Core Ideas

- A robust PTF was developed to predict water contents at −1, −10, and −158 m tension.
- Drip infiltrometer experiments were inversely modeled to predict soil hydraulic properties.
- Both $\theta(h)$ and $K(h)$ can be accurately estimated from experimental data together with PTFs.

Soil Hydraulic Properties Determined by Inverse Modeling of Drip Infiltrometer Experiments Extended with Pedotransfer Functions

Ali Mehmandoost Kotlar,* Ioannis Varvaris, Quirijn de Jong van Lier, Lis Wollesen de Jonge, Per Møldrup, and Bo V. Iversen

A transient flow experiment using automated drip infiltrometers (ADIs) was performed on soil columns (about 6 dm$^3$) large enough to incorporate macro-pore flow effects. We investigated to what extent the estimated soil hydraulic parameters obtained from inverse modeling of these experiments are reliable. A machine learning based pedotransfer function (PTF) for prediction of water content at −1, −10, and −158 m pressure head was developed. Sensitivity analysis of the van Genuchten parameters (residual and saturated water content $\theta_r$ and $\theta_s$, fitting parameters $\alpha$, $n$, and $\lambda$, and saturated hydraulic conductivity $K_s$) in soils of sandy, silty, and clayey textures showed that the temporal variation of pressure heads in ADI scenarios was not sensitive to $\theta_r$ and $\theta_s$. The other parameters were accurately estimated from numerically synthesized data. The uniqueness of the estimated parameters did not change when a bias, representing experimental error, was added to the data set. In actual columns, using the temporal and spatial pressure head data from the ADIs and the water contents in the drier range predicted by the developed PTF resulted in a precise estimation of the van Genuchten parameters. Not including the PTF water contents resulted in non-unique estimated van Genuchten parameters.

The accurate modeling of water flow and budgeting of solute and heat transport in the vadose zone based on a numerical solution of the Richards equation depends on precise knowledge of the fundamental soil hydraulic properties (SHPs): water retention [$\theta(h)$] and hydraulic conductivity [$K(h)$] (Groh et al., 2018; Weninger et al., 2018). Direct measurement of $K(h)$ is laborious, therefore $K(h)$ is most commonly indirectly derived by using the measured saturated hydraulic conductivity, the $\theta(h)$ function, and an empirical parameter related to tortuosity and connectivity, which, in its turn, exhibits a very large spatial variability (Durner et al., 1999). Consequently, any method to directly quantify $K(h)$ represents an improvement of modeling quality, but this is predominantly neglected (Weninger et al., 2018; Weller et al., 2011).

Both soil heterogeneity at different scales, from millimeters to kilometers, and the large amount of required data restrict the use of pedotransfer functions for the determination of SHPs and especially the soil water retention function [$\theta(h)$] (Graham et al., 2018). Methods to measure SHPs, including pressure plates and in situ techniques like internal drainage experiments, are commonly laborious and costly. Most of these methods are inadequate to describe water dynamics at larger scales because they require hydraulic equilibrium, restricting experiments to small sample sizes (Scharnagl et al., 2011). A proper and real estimation of SHPs requires experiments performed on sample sizes larger than the representative elementary volume under transient water flow conditions. Subsequent inverse modeling of the observed data allows hydraulic properties to be effectively lumped...
Numerical methods to inversely model transient flow methods under diverse boundary conditions have recently gained considerable attention (Li et al., 2018; Rashid et al., 2015; Arora et al., 2011). In the upward infiltration method, a constant upward flux from the bottom is established as a boundary condition to obtain SHPs in the wetting branch. This method was proposed by Hudson et al. (1996) and modified through an applied constant bottom suction, allowing the cumulative flux data to be included as an auxiliary variable in the objective function (Young et al., 2002). The latest improvement in the upward infiltration method was to impose multiple tensions at the lower boundary (Moret-Fernández et al., 2016) and estimate SHPs without using tensiometers but small ring soil cores (100 cm³).

Another boundary condition is established in evaporation scenarios, initially proposed on horizontal (Gardner and Miklich, 1962) and vertical columns (Wind, 1969) and now frequently used for the simultaneous determination of retention and conductivity functions (Šimůnek et al., 1998; Romano and Santini, 1999; Schelle et al., 2011). These methods tend to fail in providing conductivities under near-saturated conditions.

Outflow experiments inversely model the measured outflow of water from a soil sample under established pressure or suction conditions. For instance, inverse modeling of one-step outflow experiments (Kool et al., 1985) supplied with further information of water content (van Dam et al., 1992) was improved by multistep outflow experiments (van Dam et al., 1994).

