

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).

66 Therefore, it is indispensable to estimate WRCs and HCFs using time-series data
67 from field experiments and the inverse solution of the RRE. Commonly, the inverse prob-
68 lem requires finding the parameters of the constitutive relationships that best describe
69 observed time-series data. In principle, it is possible to fit WRCs and HCFs indepen-
70 dently, albeit at the expense of significant increase in the tunable parameters. Several
71 studies also employed free-form functions to estimate WRCs and HCFs (Bitterlich et al.,
72 2004; Iden & Durner, 2007). Inverse methods for characterizing soil hydraulic proper-
73 ties often involve the repeated solution of the forward problem, which requires knowl-
74 edge of the relevant initial and boundary conditions of the RRE. Global optimization
75 algorithm (Durner et al., 2008) and Gaussian processes (Rai & Tripathi, 2019) are other
76 approaches used to find the best-fitted constitutive relationships.

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

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

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

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

112 The robustness of using monotonic neural networks to represent WRCs and HCFs
113 in the PINNs is demonstrated by comparing the results with those from the PINNs that
114 lacks the monotonicity constraints. The performance of the framework was further tested
115 by introducing varying degrees of noise to the synthetic volumetric water content data,
116 altering the spacing between the locations at which volumetric water content data were
117 sampled, using different initial weight parameters of the neural networks. The general-

118 ization capability of the framework was investigated by training the PINNs with volu-
 119 metric water content data for three soils (sandy loam, loam, and silt loam soil) and for
 120 two different scenarios of the upper boundary condition. Finally, we show the potential
 121 application of the PINNs to estimate soil water flux density using only an array of soil
 122 moisture sensors.

123 2 Background

124 2.1 Richardson-Richards Equation

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

$$\frac{\partial \theta}{\partial t} = -\frac{\partial q}{\partial z}, \quad (1)$$

128 where θ is volumetric water content [$\text{L}^3 \text{L}^{-3}$]; t is time [T]; z is vertical coordinate (pos-
 129 itive upward) [L]; q is soil water flux density [L T^{-1}]. Soil water flux density q is related
 130 to matric potential of water in the soil ψ [L] through the Buckingham-Darcy law (Buckingham,
 131 1907):

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

132 where K is hydraulic conductivity [L T^{-1}]. The two equations (Equation (1) and (2))
 133 are combined to derive the Richardson-Richards equation (RRE) (Richardson, 1922; Richards,
 134 1931):

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

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

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

138 The functions $\theta(\psi)$ and $K(\psi)$ are called constitutive relationships of the RRE and re-
 139 ferred to as water retention curves (WRCs) and hydraulic conductivity functions (HCFs),
 140 respectively. WRCs and HCFs are commonly expressed by parametric models (e.g., Brooks
 141 and Corey (1964); van Genuchten (1980); Durner (1994); Kosugi (1996); Tuller and Or
 142 (2001); Assouline (2006)). The WRCs and HCFs for three types of soil (sandy loam, loam,
 143 and silt loam soil) using the Mualem-van Genuchten model (van Genuchten, 1980) are
 144 shown in Figure 1. As shown in the figure, both WRCs and HCFs are monotonically in-
 145 creasing functions with respect to matric potential ψ , which is an accepted physical prin-
 146 ciple of water movement in soils. The monotonicity of WRCs and HCFs will be employed
 147 to design the architecture of the neural networks in this study later on.

148 2.2 Feedforward Neural Networks

149 A standard fully-connected feedforward neural network with three layers (one hid-
 150 den layer) is introduced here for readers who are not well versed in the topic. The read-
 151 ers should refer to textbooks (e.g., Goodfellow et al. (2016)) for more general explana-
 152 tions.

153 Given a training dataset $\{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}$, where superscript (i) denotes the i th train-
 154 ing 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
 155 vector for the size of the output n_y , a neural network \hat{f} is a mathematical function map-
 156 ping the input vector $\mathbf{x}^{(i)}$ to predicted output vector $\hat{\mathbf{y}}^{(i)} \in \mathbb{R}^{n_y}$:

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

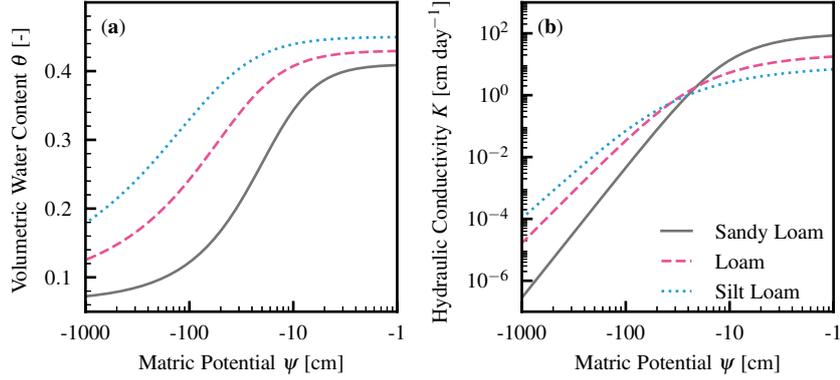


Figure 1. Constitutive relationships for three types of soil (sandy loam, loam, and silt loam soil) generated by using the Mualem-van Genuchten model (van Genuchten, 1980). (a) Water retention curves (WRCs). (b) Hydraulic conductivity functions (HCFs).

157 The hat operator represents prediction throughout the paper. The inside of the neural
 158 network \hat{f} is commonly represented by layers of units (or neurons), as shown in Figure
 159 2. Herein, $\mathbf{a}^{[L]} \in \mathbb{R}^{n^{[L]}}$ denotes the vector value for the L th layer of the neural network,
 160 where the L th layer is composed of $n^{[L]}$ units. To calculate the predicted output vector
 161 $\hat{\mathbf{y}}^{(i)}$, the input vector $\mathbf{x}^{(i)}$ is entered in the first layer:

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

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

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

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

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

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

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

171 where the number of the units in the last layer $n^{[3]}$ is equal to n_y . In this study, the sig-
 172 moid function (Figure 2 (c)) and the exponential function (Figure 2 (d)) are used as out-
 173 put functions.

