Research papers

Predicting near-saturated hydraulic conductivity in urban soils

Jinshi Jian a,b, Alexey Shiklomanov c, William D. Shuster d, Ryan D. Stewart b, * 

a Pacific Northwest National Laboratory-University of Maryland Joint Global Change Research Institute, 5825 University Research Court, Suite 3500, College Park, MD, United States 
b School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States 
c NASA Goddard Space Flight Center, 8800 Greenbelt Rd., Building 33, Greenbelt, MD 20771, United States 
d Department of Civil and Environmental Engineering, Wayne State University, Detroit, MI, United States

ARTICLE INFO

This manuscript was handled by C. Corradini, Editor-in-Chief, with the assistance of Renato Morbidelli, Associate Editor

Keywords: Infiltration Artificial neural network Random forest Pedotransfer function Soil science Urban hydrology

ABSTRACT

Pedotransfer functions (PTFs) provide point predictions of soil hydraulic properties from more readily measured soil characteristics, yet uncertainties and biases in measurement methods, sampling distributions, and boundary conditions can limit accuracy when estimating near-saturated hydraulic conductivity ($K_{ns}$). These limitations may be particularly problematic in understudied urban landscapes that often contain altered hydraulic properties. To better treat deficiencies in PTF performance, we addressed three objectives, which were to: 1) develop PTFs to predict urban $K_{ns}$; 2) assess bulk density and coarse fragments as explanatory variables; and 3) evaluate the predictive capability of these PTFs by comparing their output to measured hydraulic conductivity values from three other studies of urban soil hydraulics. We used artificial neural networks (ANN) and random forest (RF) approaches to predict urban $K_{ns}$ with the training dataset including 307 tension infiltrometer tests and other measurements drawn from urban soil assessments in 11 U.S. cities. The PTFs utilized a hierarchy of inputs, starting with percentage sand, silt, clay, and then adding percentage coarse fragments and bulk density. The ANN models performed similar to the RF models, and all models exhibited similar or better predictive performance as models results collected from published articles. The inclusion of bulk density or coarse fragments did not improve accuracy over soil texture alone. Possible reasons for this result include low correlation between $K_{ns}$ and bulk density and the exclusion of large voids during flow measurements with tension infiltrometers. The models have been made available as an open-source software package to encourage adoption by users working in urban systems.

1. Introduction

Hydraulic conductivity represents a master variable in subsurface water movement, regulating infiltration, drainage, and redistribution of moisture in the soil profile (Jackson et al., 2016; Libohova et al., 2018; Stewart and Abou Najm, 2018). Nearly every mainstream model that apportions rainfall between infiltration and surface runoff requires hydraulic conductivity to be constrained (Alagna et al., 2018; Ameli et al., 2015; Li et al., 2014; Trinh et al., 2018). However, it is often time-consuming and difficult to measure this property in the field, meaning that relatively few studies have endeavored to quantify hydraulic conductivity at numerous points or locations (e.g., Braud et al., 2017; Di Prima et al., 2019; Rahmati et al., 2018).

Pedotransfer functions (PTFs) – which predict relatively difficult to measure hydraulic properties from more easily measured physical properties (such as soil texture) – are often used to estimate hydraulic conductivity at multiple points. As an example, ROSETTA (Schaap et al., 2001) is based on a PTF that reaches back to the work of Rawls et al. (1982). ROSETTA includes a hierarchical structure with flexibility in inputs, starting with relatively basic data (i.e., soil textural separates as sand, silt, clay), ascending to more involved data, such as bulk density and points along the soil water retention curve (i.e., field capacity at 1/3 bar, wilting point at 15 bar overpressure). The ROSETTA model is arguably the most popular predictive algorithm, and its inclusion in the HYDRUS software for modeling unsaturated flow has made its use customary in many hydrological studies (Schaap, 1999; Schaap et al., 2001; Simunek et al., 2008).

Many other PTFs also have been developed over the years, using

* Corresponding author. 
E-mail address: ryan.stewart@vt.edu (R.D. Stewart).

https://doi.org/10.1016/j.jhydrol.2021.126051

Received 15 October 2020; Received in revised form 8 January 2021; Accepted 27 January 2021

Available online 5 February 2021

0022-1694/© 2021 Elsevier B.V. All rights reserved.
anywhere from six to more than 5000 soil samples (e.g., Arya et al., 1981, 1999; Arya and Paris, 1981; Gimenez et al., 1997; Guber and Pachepsky, 2010; Rawls et al., 1993). Some of these models include additional characterizing information such as soil horizon type, organic matter content, topography, vegetation, and general categories of land management (Moeys et al., 2012; Sharma et al., 2006; Wosten et al., 1999). Further, the models represent a variety of techniques to estimate parameter values, including multiple regression models (Lin et al., 1999; Wosten et al., 1999) and artificial neural networks (Parasuraman et al., 2006; Schaap, 1999; Schaap et al., 1998). Different PTFs have also been integrated together into multi-model ensemble predictive tools (Guber et al., 2006). While these various PTFs tend to perform well when predicting water retention characteristics, the variability between model output and actual measurement becomes much wider when predicting hydraulic conductivity (Pachepsky and Rawls, 2003; Rawls et al., 1998).