Similar to outflow experiments, a group of methods is based on the crust method (Bouma and Baker, 1974), in which vertical flow is established under a gravitational gradient alone, without a pressure head gradient. As an alternative to the original crust method, which requires a semi-infinite soil column, automated drip infiltrometers (Iversen et al., 2004) allow similar boundary conditions in a finite geometry. In automatic drip infiltrometer (ADI) experiments, data on \( K(h) \) can be obtained for the wet range. The extension of these pressure head data with information on water content may improve the well-posedness of the inverse modeling problem (Zhang et al., 2003). The objective of this study was to analyze ADI experimental data using inverse modeling techniques and to investigate if the extension of retention data to the drier range using prediction by a Gaussian process regression pedotransfer function trained on a local data set may result in a better assessment of SHPs.

**Materials and Methods**

**Sampling and Experimental Setup**

Fifteen undisturbed soil columns (20 cm high and 20 cm in diameter, volume 6.283 dm³) were sampled from the top layer of an experimental field located in Lund, Denmark (coordinates 55.24 N, 12.29 E), which belongs to the Danish Pesticide Leaching Assessment Program (Lindhardt et al., 2001). The soil has a sandy loam texture (64% sand, 23% silt, and 13% clay, 2.5% organic matter content) (Kotlar et al., 2019a). Winter wheat was cultivated at the location during the 5 yr before sampling. In the same field, small ring samples (100 cm³) were taken for water retention analysis using a sand box for pressure heads from −0.1 to −1 m and ceramic plate equipment for pressure heads between −1.6 and −150 m. Water content for the very dry soil was determined after one night of oven drying using a WP4-T dewpoint potentiometer (METER Group).

Analogous to the studies of McKenzie et al. (2001) and Weller et al. (2011), unsaturated hydraulic conductivity was measured using automated (step flow) drip infiltrometers (Fig. 1) as reported by Iversen et al. (2004). In the ADI setup, tensiometers recorded pressure heads at five depths in the column under step flow. A suction was applied at the bottom of the sample, and the inflow was adjusted until a steady state was established in which the five tensiometers showed similar readings and flow was due to a gravitational gradient only. Subsequently, suction was increased and the process repeated. Five to eight bottom suctions were applied, varying between −0.1 to −1 m pressure head, allowing determination of \( K(h) \) in this range of pressure heads. Saturated hydraulic conductivity was independently measured using the constant-head method as described by Iversen et al. (2004).

**Generation of Synthetic Data**

To evaluate the reliability of the SHPs estimated by inverse modeling of ADI scenarios, ADI experiments similar to the...
experimental setup were simulated numerically by HYDRUS-1D for three reference soils used by Vrugt et al. (2001) (Table 1).

The HYDRUS-1D model numerically simulates the temporal and spatial changes in water content or pressure head by solving the Richards equation (Šimůnek et al., 2008):

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} \left[ K(b) \frac{\partial h}{\partial z} + K(b) \right]$$

[1]

where $\theta$ is the volumetric soil water content, $t$ is time ($s$), $z$ is the vertical space coordinate (cm), $K$ is the hydraulic conductivity (cm s$^{-1}$), and $h$ is pressure head (cm). The soil hydraulic properties were modeled using the van Genuchten–Mualem constitutive relationships (van Genuchten, 1980):

$$S_e(h) = \frac{\theta_s - \theta_r}{\theta_s - \theta_r} \left( 1 + \left| \frac{\theta_s - \theta_r}{\theta_s - \theta_r} \right| \right)^{-m}$$

[2]

$$K(h) = K_s \left[ 1 - \left( 1 - \theta / \theta_s \right)^{\lambda} \right]^{1/ \lambda}$$

[3]

where $S_e$ is the saturation degree, $\theta_s$ and $\theta_r$ are volumetric saturated and residual water contents, and $\alpha$ (cm$^{-1}$), $m = 1 - \lambda$, and $\lambda$ are fitting parameters.

The simulated scenarios were chosen to represent the ADI experiment in 20-cm-high soil columns with three tensiometers (at depths of 7, 10, and 13 cm), discretized into 100 elements. A finer grid size was chosen near the virtual tensiometers. At the bottom of the soil column, a 1-cm-thick porous plate with $K_s$ equal to 2.7 cm d$^{-1}$ was simulated with van Genuchten (1980) parameters $n = 1.001$ and $\alpha = 10^{-20}$ cm$^{-1}$, thus ensuring that it remained fully saturated for the pressure heads applied (Šimůnek et al., 2008). Boundary conditions were multistep flux [always less than $K$ (1 cm)] and zero evaporation. The bottom boundary condition was a stepwise varying pressure head, never lower than $-100$ cm.

