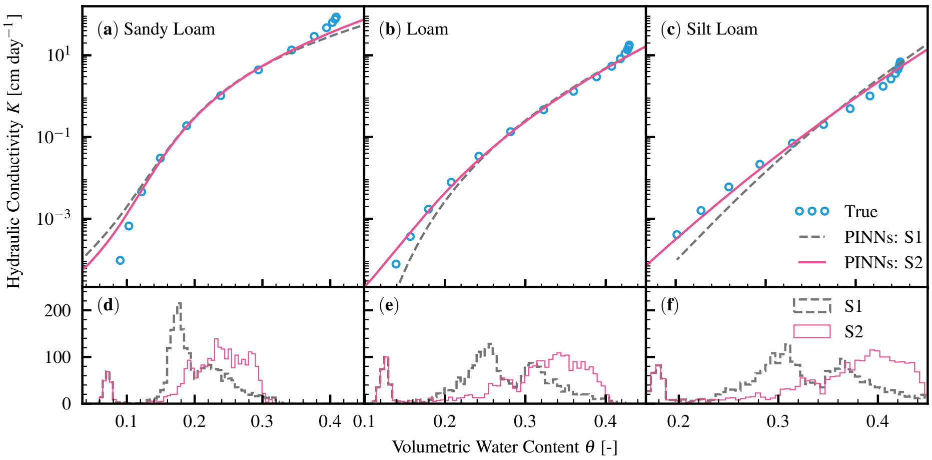

Figure 9. Comparison of true hydraulic conductivity functions (True) to the ones predicted by the PINNs with monotonicity constraints for the three soils for the two scenarios (S1: Scenario 1, S2: Scenario 2) with the histogram of the noisy training data. Hydraulic conductivity function for (**a**) sandy loam, (**b**) loam, and (**c**) silt loam. Histogram of the training data for (**d**) sandy loam, (**e**) loam, and (**f**) silt loam.

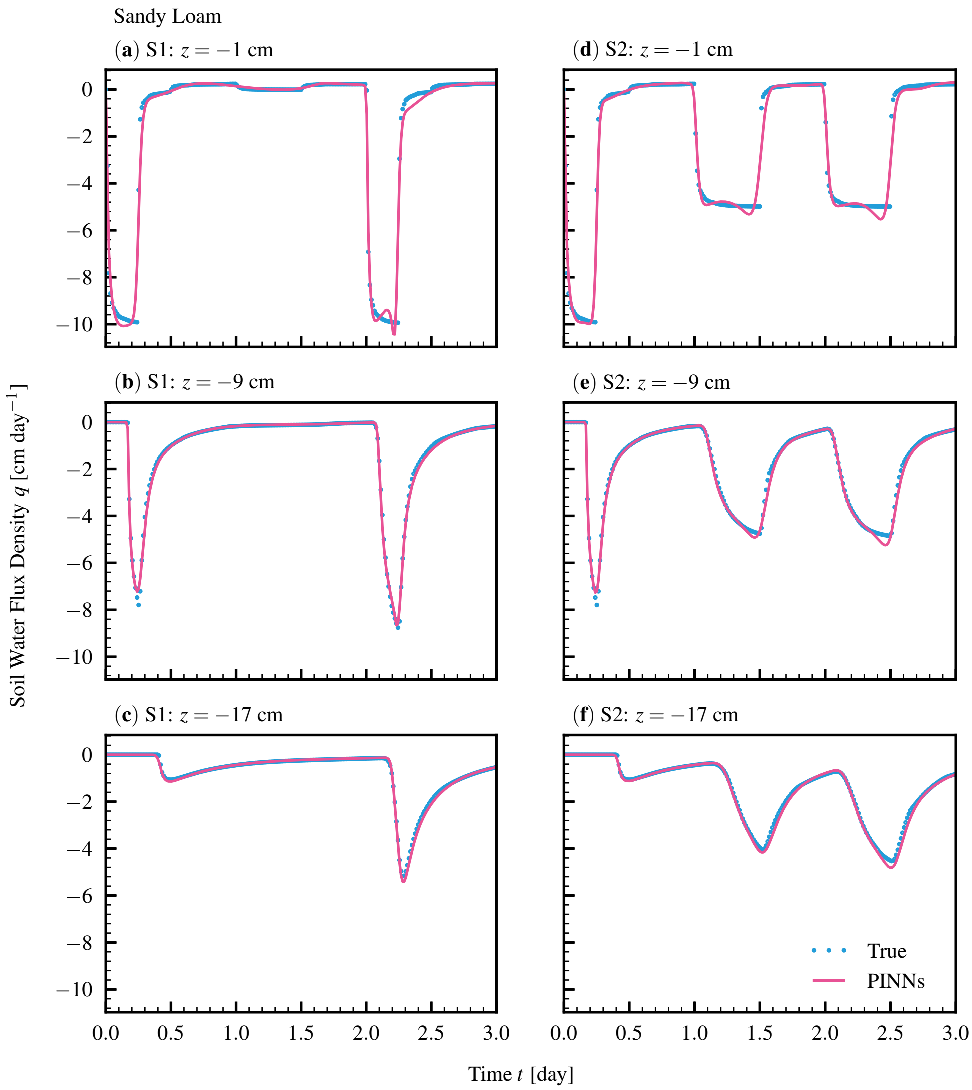

Figure 10. Estimated soil water flux density against the true one at three different depths for sandy loam soil. Scenario 1 (S1): (**a**) $z = -1$ cm, (**b**) $z = -9$ cm, and (**c**) $z = -17$ cm. Scenario 2 (S2): (**d**) $z = -1$ cm, (**e**) $z = -9$ cm, and (**f**) $z = -17$ cm.

and HCFs. To demonstrate the effectiveness of incorporating monotonicity constraints into the PINNs, we compared the performance of the PINNs between with and without monotonicity constraints. As a result, the PINNs with monotonicity constraints has a great advantage over the PINNs without monotonicity constraints in terms of its high ability to prevent overfitting and reliability of the results for noisy training data.

The PINNs, with and without monotonicity constraints, were trained using synthetic volumetric water content data for three distinct soil textures (sandy loam, loam, and silt loam) with Gaussian noise. The generalization ability of the framework was assessed in terms of its ability to estimate WRCs, HCFs, and soil water flux densities. The PINNs without monotonicity constraints could not produce satisfactory results. On the other hand, the PINNs with monotonicity constraints could estimate true soil moisture dynamics from noisy synthetic data for all types of soil. In terms of WRCs, the PINNs with monotonicity constraints could not precisely estimate the true WRCs. However, the estimated WRCs were surprisingly similar to the true ones in the middle range regardless of the fact that any matric potential data was provided. Unlike WRCs, the PINNs with monotonicity constraints could predict the HCFs well, especially for the range that is covered in the training data.

It was demonstrated that employing monotonic neural networks in the PINNs to represent WRCs and HCFs improved the ability of the PINNs to prevent overfitting the training data. Furthermore, the PINNs with monotonicity constraints is shown to have better durability against noisy data than the PINNs without monotonicity constraints.

It was illustrated that the PINNs with monotonicity constraints has a great potential to predict constitutive relationships of the RRE and soil water flux density from only noisy volumetric water content data in fields. The advantage of this method is the current PINNs framework does not need initial and boundary conditions and any information about the HCF a priori. The current framework must be tested with real experimental data for homogeneous soil in future research.

The PINNs with monotonicity constraints could estimate true soil water flux density from noisy synthetic volumetric water content data at different depths. At present, the only measurement technique for measuring soil water flux density is using heat flux sensors, which is limited to soil water flux density larger than 1 cm day^{-1}. The proposed method has the potential for determining soil water flux density over a broader range.

Acronyms

HCFs Hydraulic Conductivity Functions
PDE Partial Differential Equation
PINNs Physics-Informed Neural Networks
RRE Richardson-Richards Equation
WRCs Water Retention Curves
VWC Volumetric Water Content

Notation

:= Equal by definition
$\hat{}$ Hat indicating predicted values or functions (e.g., $\hat{y}$)
$^{(i)}$ Superscript (i) denoting ith data (e.g., $\theta^{(i)}$)
$^{[L]}$ Superscript [L] denoting Lth layer
$\mathbf{a}^{[L]} \in \mathbb{R}^{n^{[L]}}$ Vector value for the Lth layer consisting of $n^{[L]}$ units
$\boldsymbol{b}$ Bias vector
$\hat{\boldsymbol{f}}$ Neural network

$\boldsymbol{g}$ Activation function
$\boldsymbol{h}$ Output function
$\boldsymbol{K}$ Hydraulic conductivity [L T^{-1}]
$\boldsymbol{K_s}$ Mualem-van Genuchen parameter
$\boldsymbol{\mathcal{L}}$ Loss function
$\boldsymbol{l}$ Mualem-van Genuchen parameter
$\boldsymbol{N}$ Number of data points
$\boldsymbol{n}$ Mualem-van Genuchen parameter
$\boldsymbol{n_i}$ Number of size a vector, as in n_x and n_y
$\boldsymbol{n^{[L]}}$ Number of units in Lth layer of a neural network
$\boldsymbol{q}$ Soil water flux density [L T^{-1}]
$\boldsymbol{\hat{r}}$ Residual of the Richardson-Richards euqation
$\boldsymbol{S_e}$ Effective saturation
$\boldsymbol{t}$ Time [T]
$\boldsymbol{W}$ Weight matrix
$\mathbf{x} \in \mathbb{R}^{\boldsymbol{n_x}}$ Input vector for the size of the input n_x
$\mathbf{y} \in \mathbb{R}^{\boldsymbol{n_y}}$ Output vector for the size of the output n_y
$\boldsymbol{z}$ Vertical coordinate or depth (positive upward) [L]
$\boldsymbol{\alpha}$ Mualem-van Genuchen parameter [L^{-1}]
$\boldsymbol{\epsilon}$ Relative error
$\boldsymbol{\theta}$ Volumetric water content [L^3 L^{-3}]
$\boldsymbol{\theta_r}$ Mualem-van Genuchen parameter [L^3 L^{-3}]
$\boldsymbol{\theta_s}$ Mualem-van Genuchen parameter [L^3 L^{-3}]
$\boldsymbol{\psi}$ Matric potential of water in the soil [L]
$\boldsymbol{\psi_{log}}$ Matric potential in logarithmic scale

Acknowledgments

The publicly available code for physics-informed neural networks provided by Dr. Maziar Raissi (University of Colorado Boulder) was instrumental in the development of our model. We are indebted to anonymous reviewers for improving this manuscript. The dataset used in this study is available through Dyrad (https://datadryad.org/stash; doi:10.6071/M3T376).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous distributed systems.* Retrieved from `https://www.tensorflow.org/`

Assouline, S. (2006). Modeling the relationship between soil bulk density and the hydraulic conductivity function. *Vadose Zone Journal*, *5*, 697–705. doi: 10.2136/vzj2005.0084

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., & Tuller, M. (2019). Ground, proximal, and satellite remote sensing of soil moisture. *Reviews of Geophysics*, *57*, 530–616. doi: 10.1029/2018RG000618

Bitterlich, S., Durner, W., Iden, S. C., & Knabner, P. (2004). Inverse estimation of the unsaturated soil hydraulic properties from column outflow experiments using free-form parameterizations. *Vadose Zone Journal*, *3*, 971–981. doi: 10.2113/3.3.971

Brooks, R. H., & Corey, A. T. (1964). *Hydraulic properties of porous media.* Fort Collins, CO, USA.

Buckingham, E. (1907). *Studies on the movement of soil moisture* (Vol. 38). Washington, DC, USA.

Burdine, N. T. (1953). Relative permeability calculations from pore size distribution data. *Society of Petroleum Engineers*, *5*(3). doi: 10.2118/225-G

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *Journal of Scientific Computing*, *16*(5), 1190–1208. doi: 10.1137/0916069

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control Signals Systems*, *2*, 303–314. doi: 10.1007/BF02836480

Daniels, H., & Velikova, M. (2010). Monotone and partially monotone neural networks. *IEEE Transactions on Neural Networks*, *21*(6), 906–917. doi: 10.1109/TNN.2010.2044803

Degré, A., van der Ploeg, M. J., Caldwell, T., & Gooren, H. P. A. (2017). Comparison of soil water potential sensors: A drying experiment. *Vadose Zone Journal*, *16*(4), 1–8. doi: 10.2136/vzj2016.08.0067

Durner, W. (1994). Hydraulic conductivity estimation for soils with heterogeneous pore structure. *Water Resources Research*, *30*(2), 211–223. doi: 10.1029/93WR02676

Durner, W., Jansen, U., & Iden, S. C. (2008). Effective hydraulic properties of layered soils at the lysimeter scale determined by inverse modelling. *European Journal of Soil Science*, *59*, 114–124. doi: 10.1111/j.1365-2389.2007.00972.x

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*, 1257–1269. doi: 10.2136/sssaj2017.02.0058

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* Cambridge, MA, USA: The MIT Press.