174 The collection of the weight matrices $\mathbf{W} = \{\mathbf{W}^{[1]}, \mathbf{W}^{[2]}\}$ and bias vectors $\mathbf{b} =$
 175 $\{\mathbf{b}^{[1]}, \mathbf{b}^{[2]}\}$ are the parameters of the neural network, which are estimated by minimiz-
 176 ing a loss function comprising of the output vector $\mathbf{y}^{(i)}$ (training data) and the predicted

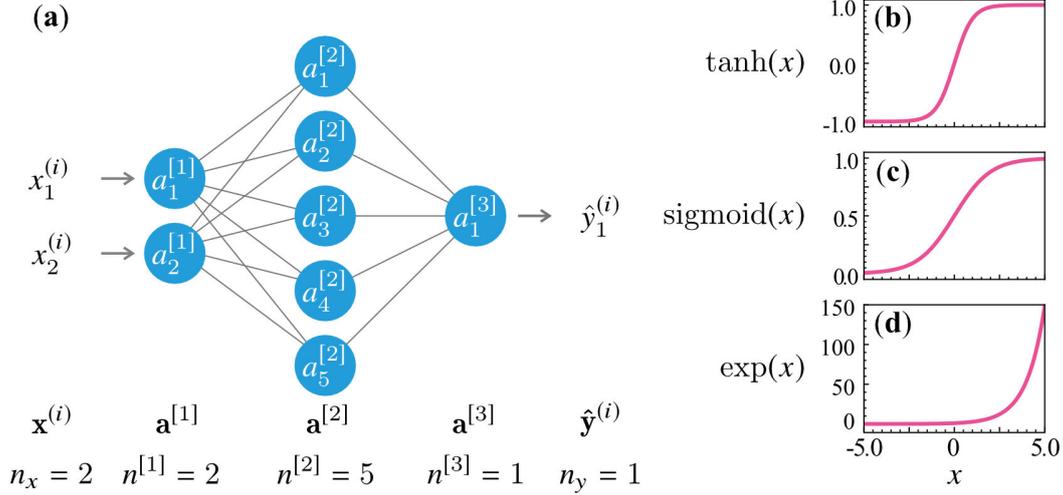


Figure 2. A fully-connected feedforward neural network consisting of three layers (one hidden layer) with activation and output functions. (a) A fully-connected feedforward neural network consisting of the input layer with two units, the hidden layer with five units, and the output layer with one unit.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

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

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

184 In the next section, three fully-connected feedforward neural networks are combined
 185 to construct physics-informed neural networks (PINNs) for the RRE, and the loss func-
 186 tion for the PINNs framework is defined to estimate WRCs and HCFs from volumet-
 187 ric water content measurements.

188 3 Methods

189 3.1 Physics-Informed Neural Networks with Monotonicity Constraints 190 for RRE

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

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

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

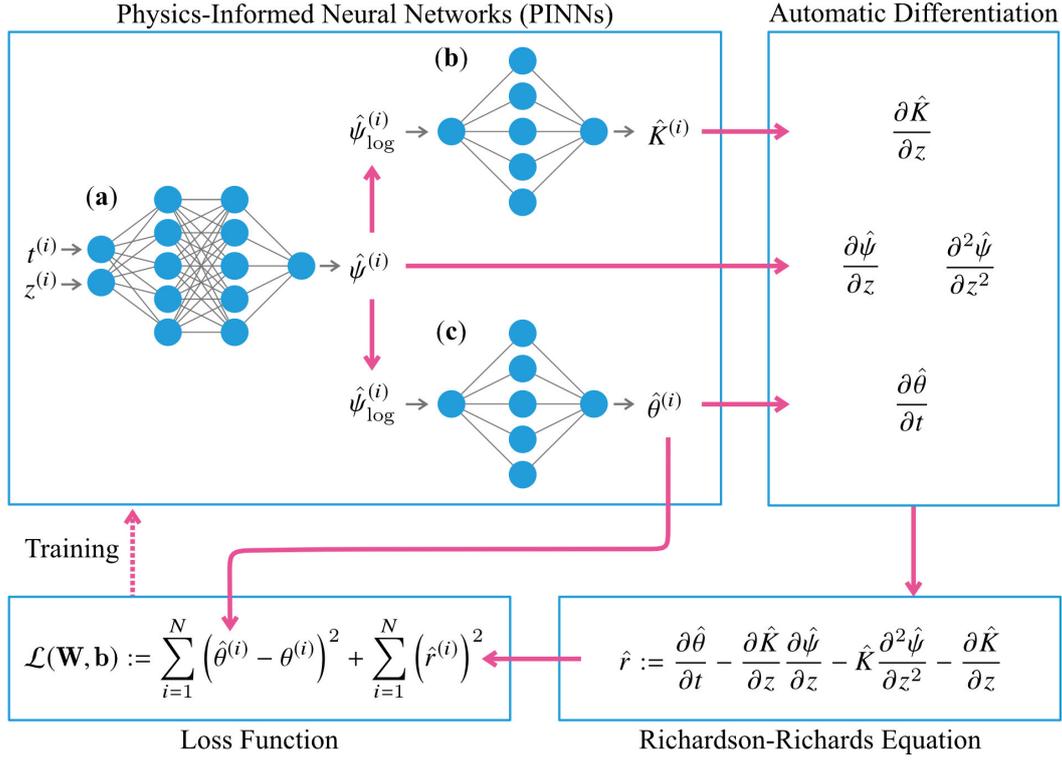


Figure 3. Physics-informed neural networks (PINNs) for the Richardson-Richards equation consisting of three fully-connected feedforward neural networks to predict (a) matric potential $\hat{\psi}$, (b) hydraulic conductivity \hat{K} , and (c) volumetric water content $\hat{\theta}$. The number of layers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

200 where \mathbf{W}_ψ and \mathbf{b}_ψ 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.

205 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}_θ , \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:

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

211 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:

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

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

214 The tanh function is used as the activation function for both neural networks. The output functions for \hat{f}_θ and \hat{f}_K are the sigmoid function and the exponential function, re-

216 spectively to ensure predicted volumetric water content between 0 and 1 and positive
217 predicted hydraulic conductivity (see Figure 2 (c) and (d)).

218 To embrace the monotonicity of WRCs and HCFs, the weight parameters \mathbf{W}_ψ and
219 \mathbf{W}_K are constrained to be non-negative so that \hat{f}_θ and \hat{f}_K are monotonically increas-
220 ing functions with respect to the predicted matric potential $\hat{\psi}$ (Daniels & Velikova, 2010).
221 This type of neural networks is called (totally) monotonic neural networks, where the
222 output values depend monotonically on all the variables in the input vector. It is known
223 that a three-layer fully-connected feedforward neural network with non-negative weights
224 can arbitrarily approximate any monotonic scalar functions (Daniels & Velikova, 2010).
225 Readers interested in monotonic neural networks should refer to Daniels and Velikova
226 (2010), where various types of monotonic neural networks are explained.

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

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

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

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

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

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

244 To calculate the residual of the RRE \hat{r} at all the data points, the derivatives (i.e.,
245 $\frac{\partial \hat{\theta}}{\partial t}$, $\frac{\partial \hat{\psi}}{\partial z}$, $\frac{\partial^2 \hat{\psi}}{\partial z^2}$, $\frac{\partial \hat{K}}{\partial z}$) are evaluated at the data points by using automatic differentiation (Nocedal
246 & Wright, 2006). It should be noted that the residual of the RRE \hat{r} can be evaluated
247 at any point in the domain (called collocation points). However, we forced the colloca-
248 tion points to be the same as the measurement locations.