### Sensitivity Analysis

To study the parameter sensitivity of the described ADI scenario and to evaluate if insensitive parameters might be removed (Ritter et al., 2003; Lambot et al., 2004), one-parameter analyses as well as two-parameter analyses were performed using the same scenarios described for synthetic data generation. In the one-parameter analysis, 100 forward simulations (Monte Carlo) were performed by changing one of the parameters $\theta_\ell$, $\theta_s$, $\alpha$, $n$, $\lambda$, or $K_s$ within an interval around the respective true value. Similarly, in the two-parameter analysis, parameter pairs $\alpha - n$, $\alpha - K_s$, $\alpha - \lambda$, $n - K_s$, and $n - \lambda$ were simultaneously changed, resulting in $100 \times 100 = 10,000$ realizations. The two-parameter analyses allowed assessment of the corresponding response surfaces. In both one- and two-parameter analyses, the remaining parameters were kept at their true value.

Two statistical indicators were used to compare the Monte Carlo realizations of simulated pressure head readings ($h_{sim}$) with the pressure heads simulated with the reference values ($h_{ref}$): the root mean square error (RMSE) and the Nash–Sutcliffe efficiency (NSE), where $j$ is the number of samples:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N} (h_{sim} - h_{ref})^2}{N}}$$

[4]

$$NSE = 1 - \frac{\sum_{j=1}^{N} (h_{sim} - h_{ref})^2}{\sum_{j=1}^{N} (h_{ref} - \bar{h}_{ref})^2}$$

[5]

### Inverse Modeling

The inverse modeling to obtain the SHP functions $h(b)$ and $K(b)$ aimed to minimize the following objective function $F(\theta)$:

$$F(\theta) = \sum_{j=1}^{N} \left\{ \| \theta_{k,obs}(z_j, t_j) - \theta_{k,sim}(z_j, t_j) \|_2^2 \right\}$$

[6]

where $\theta_{k,obs}$ and $\theta_{k,sim}$ are the observed and simulated values, respectively, of a target parameter (e.g., pressure head) at depth $z_j$ and time $t_j$. Minimization of the objective function was performed using HYDRUS-1D (Šimůnek et al., 2008). HYDRUS-1D uses the Levenberg–Marquardt non-linear minimization method, a local gradient-type search algorithm, as opposed to global search algorithms that search the entire parameter space. Local search algorithms are generally sensitive to the initial parameter estimates (Šimůnek et al., 2005; Kellners et al., 2005).

To reduce the number of parameters to be estimated by inverse modeling, parameters $\theta_\ell$ and $\theta_s$ were fixed at experimentally observed values. Saturated water content was calculated based on bulk density and particle density (2650 kg m$^{-3}$). Estimated water content at pH = 5 using a WP4 psychrometric tensiometer was used as the residual water content. A stochastic bias was introduced, applying a 1-cm noise to the observed pressure heads to

![Table 1. Soil hydraulic parameters for reference soils (extracted from Vrugt et al., 2001).](image-url)
evaluate the sensitivity to experimental errors (Kool and Parker, 1988; Peters and Durner, 2008).

In this study, two data sets were used in inverse modeling to estimate the SHPs: (i) the traditional ADI data set consisting of measured pressure head data from the three most central tensiometers (eliminating the top and bottom tensiometers, thus avoiding possible boundary disturbance); and (ii) the ADI data set consisting of measured pressure head data from the three most central tensiometers extended with water contents predicted by a pedotransfer function.

The testing of inverse modeling and obtained parameters was accomplished by comparison of \( K(b) \) observed in ADI experiments and simulated by estimated parameters using the RMSE and the mean square percentage error (MSPE):

\[
MSPE = \frac{100\%}{n} \sum_{i=1}^{n} \left( \frac{K_{\text{sim},i} - K_{\text{obs},i}}{K_{\text{obs},i}} \right)^2
\]

where \( K_{\text{obs},i} \) is the \( i \)th hydraulic conductivity measured in the ADI experiment, \( K_{\text{sim},i} \) is the corresponding \( K(b) \) calculated using parameters obtained by inverse modeling, and \( n \) is the number of observations. For the cases where water contents predicted by PTF and simulated ones obtained from the van Genuchten parameters or the measured ones from small rings were compared with each other, the RMSE was the evaluation criterion.

**Pedotransfer Function for Water Contents**

A machine learning based pedotransfer function (PTF) was developed by Gaussian process regression (GPR) to obtain retention data for the dry range. Gaussian process regression uses nearest neighbors, considering the distance between neighbors based on covariance (or kernel) function. Closeness or similarity between two points (distance) is given by kernel functions (Rasmussen and Williams, 2006). Kernel similarities between a test point and each point of the training data are found to predict the target of the test point, thus kernel values of far points approach zero. Briefly, the Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( y_{\text{ts}} \) and \( y_{\text{ss}} \) are training and test targets (e.g., here water content points) and \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ss}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ts}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ts}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ts}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ts}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ts}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ts}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ts}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ts}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ts}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]

where \( K_{\text{ss}} \) is the covariance of the training data, \( K_{\text{ts}} \) of the test data, and \( K_{\text{ts}} \) between test and training data. Considering a Gaussian likelihood function, the predictive mean \( \hat{y}_{\text{ts}} \) for a given test point \( x_{\text{ts}} \) is

\[
\hat{y}_{\text{ts}} = K_{\text{ts}}^T K_{\text{ss}}^{-1} y_{\text{ss}}
\]
heads observed at the tensiometers. For the clay soil, the model is more sensitive to $n$. Sensitivity to $\lambda$ was not significant for the clay soil (Table 2), but for the studied Danish soils, with sandy loam texture, $\lambda$ was maintained as an optimization parameter.