Current PTFs are typically trained using hydraulic conductivity values collected from soil cores analyzed under laboratory conditions (Clapp and Hornberger, 1978; Rawls et al., 1998), which often do not reflect values measured in the field (Reynolds et al., 2002a, 2002b; Stewart et al., 2016). This discrepancy can arise through several means. First, core sampling is a destructive process carrying with it artifacts like structural cracks and disturbance of soil pore structure. Second, soil cores are uniformly saturated in the laboratory prior to analysis, which diverges from typical field-saturated or near-saturated conditions where air still exists in trapped pockets or large pores (Reynolds et al., 2002a, 2002b). Due to imperfect filling of the core volume, soil cores can also allow water to short-circuit the soil matrix, creating bypass flow between the soil sample and core wall. This process exaggerates hydraulic conductivity values, though the effect can diminish with longer samples (Anderson and Bouma, 1973). Nonetheless, such inconsistencies point to the need for PTFs based on field-relevant hydraulic conductivity values, following examples such as Jarvis et al. (2002) and Yang et al. (2019).

The need for consistent and relevant PTFs is even more critical in urban soils, where the paucity of hydraulic conductivity data is particularly acute. These soils are often borne of cutting, filling, and generally re-arranging the soil profile and its constituent soil textures (Herrmann...
et al., 2018). As such, urban soils often do not resemble their agricultural, forested, and undisturbed counterparts. The disturbances brought on by urbanization can alter properties such as bulk density (BD), which often has a negative correlation to hydraulic conductivity (Gregory et al., 2006; Hamilton and Waddington, 1999). Likewise, coarse fragments (e.g., stones, concrete, bricks) are often incorporated into urban soils during construction or demolition. Laboratory studies have examined how these materials affect infiltration rates (Naseri et al., 2019) and water retention in urban soils (Barbu and Ballestero, 2015). However, the role of coarse fragments in restricting or facilitating matrix flow in soils is divergent. While not specific to urban soils, field-based studies have shown that stones in soil can block pathways otherwise open to flow and thus decrease hydraulic conductivity (Hlavacikova and Novak, 2014). Other studies determined that stony soils can be better aggregated and macroporous (Cerdà, 2001; Poesen and Ingelmo-Sanchez, 1992), and as a result can have relatively high hydraulic conductivity (Seckers et al., 2016; Lilly, 2000). Thus, it appears that the type of coarse fragment(s), the predominant soil texture of the surrounding matrix, and packing (e.g., bulk density) all affect hydraulic conductivity.

In this study we used soil physical datasets measured on urban soils to develop and test the efficacy of new PTF models for estimating near-saturated hydraulic conductivity ($K_h$). We approached our study with three sequential objectives: 1) develop a set of PTFs that accurately predict $K_h$ in surficial urban soils; 2) test the sensitivity of the models to the independent variables of soil bulk density and percentage coarse fragments; and 3) evaluate the predictive capability of these models to predict hydraulic conductivity measured in other urban areas. We leveraged a novel urban soils dataset collected from 11 U.S. cities (Herrmann et al., 2018; Stewart et al., 2019), with $n = 307$ observations of $K_h$ serving to train the PTFs. The last objective relied on hydraulic conductivity values reported in three other studies ($n = 20$).

2. Methods

2.1. Near-saturated hydraulic conductivity ($K_h$) data from urban soils

Soil samples were collected and field measurements were taken from surficial urban soil profiles located in 11 cities within the United States: Atlanta, GA; Camden, NJ; Cincinnati, OH; Cleveland, OH; Detroit, MI; New Orleans, LA; Omaha, NE; Phoenix, AZ; Portland, ME; Tacoma, WA; and San Juan, PR (Fig. 1). Sites represented 10 of the 12 major soil orders in the United States Department of Agriculture (USDA) soil taxonomy (Soil Survey Staff, 2014). The soils assessed were representative of all twelve USDA soil textures, though the sandy clay, silt, and silt loam textures had fewer samples than the others (Fig. 1). Note that the urban datasets are available for download through a repository maintained by the US Government (https://catalog.data.gov/harvest/about/epa-sciencehub), and also have been extensively discussed in Herrmann et al. (2017), Herrmann et al. (2018), Schiffman and Shuster (2019), Shuster et al. (2014), Shuster et al. (2015), and Stewart et al. (2019).