He, Q., Barajas-Solano, D., Tartakovsky, G. D., Alexandre, M., & Tartakovsky, A. M. (2020). Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Advances in Water Resources*, *141*. doi: 10.1016/j.advwatres.2020.103610

Hopmans, J. W., Šimůnek, J., Romano, N., & Durner, W. (2002). Simultaneous determination of water transmission and retention properties. Inverse methods. In J. H. Dane & G. C. Topp (Eds.), *Methods of soil analysis, part 4, physical methods* (pp. 963–1008). Madison, WI, USA: Soil Science Society of America.

Iden, S. C., & Durner, W. (2007). Free-form estimation of the unsaturated soil hydraulic properties by inverse modeling using global optimization. *Water Resources Research*, *43*, 1–12. doi: 10.1029/2006WR005845

Kamai, T., Tuli, A., Kluitenberg, G. J., & Hopmans, J. W. (2008). Soil water flux density measurements near 1 cm d1 using an improved heat pulse probe design. *Water Resources Research*, *44*, 1–12. doi: 10.1029/2008wr007036

Kamai, T., Tuli, A., Kluitenberg, G. J., & Hopmans, J. W. (2010). Correction to "Soil water flux density measurements near 1 cm d1 using an improved heat pulse probe design". *Water Resources Research*, *46*, 1–5. doi: 10.1029/2010WR009423

Kingma, D. P., & Ba, J. B. (2014). Adam: A method for stochastic optimization. *arXiv preprint*, 1–15. Retrieved from `https://arxiv.org/pdf/1412.6980.pdf`

Kosugi, K. (1996). Lognormal distribution model for unsaturated soil hydraulic properties. *Water Resources Research*, *32*(9), 2697–2703. doi: 10.1029/96WR01776

Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, *12*(3), 513–522. doi: 10.1029/WR012i003p00513

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization.* New York, NY, USA: Springer. doi: 10.1007/978-0-387-40065-5

Rai, P. K., & Tripathi, S. (2019). Gaussian process for estimating parameters of partial differential equations and its application to the Richards equation. *Stochastic Environmental Research and Risk Assessment*, *33*, 1629–1649. doi: 10.1007/s00477-019-01709-8

Raissi, M., & Karniadakis, G. E. (2018). Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, *357*, 125–141. doi: 10.1016/j.jcp.2017.11.039

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. doi: 10.1016/j.jcp.2018.10.045

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, *1*, 318–333. doi: 10.1063/1.1745010

Richardson, L. F. (1922). *Weather prediction by numerical process.* Cambridge, United Kingdom: Cambridge University Press.

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., ... Wendroth, O. (2008). Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, *7*, 358–389. doi: 10.2136/vzj2007.0143

Sheng, W., Zhou, R., Sadeghi, M., Babaeian, E., Robinson, D. A., Tuller, M., & Jones, S. B. (2017). A TDR array probe for monitoring near-surface soil moisture distribution. *Vadose Zone Journal*, *16*(4), 1–8. doi: 10.2136/vzj2016.11.0112

Šimůnek, J., Šejna, M., Saito, H., Sakai, M., & van Genuchten, M. T. (2013). *The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably saturated media, Version 4.17* (Tech. Rep.). Riverside, CA, USA: Department of Environmental Sciences, University of California Riverside, Riverside.

Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., & Barajas-Solano, D. (2020). Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems. *Water Resources Research*. doi: 10.1029/2019wr026731

Tuller, M., & Or, D. (2001). Hydraulic conductivity of variably saturated porous media: Film and corner flow in angular pore space. *Water Resources Research*, *37*(5), 1257–1276. doi: 10.1029/2000WR900328

van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America*, *44*, 892–898. doi: 10.1016/j.pan.2017.07.214

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... Contributors, S. . . (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. doi: https://doi.org/10.1038/s41592-019-0686-2

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. *Wiley Interdisciplinary Reviews: Water*, *6*, 1–23. doi: 10.1002/wat2.1364

Physics-informed neural networks with monotonicity constraints for Richardson-Richards equation: Estimation of constitutive relationships and soil water flux density from volumetric water content measurements

Toshiyuki Bandai [1], Teamrat A. Ghezzehei [1]

[1]Life and Environmental Science Department, University of California, Merced, 5200, Lake Rd, Merced, CA, 95343, USA

Key Points:

- Hydraulic conductivity functions were precisely estimated from only volumetric water content data using the proposed framework.
- Soil water flux density was accurately derived from the trained physics-informed neural networks.
- Incorporating monotonic neural networks to represent constitutive relationships in physics-informed neural networks prevents overfitting.

Corresponding author: Toshiyuki Bandai, `tbandai@ucmerced.edu`

Abstract

Water retention curves (WRCs) and hydraulic conductivity functions (HCFs) are critical soil-specific characteristics necessary for modeling the movement of water in soils using the Richardson-Richards equation (RRE). Well-established laboratory measurement methods of WRCs and HCFs are not usually unsuitable for simulating field-scale soil moisture dynamics because of the scale mismatch. Hence, the inverse solution of the RRE is used to estimate WRCs and HCFs from field measured data. Here, we propose a physics-informed neural networks (PINNs) framework for the inverse solution of the RRE and the estimation of WRCs and HCFs from only volumetric water content (VWC) measurements. Unlike conventional inverse methods, the proposed framework does not need initial and boundary conditions. The PINNs consists of three linked feedforward neural networks, two of which were constrained to be monotonic functions to reflect the monotonicity of WRCs and HCFs. Alternatively, we also tested PINNs without monotonicity constraints. We trained the PINNs using synthetic VWC data with artificial noise, derived by a numerical solution of the RRE for three soil textures. The PINNs were able to reconstruct the true VWC dynamics. The monotonicity constraints prevented the PINNs from overfitting the training data. We demonstrated that the PINNs could recover the underlying WRCs and HCFs in non-parametric form, without a need for initial guess. However, the reconstructed WRCs at near-saturation–which was not fully represented in the training data–was unsatisfactory. We additionally showed that the trained PINNs could estimate soil water flux density with a broader range of estimation than the currently available methods.

1 Introduction

Accurate prediction of soil moisture dynamics is vital for many applications, including weather forecasts, agricultural water management, and prediction of natural disasters, such as landslides and floods, and drought (Robinson et al., 2008; Babaeian et al., 2019). Notably, detailed information about near-surface soil moisture dynamics is essential for land surface modeling and remote sensing applications.

Mathematically, soil moisture dynamics is described by a non-linear partial differential equation (PDE), commonly referred to as the Richardson-Richards equation (RRE) (Richardson, 1922; Richards, 1931). The RRE is composed of the continuity equation and the Buckingham-Darcy law (Buckingham, 1907) and consists of three primary variables: matric potential ψ, volumetric water content θ, and hydraulic conductivity K. The latter two variables are commonly expressed as functions of matric potential using water retention curves (WRCs) and hydraulic conductivity functions (HCFs), respectively. Furthermore, the two soil hydraulic functions (also referred to as constitutive relationships) are often treated as interdependent by employing conceptual models of unsaturated flow, such as the bundle of capillaries (Mualem, 1976; Burdine, 1953) or angular-pores and slits model (Tuller & Or, 2001). These assumptions simplify soil water dynamics models by allowing WRCs and HCFs to be expressed using a shared set of parameters. Several parametric models have been proposed to describe soil hydraulic functions (Brooks & Corey, 1964; van Genuchten, 1980; Durner, 1994; Kosugi, 1996; Tuller & Or, 2001; Assouline, 2006).

The constitutive relationships embody the characteristic features of soil pore network and are the manifestation of the interactions between soil texture and structure. Hence, the reliability of simulated soil water dynamics largely depends on the accuracy of these soil hydraulic functions (Farthing & Ogden, 2017; Zha et al., 2019). Although well-established laboratory methods for characterizing WRCs and HCFs are available, their direct application for field-scale simulations is typically unsatisfactory because of the scale mismatch as well as sampling and measurement artifacts (Hopmans et al., 2002).

Therefore, it is indispensable to estimate WRCs and HCFs using time-series data from field experiments and the inverse solution of the RRE. Commonly, the inverse problem requires finding the parameters of the constitutive relationships that best describe observed time-series data. In principle, it is possible to fit WRCs and HCFs independently, albeit at the expense of significant increase in the tunable parameters. Several studies also employed free-form functions to estimate WRCs and HCFs (Bitterlich et al., 2004; Iden & Durner, 2007). Inverse methods for characterizing soil hydraulic properties often involve the repeated solution of the forward problem, which requires knowledge of the relevant initial and boundary conditions of the RRE. Global optimization algorithm (Durner et al., 2008) and Gaussian processes (Rai & Tripathi, 2019) are other approaches used to find the best-fitted constitutive relationships.

Here, we propose a deep-learning framework for the inverse solution of the time-dependent RRE and the estimation of both WRCs and HCFs, with fewer assumptions and constraints than approaches described above. The method is based on physics-informed neural networks (PINNs) developed by Raissi et al. (2019). PINNs employs the universal approximation capability of neural networks (Cybenko, 1989) to approximate the solution of PDEs. The neural networks' parameters are trained by minimizing the sum of data-fitting error and the residual of the PDEs simultaneously. This simultaneous fitting enables PINNs to learn the dynamics of the system from measurement data and known physics. This novel PINNs approach has shown promising successes in computational physics (Raissi & Karniadakis, 2018; Raissi et al., 2019; Tartakovsky et al., 2020; He et al., 2020). Notably, Tartakovsky et al. (2020) employed PINNs to determine the hydraulic conductivity function of an unsaturated homogeneous soil from synthetic matric potential data based on the two-dimensional time-independent RRE. In this study, we coupled the PINNs framework with two additional monotonic neural networks (Daniels & Velikova, 2010) to describe the known monotonicity of WRCs and HCFs.

Although matric potential is the variable of choice for training purposes, the range and accuracy of matric potential sensors are still limited (Degré et al., 2017). Therefore, the proposed approach uses only volumetric water content time-series data. There are numerous fully developed methods to measure volumetric water content in fields, including the TDR-array probe (Sheng et al., 2017) and the heat-pulse method .

Unlike conventional inverse methods, this proposed approach does not require the repeated solution of the forward problem. Instead, it simultaneously learns (1) the physics of soil water dynamics as defined by the RRE and the monotonicity of the constitutive relationships and (2) the volumetric water content time-series data. The simultaneous learning eliminates the critical shortcomings of conventional inverse approaches, including (1) the need for initial and boundary conditions to solve the forward problems; (2) the dependence of the optimization algorithms on good prior approximations of WRCs and HCFs; and (3) the need to define the shapes of WRCs and HCFs and their interdependence a priori.

In this study, we generated synthetic training data by forward modeling of the RRE using HYDRUS-1D (Šimůnek et al., 2013). Using synthetic data has distinct advantages for testing this novel inverse-solution framework. First, it eliminates the uncertainties of field conditions that equally affect other inverse methods. Second, the synthetic data provide information that is not typically available in routine field measurements, including matric potential and soil water flux density at every location and time.

The robustness of using monotonic neural networks to represent WRCs and HCFs in the PINNs is demonstrated by comparing the results with those from the PINNs that lacks the monotonicity constraints. The performance of the framework was further tested by introducing varying degrees of noise to the synthetic volumetric water content data, altering the spacing between the locations at which volumetric water content data were sampled, using different initial weight parameters of the neural networks. The generalization capability of the framework was investigated by training the PINNs with volumetric water content data for three soils (sandy loam, loam, and silt loam soil) and for two different scenarios of the upper boundary condition. Finally, we show the potential application of the PINNs to estimate soil water flux density using only an array of soil moisture sensors.