The low sensitivity of $\theta_r$ and $\theta_s$, added to the fact that they can be experimentally measured, made it advantageous to fix these parameters at their true values instead of inversely modeling them. Therefore, only the parameters $\alpha$, $n$, $\lambda$, and $K_s$ were optimized to obtain the soil water retention and hydraulic conductivity functions.

Results of the two-parameter sensitivity analyses are shown in Fig. 2. For $n$ and $\alpha$, any change in $\alpha$ leads to large variation in the objective function.
function; however, \( n \) hardly affects the objective function and contours are parallel to the \( n \) axis (Fig. 2a). As a result, \( n \) is not sensitive to \( \alpha \) and the forward problem of ADI is highly sensitive to \( \alpha \). This makes estimation of these parameters cumbersome and non-unique when only pressure head data are used. The same applies to \( \lambda \) (Fig. 2b), which is difficult to predict because the objective function approaches its minimum value when \( \alpha \) is close to its reference (0.0191 cm\(^{-1}\)).

An increase in \( \alpha \) together with an increase in \( K_s \) leads to similar values for the objective function (Fig. 2c). Consequently, the forward problem of the ADI experiment has a high sensitivity to either \( \alpha \) or \( K_s \). There is a larger sensitivity to \( n-\lambda \), \( n-K_s \), and \( K_s-\lambda \) (Fig. 2d, 2e, and 2f, respectively). Observing the \( n-\lambda \) response surface, the objective function reaches a low value for \( n \) between 1.1 and 1.2, almost independent of \( \lambda \) between \(-3\) and \(3\). Low values for the objective function also occur when \( \lambda \) is between \(-1.2\) and \(-2\) and \( n \) varies from \(1.3\) to \(2.5\). For the case of \( n-K_s \) (Fig. 2e), there is a large area of insensitivity, especially when both parameters are above their reference values (1.159 and 13.8 cm d\(^{-1}\), respectively). There is no specific pattern for the sensitivity of the problem to the simultaneous variation of \( K_s \) and \( \lambda \) (Fig. 2f); however, low \( K_s \) with positive \( \lambda \) and high \( K_s \) with negative \( \lambda \) causes the objective function to increase.

**Inverse Modeling Using Synthetic Data**

The synthetic data obtained by forward modeling of a hypothetical ADI experiment on the three soil types from Table 1 are shown in Fig. 3a. As expected, a decrease in pressure head is faster and larger in the sandy soil. Temporal variation of the pressure head measured by tensiometers is the typical output of ADI experiments (Fig. 3a) and is used in inverse modeling to predict SHPs. Figure 3b shows a simulated measurement error imposed as a stochastic bias added to the numerically synthesized pressure head data (ADI output) for the silt soil.

The VG parameters estimated with and without added bias are shown in Table 3 together with their true values. The robustness

![Fig. 3. (a) Simulated pressure head with time at three depths (7, 10, and 13 cm) in simulated automatic drip infiltrometer experiments for three soil types, and (b) example of the imposed stochastic bias added to the pressure heads for the silt soil.](image)

<table>
<thead>
<tr>
<th>Soil</th>
<th>Scenario</th>
<th>van Genuchten parameters†</th>
<th>( K_s )</th>
<th>( \lambda )</th>
<th>RMSE, cm‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sand true value</td>
<td>0.0214</td>
<td>2.075</td>
<td>15.56</td>
<td>0.039</td>
<td>–</td>
</tr>
<tr>
<td>estimated</td>
<td>0.0214</td>
<td>2.075</td>
<td>15.54</td>
<td>0.038</td>
<td>0.002</td>
</tr>
<tr>
<td>estimated with bias</td>
<td>0.0213</td>
<td>2.10 (0.3)§</td>
<td>15.06</td>
<td>0.000 (0.006)</td>
<td>1.013</td>
</tr>
<tr>
<td>Silt true value</td>
<td>0.0160</td>
<td>1.370</td>
<td>6.00</td>
<td>0.5</td>
<td>–</td>
</tr>
<tr>
<td>estimated</td>
<td>0.0172</td>
<td>1.361</td>
<td>6.678</td>
<td>–0.001</td>
<td>0.018</td>
</tr>
<tr>
<td>estimated with bias</td>
<td>0.0171</td>
<td>1.363</td>
<td>6.610</td>
<td>0.000</td>
<td>0.052</td>
</tr>
<tr>
<td>Clay true value</td>
<td>0.0191</td>
<td>1.152</td>
<td>13.80</td>
<td>–1.384</td>
<td>–</td>
</tr>
<tr>
<td>estimated</td>
<td>0.0191</td>
<td>1.152</td>
<td>13.80</td>
<td>–1.384</td>
<td>0.000</td>
</tr>
<tr>
<td>estimated with bias</td>
<td>0.0182</td>
<td>1.154</td>
<td>12.83 (1.27)</td>
<td>–0.975 (0.333)</td>
<td>0.036</td>
</tr>
</tbody>
</table>