Soil samples were collected from each profile using a 6 cm diameter by 130 cm length hammer-action coring cylinder. The cores were then separated by diagnostic horizon for further analyses, including sand, silt, and clay percentages (%Sand, %Silt, and %Clay) using the hydrometer method (Gee and Bauder, 1986), and volumetric percentage of coarse fragments (%Rock) as the volume of coarse fragments per horizon identified in the soil core. In this study we used data from the uppermost (surficial) horizon for each profile, as these horizons were most pertinent to the surface-based infiltration tests we used (as described below). In total we had data for 307 horizons, with $n = 238$ soil profiles including estimates for %Rock (thus representing 79% of the total dataset). Soil bulk density was also measured on undisturbed soil cores for $n = 67$ surficial horizons, representing 22% of the total dataset. All sampled horizons were analyzed for their diagnostic information, including horizon type, thickness, and whether or not they were formed by anthropogenic-transported material. In total, 77% of the studied horizons (235 out of 307) were characterized as having evidence of anthropogenically transported materials, while the remaining 23% of horizons were deemed to not be of anthropogenic origin. Further, 84% of the sampled layers (256 out of 307) were classified as A horizons, with 1 exposed B horizon, 38 exposed C horizons, 3 profiles deemed to be in process of transitioning from C to A horizons, and 9 horizons that were not classified.

We measured surface infiltration rates using a mini-disk tension infiltrometer (METER Group, Pullman, WA, USA). For each soil profile, four measurements were collected along a transect. At each point, the surface was cleared of above-ground biomass and any other loose debris. The infiltrometer was placed directly on the soil surface with the source pressure head set to $h = -2$ cm. Infiltrated water volumes were measured at 0.5, 1, 2, 3, 4, 5, 10, 15, 20, and 25 min. The infiltration data were fitted to a two-term infiltration model (Zhang, 1997) with the hydraulic capillarity parameter ($a^*$) estimated based on the measured soil texture for each surface horizon. The model produced estimates for hydraulic conductivity at $-2$ cm tension for each test; the hydraulic conductivity values from the four measurements were then averaged to produce a site-level estimate of $K_h$. All $K_h$ values were transformed using the natural logarithm prior to use in the PTFs.

2.2. $K_h$ modeling

The artificial neural network (ANN) model included input, hidden, bias, and output layers, with each layer connected to the next via neurons (noting that each neuron computes the weighted average of its input through nonlinear functions; Fig. A1). Input layers included %Sand, %Silt, %Clay, and/or %Rock, and the output layer was predicted $K_h$. We varied the number of hidden layers from 1 to 3, with each layer possessing 2 to 6 neurons. As ANN model performance was not sensitive to the number of neurons or hidden layers, we used 2 hidden layers, with 5 neurons for the first hidden layer and 3 for the second hidden layer. ANN modeling was done with the R package “neuralnet” (Fritsch et al., 2019).

The random forest (RF) algorithm is a type of regression calculation that allows both numeric and categorical input parameters. We set the number of trees to 100 (noting that each tree in a RF model represents a single decision tree algorithm) and the number of variables in each tree to 2 (Fig. A2a). RF modeling was conducted using the R package “randomForest” (Liaw and Wiener 2002). At first, %Sand, %Silt, %Clay, %Rock, and BD were used as predictors; however, a preliminary analysis revealed that adding BD into the ANN and RF models did not significantly improve the model performance (Table A1). In addition, there were only 67 soil samples with measured BD, so we omitted that parameter as a predictor in subsequent models. After removing BD, all models used %Sand, %Silt, and %Clay as input parameters, while the ANN-with-rock and RF-with-rock models used %Rock as an additional predictor.

2.3. Model evaluation

The importance of individual input variables was tested by quantifying the increase in model mean square error (MSE) as each variable was removed. This analysis determined that %Sand was the most important variable, following by %Clay, %Silt, and %Rock (Fig. A2b). The node purity also showed %Sand to be the most important factor (Fig. A2c). For each ANN or RF model, we performed random sub-sampling cross-evaluation using a non-parametric bootstrapping procedure, wherein we randomly selected 80% of the $K_h$ values as a training dataset and used the remaining 20% of values as a cross-evaluation dataset (Jian et al., 2018). We repeated this process 500 times to generate a distribution of fits for each model. During model usage, a given input was processed using each individual fit to generate a total of 500 $K_h$ predictions, with the final output represented as a mean and standard deviation.
Table 1
Evaluation data collected from three studies used to evaluate pedotransfer function (PTF) performance. Texture = soil texture; BD = bulk density; K = hydraulic conductivity. NA indicates data not available.