2 Background

2.1 Richardson-Richards Equation

We consider one-dimensional liquid water flow in a homogeneous rigid soil and ignore water vapor, sink term, and hysteresis. The mass balance of water in the soil leads to the continuity equation:

$$\frac{\partial \theta}{\partial t} = -\frac{\partial q}{\partial z}, \tag{1}$$

where θ is volumetric water content [L^3 L^{-3}]; t is time [T]; z is vertical coordinate (positive upward) [L]; q is soil water flux density [L T^{-1}]. Soil water flux density q is related to matric potential of water in the soil ψ [L] through the Buckingham-Darcy law (Buckingham, 1907):

$$q = -K\left(\frac{\partial \psi}{\partial z} + 1\right), \tag{2}$$

where K is hydraulic conductivity [L T^{-1}]. The two equations (Equation (1) and (2)) are combined to derive the Richardson-Richards equation (RRE) (Richardson, 1922; Richards, 1931):

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K\left(\frac{\partial \psi}{\partial z} + 1\right)\right]. \tag{3}$$

To solve the RRE, matric potential ψ is commonly treated as the primary variable that is dependent on t and z, and volumetric water content θ and hydraulic conductivity K are parameterized through matric potential ψ, as in

$$\frac{\partial \theta(\psi(t,z))}{\partial t} = \frac{\partial}{\partial z}\left[K(\psi(t,z))\left(\frac{\partial \psi(t,z)}{\partial z} + 1\right)\right]. \tag{4}$$

The functions $\theta(\psi)$ and $K(\psi)$ are called constitutive relationships of the RRE and referred to as water retention curves (WRCs) and hydraulic conductivity functions (HCFs), respectively. WRCs and HCFs are commonly expressed by parametric models (e.g., Brooks and Corey (1964); van Genuchten (1980); Durner (1994); Kosugi (1996); Tuller and Or (2001); Assouline (2006)). The WRCs and HCFs for three types of soil (sandy loam, loam, and silt loam soil) using the Mualem-van Genuchen model (van Genuchten, 1980) are shown in Figure 1. As shown in the figure, both WRCs and HCFs are monotonically increasing functions with respect to matric potential ψ, which is an accepted physical principle of water movement in soils. The monotonicity of WRCs and HCFs will be employed to design the architecture of the neural networks in this study later on.

2.2 Feedforward Neural Networks

A standard fully-connected feedforward neural network with three layers (one hidden layer) is introduced here for readers who are not well versed in the topic. The readers should refer to textbooks (e.g., Goodfellow et al. (2016)) for more general explanations.

Given a training dataset $\{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}$, where superscript (i) denotes the ith training data; $\mathbf{x}^{(i)} \in \mathbb{R}^{n_x}$ is input vector for the size of the input n_x, $\mathbf{y}^{(i)} \in \mathbb{R}^{n_y}$ is output vector for the size of the output n_y, a neural network $\hat{f}$ is a mathematical function mapping the input vector $\mathbf{x}^{(i)}$ to predicted output vector $\hat{\mathbf{y}}^{(i)} \in \mathbb{R}^{n_y}$:

$$\hat{\mathbf{y}}^{(i)} = \hat{f}(\mathbf{x}^{(i)}). \tag{5}$$

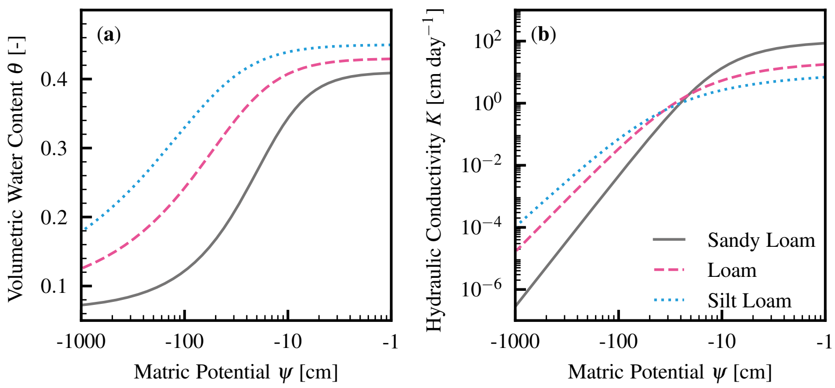

Figure 1. Constitutive relationships for three types of soil (sandy loam, loam, and silt loam soil) generated by using the Mualem-van Genuchten model (van Genuchten, 1980). (**a**) Water retention curves (WRCs). (**b**) Hydraulic conductivity functions (HCFs).

The hat operator represents prediction throughout the paper. The inside of the neural network $\hat{f}$ is commonly represented by layers of units (or neurons), as shown in Figure 2. Herein, $\mathbf{a}^{[L]} \in \mathbb{R}^{n^{[L]}}$ denotes the vector value for the Lth layer of the neural network, where the Lth layer is composed of $n^{[L]}$ units. To calculate the predicted output vector $\hat{\mathbf{y}}^{(i)}$, the input vector $\mathbf{x}^{(i)}$ is entered in the first layer:

$$\mathbf{a}^{[1]} = \mathbf{x}^{(i)}, \tag{6}$$

where the number of units in the first layer $n^{[1]}$ is equal to n_x. Then, the value for the jth unit of the second layer $\mathbf{a}^{[2]}$ is calculated from all the units in the previous layer (i.e., the first layer) with the weight matrix $\mathbf{W}^{[1]}$ and bias vector $\mathbf{b}^{[1]}$ for the first layer in the following way:

$$a_j^{[2]} = g^{[1]} \left(\sum_{k=1}^{n^{[1]}} W_{j,k}^{[1]} a_k^{[1]} + b_j^{[1]} \right), \tag{7}$$

where $g^{[1]}$ is a non-linear activation function for the first layer, such as the hyperbolic tangent function (tanh) shown in Figure 2 (**b**). The jth unit of the third layer is computed from all the units of the second layer (hidden layer):

$$a_j^{[3]} = \sum_{k=1}^{n^{[2]}} W_{j,k}^{[2]} a_k^{[2]} + b_j^{[2]}. \tag{8}$$

Finally, the predicted output vector $\hat{\mathbf{y}}^{(i)}$ is derived from the last layer with an output function h:

$$\hat{y}_j^{(i)} = h(a_j^{[3]}), \tag{9}$$

where the number of the units in the last layer $n^{[3]}$ is equal to n_y. In this study, the sigmoid function (Figure 2 (**c**)) and the exponential function (Figure 2 (**d**)) are used as output functions.

The collection of the weight matrices $\mathbf{W} = \{\mathbf{W}^{[1]}, \mathbf{W}^{[2]}\}$ and bias vectors $\mathbf{b} = \{\mathbf{b}^{[1]}, \mathbf{b}^{[2]}\}$ are the parameters of the neural network, which are estimated by minimizing a loss function comprising of the output vector $\mathbf{y}^{(i)}$ (training data) and the predicted

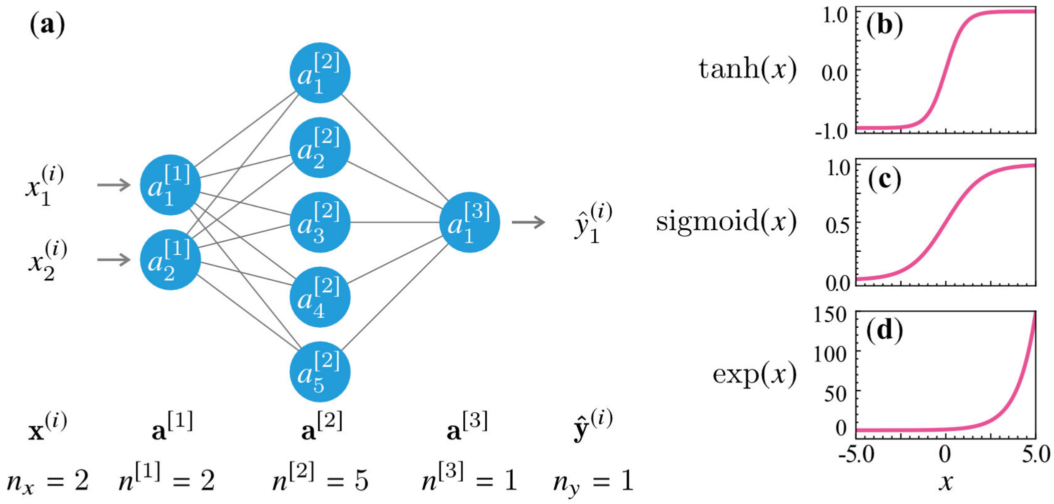

Figure 2. A fully-connected feedforward neural network consisting of three layers (one hidden layer) with activation and output functions. (**a**) A fully-connected feedforward neural network consisting of the input layer with two units, the hidden layer with five units, and the output layer with one unit. (**b**) Hyperbolic tangent function. (**c**) Sigmoid function. (**d**) Exponential function.

output vector $\hat{\mathbf{y}}^{(i)}$. The definition of the loss function varies depending on the purpose of the training, and the loss function used in this study is defined in Equation (14).

It is well known that a feedforward neural network with more hidden layers has a better capability of function approximation (Goodfellow et al., 2016), and such a neural network with more than two hidden layers is called a deep neural network. In such a case, a unit of a hidden layer is computed from all the units of the previous hidden layer in the same way explained above (Equation (7)).

In the next section, three fully-connected feedforward neural networks are combined to construct physics-informed neural networks (PINNs) for the RRE, and the loss function for the PINNs framework is defined to estimate WRCs and HCFs from volumetric water content measurements.

3 Methods

3.1 Physics-Informed Neural Networks with Monotonicity Constraints for RRE

Physics-informed neural networks (PINNs) has been proposed as a deep learning framework to derive the forward and inverse solution of PDEs (Raissi et al., 2019). In this study, PINNs was used to derive the inverse solution of the RRE and the constitutive relationships (i.e., WRCs and HCFs) from a set of volumetric water content time-series data measured at different depths in soils $\{t^{(i)}, z^{(i)}, \theta^{(i)}\}_{i=1}^{i=N}$, where N is the number of measurement data.

PINNs for the RRE was constructed using three fully-connected feedforward neural networks, as shown in Figure 3. The neural network $\hat{f}_\psi$ (Figure 3 (**a**)) is a function mapping from time t and vertical coordinate z into predicted matric potential $\hat{\psi}$:

$$\hat{\psi}^{(i)} = \hat{f}_\psi(t^{(i)}, z^{(i)}; \mathbf{W}_\psi, \mathbf{b}_\psi), \tag{10}$$

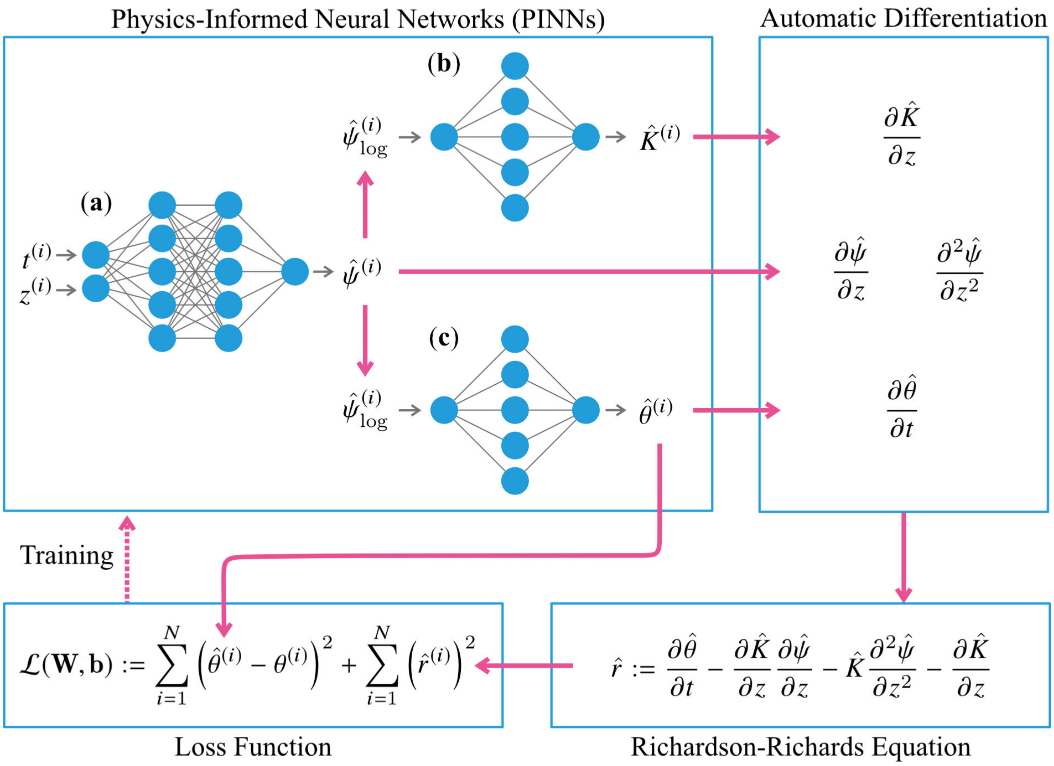

Figure 3. Physics-informed neural networks (PINNs) for the Richardson-Richards equation consisting of three fully-connected feedforward neural networks to predict (**a**) matric potential $\hat{\psi}$, (**b**) hydraulic conductivity $\hat{K}$, and (**c**) volumetric water content $\hat{\theta}$. The number of layers and units in the figure is not actual.

where $\mathbf{W}_\psi$ and $\mathbf{b}_\psi$ are the collection of weight and bias parameters in the neural network. The hyperbolic tangent function (Figure 2 (**b**)) is used for the activation function, as recommended by Raissi et al. (2019). As for the output function, the negative exponential function (i.e., $-\exp(x)$, see Figure 2 (**d**)) is used to force the predicted matric potential to be negative.