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

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

252 The optimization problem was solved by the Adam algorithm (Kingma & Ba, 2014) fol-
253 lowed by the L-BFGS-B algorithm (Byrd et al., 1995). This two-step training procedure
254 has been reported to be effective to train PINNs (Raissi et al., 2019; He et al., 2020). In
255 our implementation, the default settings of the Adam optimizer in TensorFlow (Abadi
256 et al., 2015) was used until 300,000 iterations finished. Then, the L-BFGS-B optimizer
257 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

258 $maxfun = 50,000$, $ftol = 2.220446049250313 \times 10^{-16}$, and the default values for the
 259 other parameters was applied to achieve the convergence of the loss function. The in-
 260 vestigation on the hyperparameters of those optimization algorithms is beyond the scope
 261 of the paper. This PINNs framework for the RRE was implemented through TensorFlow
 262 1.14 (Abadi et al., 2015), and the source code is available on https://github.com/ToshiyukiBandai/PINNs_RRE.

263 3.2 Synthetic Data Generated by HYDRUS-1D

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

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

$$\frac{\partial \psi}{\partial z} = 0. \quad (17)$$

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

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

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

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

279 where θ_r , θ_s , α , n , K_s , and l are the Mualem-van Genuchten fitting parameters; $m = 1 -$
 280 $1/n$; and the effective saturation S_e is defined as

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r}. \quad (20)$$

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

Table 2. The Mualem-van Genuchten 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

283

3.3 Determination of Architecture of Neural Networks

284

285

286

287

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.

288

289

290

291

292

293

First, we set the number of hidden layers and units of the two neural networks, \hat{f}_θ 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}_ψ (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.

294

295

296

Second, the number of hidden layers and units of the other two neural networks, \hat{f}_θ 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.

297

298

299

300

301

302

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.

303

304

305

306

307

308

309

310

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

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

311

312

313

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.

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

316 Because the results of training PINNs were affected by the initial values of the weight
 317 parameters of the neural networks determined by Xavier initialization (Glorot & Ben-
 318 gio, 2010), three different random seeds were used in the code, and three replicates were
 319 obtained for each of those combinations of the number of hidden layers and units. As
 320 a result, 63 trainings for \hat{f}_ψ (Figure 3 (a)) and 243 ones for \hat{f}_θ (Figure 3 (c)) and \hat{f}_K (Fig-
 321 ure 3 (b)) were conducted to determine their architecture for the PINNs both with and
 322 without monotonicity.

323 3.4 Application of PINNs to Various Datasets

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

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

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

339 Those training data were applied to the PINNs with and without monotonicity con-
 340 straints, and the results were evaluated in terms of relative errors defined in Equation
 341 (21). For all the cases above, five different random seeds were set in the code to inves-
 342 tigate the effects of neural network initialization on the results.

343 4 Results and Discussions

344 4.1 Architecture of Neural Networks in PINNs

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

356 The lowest arithmetic mean of relative error was observed when the number of hid-
 357 den layers is 4 for volumetric water content θ , 6 for hydraulic conductivity K , and 8 for
 358 soil water flux density q when the number of units is 40. Clear trends were not obtained
 359 for relative error for matric potential ϵ_ψ . Because relative error for soil water flux den-
 360 sity q reflects the predictive accuracy of the PINNs for both matric potential ψ and hy-

361 hydraulic conductivity K fields, we set the neural network for the predicted matric potential \hat{f}_ψ to 8 layers with 40 units.
362

363 For the PINNs without monotonicity constraints (right column), the architecture
364 of the neural network for the predicted matric potential \hat{f}_ψ was set to 6 hidden layers
365 with 40 units, which coincides with the lowest relative error of soil water flux density ϵ_q .
366 We observed a non-linear correlation between the number of hidden layers and relative
367 error for volumetric water content θ , hydraulic conductivity K , and soil water flux den-
368 sity q ; relative error reached the lowest when the number of hidden layers was 6 and in-
369 creased again. This is clear evidence that the PINNs without monotonicity was overfit-
370 ting the training data. On the other hand, such a non-linear behavior was minimized for
371 the PINNs with monotonicity constraints, which means imposing monotonicity can pre-
372 vent the PINNs from overfitting the training data. In addition, the variability of rela-
373 tive errors between different initializations of the neural networks was lower for the PINNs
374 with monotonic neural networks than the PINNs with non-monotonic neural networks.
375 This further demonstrates the benefit of the monotonicity constraints in improving the
376 stability and reliability of the training.

377 After determining the architecture of the neural network for the predicted matric
378 potential \hat{f}_ψ , the number of hidden layers and units for the other two neural networks,
379 \hat{f}_θ and \hat{f}_K , was varied. We did not observe clear trends of relative error for different neu-
380 ral network architectures for the PINNs with and without monotonicity (see Table S1
381 and Figure S1 in the supplementary information). However, the performance of the PINNs
382 without monotonicity constraints was much more sensitive to the neural network archi-
383 tecture. This implied that incorporating monotonicity constraints stabilized the train-
384 ing, which enabled us to determine the neural network structure easier than the PINNs
385 without monotonicity constraints. As a result, the architecture of the two neural net-
386 works was set as follows: 1 hidden layer with 40 units for the PINNs with monotonic-
387 ity and 3 hidden layers with 40 units for without monotonicity for the neural network
388 for the predicted volumetric water content \hat{f}_θ ; 3 hidden layer with 40 units for PINNs
389 with monotonicity and 2 hidden layers with 20 units for without monotonicity for the
390 neural network for the predicted hydraulic conductivity \hat{f}_K .

391 4.2 Effect of Noise and Sparsity of Training Data

392 To investigate the effect of measurement noise on the performance of the PINNs,
393 Gaussian noise with mean of zero and different values of standard deviation (0, 0.005,
394 0.01, 0.02) was added to the synthetic volumetric water content data (see Section 3.4),
395 which was used to train the PINNs with and without monotonicity constraints. Figure
396 5 (a) shows relative error for soil water flux density ϵ^q for different values of noise added
397 to the true volumetric water content data. For the PINNs with and without monotonic-
398 ity constraints, relative error increased with the standard deviation of noise, although
399 the effect of the noise was substantially lower for the PINNs with monotonicity. On the
400 other hand, the PINNs without monotonicity constraints exhibited consistently large rel-
401 ative error for all levels of noise. These observations indicate that monotonicity constraints
402 are critical for ensuring stability and reliability when fitting noisy data. Therefore, PINNs
403 without monotonic neural networks is not practically feasible for field applications.