<table>
<thead>
<tr>
<th>Texture</th>
<th>% Sand</th>
<th>% Silt</th>
<th>% Clay</th>
<th>BD (g cm$^{-2}$)</th>
<th>K (cm h$^{-1}$)</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clay</td>
<td>44</td>
<td>37</td>
<td>19</td>
<td>NA</td>
<td>1.80</td>
<td>Hamilton and Waddington (1999)</td>
</tr>
<tr>
<td>Silt loam</td>
<td>14</td>
<td>63</td>
<td>23</td>
<td>NA</td>
<td>2.20</td>
<td>Hamilton and Waddington (1999)</td>
</tr>
<tr>
<td>Silt loam</td>
<td>24</td>
<td>47</td>
<td>29</td>
<td>NA</td>
<td>2.30</td>
<td>Hamilton and Waddington (1999)</td>
</tr>
<tr>
<td>Silt loam</td>
<td>27</td>
<td>51</td>
<td>22</td>
<td>NA</td>
<td>2.70</td>
<td>Hamilton and Waddington (1999)</td>
</tr>
<tr>
<td>Silt loam</td>
<td>42</td>
<td>33</td>
<td>25</td>
<td>NA</td>
<td>2.90</td>
<td>Hamilton and Waddington (1999)</td>
</tr>
<tr>
<td>Silt loam</td>
<td>32</td>
<td>43</td>
<td>25</td>
<td>NA</td>
<td>4.90</td>
<td>Hamilton and Waddington (1999)</td>
</tr>
<tr>
<td>Silt loam</td>
<td>32</td>
<td>39</td>
<td>29</td>
<td>NA</td>
<td>5.10</td>
<td>Hamilton and Waddington (1999)</td>
</tr>
<tr>
<td>Silt loam</td>
<td>18</td>
<td>60</td>
<td>22</td>
<td>NA</td>
<td>8.50</td>
<td>Hamilton and Waddington (1999)</td>
</tr>
<tr>
<td>Silt loam</td>
<td>22</td>
<td>57</td>
<td>21</td>
<td>NA</td>
<td>9.80</td>
<td>Hamilton and Waddington (1999)</td>
</tr>
<tr>
<td>Clay</td>
<td>30</td>
<td>41</td>
<td>29</td>
<td>NA</td>
<td>10.0</td>
<td>Hamilton and Waddington (1999)</td>
</tr>
<tr>
<td>Sand</td>
<td>94</td>
<td>3</td>
<td>3</td>
<td>NA</td>
<td>3.00</td>
<td>Pit et al. (1999)</td>
</tr>
<tr>
<td>Clay</td>
<td>15</td>
<td>15</td>
<td>70</td>
<td>NA</td>
<td>0.70</td>
<td>Pit et al. (1999)</td>
</tr>
</tbody>
</table>

As a test for the broader applicability of the PTFs, we also collected and published hydraulic conductivity data from other urban soils studies. We conducted our literature search through ISI Web of Science, using the keywords “urban soil hydraulic conductivity” or “urban soil infiltration” to identify relevant studies. Papers from peer reviewed journals, theses, and dissertations were included. From a total of more than 50 papers, we used the following criteria to identify appropriate sources: (1) data came from field measurements made on urban soils; (2) the study reported hydraulic conductivity or steady-state one-dimensional infiltration rates, which we assumed equivalent to hydraulic conductivity (Philip, 1969; Stewart and Abou Najm, 2018); (3) %Sand, %Silt, and %Clay were reported; (4) the study location or field conditions were described; and (5) the study was located within the USA. With these constraints, 20 data points were extracted from three papers to form the evaluation dataset; hydraulic conductivity values ranged from 0.4 to 18.8 cm h$^{-1}$ (Table 1).

Near-saturated hydraulic conductivity values predicted by the models were compared to measured values using both training (n = 307) and evaluation (n = 20) data. Here, we used six different statistics to assess model performance: intercept, slope, adj $R^2$, ME, RMSE, and d; Eqs. (1)–(4). First, we fitted ordinary least squares regression between the predicted and measured $K_n$ values (using the natural log-transformed data as described above); the slope, intercept and adjusted $R^2$ (adj $R^2$) of the regression lines were used to assess the ability of each model to predict $K_n$ compared to observations.

The adj $R^2$ was calculated as:

$$adj R^2 = 1 - \left[\frac{(1 - R^2)(n-1)}{n-k-1}\right]$$

where $R^2$ is the coefficient of determination, n is the total number of observations, and k is the total number of independent variables in the model. The adj $R^2$ was used to quantify how much of the variability in the response variable was explained by the predictor(s).

Model fits were also analyzed using: (1) mean error, ME; (2) root mean square error, RMSE; and (3) index of agreement, d (Yang et al., 2014). ME was calculated as:

$$ME = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n}$$

where $\hat{y}_i$ represents the ith predicted $K_n$ value and $y_i$ represents the ith measured $K_n$ value.

For interpretation, $ME \approx 0$ indicated no bias between measured data and simulation results, $ME > 0$ indicated that the predicted values were overestimated relative to measured data, and $ME < 0$ indicated that predicted values underestimated the observed data.

RMSE was calculated as:

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

where smaller RMSE values indicated better model performance. The metric $d$ was calculated as:

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

where $\bar{y}$ represents the average of all measured $K_n$ values. For interpretation, $d \geq 0.90$ was considered to indicate excellent agreement between measured and predicted values, 0.80 $< d < 0.90$ to indicate good agreement between measured and predicted values, 0.70 $< d < 0.80$ to show moderate agreement between measured and predicted values, and $d < 0.70$ to reveal poor agreement between measured and predicted values (Yang et al., 2014).