The predicted matric potential $\hat{\psi}^{(i)}$ is used to estimate volumetric water content $\hat{\theta}^{(i)}$ and hydraulic conductivity $\hat{K}^{(i)}$ through two distinct neural networks $\hat{f}_\theta$, $\hat{f}_K$ (Figure 3 (**c**) and (**b**), respectively). In other words, the two neural networks are used to represent the WRC and HCF for a given soil. Since WRCs and HCFs become simpler if matric potential is plotted in logarithmic scale, as in Figure 1, the predicted matric potential is converted into logarithmic scale by the following transformation:

$$\hat{\psi}^{(i)}_{\log} = -\log_e(-\hat{\psi}^{(i)}). \tag{11}$$

Then, the predicted matric potential in logarithmic scale $\hat{\psi}^{(i)}_{\log}$ is used as the input value for the two neural networks to represent WRCs and HCFs:

$$\hat{\theta}^{(i)} = \hat{f}_\theta(\hat{\psi}^{(i)}_{\log}; \mathbf{W}_\theta, \mathbf{b}_\theta), \tag{12}$$

$$\hat{K}^{(i)} = \hat{f}_K(\hat{\psi}^{(i)}_{\log}; \mathbf{W}_K, \mathbf{b}_K). \tag{13}$$

The tanh function is used as the activation function for both neural networks. The output functions for $\hat{f}_\theta$ and $\hat{f}_K$ are the sigmoid function and the exponential function, respectively to ensure predicted volumetric water content between 0 and 1 and positive predicted hydraulic conductivity (see Figure 2 (**c**) and (**d**)).

To embrace the monotonicity of WRCs and HCFs, the weight parameters $\mathbf{W}_\psi$ and $\mathbf{W}_K$ are constrained to be non-negative so that $\hat{f}_\theta$ and $\hat{f}_K$ are monotonically increasing functions with respect to the predicted matric potential $\hat{\psi}$ (Daniels & Velikova, 2010). This type of neural networks is called (totally) monotonic neural networks, where the output values depend monotonically on all the variables in the input vector. It is known that a three-layer fully-connected feedforward neural network with non-negative weights can arbitrarily approximate any monotonic scalar functions (Daniels & Velikova, 2010). Readers interested in monotonic neural networks should refer to Daniels and Velikova (2010), where various types of monotonic neural networks are explained.

Incorporating monotonicity constraints in the neural networks representing WRCs and HCFs honors the physical nature of the movement of water in all soils. This approach is similar to the free-form approach (Bitterlich et al., 2004; Iden & Durner, 2007), where cubic Hermite interpolation was used to approximate WRCs and HCFs. Unlike their studies, our monotonic neural network approach does not assume predetermined saturated water content and saturated hydraulic conductivity because they are not easily available in field applications.

The collection of the parameters in the three neural networks $\mathbf{W} = \{\mathbf{W}_\psi, \mathbf{W}_\theta, \mathbf{W}_K\}$ and $\mathbf{b} = \{\mathbf{b}_\psi, \mathbf{b}_\theta, \mathbf{b}_K\}$ are identified by minimizing a loss function defined as

$$\mathcal{L}(\mathbf{W}, \mathbf{b}) := \sum_{i=1}^{N} (\hat{\theta}^{(i)} - \theta^{(i)})^2 + \sum_{i=1}^{N} (\hat{r}^{(i)})^2, \tag{14}$$

where $\hat{r}$ is the residual of the RRE defined as

$$\hat{r} := \frac{\partial \hat{\theta}}{\partial t} - \frac{\partial}{\partial z}\left[\hat{K}\left(\frac{\partial \hat{\psi}}{\partial z} + 1\right)\right] = \frac{\partial \hat{\theta}}{\partial t} - \frac{\partial \hat{K}}{\partial z}\frac{\partial \hat{\psi}}{\partial z} - \hat{K}\frac{\partial^2 \hat{\psi}}{\partial z^2} - \frac{\partial \hat{K}}{\partial z}. \tag{15}$$

The first term of the loss function (Equation (14)) represents the fitting error of volumetric water content, and the second term represents the contraint by the RRE. This simultaneous learning enables the PINNs to learn the dynamics of water in soils from both volumetric water content data and knowledge in soil physics (i.e., the RRE). In the other studies on PINNs (e.g. Raissi et al. (2019); Tartakovsky et al. (2020); He et al. (2020)), the boundary and initial conditions of PDEs are also included in the loss function. However, we omitted these terms because they are difficult to obtain in real applications.

To calculate the residual of the RRE $\hat{r}$ at all the data points, the derivatives (i.e., $\frac{\partial \hat{\theta}}{\partial t}$, $\frac{\partial \hat{\psi}}{\partial z}$, $\frac{\partial^2 \hat{\psi}}{\partial z^2}$, $\frac{\partial \hat{K}}{\partial z}$) are evaluated at the data points by using automatic differentiation (Nocedal & Wright, 2006). It should be noted that the residual of the RRE $\hat{r}$ can be evaluated at any point in the domain (called collocation points). However, we forced the collocation points to be the same as the measurement locations.

Before training the PINNs, the weight parameters $\mathbf{W}$ are initialized through Xavier initialization (Glorot & Bengio, 2010), and the bias parameters $\mathbf{b}$ are all set to zero. Then, these parameters $\mathbf{W}$ and $\mathbf{b}$ are trained by minimizing the loss function:

$$\min_{\mathbf{W}, \mathbf{b}} \mathcal{L}(\mathbf{W}, \mathbf{b}). \tag{16}$$

The optimization problem was solved by the Adam algorithm (Kingma & Ba, 2014) followed by the L-BFGS-B algorithm (Byrd et al., 1995). This two-step training procedure has been reported to be effective to train PINNs (Raissi et al., 2019; He et al., 2020). In our implementation, the default settings of the Adam optimizer in TensorFlow (Abadi et al., 2015) was used until 300,000 iterations finished. Then, the L-BFGS-B optimizer from Scipy (Virtanen et al., 2020) with $maxcor = 50$, $maxls = 50$, $maxiter = 50,000$,

Table 1. Two scenarios of surface water flux density [cm day^{-1}] (positive upward) were applied to generate synthetic data using HYDRUS-1D (Šimůnek et al., 2013).

Time (day)	Scenario 1	Scenario 2
0.25	-10	-10
0.50	0	0
1.0	0.3	0.3
1.5	0	-5
2.0	0.3	0.3
2.25	-10	-5
2.5	0	-5
3.0	0.3	0.3

$maxfun = 50,000$, $ftol = 2.220446049250313 \times 10^{-16}$, and the default values for the other parameters was applied to achieve the convergence of the loss function. The investigation on the hyperparameters of those optimization algorithms is beyond the scope of the paper. This PINNs framework for the RRE was implemented through TensorFlow 1.14 (Abadi et al., 2015), and the source code is available on https://github.com/ToshiyukiBandai/PINNs_RRE.

3.2 Synthetic Data Generated by HYDRUS-1D

To develop and assess the PINNs framework for the RRE, synthetic soil moisture data were generated by using HYDRUS-1D (Šimůnek et al., 2013). The synthetic data was used for two purposes: (1) to determine the architecture of the neural networks (i.e., the number of hidden layers and units; Section 3.3) (Section 3.3); (2) to investigate the the generalization capability of the PINNs (Section 3.4).

In the HYDRUS-1D simulation, soil moisture dynamics for three days in the 100 cm of homogeneous three soils with different textures (sandy loam, loam, and silt loam soil) were simulated. The soil column was uniformly discretized at a 0.1 cm interval. The initial matric potential was set at -1000 cm for all the depths. The bottom boundary condition was the Neumann boundary condition:

$$\frac{\partial \psi}{\partial z} = 0. \tag{17}$$

The upper boundary was set to the atmospheric upper boundary condition, where two different scenarios of time-dependent surface flux density were applied (see Table 1).

The Mualem-van Genuchen model was used to parameterize WRCs and HCFs in the HYDRUS-1D simulation (van Genuchten, 1980):

$$\theta(\psi) = \theta_r + \frac{\theta_s - \theta_r}{(1 + (-\alpha\psi)^n)^m}, \tag{18}$$

$$K(\theta(\psi)) = K_s S_e^l (1 - (1 - S_e^{1/m})^m)^2, \tag{19}$$

where θ_r, θ_s, α, n, K_s, and l are the Mualem-van Genuchen fitting parameters; $m = 1 - 1/n$; and the effective saturation S_e is defined as

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r}. \tag{20}$$

The Mualem-van Genuchen fitting parameters for the three soils used in this study are summarized in Table 2.

Table 2. The Mualem-van Genuchen fitting parameters for three types of soils (van Genuchten, 1980).

Parameters	Sandy Loam	Loam	Silt Loam
θ_r [cm^3 cm^{-3}]	0.065	0.078	0.067
θ_s [cm^3 cm^{-3}]	0.41	0.43	0.45
α [cm^{-1}]	0.075	0.036	0.02
n [-]	1.89	1.56	1.41
K_s [cm day^{-1}]	106.1	24.96	10.8
l [-]	0.5	0.5	0.5

3.3 Determination of Architecture of Neural Networks

It is known that the architecture of feedforward neural networks (i.e., the number of hidden layers and units) influences their performance. Therefore, the number of hidden layers and units for the three neural networks in the PINNs was determined empirically in two steps.

First, we set the number of hidden layers and units of the two neural networks, $\hat{f}_\theta$ for volumetric water content (Figure 3 (**c**)) and $\hat{f}_K$ for hydraulic conductivity (Figure 3 (**b**)), to 1 hidden layer with 20 units and varied the number of hidden layers and units of the neural network for the predicted matric potential $\hat{f}_\psi$ (Figure 3 (**a**)). Seven different numbers of hidden layers (2, 4, 6, 8, 9, 10, 11) and three different numbers of units (10, 20, 40) were tested.

Second, the number of hidden layers and units of the other two neural networks, $\hat{f}_\theta$ and $\hat{f}_K$, was varied. Three different numbers of layers (1, 2, 3) and units (10, 20, 40) were tested for each neural network.

To determine the architecture of the neural networks in the PINNs, the synthetic data for sandy loam soil for Scenario 1 were used (see Section 3.2). As training data, volumetric water content was sampled every 0.012 day (i.e., 251 data points for a depth) at 10 equally spaced different depths within the top of the 20 cm of the soil column (z = -1, -3, -5, -7, -9, -11, -13, -15, -17, -19 cm) because our study is focused on soil moisture dynamics in near-surface soils.

To evaluate the performance of the PINNs, we compared the predicted and true volumetric water content, matric potential, hydraulic conductivity, and soil water flux density. The predicted soil water flux density $\hat{q}$ was derived using the Buckingham-Darcy law (Equation (2)) with the estimated hydraulic conductivity $\hat{K}$ and the gradient of the predicted matric potential $\partial\hat{\psi}/\partial z$. We quantified the prediction error over the time $t \in [0, 3]$ day with an interval of 0.012 days and the spatial domain $z \in (-20, 0]$ cm with an interval of 0.1 cm for all the four variables in terms of the relative L_2 errors ϵ^γ for $\gamma = \theta, \psi, K, q.$, defined as

$$\epsilon^\gamma := \frac{\sum_{t\in[0,3]}\sum_{z\in(-20,0]}(\hat{\gamma}(t,z) - \gamma(t,z))^2}{\sum_{t\in[0,3]}\sum_{z\in(-20,0]}\gamma(t,z)^2} \tag{21}$$

To demonstrate the effectiveness of including monotonic neural networks in the PINNs, we also trained the PINNs without monotonicity constraints (i.e., standard feedforward neural networks are used to represent WRCs and HCFs) with the same training data.