† \( \theta_r \), residual volumetric water content; \( \theta_s \), saturated volumetric water content; \( \alpha \), \( n \), and \( \lambda \), fitting parameters; \( K_s \), saturated hydraulic conductivity.
‡ RMSE of pressure heads between scenario with true values and scenario with estimated values or estimated values with bias.
§ Standard deviation in parentheses if \( \geq 0.5\% \) of average value.
of the proposed experiment to estimate SHPs is confirmed, especially for sandy and clayey soils, with reference parameters equal or very close to the estimated parameters (Table 3). As mentioned by Peters and Durner (2008), prediction of the hydraulic conductivity function is the most crucial part.

Adding a small bias to the input data (as illustrated in Fig. 3b) introduced some uncertainties in the estimated parameters, as shown in Table 3, except for the silty soil parameters. Prediction of $\lambda$ is cumbersome, occasionally with major differences between simulations with and without bias (sand soil) as well as large CVs (clay soil).

Pedotransfer Function

The developed GPR-PTF did not appear to be robust for the prediction of water contents at pF 1, but at pF 2, 3, and 4.2 well-trained GPR-PTFs were obtained for the training data set with, on average, RMSE and $R^2$ values of 0.015 and 0.99 for pF 2, 3, and 4.2, respectively, and the results of the testing stage are shown in Table 4. Feature selection of GPR based on kernel parameters allows elimination of predictors without a relevant effect on the response prediction (Kotlar et al., 2019a). Therefore, to predict $\theta_{pF2}$, $\theta_{pF3}$ and $\theta_{pF4.2}$, two components of textural data (sand, sand and clay, or sand and silt content), organic matter content, and/or bulk density were important predictors. Their predictive role was of similar importance for $\theta_{pF1}$ and $\theta_{pF3}$ (Table 4); however, the silt fraction plays a more predominant role in the prediction of $\theta_{pF4.2}$.

Given the observed performance of the four PTFs, values of $\theta_{pF2}$, $\theta_{pF3}$, and $\theta_{pF4.2}$ for each experimental ADI soil column were predicted with the trained GPR-PTF and used as additional data for the drier part of the inverse simulation. The use of both pressure heads and water content data improves the well-posedness of the inverse problem (Kool and Parker, 1988).

Considering the bad performance of the respective PTF with $R^2$ and RMSE of 0.5 and 0.035 (Table 4), $\theta_{pF1}$ data were not used. This, however, was acceptable because many data from the ADI experiments were already available for very wet soil conditions. The average and CV of the measured soil parameters and estimated water contents for the 15 soil columns are given in Table 5.

Inverse Modeling of Actual Measurements

Addition of the water content data obtained from the developed PTF to the pressure head data from the ADI experiments resulted in proper estimation of VG parameters, as illustrated in Table 6. Although the columns were sampled from the same field, there is an extensive variation in VG parameters except for $n$ (with a standard deviation of 0.08). Minimum and maximum values for $\alpha$ are 0.014 and 0.052 cm$^{-1}$ and for $\lambda$ are $-1$ to 24.4.

There was no agreement between values of observed and simulated $K_s$. It should be remembered that $K_s^{\text{sim}}$ values were obtained by inverse modeling from unsaturated columns, so they refer to extrapolation and are meant to be used for the prediction of unsaturated $K$ values. The $K_s^{\text{obs}}$ values (Table 6) were obtained from actual measurements in the laboratory. Similar findings were discussed by Pinheiro et al. (2018). Optimized parameters should be used only in the range where they were determined, and any extrapolation outside that range (as toward saturation) will be associated with a high level of uncertainty. Furthermore, there is spatial variability of $K_s$ across small distances due to the variability and connectedness of macropores formed by root channels and earthworms or variability in bulk density (compaction) highlighted by Ghanbarian et al. (2017) and García-Gutiérrez et al. (2018). According to Table 6, the columns with the largest bulk density ($>1.6$ g cm$^{-3}$) were Columns 2, 3, and 15, corresponding to the lowest measured $K_s$. The highest values of $K_s$ (9800, 5386, and 4068 cm d$^{-1}$) belong to Columns 13, 4, and 12, respectively, with an average bulk density of 1.45 g cm$^{-3}$.