We also compared the $R^2$ and RMSE values corresponding to the models developed in this study versus those same metrics reported for 41 other PTF models in three previous publications (Table A2). Note that neither the PTFs nor the test data in those models were specific to urban soils, yet we included this comparison as a way to provide context for the performance of the PTFs developed in this study versus others. For our PTFs we included both training and evaluation dataset metrics.

All data analyses were carried out using R software (Version 3.5.2, R Core Team, 2019).

3. Results

3.1. Training data

We first analyzed the relationship between measured values for $K_n$, BD, and %Rock. These results showed that $K_n$ did not have any notable trends with BD or %Rock (Fig. 2).

We next analyzed the performance of the PTF models (Fig. 3). The training data analysis indicated that the ANN model explained approximately 50% of $K_n$ variability (adj $R^2 = 0.53$; Table 2). The ANN-no-rock model had relatively low bias, with $ME = 0.17$ log (cm h$^{-1}$), and low error, with $RMSE = 1.03$ log(cm h$^{-1}$). The RF-no-rock model had a slightly lower correlation coefficient as the ANN model (Table 2), with adj $R^2 = 0.45$ and $RMSE = 1.04$ log (cm h$^{-1}$). The models that accounted for coarse rock fragments had similar amounts of bias as the no-rock models, with $ME = 0.23$ log (cm h$^{-1}$) for the ANN model and 0.26 log (cm h$^{-1}$) for the RF model. However, the with-rock models had lower adj $R^2$ and $d$ values than the models without rock fragments. Linear regressions between predicted and field-measured $K_n$ had slopes ranging from 0.61 (ANN-with-rock) to 1.00 (RF-no-rock), and intercepts all $< 0.30$ log (cm h$^{-1}$).
3.2. Model evaluation

When tested against the evaluation dataset, the ANN and RF models explained between 27% and 33% of variability in K_n, as determined using \( \text{adj} R^2 \) (Table 3). Both models had poor agreement with predicted K_n (\( d < 0.70 \)). The intercepts of both regression lines were 1.34–1.35 log (cm h\(^{-1}\)), while the slopes were both < 0.4, indicating over-estimation of K_n for small values and under-estimation for larger values of K_n (Table 3 and Fig. 4).

We also compared the \( R^2 \) and RMSE values for the models developed in this study versus 41 other PTF models developed for non-urban soils. The range of \( R^2 \) values from the present study were similar to or higher than those associated with the other PTF models (Fig. 5a). The RMSE values from our models were lower than most generated by the other PTFs (Fig. 5b). Therefore, while our models had low to moderate accuracy for the evaluation dataset, their performance was similar to or better than other published PTF models.

4. Discussion

Where field-based measurement campaigns are not feasible, reliable, properly constrained pedotransfer functions (PTFs) can be useful to predict near-saturated hydraulic conductivity (K_n) values. In the present study, we synthesized a novel dataset for urban soil hydraulic and physical parameters into a new set of PTFs. As one notable feature, our training dataset incorporated \textit{in situ} assessment methods that were designed to produce consistent and authentic estimates of hydraulic conductivity under near-saturated conditions. As another, we tested both artificial neural network (ANN) and random forest (RF) techniques. These two machines learning techniques differ in their structures and fitting algorithms, but produced comparable results for most instances.

We started our study with the principle that useful PTFs include easy to measure or readily accessible input data and provide predictions with “order-of-magnitude” accuracy. To the first criterion, the models only used %Sand, %Silt, and %Clay, bulk density (BD), and coarse fragments (%Rock) as possible inputs. These inputs represent information that can be found almost everywhere in the conterminous United States (NCSS, 2019) and are commonly measured in other locations and studies. To the second criterion, we assessed PTF accuracy using three summary statistics and linear regression between predicted versus measured K_n values.

When considering all of the evaluation statistics together, the ANN models had similar performance as the RF models for both the training (Table 2) and evaluation data (Table 3). All models had low to moderate...
adj R² values (0.25–0.53), meaning that they explained only 25–53% of the observed variability in Kₙ. However, the models tended to produce predicted Kₙ estimates that were of the same magnitude as the measured values. At the same time, the adj R² values determined here were similar to those recorded in other studies (Fig. 5), and thereby emphasize the limitations of using PTFs to predict Kₙ and other hydraulic properties.

The PTF models developed in this study performed better for the training dataset than the evaluation dataset, with higher ME and RMSE values and lower d values. One possible reason for this discrepancy is that the two datasets were collected using different infiltration tests: the training data were collected using tension infiltrometers, while the evaluation data tests were all conducted using double ring infiltration tests. These experiments differ in their boundary condition (positive versus negative pressure heads) and the type of pores that they potentially activate (macropores versus matrix pores). Perhaps for this reason, the PTFs predicted Kₙ values that were often lower than the observed hydraulic conductivity values (Fig. 4). Specifically, it is possible that the double ring tests activated all (matrix and macroporous) flow paths than would not be captured by the tension infiltrometer.