The architecture of the three neural networks in the PINNs without monotonicity was also determined in the same way as above.

Because the results of training PINNs were affected by the initial values of the weight parameters of the neural networks determined by Xavier initialization (Glorot & Bengio, 2010), three different random seeds were used in the code, and three replicates were obtained for each of those combinations of the number of hidden layers and units. As a result, 63 trainings for $\hat{f}_\psi$ (Figure 3 (**a**)) and 243 ones for $\hat{f}_\theta$ (Figure 3 (**c**)) and $\hat{f}_K$ (Figure 3 (**b**)) were conducted to determined their architecture for the PINNs both with and without monotonicity.

3.4 Application of PINNs to Various Datasets

Different types of data were prepared by using HYDRUS-1D to assess the performance of the PINNs with and without monotonicity constraints. First, we investigated the effect of noise in the training data. To this end, Gaussian noise with the mean of zero and four different values of standard deviation (0, 0.005, 0.01, 0.02) was added to the sampled volumetric water content for sandy loam soil for Scenario 1.

Next, the effect of the sparsity of the training data was studied by using volumetric water content data for sandy loam soil for Scenario 1 without adding noise. We considered three cases for the number of depths at which volumetric water content were sampled: 10 ($z = -1, -3, -5, -7, -9, -11, -13, -15, -17, -19$ cm), 5 ($z = -1, -5, -9, -13, -17$ cm) and 3 ($z = -1, -9, -17$ cm).

Lastly, volumetric water content data for three different types of soils (sandy loam, loam, and silt loam soil) with the two different scenarios of upper boundary condition (see Table 1) were generated. Gaussian noise with the mean of zero and the standard deviation of 0.005 was added to the synthetic data to reflect measurement noise encountered in field applications.

Those training data were applied to the PINNs with and without monotonicity constraints, and the results were evaluated in terms of relative errors defined in Equation (21). For all the cases above, five different random seeds were set in the code to investigate the effects of neural network initialization on the results.

4 Results and Discussions

4.1 Architecture of Neural Networks in PINNs

To determine the number of hidden layers and units of the three neural networks in the PINNs with and without monotonicity, various combinations of layers and units were tested. Figure 4 shows relative error ϵ defined in Equation (21) for volumetric water content θ, matric potential ψ, hydraulic conductivity K, and soil water flux density q for different numbers of hidden layers and units for the neural network $\hat{f}_\psi$ (Figure 3 (**a**)) of the PINNs with and without monotonicity while the architecture of the other two neural networks are fixed (1 hidden layer with 20 units). For the PINNs with monotonic neural networks (left column), relative error for volumetric water content ϵ^θ, hydraulic conductivity ϵ^K, and soil water flux density ϵ^q decreased with the increase in number of units, with 40 units resulting in the lowest error (the Pearson correlation coefficient is provided in Table S1 in the supplementary information.).

The lowest arithmetic mean of relative error was observed when the number of hidden layers is 4 for volumetric water content θ, 6 for hydraulic conductivity K, and 8 for soil water flux density q when the number of units is 40. Clear trends were not obtained for relative error for matric potential ϵ_ψ. Because relative error for soil water flux density q reflects the predictive accuracy of the PINNs for both matric potential ψ and hydraulic conductivity K fields, we set the neural network for the predicted matric potential $\hat{f}_\psi$ to 8 layers with 40 units.

For the PINNs without monotonicity constraints (right column), the architecture of the neural network for the predicted matric potential $\hat{f}_\psi$ was set to 6 hidden layers with 40 units, which coincides with the lowest relative error of soil water flux density ϵ_q. We observed a non-linear correlation between the number of hidden layers and relative error for volumetric water content θ, hydraulic conductivity K, and soil water flux density q; relative error reached the lowest when the number of hidden layers was 6 and increased again. This is clear evidence that the PINNs without monotonicity was overfitting the training data. On the other hand, such a non-linear behavior was minimized for the PINNs with monotonicity constraints, which means imposing monotonicity can prevent the PINNs from overfitting the training data. In addition, the variability of relative errors between different initializations of the neural networks was lower for the PINNs with monotonic neural networks than the PINNs with non-monotonic neural networks. This further demonstrates the benefit of the monotonicity constraints in improving the stability and reliability of the training.

After determining the architecture of the neural network for the predicted matric potential $\hat{f}_\psi$, the number of hidden layers and units for the other two neural networks, $\hat{f}_\theta$ and $\hat{f}_K$, was varied. We did not observe clear trends of relative error for different neural network architectures for the PINNs with and without monotonicity (see Table S1 and Figure S1 in the supplementary information). However, the performance of the PINNs without monotonicity constraints was much more sensitive to the neural network architecture. This implied that incorporating monotonicity constraints stabilized the training, which enabled us to determine the neural network structure easier than the PINNs without monotonicity constraints. As a result, the architecture of the two neural networks was set as follows: 1 hidden layer with 40 units for the PINNs with monotonicity and 3 hidden layers with 40 units for without monotonicity for the neural network for the predicted volumetric water content $\hat{f}_\theta$; 3 hidden layer with 40 units for PINNs with monotonicity and 2 hidden layers with 20 units for without monotonicity for the neural network for the predicted hydraulic conductivity $\hat{f}_K$.

4.2 Effect of Noise and Sparsity of Training Data

To investigate the effect of measurement noise on the performance of the PINNs, Gaussian noise with mean of zero and different values of standard deviation (0, 0.005, 0.01, 0.02) was added to the synthetic volumetric water content data (see Section 3.4), which was used to train the PINNs with and without monotonicity constraints. Figure 5 (**a**) shows relative error for soil water flux density ϵ^q for different values of noise added to the true volumetric water content data. For the PINNs with and without monotonicity constraints, relative error increased with the standard deviation of noise, although the effect of the noise was substantially lower for the PINNs with monotonicity. On the other hand, the PINNs without monotonicity constraints exhibited consistently large relative error for all levels of noise. These observations indicate that monotonicity constraints are critical for ensuring stability and reliability when fitting noisy data. Therefore, PINNs without monotonic neural networks is not practically feasible for field applications.

The number of measurement locations at which simulated volumetric water content was sampled data were varied from 10 to 5 and 3 to investigate the effect of the sparsity of the training data. Figure 5 (**b**) illustrates that smaller relative error for soil water flux density ϵ^q was observed for denser training data. Although PINNs have been shown to be effective for sparse training data (Raissi et al., 2019; Tartakovsky et al., 2020), the PINNs for this application needs dense volumetric water content measurements (e.g., 2 cm interval).

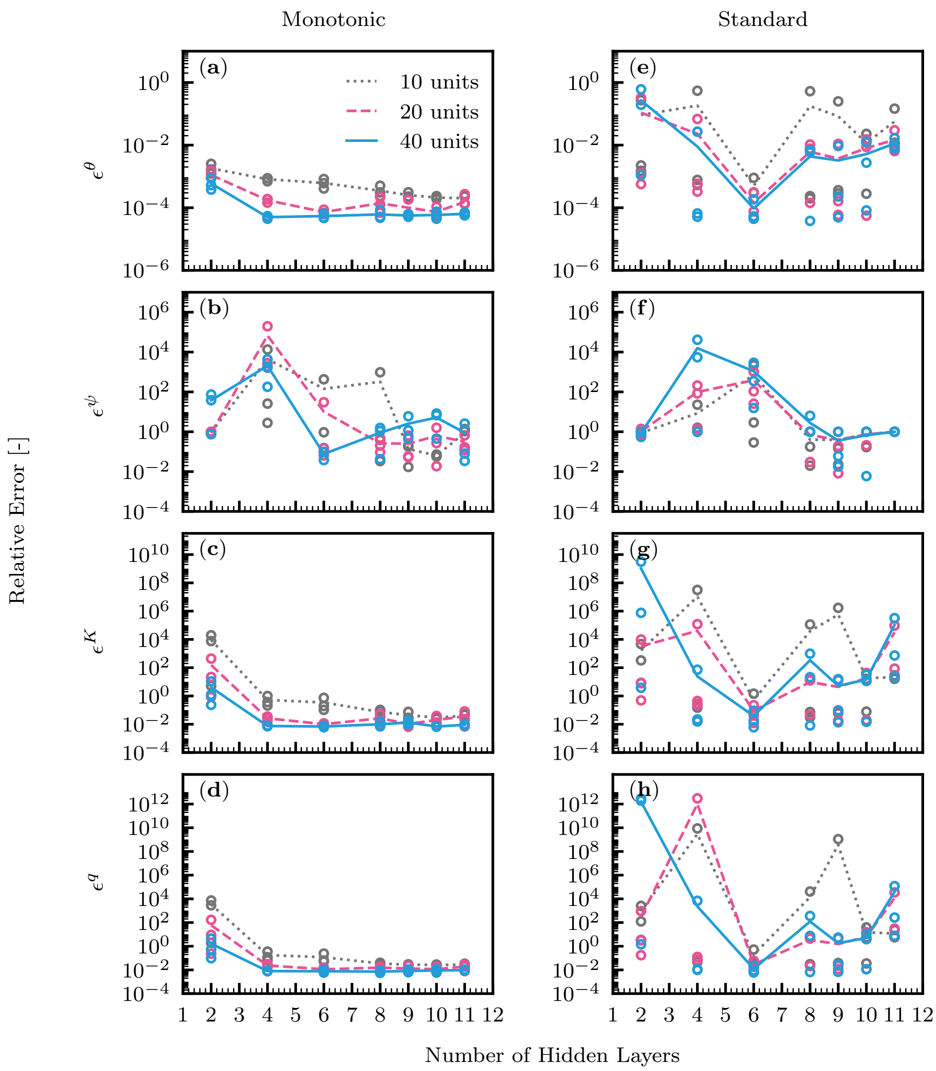

Figure 4. Relative error ϵ for volumetric water content θ, matric potential ψ, hydraulic conductivity K, and soil water flux density q for different numbers of hidden layers and units in the neural network for the predicted matric potential $\hat{f}_\psi$ (Figure 3 (**a**)); with (left column) and without monotonicity (right column). The architecture of the other two neural networks are set to 1 hidden layer with 20 units each. The lines represent the arithmetic mean of five replicates for each neural network architecture.

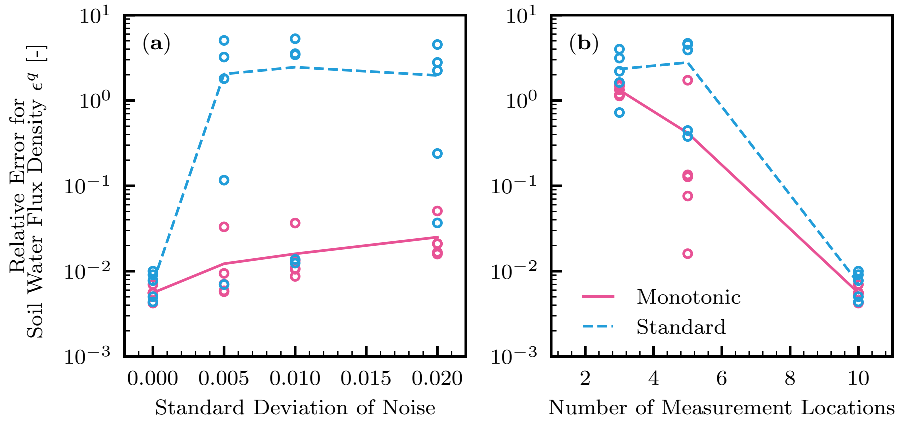

Figure 5. Relative error for soil water flux density ϵ^q for different values of standard deviation of noise (**a**) and measurement locations at which synthetic volumetric water content data were sampled (**b**). The number of measurement locations was varied from 10 (z = -1, -3, -5, -7, -9, -11, -13, -15, -17, -19 cm) to 5 ($z = -1, -5, -9, -13, -17$ cm) and 3 ($z = -1, -9, -17$ cm). The lines represent the arithmetic mean of five replicates.

.