When using only pressure head values in inverse modeling, no reliable SHPs were obtained. Values of RMSE for $\theta(h)$ are very high without using the GPR-PTF and reduce to more acceptable values when including the GPR-PTF water contents (Table 7). Because ADI data covered the wet range only, by excluding water content data no convergence was obtained for $K(h)$ parameters, and the resulting retention parameters showed high errors in water content prediction (Table 7). In the predictions without GPR-PTF, unrealistic parameter prediction was frequently observed, and the inclusion of some water content values ($\theta_{pF2}$, $\theta_{pF3}$, $\theta_{pF4.2}$) results in proper estimation of VG parameters, as illustrated in Table 6, except for the silty soil parameters. Prediction of $\lambda$ is cumbersome, occasionally with major differences between simulations with and without bias (sand soil) as well as large CVs (clay soil).

### Table 4. Statistical indicators of performance of the Gaussian process regression pedotransfer functions for the prediction of volumetric water contents at pressure heads $-0.1, -1, -10, \text{and} -158 \text{m}$ ($\theta_{pF1}, \theta_{pF2}, \theta_{pF3}, \text{and} \theta_{pF4.2}$, respectively) for testing data and respective predictors (bulk density [BD], organic matter content [OM], and sand, silt, and clay contents).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>$\theta_{pF1}$</th>
<th>$\theta_{pF2}$</th>
<th>$\theta_{pF3}$</th>
<th>$\theta_{pF4.2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.534</td>
<td>0.952</td>
<td>0.894</td>
<td>0.946</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.035</td>
<td>0.019</td>
<td>0.026</td>
<td>0.012</td>
</tr>
<tr>
<td>Predictors (in prediction)</td>
<td>BD (1.0)</td>
<td>BD (0.34)</td>
<td>BD (0.40)</td>
<td>silt (0.79)</td>
</tr>
<tr>
<td></td>
<td>sand (0.34)</td>
<td>sand (0.35)</td>
<td>OM (0.32)</td>
<td>clay (0.25)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sand (0.01)</td>
</tr>
</tbody>
</table>

### Table 5. Average and coefficient of variation (CV) of measured soil physical and hydraulic properties (sand, silt, and clay contents, organic matter content [OM], bulk density [BD], and saturated hydraulic conductivity [$K_s$]) and pedotransfer function-estimated water contents ($\theta$) at pF2, 3, and 4.2 for 15 soil columns.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sand</th>
<th>Silt</th>
<th>Clay</th>
<th>OM</th>
<th>BD</th>
<th>$K_s$</th>
<th>$\theta_{pF2}$</th>
<th>$\theta_{pF3}$</th>
<th>$\theta_{pF4.2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>61.8</td>
<td>22.3</td>
<td>13.4</td>
<td>2.5</td>
<td>1.5</td>
<td>2.88</td>
<td>0.30</td>
<td>0.21</td>
<td>0.09</td>
</tr>
<tr>
<td>CV, %</td>
<td>3.5</td>
<td>3.9</td>
<td>9.1</td>
<td>7.5</td>
<td>4.9</td>
<td>122</td>
<td>3.9</td>
<td>7.0</td>
<td>8.4</td>
</tr>
</tbody>
</table>
is required for proper estimation of $\theta(h)$ and $K(h)$. Table 7 also evaluates the accuracy of simulated $K(h)$ by Eq. [3] in terms of the RMSE and MSPE using the parameters in Table 6 compared with the observed values of $K(h)$. Small values of $K(h)$ or the dry-end tail of the $K(h)$ function may be overestimated by the obtained SHPs. The RMSE is less able to detect this because absolute values of errors in the dry zone are small. The relative indicator MSPE expressed this more adequately.

Small values of the RMSE do not always indicate a proper prediction of $K(h)$. For example, Columns 3, 7, and 10 correspond to similar RMSE values but resulted in very different values of MSPE (126, 1519, and 131, respectively). Similarly, the lowest RMSE value belongs to Column 6, with MSPE equal to 260.

Figure 4 shows the retention and hydraulic conductivity functions obtained in Soil Columns 7 and 12, the ones with the highest and lowest deviations, respectively, between measured values of $K(h)$ and those obtained by inverse modeling. In the retention graphs, the dashed line is obtained using the VG parameters from Table 6. In some cases, larger deviations occur between measured values of $K(h)$ and those obtained by inverse modeling in the drier part. This is the case, for example, for Soil Column 7, indicated in Fig. 4a by the red circle.

Finally, average and standard deviations of water retention data ($q_s$, $q_{pF1}$, $q_{pF2}$, $q_{pF3}$, $q_{pF4.2}$) measured by steady-state methods on small rings with those obtained from inverse modeling of ADI+PTF performed on large columns are compared in Fig. 5. The difference between measured and simulated water

---

**Table 6.** Estimated parameters obtained from inverse modeling of automatic drip infiltrometer experiments in 15 soil columns, including Gaussian process regression pedotransfer function water retention $[\theta(h)]$ values.