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

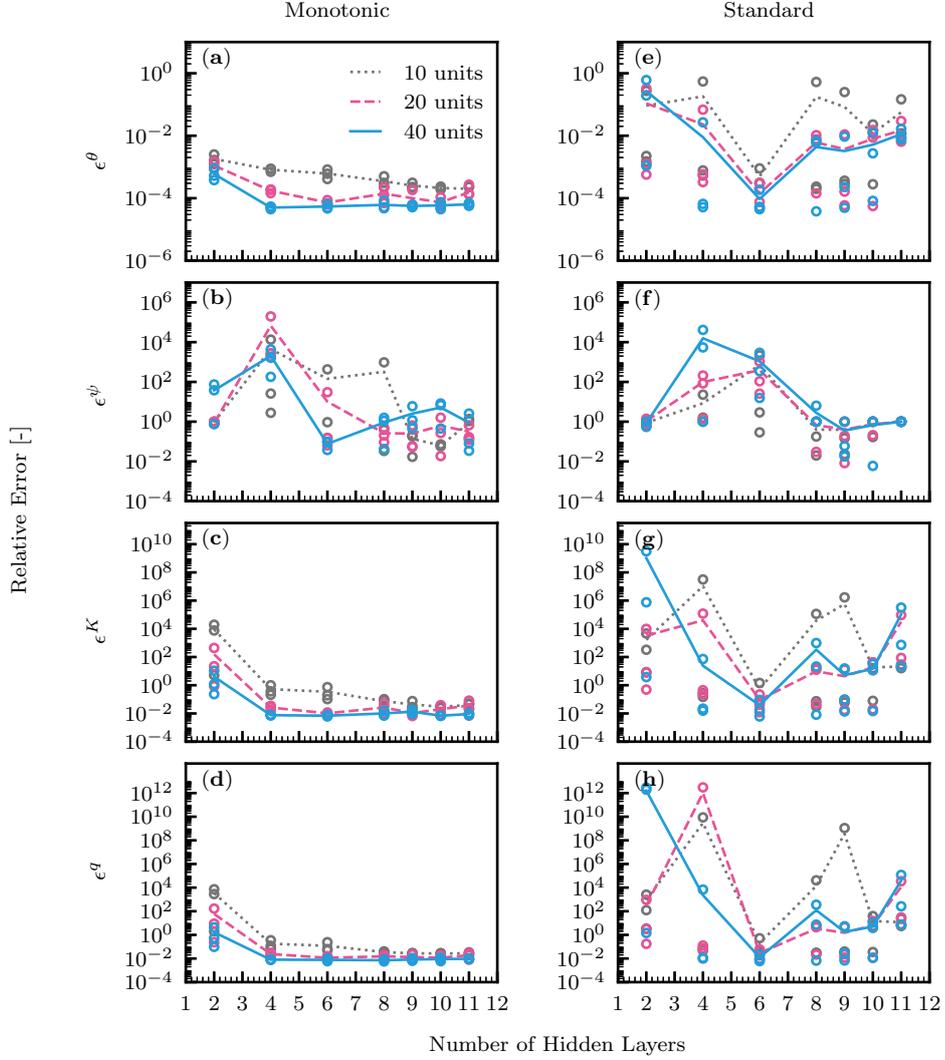


Figure 4. Relative error ϵ for volumetric water content θ , matric potential ψ , hydraulic conductivity K , and soil water flux density q for different numbers of hidden layers and units in the neural network for the predicted matric potential \hat{f}_ψ (Figure 3 (a)); with (left column) and without 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.

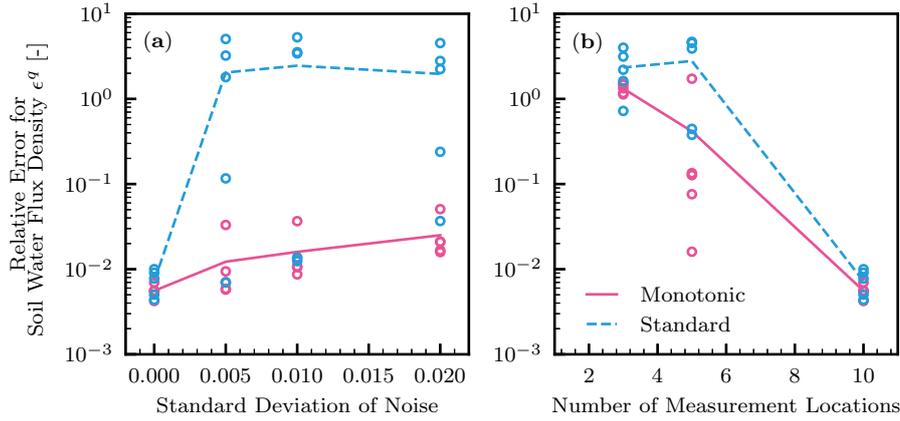


Figure 5. Relative error for soil water flux density ϵ^q for different values of standard deviation of noise (a) and measurement locations at which synthetic volumetric water content data were sampled (b). The number of measurement locations was varied from 10 ($z = -1, -3, -5, -7, -9, -11, -13, -15,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

411

4.3 Generalization Capability of PINNs

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

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 .

427

4.3.1 Volumetric water content

428

429

430

431

432

433

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).

434

435

436

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 (\pm standard deviation)) 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	$9.69(\pm 8.30) \times 10^{-1}$	$2.32(\pm 4.30) \times 10^3$	$7.84(\pm 8.66) \times 10^3$

437 ics. For the same reason, larger errors were observed just after the initial condition ($t =$
 438 0 day). Also, the PINNs could not reproduce the true volumetric water content at depths
 439 that are not covered in the training data (i.e., near the surface and lower than $z = -19$
 440 cm). This means the PINNs could not extrapolate the volumetric water content data
 441 while it could interpolate. Similar trends were observed for the other two soils (see Fig-
 442 ure S2 and S3 in the information).

443 **4.3.2 Residual of RRE**

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

459 **4.3.3 Water Retention Curves**

460 Predicting matric potential from the noisy volumetric water content corresponds
 461 to estimating WRCs, which is one of the primary goals of the study. The PINNs with
 462 monotonicity constraints could not precisely predict the WRCs for the three soils, as shown
 463 in Figure 8. Especially, the prediction was not satisfactory for low and high volumetric
 464 water content, where the training data points were not provided. This suggests the dif-
 465 ficulty in representing the two characteristics of WRCs by using a monotonic neural net-
 466 work: monotonicity and well defined upper and lower limits (saturation and dryness, re-
 467 spectively). This weakness of the current PINNs needs to be fixed in future research. Nev-
 468 ertheless, the predicted WRCs were surprisingly similar to the true WRCs in the mid-
 469 dle range regardless of the fact that any actual value of matric potential was not used
 470 to train the PINNs.