Numerous studies have compared methods for measuring infiltration (Braud et al., 2017; Köhne et al., 2011; Reynolds et al., 2000; Verbist et al., 2010) and discussed advantages and tradeoffs associated with each. Based on such consideration, we relied on tension infiltration measurements to determine Kₒ for several reasons. For one, the test itself was simpler (e.g., easier to set-up and run, requiring less water), thereby making it possible to measure Kₒ at four points per plot and at thousands of points during the larger study. Two, double ring infiltrometers can suffer from lateral flow effects unless the outer (buffer) ring is sufficiently large (e.g., >0.8 m; Lai and Ren, 2007), which only exacerbates set-up and water supply issues. These lateral flow issues can be particularly pronounced in the presence of subsurface restrictive layers. Three,

![Graph](image_url)

**Table 2**
Summary of model evaluation results for the training data. Slopes and intercepts come from linear regression between modeled and measured Kᵣ. Adj R² = adjusted R²; ME = mean error; RMSE = root mean square error; d = index of agreement.

<table>
<thead>
<tr>
<th>Model</th>
<th>Intercept log (cm h⁻¹)</th>
<th>Slope</th>
<th>Adj R²</th>
<th>ME log (cm h⁻¹)</th>
<th>RMSE log (cm h⁻¹)</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training data (n = 307)</td>
<td>ANN-no-rock</td>
<td>0.18</td>
<td>0.70</td>
<td>0.53</td>
<td>0.17</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>RF-no-rock</td>
<td>0.29</td>
<td>1.00</td>
<td>0.45</td>
<td>0.29</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>ANN-with-rock</td>
<td>0.19</td>
<td>0.61</td>
<td>0.25</td>
<td>0.23</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>RF-with-rock</td>
<td>0.25</td>
<td>0.92</td>
<td>0.28</td>
<td>0.26</td>
<td>1.11</td>
</tr>
</tbody>
</table>

**Table 3**
Summary of model evaluation results for the ANN-no-rock and RF-no-rock models using the evaluation data. Slope and intercept come from linear regression between modeled and measured hydraulic conductivity; Adj R² = adjusted R²; ME = mean error; RMSE = root mean square error; d = index of agreement.

<table>
<thead>
<tr>
<th>Model</th>
<th>Intercept log (cm h⁻¹)</th>
<th>Slope</th>
<th>Adj R²</th>
<th>ME log (cm h⁻¹)</th>
<th>RMSE log (cm h⁻¹)</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation data (n = 20)</td>
<td>ANN-no-rock</td>
<td>1.35</td>
<td>0.30</td>
<td>0.33</td>
<td>1.36</td>
<td>2.28</td>
</tr>
<tr>
<td></td>
<td>RF-no-rock</td>
<td>1.34</td>
<td>0.39</td>
<td>0.27</td>
<td>1.32</td>
<td>1.89</td>
</tr>
</tbody>
</table>
ring infiltrometers can disturb the soil surface during insertion and measurement, which can influence surface properties (Zhang and Li, 2020). Four, near-saturated flow often occurs even under ponded surface conditions, as real-world factors such as trapped pockets of air, dead-end pores, and variation in soil wettability induce a “field-saturated” condition (Reynolds et al., 2002a, 2002b). As a result, certain studies have found field-saturated hydraulic conductivity and $K_n$ to be comparable to each other (e.g., Alagna et al., 2016; Bodhinayake et al., 2004). Finally, the near-saturated conditions maintained by at low tension (e.g., $-2 \text{ cm}$) have been shown to reduce experimental uncertainty relative to ponded measurements (Verbist et al., 2013), decrease non-uniform wetting (Schifman and Shuster, 2019), while still allowing water to move through many hydraulically active macropores (Cey and Rudolph, 2009). Based on the applied tension, water likely moved through pores approximately smaller than $1 \text{ mm}$ in diameter (Dohnal et al., 2010).

Here we note that other instruments such as Guelph permeameters (Braud et al., 2017; Comino et al., 2016; Lilly, 2000) and single ring infiltrometers (Concialdi et al., 2020; Di Prima et al., 2019; Lozano-Baez et al., 2018; Stewart et al., 2015) can also characterize hydraulic conductivity. These devices typically require considerably less effort and time than double ring infiltrometers, albeit with different boundary conditions and more potential for soil disturbance than tension infiltrometers (Lilly, 2000). We also note that the infiltration tests used here primarily interrogated the uppermost soil layers, and therefore can be considered to represent surficial urban soils. Use of subsurface tests, such as borehole permeameter tests, may help generate PTFs that are more applicable to deeper soil layers (e.g., Stewart et al., 2019).