<table>
<thead>
<tr>
<th>Column</th>
<th>Parameter†</th>
<th>$\theta_s$</th>
<th>$\alpha$</th>
<th>$n$</th>
<th>$K_s^{\text{sim}}$</th>
<th>$K_s^{\text{obs}}$</th>
<th>$\lambda$</th>
<th>BD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{cm}^{-1}$</td>
<td></td>
<td></td>
<td></td>
<td>$\text{cm d}^{-1}$</td>
<td>$\text{cm d}^{-1}$</td>
<td></td>
<td>$\text{g m}^{-3}$</td>
</tr>
<tr>
<td>1</td>
<td>0.442</td>
<td>0.0360</td>
<td>1.342</td>
<td></td>
<td>186.8 (22.4)†</td>
<td>532</td>
<td>5.9 (0.2)</td>
<td>1.48</td>
</tr>
<tr>
<td>2</td>
<td>0.381</td>
<td>0.0249</td>
<td>1.338</td>
<td></td>
<td>8.1 (0.2)</td>
<td>43</td>
<td>−0.02</td>
<td>1.64</td>
</tr>
<tr>
<td>3</td>
<td>0.381</td>
<td>0.0512</td>
<td>1.263</td>
<td></td>
<td>21.60</td>
<td>60</td>
<td>0.01</td>
<td>1.64</td>
</tr>
<tr>
<td>4</td>
<td>0.46</td>
<td>0.0515</td>
<td>1.269</td>
<td></td>
<td>29.6 (0.3)</td>
<td>5386.2</td>
<td>0.66</td>
<td>1.43</td>
</tr>
<tr>
<td>5</td>
<td>0.457</td>
<td>0.0141</td>
<td>1.552</td>
<td></td>
<td>5.9</td>
<td>2146</td>
<td>24.38 (1.47)</td>
<td>1.44</td>
</tr>
<tr>
<td>6</td>
<td>0.437</td>
<td>0.0391</td>
<td>1.275</td>
<td></td>
<td>7.1 (0.3)</td>
<td>845</td>
<td>0.001</td>
<td>1.49</td>
</tr>
<tr>
<td>7</td>
<td>0.479</td>
<td>0.0491</td>
<td>1.247</td>
<td></td>
<td>32.5 (0.2)</td>
<td>2680</td>
<td>−1.03</td>
<td>1.38</td>
</tr>
<tr>
<td>8</td>
<td>0.430</td>
<td>0.0360</td>
<td>1.372</td>
<td></td>
<td>9.6 (0.3)</td>
<td>494</td>
<td>0.014 (2.35)</td>
<td>1.51</td>
</tr>
<tr>
<td>9</td>
<td>0.415</td>
<td>0.0249</td>
<td>1.329</td>
<td></td>
<td>3.0 (0.2)</td>
<td>1447</td>
<td>0.02</td>
<td>1.55</td>
</tr>
<tr>
<td>10</td>
<td>0.453</td>
<td>0.0397</td>
<td>1.281</td>
<td></td>
<td>24.1 (1.8)</td>
<td>1735</td>
<td>−0.04</td>
<td>1.45</td>
</tr>
<tr>
<td>11</td>
<td>0.445</td>
<td>0.0395</td>
<td>1.298</td>
<td></td>
<td>75.6 (4.0)</td>
<td>834</td>
<td>1.14 (0.12)</td>
<td>1.47</td>
</tr>
<tr>
<td>12</td>
<td>0.456</td>
<td>0.0407</td>
<td>1.294</td>
<td></td>
<td>21.3 (1.7)</td>
<td>4068</td>
<td>3.10 (0.24)</td>
<td>1.44</td>
</tr>
<tr>
<td>13</td>
<td>0.445</td>
<td>0.0405</td>
<td>1.245</td>
<td></td>
<td>24.66 (1.0)</td>
<td>9829</td>
<td>−0.002 (0.24)</td>
<td>1.47</td>
</tr>
<tr>
<td>14</td>
<td>0.415</td>
<td>0.0379</td>
<td>1.251</td>
<td></td>
<td>20.9 (1.4)</td>
<td>373</td>
<td>0.001 (0.04)</td>
<td>1.55</td>
</tr>
<tr>
<td>15</td>
<td>0.392</td>
<td>0.0520</td>
<td>1.239</td>
<td></td>
<td>25.6</td>
<td>32</td>
<td>0.023</td>
<td>1.61</td>
</tr>
</tbody>
</table>

† $\theta_s$, saturated volumetric water content; $\alpha$, $n$, and $\lambda$, fitting parameters; $K_s$, saturated hydraulic conductivity; BD, bulk density.

‡ Standard deviations in parentheses if $\geq 0.5\%$ of true value.