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

479 **4.3.4 Hydraulic Conductivity Functions**

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

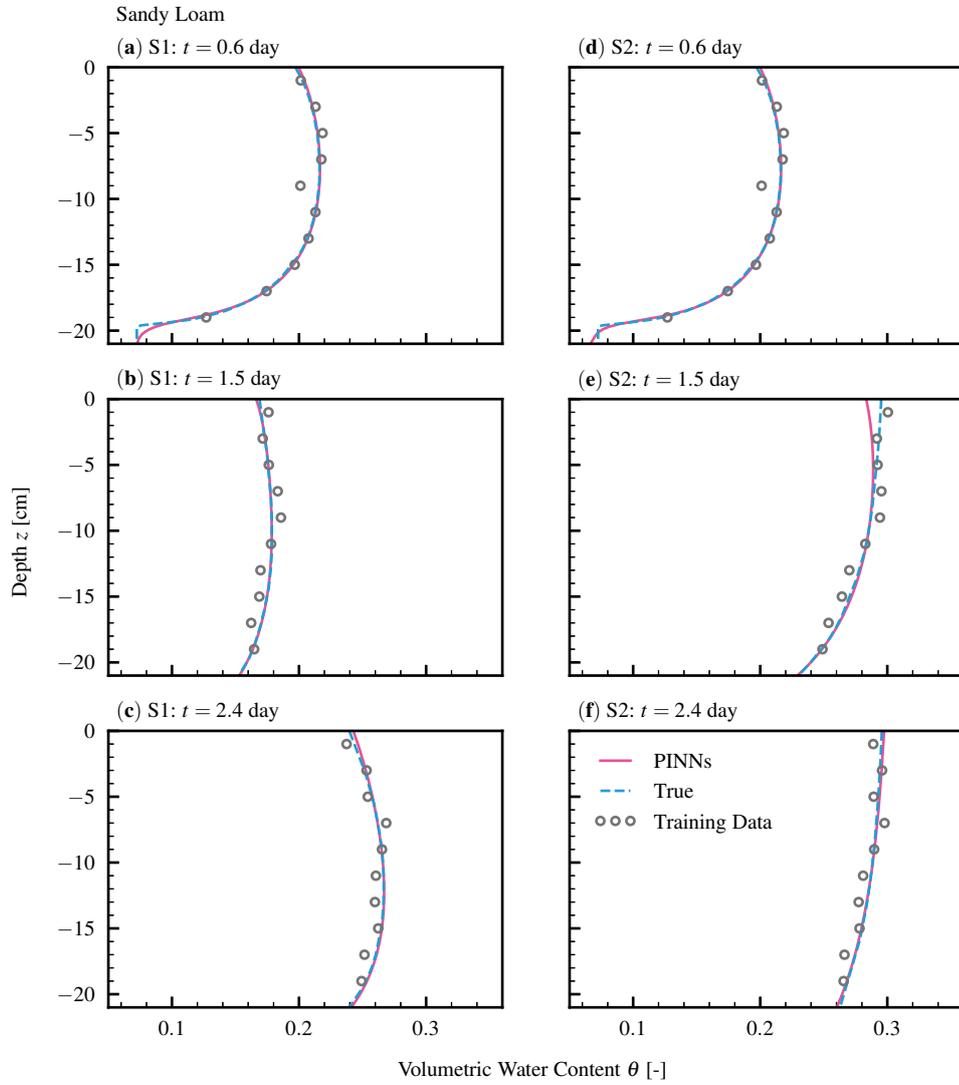


Figure 6. Predicted volumetric water content (PINNs) and noisy synthetic training data (Training Data) for sandy loam soil for the two scenarios at three different times. The dotted lines represent the synthetic data before adding the noise (True). Scenario 1 (S1): (a) $t = 0.6$ day, (b) $t = 1.5$ day,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

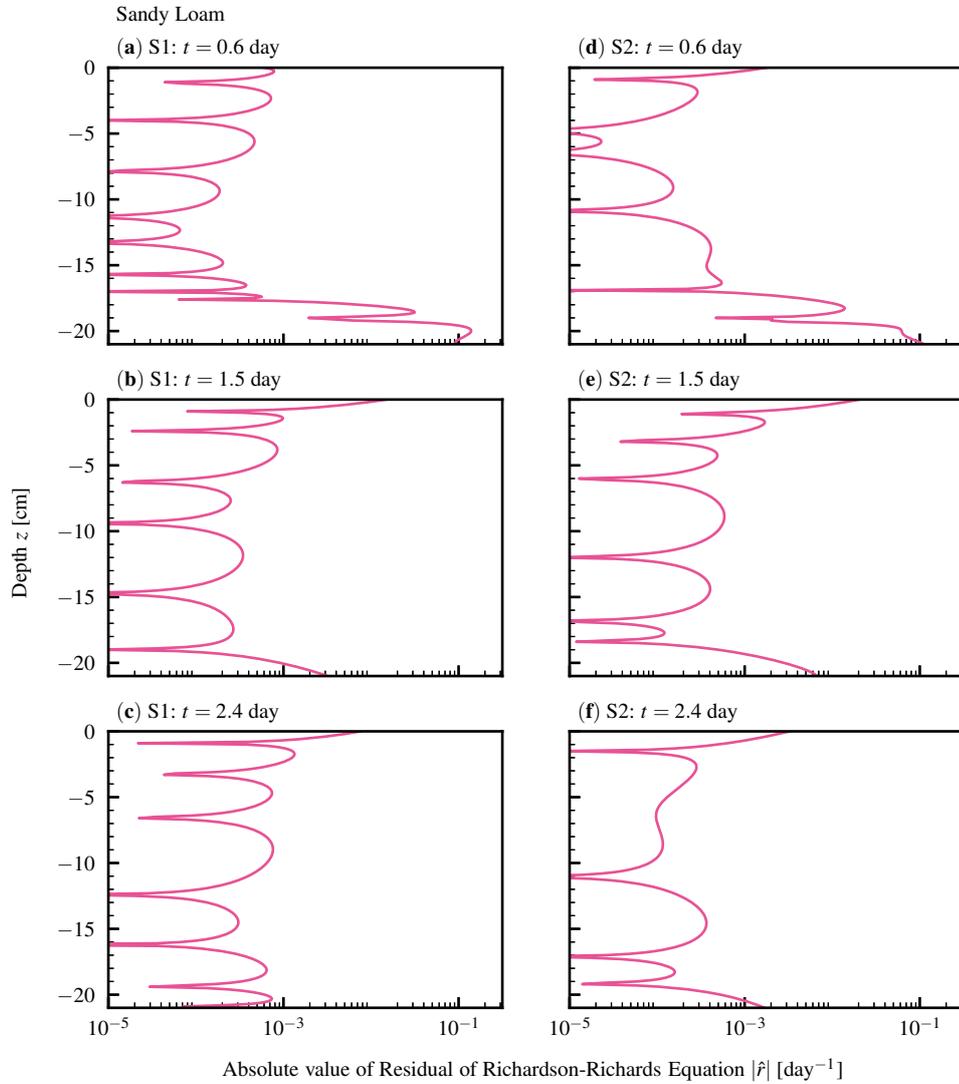


Figure 7. The absolute value of the residual of the Richardson-Richards equation at three different times for sandy loam soil for the two scenarios. Scenario 1 (S1): (a) $t = 0.6$ day, (b) $t = 1.5$ day, and (c) $t = 2.4$ day. Scenario 2 (S2): (d) $t = 0.6$ day, (e) $t = 1.5$, and (f) $t = 2.4$ day.