4.3 Generalization Capability of PINNs

The generalization capability of the PINNs with and without monotonicity constraints was assessed with noisy synthetic volumetric water content data generated by HYDRUS-1D for three types of soils (sandy loam, loam, silt loam soil) with two different scenarios of upper boundary conditions (see Table 1). Table 3 shows relative error for volumetric water content ϵ^θ, matric potential ϵ^ψ, hydraulic conductivity ϵ^K, and soil water flux density ϵ^q. The PINNs without monotonicity constraints could not produce satisfactory results, which is shown by the large values of relative error for hydraulic conductivity ϵ^K and soil water flux density ϵ^q for both scenarios. This is mainly caused by the noise in the training data, which was indicated in Figure 5 (**a**). Also, poor generalization capability of the PINNs without monotonicity constraints is implied by the fact that higher relative error was observed for loam and silt loam soil. Therefore, in the following sections, we focus on the results of the PINNs with monotonicity constraints. While the trainings were conducted with five different random seeds initializing the weight parameters of the neural networks, we provide the results that show medium performance in terms of relative error for soil water flux density ϵ^q.

4.3.1 Volumetric water content

Figure 6 shows predicted volumetric water content by the PINNs with monotonicity constraints from noisy training data for sandy loam soil for the two scenarios. The PINNs could precisely capture the true distribution of soil moisture from the training data with the noise (standard deviation of 0.005). The PINNs could capture the distribution well for the other two soils as well (shown in Figure S2 and S3 in the supporting information).

Larger errors were observed when the upper boundary condition changed abruptly (e.g., $t = 1.5$ day for Scenario 2 in Figure 6 (**e**)). This indicated that the neural networks used in the study could not represent such a sharp change in soil moisture dynam-

Table 3. Relative error (arithmtic mean (± standard devation)) for volumetric water content ϵ^θ, matric potential ϵ^ψ, hydraulic conductivity ϵ^K, and soil water flux density ϵ^q for the PINNs with and without monotonicity constraints trained by noisy volumetric water content data for three soils (sandy loam, loam, silt loam soil) for two scenarios (Scenario 1 and 2). The arithmetic mean and standard deviation were calculated from five replicates.

Relative Error	Sandy Loam	Loam	Silt Loam
		PINNs with monotonicity constraints	
		Scenario 1	
ϵ^θ	$1.05(\pm0.75)\times10^{-4}$	$4.02(\pm0.26)\times10^{-5}$	$3.60(\pm0.48)\times10^{-5}$
ϵ^ψ	$4.21(\pm0.38)\times10^{-1}$	$6.79(\pm5.99)\times10^{2}$	$1.14(\pm1.17)\times10^{1}$
ϵ^K	$3.01(\pm4.78)\times10^{-2}$	$3.34(\pm0.63)\times10^{-2}$	$2.87(\pm0.23)\times10^{-1}$
ϵ^q	$1.22(\pm1.05)\times10^{-2}$	$1.55(\pm0.32)\times10^{-2}$	$2.27(\pm0.25)\times10^{-2}$
		Scenario 2	
ϵ^θ	$4.89(\pm0.34)\times10^{-5}$	$3.03(\pm0.30)\times10^{-5}$	$3.66(\pm2.51)\times10^{-5}$
ϵ^ψ	$4.19(\pm0.43)\times10^{-1}$	$9.42(\pm18.0)\times10^{-1}$	$1.17(\pm1.09)$
ϵ^K	$5.33(\pm0.70)\times10^{-3}$	$2.47(\pm0.74)\times10^{-2}$	$5.18(\pm3.98)\times10^{-1}$
ϵ^q	$5.48(\pm0.53)\times10^{-3}$	$1.01(\pm0.09)\times10^{-2}$	$3.49(\pm2.55)\times10^{-2}$
		PINNs without monotonicity constraints	
		Scenario 1	
ϵ^θ	$2.38(\pm2.27)\times10^{-3}$	$8.38(\pm9.01)\times10^{-4}$	$7.25(\pm5.80)\times10^{-4}$
ϵ^ψ	$1.13(\pm0.58)$	$1.19(\pm2.14)\times10^{1}$	$4.46(\pm7.14)$
ϵ^K	$5.98(\pm5.70)$	$1.08(\pm1.33)\times10^{5}$	$1.54(\pm1.36)\times10^{5}$
ϵ^q	$2.04(\pm1.92)$	$1.30(\pm1.61)\times10^{4}$	$1.15(\pm1.02)\times10^{4}$
		Scenario 2	
ϵ^θ	$1.50(\pm1.23)\times10^{-3}$	$3.13(\pm3.11)\times10^{-4}$	$3.19(\pm2.50)\times10^{-4}$
ϵ^ψ	$2.76(\pm3.75)$	$3.62(\pm5.26)$	$2.30(\pm2.74)$
ϵ^K	$2.02(\pm1.74)$	$1.11(\pm2.05)\times10^{4}$	$5.95(\pm6.51)\times10^{4}$
ϵ^q	$9.69(\pm8.30)\times10^{-1}$	$2.32(\pm4.30)\times10^{3}$	$7.84(\pm8.66)\times10^{3}$

ics. For the same reason, larger errors were observed just after the initial condition ($t =$ 0 day). Also, the PINNs could not reproduce the true volumetric water content at depths that are not covered in the training data (i.e., near the surface and lower than $z = -19$ cm). This means the PINNs could not extrapolate the volumetric water content data while it could interpolate. Similar trends were observed for the other two soils (see Figure S2 and S3 in the information).

4.3.2 Residual of RRE

The PINNs minimizes the data fitting error, as well as the residual of the RRE defined by Equation (15). The absolute value of the residual of the RRE for sandy loam soil at three times for the two scenarios is shown in Figure 7. The values in the spatial domain were small (less than 10^{-3}), which means the RRE was satisfied in the spatial domain of interest (i.e., (-20cm, 0cm]). Larger deviations from zero were observed near the surface and lower than the lowest virtual sensor ($z = -19cm$). This corresponds to the fact that the collocation points at which the residual of the RRE is evaluated were set to the measurement locations. This error may be minimized by distributing more collocation points in the spatial domain, including near the surface. Tartakovsky et al. (2020) reported that the accuracy of the PINNs improved if larger numbers of collocation points were provided. The drawback of increasing the number of collocation points is increased in computational demand. Further investigations are needed for seeking an efficient strategy to distribute the collocation points to achieve a better performance of the PINNs. The results for the other soils are provided in Figure S4 and S5 in the supporting information.

4.3.3 Water Retention Curves

Predicting matric potential from the noisy volumetric water content corresponds to estimating WRCs, which is one of the primary goals of the study. The PINNs with monotonicity constraints could not precisely predict the WRCs for the three soils, as shown in Figure 8. Especially, the prediction was not satisfactory for low and high volumetric water content, where the training data points were not provided. This suggests the difficulty in representing the two characteristics of WRCs by using a monotonic neural network: monotonicity and well defined upper and lower limits (saturation and dryness, respectively). This weakness of the current PINNs needs to be fixed in future research. Nevertheless, the predicted WRCs were surprisingly similar to the true WRCs in the middle range regardless of the fact that any actual value of matric potential was not used to train the PINNs.

How does the PINNs with monotonicity constraints learn WRCs from only volumetric water content data? A possible explanation is that matric potential is estimated from the gradient of matric potential $\partial\hat{\psi}/\partial z$, which is calculated in the residual of the RRE $\hat{r}$. Also, a matric potential of zero at saturation is implied by forcing matric potential to be negative while imposing the monotonically increasing relationship between matric potential and volumetric water content. These two explanations partly support the possibility that the PINNs with monotonicity constraints can predict WRCs from only volumetric water content if sufficient numbers and quality of training data are given.

4.3.4 Hydraulic Conductivity Functions

The estimated HCFs for the three soils for the two scenarios are shown in Figure 9. It should be noted that hydraulic conductivity is plotted against volumetric water content, not matric potential, as in Figure 1, because the estimated values of matric potential do not match the actual values, unlike volumetric water content.

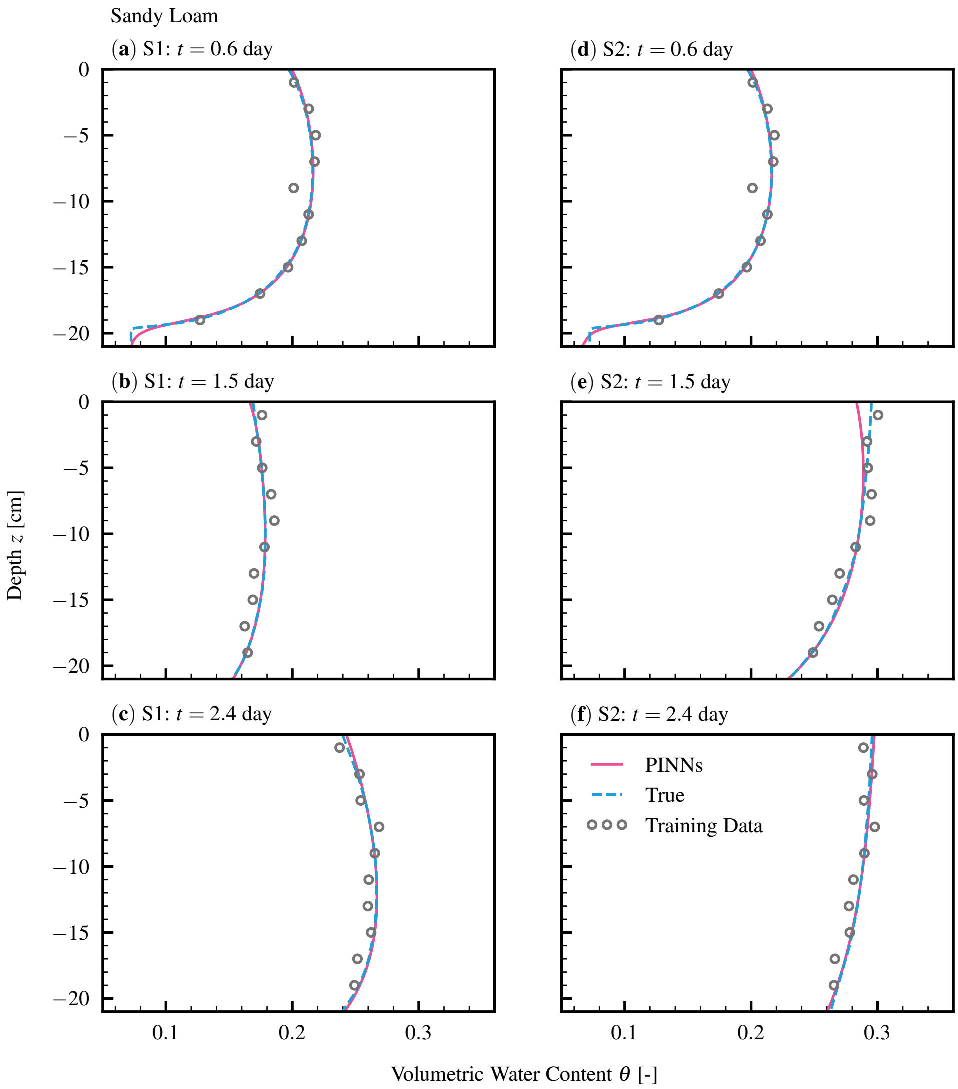

Figure 6. Predicted volumetric water content (PINNs) and noisy synthetic training data (Training Data) for sandy loam soil for the two scenarios at three different times. The dotted lines represent the synthetic data before adding the noise (True). Scenario 1 (S1): (**a**) $t = 0.6$ day, (**b**) $t = 1.5$ day, and (**c**) $t = 2.4$ day. Scenario 2 (S2): (**d**) $t = 0.6$ day, (**e**) $t = 1.5$, and (**f**) $t = 2.4$ day.