We originally hypothesized that including %Rock and BD as input variables would improve $K_n$ predictions. However, our analysis determined that the ANN and RF models both performed slightly better when constrained using only textural components (i.e., %Sand, %Silt, %Clay) as inputs. This result was unexpected, as the %SSC models had fewer degrees of freedom compared to the other models. It is possible that BD data had little effect on the models due in part to the way in which we constrained BD. Specifically, only 67 (out of 307) soil samples included BD values, and our comparison between BD and $K_n$ did not reveal any significant relationship. It is also possible that our $K_n$ measurements, by taking place at the surface, avoided subsurface layers in which compaction can commonly be found (Herrmann et al., 2018). We also note that a newly updated PTF, developed using European soil samples, determined BD to be only the eighth most important variable in predicting soil hydraulic conductivity (Szabó et al., 2020).

Fig. 4. Measured hydraulic conductivity ($K$) versus near-saturated hydraulic conductivity ($K_n$) predicted by the artificial neural network (ANN) and random forest (RF) models for the evaluation data ($n = 20$), with $K_n$ predicted using the: (a) ANN model (no rock inputs) and (b) RF model (no rock inputs). Dashed lines represent the 1:1 relationship and each solid line with light-blue range represents an ordinary least squares regression bootstrapped range. The black dots represent the mean predicted values from 500 model runs, and the error bars indicate standard deviations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Histograms show the comparisons between $R^2$ and RMSE values of the models developed in this study versus those reported for 41 pedotransfer functions reported in other papers (for details please see Table A2). Note that all RMSE values were transformed using the natural logarithm.
Coarse fragments were quantified in the same locations as $K_n$ with measured volumetric contents ranging from 0 to 90% (Fig. 2). Therefore, it is unlikely that the lack of predictive power of that variable was caused by data mismatches or limitations, with the caveat that coarse percentage were quantified using 5 cm diameter cores and did not include any very large fragments. In our study, $K_n$ was measured in the field using tension infiltrometers, which, as discussed above, may have prevented flow from entering any large voids surrounding the coarse fragments. We nonetheless contend that the hydraulic conditions imposed by the tension infiltrometer provide a more consistent (e.g., Verbis et al., 2013) and realistic assessment of water movement in surficial soils. We also note that previous studies (e.g., Beckers et al., 2016; Cerda, 2001; Hlavacikova and Novak, 2014) failed to find consistent relationships between coarse fragment content and hydraulic conductivity.

Despite the aforementioned discrepancies, the models developed in this study predicted $K_n$ values that closely matched the values measured in the field. These results therefore suggest that PTF performance will improve when those functions are developed using field-derived data taken from representative landscapes. In this study we focused on $K_n$ for urban soils, due to the lack of models developed for those soils and the rapid rate of urbanization seen in many regions of the world. The ability of both ANN and RF models to provide order-of-magnitude $K_n$ estimates for different soil textures and conditions suggests that these approaches warrant broader usage, particularly as many landscapes become increasingly urbanized. Nonetheless, we recognize that the models can be improved with more data points and analyses, and therefore have made the models and data available to the public (for more details please refer to Section 6. Data and code availability). We encourage others to continue to develop and refine these models with their own data.

5. Summary and conclusions

Pedotransfer functions (PTFs) can be useful for predicting hydraulic conductivity in instances where direct measurements are not possible. However, previous PTFs have been derived primarily from agricultural, forested, or undisturbed soils, and therefore may not properly represent urban soils. To address this shortcoming, we used measurements conducted in 11 U.S. cities to develop new PTFs for predicting near-saturated hydraulic conductivity ($K_n$) in urban soils. As inputs the models included %Sand, %Silt, %Clay, volumetric percentage of coarse fragments such as rock, and bulk density. Our results show that the models had good performance when trained with soil textural separates (i.e., %Sand, %Silt, %Clay). However, bulk density and percentage coarse fragments offered little to no improvement in model performance, likely due to the lack of any consistent correlation between those variables and $K_n$. Also, the $K_n$ values were estimated using tension infiltration tests, which offered consistent and easy-to-perform measurements but may not have fully captured the contributions of large macropores and other heterogeneities (e.g., subsurface compaction layers). The PTFs developed here should therefore be considered to best represent the hydraulic behavior of surficial urban soils under near-saturated conditions.

Finally, the PTFs developed in this study will likely be improved with the addition of more data from a greater number of urban areas and soil conditions. To this end, we have disseminated the models through an open-source package, and encourage others to use and further refine these PTF models to better represent urban soil hydraulics.