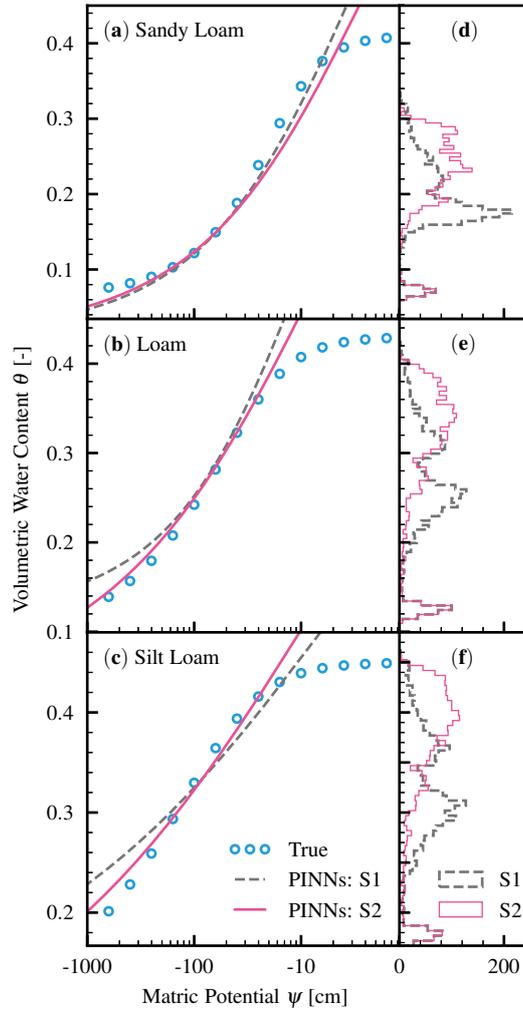


Figure 8. Comparison of true water retention curves (True) to the ones predicted by the PINNs with monotonicity constraints for the three soils for the two scenarios (S1: Scenario 1, S2: Scenario 2) with the histogram of the noisy training data. Water retention curve for (a) sandy loam, (b) loam, and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

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

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

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

505 **4.3.5 Soil Water Flux Density**

506 In this section, we will show that the current PINNs framework with monotonic-
 507 ity constraints can be used to estimate soil water flux density from noisy volumetric wa-
 508 ter content data.

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

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

526 **5 Summary and Conclusions**

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

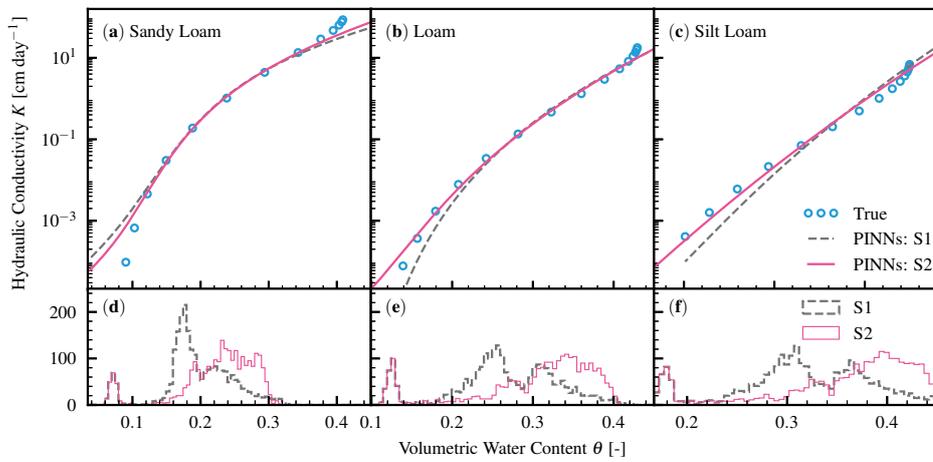


Figure 9. Comparison of true hydraulic conductivity functions (True) to the ones predicted by the PINNs with monotonicity constraints for the three soils for the two scenarios (S1: Scenario 1, S2: Scenario 2) with the histogram of the noisy training data. Hydraulic conductivity function for (a) sandy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