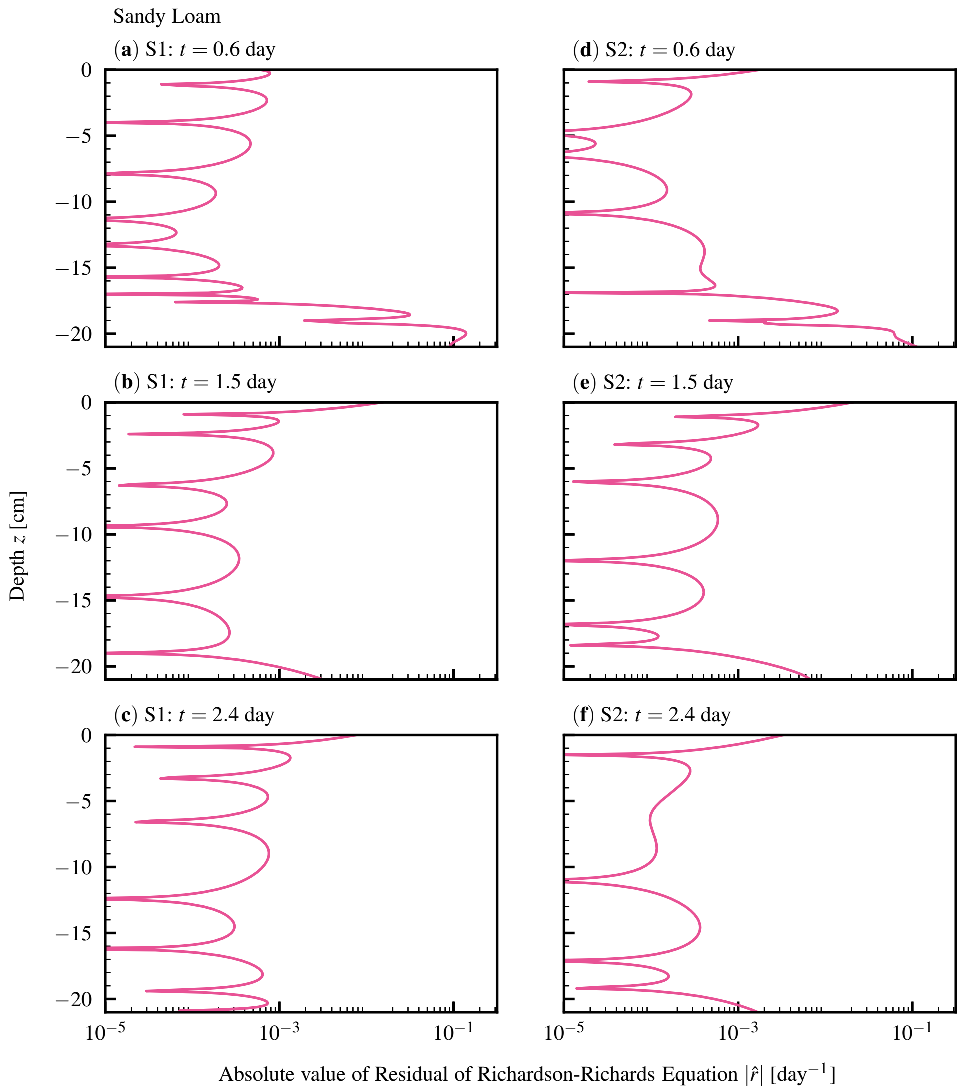

Figure 7. The absolute value of the residual of the Richardson-Richards equation at three different times for sandy loam soil for the two scenarios. Scenario 1 (S1): (**a**) $t = 0.6$ day, (**b**) $t = 1.5$ day, and (**c**) $t = 2.4$ day. Scenario 2 (S2): (**d**) $t = 0.6$ day, (**e**) $t = 1.5$, and (**f**) $t = 2.4$ day.

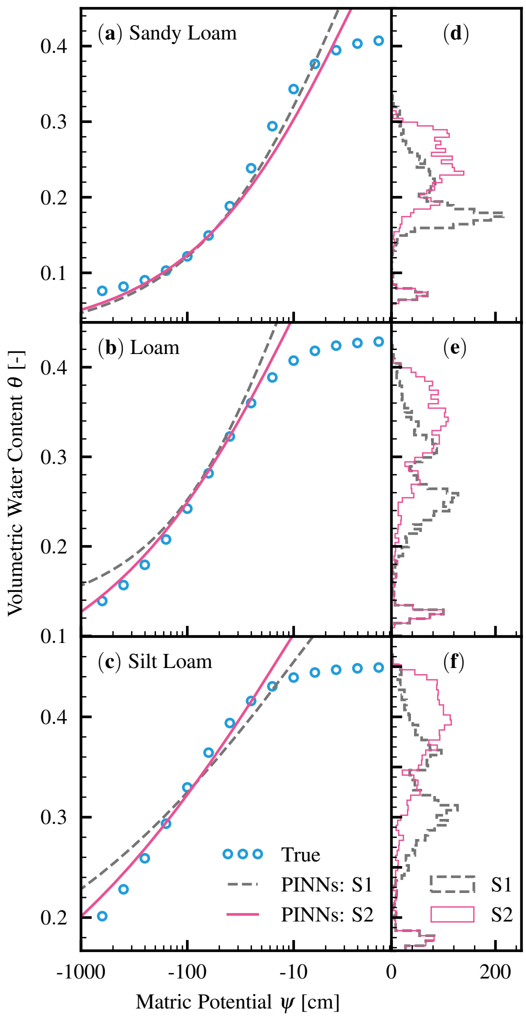

Figure 8. Comparison of true water retention curves (True) to the ones predicted by the PINNs with monotonicity constraints for the three soils for the two scenarios (S1: Scenario 1, S2: Scenario 2) with the histogram of the noisy training data. Water retention curve for (**a**) sandy loam, (**b**) loam, and (**c**) silt loam. Histogram of the training data for (**d**) sandy loam, (**e**) loam, and (**f**) silt loam.

The PINNs with monotonicity constraints could estimate the HCFs, especially for the range of the volumetric water content that is covered in the training data. On the other hand, the PINNs could not precisely extrapolate the HCFs; dryness and near saturation. As for a drier range of HCFs, although some of the training data are distributed in the range, they did not contribute to the learning of the HCFs. This is caused by the fact that these data correspond to the initial volumetric content, which increased rapidly due to the prescribed upper boundary conditions, and the PINNs could not capture the abrupt change well.

Hydraulic conductivity was estimated through minimizing the residual of the RRE, which contains hydraulic conductivity (see Equation (15)). Tartakovsky et al. (2020) reported that HCFs could be estimated from matric potential measurements using PINNs with the time-independent RRE. Considering our result and their findings, we conclude that hydraulic conductivity can be estimated from only either volumetric water content or matric potential.

The advantage of the PINNs approach over the other studies to estimate HCFs was that we did not assume any information about HCFs a priori, such as saturated water content and saturated hydraulic conductivity. Also, the neural network for HCFs is separated from WRCs, which prevents the error in WRCs from propagating into HCFs. Considering these advantages, we conclude that the current framework of PINNs for the RRE is a powerful way to estimate HCFs from only volumetric water content data, which has never been attained to the best of our knowledge.

4.3.5 Soil Water Flux Density

In this section, we will show that the current PINNs framework with monotonicity constraints can be used to estimate soil water flux density from noisy volumetric water content data.

The comparison of the estimated soil water flux density to the true one calculated by HYDRUS-1D at three different depths ($z = -1, -9, -17$ cm) for sandy loam soil for the two scenarios is shown in Figure 10. It was found that the PINNs with monotonicity constraints could estimate soil water flux density from noisy volumetric water content measurements. Larger errors were observed at wetting fronts and near the surface, where soil water flux density changed abruptly. Although larger relative error was observed for loam and silt loam (see Table 3), especially for Scenario 1, the PINNs with monotonicity constraints could reasonably capture the trend of soil water flux density, which is shown in Figure S6 and S7 in the supporting information.

The advantage of this approach over the available heat pulse method (Kamai et al., 2008, 2010) is that this method can estimate soil water flux density lower than 1 cm day^{-1} (see Figure S8, S9, and S10 in the supporting information). Because continuous measurement of volumetric water content at different depths is becoming popular with an advanced TDR array (Sheng et al., 2017), this PINNs approach can be used to estimate soil water flux density in fields. This finding has a significant implication in the application of land surface modeling, where soil water flux density near the surface is critical.

5 Summary and Conclusions

A framework of estimating soil hydraulic functions or constitutive relationships of the Richardson-Richards equation (RRE) (i.e., water retention curves (WRCs) and hydraulic conductivity functions (HCFs)) from noisy volumetric water content measurements was proposed using physics-informed neural networks (PINNs). The PINNs for the RRE was designed by endowing the neural networks with the monotonicity of WRCs

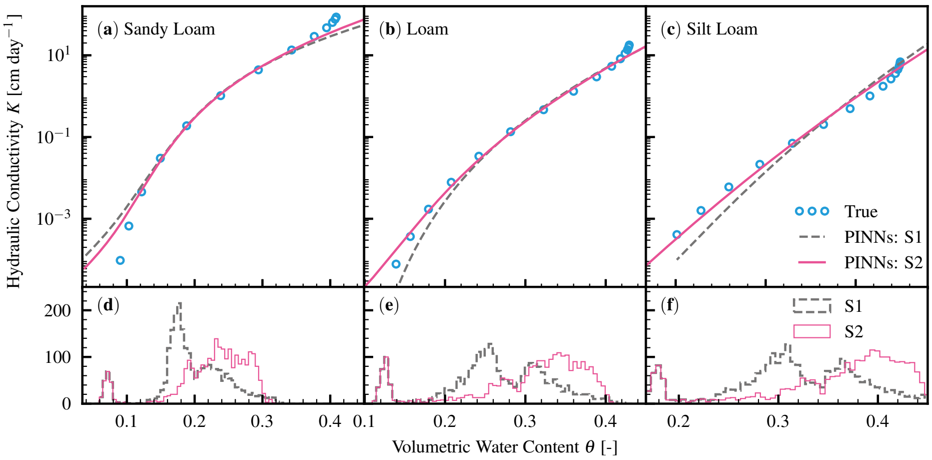

Figure 9. Comparison of true hydraulic conductivity functions (True) to the ones predicted by the PINNs with monotonicity constraints for the three soils for the two scenarios (S1: Scenario 1, S2: Scenario 2) with the histogram of the noisy training data. Hydraulic conductivity function for (**a**) sandy loam, (**b**) loam, and (**c**) silt loam. Histogram of the training data for (**d**) sandy loam, (**e**) loam, and (**f**) silt loam.

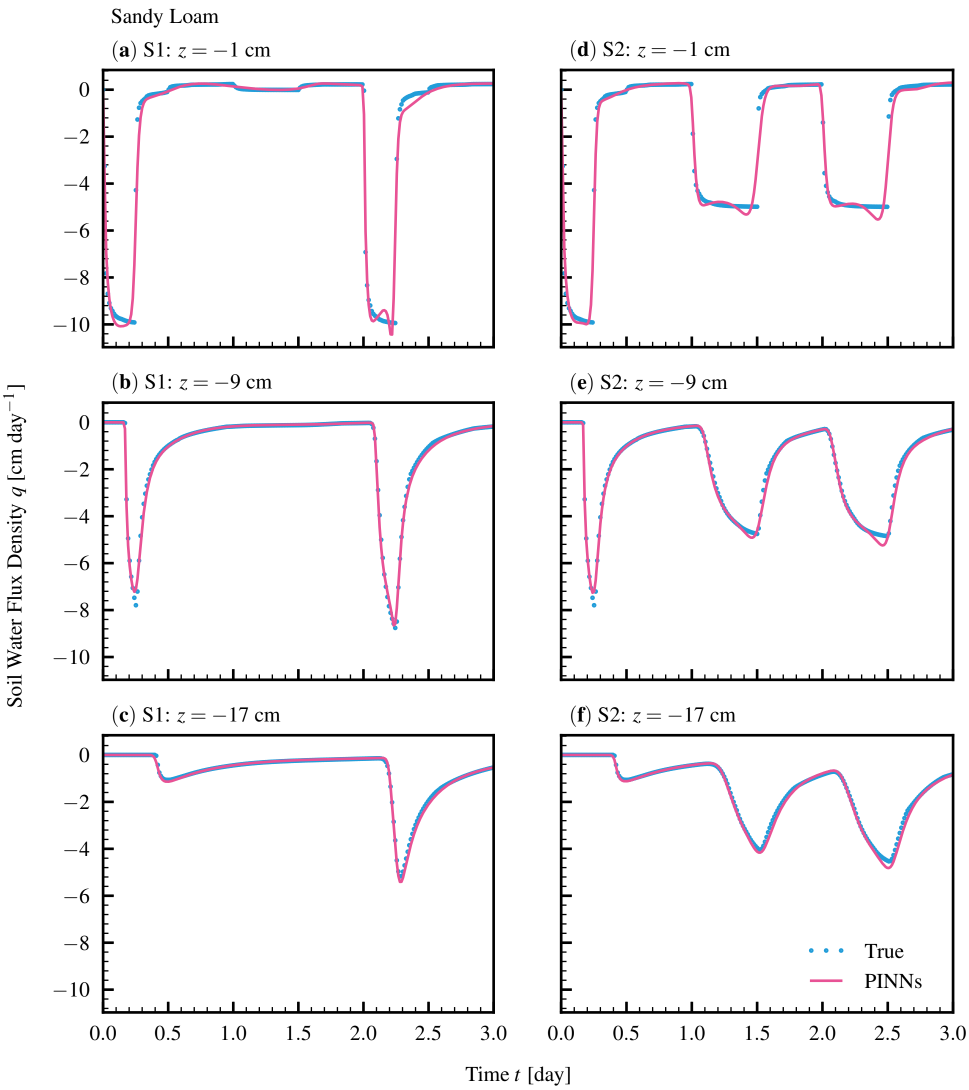

Figure 10. Estimated soil water flux density against the true one at three different depths for sandy loam soil. Scenario 1 (S1): (**a**) $z = -1$ cm, (**b**) $z = -9$ cm, and (**c**) $z = -17$ cm. Scenario 2 (S2): (**d**) $z = -1$ cm, (**e**) $z = -9$ cm, and (**f**) $z = -17$ cm.

and HCFs. To demonstrate the effectiveness of incorporating monotonicity constraints into the PINNs, we compared the performance of the PINNs between with and without monotonicity constraints. As a result, the PINNs with monotonicity constraints has a great advantage over the PINNs without monotonicity constraints in terms of its high ability to prevent overfitting and reliability of the results for noisy training data.