6. Data and code availability

We developed a publicly available package to disseminate the ANN and RF models used in this study. These models have been compiled together as an R package (version 3.5.3), which is available at https://github.com/jinshijian/UrbanK. We therefore encourage researchers, practitioners, and others to access and use these PTF models for hydraulic characterization of urban soils. Within the UrbanK repository, the file ‘AllCities_Victoria_RDS_rock_bd.csv’ under ‘extdata’ folder holds all raw data used in this study to develop the ANN and RF models; ‘fit_models.R’ under ‘scripts’ folder is the R code to develop the ANN and RF models (i.e., the four models shown in Table 2, along with the bulk density models shown Table A2); and the ‘vignettes’ folder holds the R markdown files to perform the analysis and generate all results for this study.

Author contributions

Jinshi Jian and Ryan Stewart conceived and designed the primary analysis, and led the drafting of this manuscript. Alexey Shiklomanov developed the ANN and RF models and compiled these models together as an R package. William D. Shuster oversaw the collection and analysis of soil samples and performed the infiltration measurements, and provided feedback and insights in all phases of model and manuscript development.

CRediT authorship contribution statement

Jinshi Jian: Conceptualization, Data curation, Formal analysis, Software, Writing - original draft. Alexey Shiklomanov: Software, Formal analysis, Writing - review & editing. William D. Shuster: Investigation, Resources, Writing - review & editing. Ryan D. Stewart: Conceptualization, Resources, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Funding was provided in part by the U.S. Department of Agriculture National Resources Conservation Service Virginia Agricultural Experiment Station and the Hatch Program of the National Institute of Food and Agriculture, U.S. Department of Agriculture. Jinshi Jian was partially supported by the US Department of Energy, Office of Science, Biological and Environmental Research as part of the Terrestrial Ecosystem Sciences Program. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. Alexey was supported by the NSF Award under no. 1655095.

Appendix

Artificial neural network (ANN) models typically have at least 3 layers of neurons, including input layer, hidden layer(s), and output layers, each of which is connected to the neurons in the next layer (Parasuraman et al., 2006; Schaap et al., 1998). The input layer acts as the receiver of unweighted input data. Hidden layers exist between input layers and output layers; their function is to process the inputs obtained by its previous layer by activation function applied on it. The output layer is the last layer of neurons that produces given outputs for the program.

An important but difficult challenge is determining the optimal number of hidden layers. The number of hidden layers depends on the degree of complexity of the problem and the degree of accuracy required, with 1 to 5 hidden layers enough to resolve most problems (Murata et al., 1994). In this study, we changed the number of hidden layers from 1 to 3, with each layer having 2–6 neurons. We found that the model performance was not sensitive to the number of hidden layers or the number of neurons. We thus set the ANN model to have 2 hidden...
Random forest (RF) models were also used in this study. RF is an ensemble regression tree algorithm that can deal with large number of features, including categorical variables. RF only requires setting two free parameters, including the number of trees and number of variables randomly sampled as candidates at each split. The performance of a RF model usually is not sensitive to the values of number of trees and

Fig. A1. Schematic showing configuration of the artificial neural network (ANN) model. I1 = %Sand, I2 = %Silt, and I3 = %Clay. H1-H5 are the neurons in the first hidden layer, and H6-H8 are the neurons in the second hidden layer. O represents the output layer. B1-B3 represent bias terms added in each step.

Fig. A2. (a) Change of mean square error (MSE) values as the number of trees increased in the random forest (RF) model. (b) Change in MSE as each individual variable is removed from the model. (c) Change in node purity that results as each individual variable is removed from the model. Note that these panels present results from one randomly selected RF-with-rock model.
Model evaluation values (R^2 without BD included. Slopes and intercepts come from linear regression between modeled and measured K_d for publications. We compared R^2 and mean error; RMSE = root mean square error; d = index of agreement. Predictors in the ANN-no-BD and RF-no-BD models are: % Clay + %Silt + %Sand + BD; and predictors in the ANN-with-BD and RF-with-BD models are: % Clay + %Silt + %Sand + BD.

<table>
<thead>
<tr>
<th>Model</th>
<th>Intercept log (cm h^-1)</th>
<th>Slope</th>
<th>Adjusted R^2</th>
<th>ME log (cm h^-1)</th>
<th>RMSE log (cm h^-1)</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training data (n = 67)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN-no-BD</td>
<td>0.01</td>
<td>1.35</td>
<td>0.83</td>
<td>0.01</td>
<td>0.64</td>
<td>0.91</td>
</tr>
<tr>
<td>ANN-with-BD</td>
<td>0.00</td>
<td>1.34</td>
<td>0.85</td>
<td>0.00</td>
<td>0.59</td>
<td>0.93</td>
</tr>
<tr>
<td>RF-no-BD</td>
<td>0.00</td>
<td>1.00</td>
<td>0.46</td>
<td>0.00</td>
<td>0.99</td>
<td>0.80</td>
</tr>
<tr>
<td>RF-with-BD</td>
<td>0.00</td>
<td>1.00</td>
<td>0.49</td>
<td>0.00</td>
<td>0.94</td>
<td>0.82</td>
</tr>
</tbody>
</table>

References