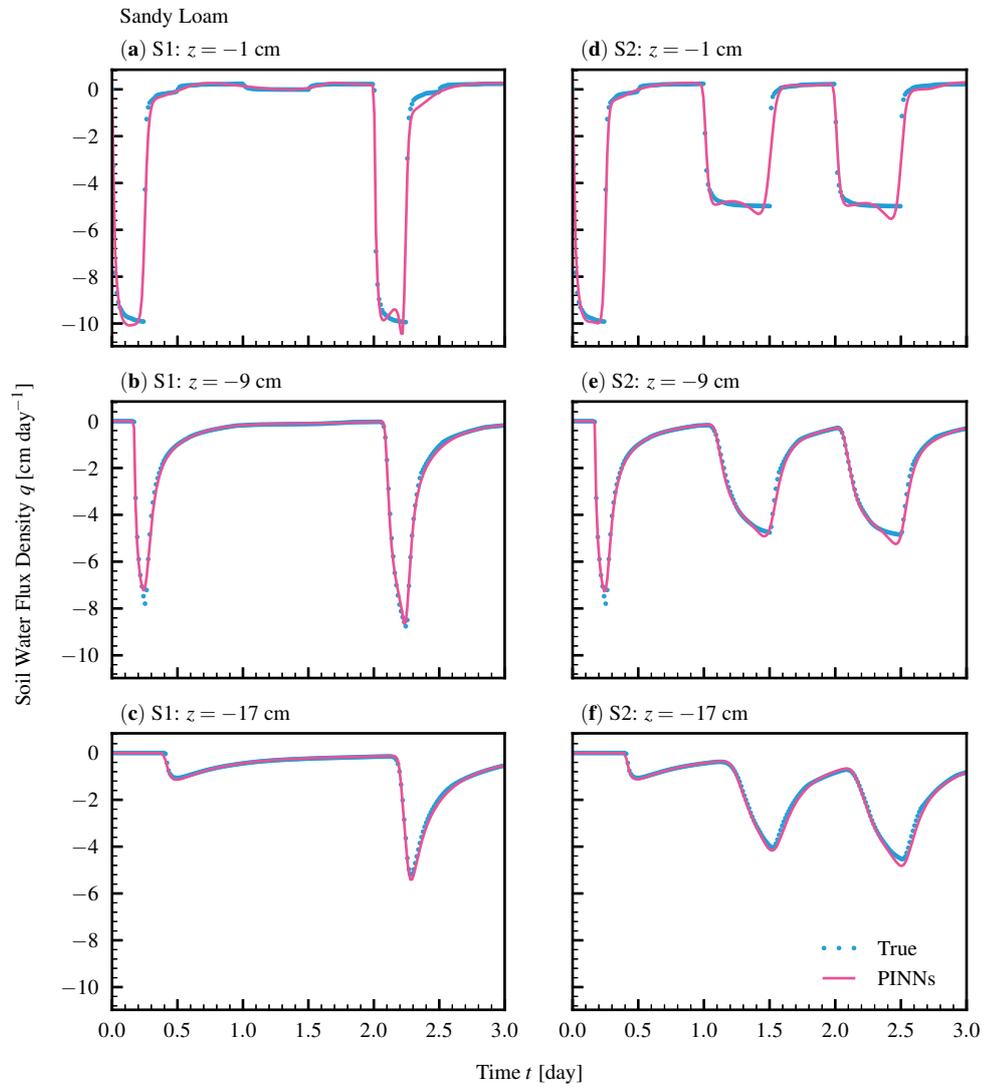


Figure 10. Estimated soil water flux density against the true one at three different depths for sandy loam soil. Scenario 1 (S1): (a) $z = -1$ cm, (b) $z = -9$ cm, and (c) $z = -17$ cm. Scenario 2 (S2): (d) $z = -1$ cm, (e) $z = -9$ cm, and (f) $z = -17$ cm.

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

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

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

553 It was illustrated that the PINNs with monotonicity constraints has a great po-
 554 tential to predict constitutive relationships of the RRE and soil water flux density from
 555 only noisy volumetric water content data in fields. The advantage of this method is the
 556 current PINNs framework does not need initial and boundary conditions and any infor-
 557 mation about the HCF a priori. The current framework must be tested with real exper-
 558 imental data for homogeneous soil in future research.

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

564 Acronyms

565 **HCFs** Hydraulic Conductivity Functions
 566 **PDE** Partial Differential Equation
 567 **PINNs** Physics-Informed Neural Networks
 568 **RRE** Richardson-Richards Equation
 569 **WRCs** Water Retention Curves
 570 **VWC** Volumetric Water Content

571 Notation

572 $:=$ Equal by definition
 573 $\hat{\cdot}$ Hat indicating predicted values or functions (e.g., \hat{y})
 574 (i) Superscript (i) denoting ith data (e.g., $\theta^{(i)}$)
 575 $[L]$ Superscript [L] denoting L th layer
 576 $\mathbf{a}^{[L]} \in \mathbb{R}^{n^{[L]}}$ Vector value for the L th layer consisting of $n^{[L]}$ units
 577 \mathbf{b} Bias vector
 578 $\hat{\mathbf{f}}$ Neural network

579	g	Activation function
580	h	Output function
581	K	Hydraulic conductivity [$L T^{-1}$]
582	K_s	Mualem-van Genuchten parameter
583	\mathcal{L}	Loss function
584	l	Mualem-van Genuchten parameter
585	N	Number of data points
586	n	Mualem-van Genuchten parameter
587	n_i	Number of size a vector, as in n_x and n_y
588	$n^{[L]}$	Number of units in L th layer of a neural network
589	q	Soil water flux density [$L T^{-1}$]
590	\hat{r}	Residual of the Richardson-Richards equation
591	S_e	Effective saturation
592	t	Time [T]
593	W	Weight matrix
594	$\mathbf{x} \in \mathbb{R}^{n_x}$	Input vector for the size of the input n_x
595	$\mathbf{y} \in \mathbb{R}^{n_y}$	Output vector for the size of the output n_y
596	z	Vertical coordinate or depth (positive upward) [L]
597	α	Mualem-van Genuchten parameter [L^{-1}]
598	ϵ	Relative error
599	θ	Volumetric water content [$L^3 L^{-3}$]
600	θ_r	Mualem-van Genuchten parameter [$L^3 L^{-3}$]
601	θ_s	Mualem-van Genuchten parameter [$L^3 L^{-3}$]
602	ψ	Matric potential of water in the soil [L]
603	ψ_{log}	Matric potential in logarithmic scale

604 Acknowledgments

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

609 References

- 610 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X.
 611 (2015). *TensorFlow: Large-scale machine learning on heterogeneous distributed*
 612 *systems*. Retrieved from <https://www.tensorflow.org/>
- 613 Assouline, S. (2006). Modeling the relationship between soil bulk density and the
 614 hydraulic conductivity function. *Vadose Zone Journal*, 5, 697–705. doi: 10
 615 .2136/vzj2005.0084
- 616 Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., & Tuller, M.
 617 (2019). Ground, proximal, and satellite remote sensing of soil moisture. *Re-*
 618 *views of Geophysics*, 57, 530–616. doi: 10.1029/2018RG000618
- 619 Bitterlich, S., Durner, W., Iden, S. C., & Knabner, P. (2004). Inverse estimation
 620 of the unsaturated soil hydraulic properties from column outflow experiments
 621 using free-form parameterizations. *Vadose Zone Journal*, 3, 971–981. doi:
 622 10.2113/3.3.971
- 623 Brooks, R. H., & Corey, A. T. (1964). *Hydraulic properties of porous media*. Fort
 624 Collins, CO, USA.
- 625 Buckingham, E. (1907). *Studies on the movement of soil moisture* (Vol. 38). Wash-
 626 ington, DC, USA.