The PINNs, with and without monotonicity constraints, were trained using synthetic volumetric water content data for three distinct soil textures (sandy loam, loam, and silt loam) with Gaussian noise. The generalization ability of the framework was assessed in terms of its ability to estimate WRCs, HCFs, and soil water flux densities. The PINNs without monotonicity constraints could not produce satisfactory results. On the other hand, the PINNs with monotonicity constraints could estimate true soil moisture dynamics from noisy synthetic data for all types of soil. In terms of WRCs, the PINNs with monotonicity constraints could not precisely estimate the true WRCs. However, the estimated WRCs were surprisingly similar to the true ones in the middle range regardless of the fact that any matric potential data was provided. Unlike WRCs, the PINNs with monotonicity constraints could predict the HCFs well, especially for the range that is covered in the training data.

It was demonstrated that employing monotonic neural networks in the PINNs to represent WRCs and HCFs improved the ability of the PINNs to prevent overfitting the training data. Furthermore, the PINNs with monotonicity constraints is shown to have better durability against noisy data than the PINNs without monotonicity constraints.

It was illustrated that the PINNs with monotonicity constraints has a great potential to predict constitutive relationships of the RRE and soil water flux density from only noisy volumetric water content data in fields. The advantage of this method is the current PINNs framework does not need initial and boundary conditions and any information about the HCF a priori. The current framework must be tested with real experimental data for homogeneous soil in future research.

The PINNs with monotonicity constraints could estimate true soil water flux density from noisy synthetic volumetric water content data at different depths. At present, the only measurement technique for measuring soil water flux density is using heat flux sensors, which is limited to soil water flux density larger than 1 cm day^{-1}. The proposed method has the potential for determining soil water flux density over a broader range.

Acronyms

HCFs Hydraulic Conductivity Functions
PDE Partial Differential Equation
PINNs Physics-Informed Neural Networks
RRE Richardson-Richards Equation
WRCs Water Retention Curves
VWC Volumetric Water Content

Notation

:= Equal by definition
$\hat{}$ Hat indicating predicted values or functions (e.g., $\hat{y}$)
$^{(i)}$ Superscript (i) denoting ith data (e.g., $\theta^{(i)}$)
$^{[L]}$ Superscript [L] denoting Lth layer
$\mathbf{a}^{[L]} \in \mathbb{R}^{n^{[L]}}$ Vector value for the Lth layer consisting of $n^{[L]}$ units
$\boldsymbol{b}$ Bias vector
$\hat{\boldsymbol{f}}$ Neural network

$\boldsymbol{g}$ Activation function
$\boldsymbol{h}$ Output function
$\boldsymbol{K}$ Hydraulic conductivity [L T^{-1}]
$\boldsymbol{K_s}$ Mualem-van Genuchen parameter
$\boldsymbol{\mathcal{L}}$ Loss function
$\boldsymbol{l}$ Mualem-van Genuchen parameter
$\boldsymbol{N}$ Number of data points
$\boldsymbol{n}$ Mualem-van Genuchen parameter
$\boldsymbol{n_i}$ Number of size a vector, as in n_x and n_y
$\boldsymbol{n^{[L]}}$ Number of units in Lth layer of a neural network
$\boldsymbol{q}$ Soil water flux density [L T^{-1}]
$\boldsymbol{\hat{r}}$ Residual of the Richardson-Richards euqation
$\boldsymbol{S_e}$ Effective saturation
$\boldsymbol{t}$ Time [T]
$\boldsymbol{W}$ Weight matrix
$\mathbf{x} \in \mathbb{R}^{\boldsymbol{n_x}}$ Input vector for the size of the input n_x
$\mathbf{y} \in \mathbb{R}^{\boldsymbol{n_y}}$ Output vector for the size of the output n_y
$\boldsymbol{z}$ Vertical coordinate or depth (positive upward) [L]
$\boldsymbol{\alpha}$ Mualem-van Genuchen parameter [L^{-1}]
$\boldsymbol{\epsilon}$ Relative error
$\boldsymbol{\theta}$ Volumetric water content [L^3 L^{-3}]
$\boldsymbol{\theta_r}$ Mualem-van Genuchen parameter [L^3 L^{-3}]
$\boldsymbol{\theta_s}$ Mualem-van Genuchen parameter [L^3 L^{-3}]
$\boldsymbol{\psi}$ Matric potential of water in the soil [L]
$\boldsymbol{\psi_{log}}$ Matric potential in logarithmic scale

Acknowledgments

The publicly available code for physics-informed neural networks provided by Dr. Maziar Raissi (University of Colorado Boulder) was instrumental in the development of our model. We are indebted to anonymous reviewers for improving this manuscript. The dataset used in this study is available through Dyrad (https://datadryad.org/stash; doi:10.6071/M3T376).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous distributed systems.* Retrieved from `https://www.tensorflow.org/`

Assouline, S. (2006). Modeling the relationship between soil bulk density and the hydraulic conductivity function. *Vadose Zone Journal*, *5*, 697–705. doi: 10.2136/vzj2005.0084

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., & Tuller, M. (2019). Ground, proximal, and satellite remote sensing of soil moisture. *Reviews of Geophysics*, *57*, 530–616. doi: 10.1029/2018RG000618

Bitterlich, S., Durner, W., Iden, S. C., & Knabner, P. (2004). Inverse estimation of the unsaturated soil hydraulic properties from column outflow experiments using free-form parameterizations. *Vadose Zone Journal*, *3*, 971–981. doi: 10.2113/3.3.971

Brooks, R. H., & Corey, A. T. (1964). *Hydraulic properties of porous media.* Fort Collins, CO, USA.

Buckingham, E. (1907). *Studies on the movement of soil moisture* (Vol. 38). Washington, DC, USA.

Burdine, N. T. (1953). Relative permeability calculations from pore size distribution data. *Society of Petroleum Engineers*, *5*(3). doi: 10.2118/225-G

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *Journal of Scientific Computing*, *16*(5), 1190–1208. doi: 10.1137/0916069

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control Signals Systems*, *2*, 303–314. doi: 10.1007/BF02836480

Daniels, H., & Velikova, M. (2010). Monotone and partially monotone neural networks. *IEEE Transactions on Neural Networks*, *21*(6), 906–917. doi: 10.1109/TNN.2010.2044803

Degré, A., van der Ploeg, M. J., Caldwell, T., & Gooren, H. P. A. (2017). Comparison of soil water potential sensors: A drying experiment. *Vadose Zone Journal*, *16*(4), 1–8. doi: 10.2136/vzj2016.08.0067

Durner, W. (1994). Hydraulic conductivity estimation for soils with heterogeneous pore structure. *Water Resources Research*, *30*(2), 211–223. doi: 10.1029/93WR02676

Durner, W., Jansen, U., & Iden, S. C. (2008). Effective hydraulic properties of layered soils at the lysimeter scale determined by inverse modelling. *European Journal of Soil Science*, *59*, 114–124. doi: 10.1111/j.1365-2389.2007.00972.x

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*, 1257–1269. doi: 10.2136/sssaj2017.02.0058

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* Cambridge, MA, USA: The MIT Press.

He, Q., Barajas-Solano, D., Tartakovsky, G. D., Alexandre, M., & Tartakovsky, A. M. (2020). Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Advances in Water Resources*, *141*. doi: 10.1016/j.advwatres.2020.103610

Hopmans, J. W., Šimůnek, J., Romano, N., & Durner, W. (2002). Simultaneous determination of water transmission and retention properties. Inverse methods. In J. H. Dane & G. C. Topp (Eds.), *Methods of soil analysis, part 4, physical methods* (pp. 963–1008). Madison, WI, USA: Soil Science Society of America.

Iden, S. C., & Durner, W. (2007). Free-form estimation of the unsaturated soil hydraulic properties by inverse modeling using global optimization. *Water Resources Research*, *43*, 1–12. doi: 10.1029/2006WR005845

Kamai, T., Tuli, A., Kluitenberg, G. J., & Hopmans, J. W. (2008). Soil water flux density measurements near 1 cm d1 using an improved heat pulse probe design. *Water Resources Research*, *44*, 1–12. doi: 10.1029/2008wr007036

Kamai, T., Tuli, A., Kluitenberg, G. J., & Hopmans, J. W. (2010). Correction to "Soil water flux density measurements near 1 cm d1 using an improved heat pulse probe design". *Water Resources Research*, *46*, 1–5. doi: 10.1029/2010WR009423

Kingma, D. P., & Ba, J. B. (2014). Adam: A method for stochastic optimization. *arXiv preprint*, 1–15. Retrieved from `https://arxiv.org/pdf/1412.6980.pdf`

Kosugi, K. (1996). Lognormal distribution model for unsaturated soil hydraulic properties. *Water Resources Research*, *32*(9), 2697–2703. doi: 10.1029/96WR01776

Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, *12*(3), 513–522. doi: 10.1029/WR012i003p00513

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization.* New York, NY, USA: Springer. doi: 10.1007/978-0-387-40065-5

Rai, P. K., & Tripathi, S. (2019). Gaussian process for estimating parameters of partial differential equations and its application to the Richards equation. *Stochastic Environmental Research and Risk Assessment*, *33*, 1629–1649. doi: 10.1007/s00477-019-01709-8

Raissi, M., & Karniadakis, G. E. (2018). Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, *357*, 125–141. doi: 10.1016/j.jcp.2017.11.039

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. doi: 10.1016/j.jcp.2018.10.045

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, *1*, 318–333. doi: 10.1063/1.1745010

Richardson, L. F. (1922). *Weather prediction by numerical process.* Cambridge, United Kingdom: Cambridge University Press.

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., … Wendroth, O. (2008). Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, *7*, 358–389. doi: 10.2136/vzj2007.0143

Sheng, W., Zhou, R., Sadeghi, M., Babaeian, E., Robinson, D. A., Tuller, M., & Jones, S. B. (2017). A TDR array probe for monitoring near-surface soil moisture distribution. *Vadose Zone Journal*, *16*(4), 1–8. doi: 10.2136/vzj2016.11.0112

Šimůnek, J., Šejna, M., Saito, H., Sakai, M., & van Genuchten, M. T. (2013). *The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably saturated media, Version 4.17* (Tech. Rep.). Riverside, CA, USA: Department of Environmental Sciences, University of California Riverside, Riverside.

Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., & Barajas-Solano, D. (2020). Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems. *Water Resources Research*. doi: 10.1029/2019wr026731

Tuller, M., & Or, D. (2001). Hydraulic conductivity of variably saturated porous media: Film and corner flow in angular pore space. *Water Resources Research*, *37*(5), 1257–1276. doi: 10.1029/2000WR900328

van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America*, *44*, 892–898. doi: 10.1016/j.pan.2017.07.214

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … Contributors, S. . . (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. doi: https://doi.org/10.1038/s41592-019-0686-2

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. *Wiley Interdisciplinary Reviews: Water*, *6*, 1–23. doi: 10.1002/wat2.1364