---

**Table 7.** Statistical indicators of root mean square error (RMSE) and mean square percentage error (MSPE) for the performance of water retention $[\theta(h)]$ and hydraulic conductivity $[K(h)]$ prediction by inverse modeling of automatic drip infiltrometer experiments in 15 soil columns including Gaussian process regression pedotransfer function (GPR-PTF) water contents and without GPR-PTF data.

<table>
<thead>
<tr>
<th>Column</th>
<th>Including GPR-PTF</th>
<th>Without GPR-PTF†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE $K(h)$</td>
<td>MSPE $K(h)$</td>
</tr>
<tr>
<td></td>
<td>$\text{cm d}^{-1}$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9.6</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>0.8</td>
<td>126</td>
</tr>
<tr>
<td>4</td>
<td>2.3</td>
<td>176</td>
</tr>
<tr>
<td>5</td>
<td>4.5</td>
<td>48</td>
</tr>
<tr>
<td>6</td>
<td>0.1</td>
<td>260</td>
</tr>
<tr>
<td>7</td>
<td>0.9</td>
<td>1519</td>
</tr>
<tr>
<td>8</td>
<td>0.5</td>
<td>93</td>
</tr>
<tr>
<td>9</td>
<td>0.5</td>
<td>13</td>
</tr>
<tr>
<td>10</td>
<td>0.8</td>
<td>131</td>
</tr>
<tr>
<td>11</td>
<td>3.5</td>
<td>54</td>
</tr>
<tr>
<td>12</td>
<td>0.0</td>
<td>5</td>
</tr>
<tr>
<td>13</td>
<td>0.5</td>
<td>50</td>
</tr>
<tr>
<td>14</td>
<td>0.9</td>
<td>65</td>
</tr>
<tr>
<td>15</td>
<td>0.3</td>
<td>14</td>
</tr>
</tbody>
</table>

† For many cases, the van Genuchten parameters obtained from the scenarios without GPR-PTF did not result in reasonable values for $K(h)$. 

---

VZJ | Advancing Critical Zone Science  
p. 8 of 11
contents for the drier soil ($\theta_{pF3}$ and $\theta_{pF4.2}$) is larger than for the wetter values. The water contents estimated by steady-state methods were higher than those inversely obtained from the ADI experiments. Differences increased to about 0.05 m$^3$ m$^{-3}$ in the drier region.

**Conclusions**

The evaluation of automated drip infiltrometer (ADI) scenarios performed here in numerical and real experiments to obtain soil hydraulic parameters from inverse modeling led to the following conclusions:

- Evaluated for three reference soils, the inverse modeling of ADI experiments showed insensitivity to residual and saturated water content. These parameters can better be measured or estimated than predicted from these experiments.

---

**Fig. 4.** Water retention [$\theta(h)$] and hydraulic conductivity functions obtained in (a) Soil Column 7, the column with the highest deviation between hydraulic conductivity $K$ measured and obtained by inverse modeling, and (b) Soil Column 12, with the lowest respective deviation. The red circle in (a) shows the high deviation for some of the observed values; PTF, pedotransfer function; MSPE, mean square percentage error.

**Fig. 5.** Average of water content $\theta$ at $pF1$, $pF2$, $pF3$, and $pF4$ obtained from steady-state methods (SSM) vs. corresponding values obtained from inverse modeling of automatic drip infiltrometer (ADI) + pedotransfer function (PTF).
• A numerical simulation of ADI scenarios showed robust prediction of soil hydraulic parameters by inverse modeling. Introduction of a random error to input data did not affect the parameter estimation notably.

• Including water contents predicted for the drier soil by a GPR-PTF for the inverse modeling ADI data from 15 large undisturbed columns collected from the same field located in Denmark, van Genuchten parameters \( \alpha, K_r, n, \) and \( \lambda \) were uniquely identified and the unsaturated hydraulic conductivities calculated from these data were in good agreement with measured \( K(\theta) \). Not including the PTF water contents resulted in non-uniquely estimated van Genuchten parameters.

Data Availability
Data from this study are available through the Dryad Data Depository (Kotlar et al., 2019d).

Acknowledgments
A.M. Kotlar is grateful to the FAPESP Foundation, Brazil, for providing financial support (scholarship) for his study (process no. 2016/18636-7). The support contributed greatly to the manuscript during the reviewing process.

References


Phosphorus Loss and Phosphorus Sensitive Water Areas (Fosforkortlaegning) and from the Danish Environment Agency financed project Mapping of Risk Areas for Phosphorus Loss and Phosphorus Sensitive Water Areas (Fosforkortlaegning) and from the Danish Pesticide Leaching Assessment Programme (pesticidvarsling.dk) is gratefully acknowledged. We also acknowledge two anonymous reviewers who contributed greatly to the manuscript during the reviewing process.


Advancing Critical Zone Science

VZJ | p. 10 of 11
from field data as an alternative to direct methods. Agric. Water Manage. 59:77–96. doi:10.1016/0378-3774(02)00160-9