- 627 Burdine, N. T. (1953). Relative permeability calculations from pore size distribution
628 data. *Society of Petroleum Engineers*, 5(3). doi: 10.2118/225-G
- 629 Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for
630 bound constrained optimization. *Journal of Scientific Computing*, 16(5), 1190–
631 1208. doi: 10.1137/0916069
- 632 Cybenko, G. (1989). Approximation by superpositions of a sigmoidal func-
633 tion. *Mathematics of Control Signals Systems*, 2, 303–314. doi: 10.1007/
634 BF02836480
- 635 Daniels, H., & Velikova, M. (2010). Monotone and partially monotone neural net-
636 works. *IEEE Transactions on Neural Networks*, 21(6), 906–917. doi: 10.1109/
637 TNN.2010.2044803
- 638 Degré, A., van der Ploeg, M. J., Caldwell, T., & Gooren, H. P. A. (2017). Compar-
639 ison of soil water potential sensors: A drying experiment. *Vadose Zone Jour-
640 nal*, 16(4), 1–8. doi: 10.2136/vzj2016.08.0067
- 641 Durner, W. (1994). Hydraulic conductivity estimation for soils with heterogeneous
642 pore structure. *Water Resources Research*, 30(2), 211–223. doi: 10.1029/
643 93WR02676
- 644 Durner, W., Jansen, U., & Iden, S. C. (2008). Effective hydraulic properties of
645 layered soils at the lysimeter scale determined by inverse modelling. *European
646 Journal of Soil Science*, 59, 114–124. doi: 10.1111/j.1365-2389.2007.00972.x
- 647 Farthing, M. W., & Ogden, F. L. (2017). Numerical solution of Richards’ equation:
648 A review of advances and challenges. *Soil Science Society of America Journal*,
649 81, 1257–1269. doi: 10.2136/sssaj2017.02.0058
- 650 Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feed-
651 forward neural networks. In *Proceedings of the thirteenth international confer-
652 ence on artificial intelligence and statistics* (pp. 249–256).
- 653 Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA,
654 USA: The MIT Press.
- 655 He, Q., Barajas-Solano, D., Tartakovsky, G. D., Alexandre, M., & Tartakovsky,
656 A. M. (2020). Physics-informed neural networks for multiphysics data assimi-
657 lation with application to subsurface transport. *Advances in Water Resources*,
658 141. doi: 10.1016/j.advwatres.2020.103610
- 659 Hopmans, J. W., Šimůnek, J., Romano, N., & Durner, W. (2002). Simultaneous
660 determination of water transmission and retention properties. Inverse methods.
661 In J. H. Dane & G. C. Topp (Eds.), *Methods of soil analysis, part 4, physical
662 methods* (pp. 963–1008). Madison, WI, USA: Soil Science Society of America.
- 663 Iden, S. C., & Durner, W. (2007). Free-form estimation of the unsaturated soil
664 hydraulic properties by inverse modeling using global optimization. *Water Re-
665 sources Research*, 43, 1–12. doi: 10.1029/2006WR005845
- 666 Kamai, T., Tuli, A., Kluitenberg, G. J., & Hopmans, J. W. (2008). Soil water
667 flux density measurements near 1 cm d1 using an improved heat pulse probe
668 design. *Water Resources Research*, 44, 1–12. doi: 10.1029/2008wr007036
- 669 Kamai, T., Tuli, A., Kluitenberg, G. J., & Hopmans, J. W. (2010). Corre-
670 ction to ”Soil water flux density measurements near 1 cm d1 using an im-
671 proved heat pulse probe design”. *Water Resources Research*, 46, 1–5. doi:
672 10.1029/2010WR009423
- 673 Kingma, D. P., & Ba, J. B. (2014). Adam: A method for stochastic optimiza-
674 tion. *arXiv preprint*, 1–15. Retrieved from [https://arxiv.org/pdf/
675 1412.6980.pdf](https://arxiv.org/pdf/1412.6980.pdf)
- 676 Kosugi, K. (1996). Lognormal distribution model for unsaturated soil hy-
677 draulic properties. *Water Resources Research*, 32(9), 2697–2703. doi:
678 10.1029/96WR01776
- 679 Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of un-
680 saturated porous media. *Water Resources Research*, 12(3), 513–522. doi: 10
681 .1029/WR012i003p00513

- 682 Nocedal, J., & Wright, S. J. (2006). *Numerical optimization*. New York, NY, USA:
683 Springer. doi: 10.1007/978-0-387-40065-5
- 684 Rai, P. K., & Tripathi, S. (2019). Gaussian process for estimating parameters of
685 partial differential equations and its application to the Richards equation.
686 *Stochastic Environmental Research and Risk Assessment*, *33*, 1629–1649. doi:
687 10.1007/s00477-019-01709-8
- 688 Raissi, M., & Karniadakis, G. E. (2018). Hidden physics models: Machine learning
689 of nonlinear partial differential equations. *Journal of Computational Physics*,
690 *357*, 125–141. doi: 10.1016/j.jcp.2017.11.039
- 691 Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural
692 networks: A deep learning framework for solving forward and inverse problems
693 involving nonlinear partial differential equations. *Journal of Computational*
694 *Physics*, *378*, 686–707. doi: 10.1016/j.jcp.2018.10.045
- 695 Richards, L. A. (1931). Capillary conduction of liquids through porous mediums.
696 *Physics*, *1*, 318–333. doi: 10.1063/1.1745010
- 697 Richardson, L. F. (1922). *Weather prediction by numerical process*. Cambridge,
698 United Kingdom: Cambridge University Press.
- 699 Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B.,
700 Knight, R., . . . Wendroth, O. (2008). Soil moisture measurement for ecolog-
701 ical and hydrological watershed-scale observatories: A review. *Vadose Zone*
702 *Journal*, *7*, 358–389. doi: 10.2136/vzj2007.0143
- 703 Sheng, W., Zhou, R., Sadeghi, M., Babaeian, E., Robinson, D. A., Tuller, M., &
704 Jones, S. B. (2017). A TDR array probe for monitoring near-surface soil
705 moisture distribution. *Vadose Zone Journal*, *16*(4), 1–8. doi: 10.2136/
706 vzj2016.11.0112
- 707 Šimůnek, J., Šejna, M., Saito, H., Sakai, M., & van Genuchten, M. T. (2013). *The*
708 *HYDRUS-1D software package for simulating the one-dimensional movement*
709 *of water, heat, and multiple solutes in variably saturated media, Version 4.17*
710 (Tech. Rep.). Riverside, CA, USA: Department of Environmental Sciences,
711 University of California Riverside, Riverside.
- 712 Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., & Barajas-
713 Solano, D. (2020). Physics-informed deep neural networks for learning pa-
714 rameters and constitutive relationships in subsurface flow problems. *Water*
715 *Resources Research*. doi: 10.1029/2019wr026731
- 716 Tuller, M., & Or, D. (2001). Hydraulic conductivity of variably saturated porous
717 media: Film and corner flow in angular pore space. *Water Resources Research*,
718 *37*(5), 1257–1276. doi: 10.1029/2000WR900328
- 719 van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic
720 conductivity of unsaturated soils. *Soil Science Society of America*, *44*, 892–
721 898. doi: 10.1016/j.pan.2017.07.214
- 722 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Courn-
723 peau, D., . . . Contributors, S. . . (2020). SciPy 1.0: Fundamental Algorithms
724 for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. doi:
725 <https://doi.org/10.1038/s41592-019-0686-2>
- 726 Zha, Y., Yang, J., Zeng, J., Tso, C. M., Zeng, W., & Shi, L. (2019). Review
727 of numerical solution of Richardson–Richards equation for variably satu-
728 rated flow in soils. *Wiley Interdisciplinary Reviews: Water*, *6*, 1–23. doi:
729 10.1002/wat2.1364