Spatial Prediction of Near Surface Soil Water Retention Functions using Hydrogeophysics and Empirical Orthogonal Functions

Justin Gibson
University of Nebraska-Lincoln, jgibson8@huskers.unl.edu

Trenton E. Franz
University of Nebraska-Lincoln, trenton.franz@unl.edu

Follow this and additional works at: https://digitalcommons.unl.edu/natrespapers
Part of the Natural Resources and Conservation Commons, Natural Resources Management and Policy Commons, and the Other Environmental Sciences Commons

https://digitalcommons.unl.edu/natrespapers/940

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Research papers

Spatial prediction of near surface soil water retention functions using hydrogeophysics and empirical orthogonal functions

Justin Gibson * , Trenton E. Franz

School of Natural Resources, University of Nebraska-Lincoln, Hardin Hall, 3310 Holdrege Street, Lincoln, NE 68583, USA

Article Info

Article history:
Received 11 October 2017
Received in revised form 13 March 2018
Accepted 17 March 2018
Available online 27 March 2018

This manuscript was handled by P. Kitanidis, Editor-in-Chief, with the assistance of J. Simunek, Associate Editor

Keywords:
Hydrogeophysics
Cosmic ray neutron probe
Soil water content
Soil hydraulic properties
Van Genuchten parameters

Abstract

The hydrological community often turns to widely available spatial datasets such as the NRCS Soil Survey Geographic database (SSURGO) to characterize the spatial variability of soil properties. When used to spatially characterize and parameterize watershed models, this has served as a reasonable first approximation when lacking localized or incomplete soil data. Within agriculture, soil data has been left relatively coarse when compared to numerous other data sources measured. This is because localized soil sampling is both expensive and time intense, thus a need exists in better connecting spatial datasets with ground observations. Given that hydrogeophysics is data-dense, rapid, non-invasive, and relatively easy to adopt, it is a promising technique to help dovetail localized soil sampling with spatially exhaustive datasets. In this work, we utilize two common near surface geophysical methods, cosmic-ray neutron probe and electromagnetic induction, to identify temporally stable spatial patterns of measured geophysical properties in three 65 ha agricultural fields in western Nebraska. This is achieved by repeat geophysical observations of the same study area across a range of wet to dry field conditions in order to evaluate with an empirical orthogonal function. Shallow cores were then extracted within each identified zone and water retention functions were generated in the laboratory. Using EOF patterns as a covariate, we quantify the predictive skill of estimating soil hydraulic properties in areas without measurement using a bootstrap validation analysis. Results indicate that sampling locations informed via repeat hydrogeophysical surveys, required only five cores to reduce the cross-validation root mean squared error by an average of 64% as compared to soil parameters predicted by a commonly used benchmark, SSURGO and ROSETTA. The reduction to five strategically located samples within the 65 ha fields reduces sampling efforts by up to ~90% as compared to the common practice of soil grid sampling every 1 ha.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Soil spatial datasets are important in different contexts. For instance, the hydrological community often turns to widely available datasets such as the NRCS Soil Survey Geographic database (SSURGO) (Soil Survey Staff, 2016) to characterize the spatial variability of soil across a field, watershed, or landscape of interest. When used to spatially characterize and parameterize watershed models (e.g. Soil and Water Assessment Tool, Neitsch et al., 2002), this approach has served as a reasonable first approximation when lacking localized soil data. Within agriculture, soil information plays a key role in the effort to increase efficiency in water and nutrient use, and in this effort an ever-increasing amount of information is being collected by farming implements (e.g. seed planters, weed sprayers, and yield monitors). However, in both use cases (watershed modeling and farming operations) soil information is often either left qualitative or informed from SSURGO as localized soil sampling is both expensive and time intense, particularly as average farm size continues to increase in the USA. Given that hydrogeophysical methods are data-dense, rapid, non-invasive, and relatively easy to adopt, they are valuable approaches to help dovetail localized soil sampling with spatially exhaustive datasets (Binley et al., 2015). Indeed, much work has been done to refine large scale surveys (i.e. SSURGO) as well as to identify and delineate smaller scale soil units (Doolittle and Brevik, 2014; Parsekian et al., 2015).

In order to parameterize watershed models one common practice is to combine SSURGO data (e.g. texture and bulk density) with a pedotransfer function (PTF) like ROSETTA (Schaap et al., 2001) to generate the required soil hydraulic parameters. While serving as a reasonable first approximation, this can be problematic for several reasons. First, soil properties provided by SSURGO (e.g. texture and bulk density) often are sourced from a limited number of soil cores...
extracted within a county. Land use and other local factors are well known to impact soil properties on the field to subfield level (Blanco-Canqui and Lal, 2007; Gambardella et al., 1994). Second, SSURGO zones are often delineated with covariates that are not necessarily causally linked with soil hydraulic properties. For instance, vegetation differences observed from aerial photographs are not necessarily the result of soil hydraulic properties driving differences in soil water content (SWC) but rather differences may be due to soil chemical properties (cation exchange capacity, pH, etc.). Lastly, soil properties may be gradational within a SSURGO zone due to topography-driven soil formational processes (Moore et al., 1993) as opposed to steep transitions.

Soil water content is well-known to govern key hydrological processes (runoff, infiltration, irrigation, drainage, etc.). Within agriculture, SWC is being aggressively monitored and managed (Irmak et al., 2010) across large areas and on a spatial scale finer than most current watershed models (Neitsch et al., 2002). This is in part due to water conservation regulation motivating farm management operations to reduce irrigation pumping volumes (Butler et al., 2016).

In need of finer scale information, additional information such as apparent electrical conductivity (ECa) and topography are commercially produced and/or utilized to create irrigation management zones for producers. In areas with similar ECa or elevation, soil properties are assumed to be reasonably similar and efforts are focused on sampling areas with variations. However, these covariates often produce noisy relationships with SWC patterns in part due to ECa being a function of not just SWC but also soil physical properties, solute concentration, and temperature (Haghdadi et al., 2015; Rodríguez-Pérez et al., 2011; Samouelian et al., 2005; Zhu et al., 2010). Mapping ECa over a range of temperature and SWC can lead to different maps due to these confounding factors (McCutcheon et al., 2006; Zhu et al., 2010). Despite this, the industry standard remains to produce one map per field. Here, we survey over a range of conditions, and then statistically contextualize the observed relative differences in the measured geophysical property throughout the field.

Based on our knowledge that soil migration/formation processes and soil water content redistribution often follow the same topographic gradients (albeit on different timescales) (Minasny and McBratney, 2016), we hypothesize that the time series of high-resolution geophysical measurements will provide the opportunity to derive high-resolution spatial maps of soil hydraulic properties which may later be used in more accurate quantitative modeling of soil water fluxes. Previous work has shown SWC patterns to be good predictors of soil physical properties (Korres et al., 2010; Pedrera-Parrilla et al., 2016), and while these soil physical properties are often correlated with soil hydraulic parameters (Patil and Singh, 2016; Vereecken et al., 2010; Wosten et al., 2001) a gap exists in explicitly linking spatial SWC patterns and soil hydraulic properties. Addressing this gap will likely better constrain flux estimates as a wide range of fluxes can occur within a single soil textural class (Groenendyk et al., 2015).

In order to approximate and quantify the spatial pattern of the time history of SWC, we utilize two common hydrogeophysical methods; SWC measured via a cosmic ray neutron probe (CRNP), and apparent bulk electrical conductivity (ECa) measured by electromagnetic induction (EMI). These measurements were taken over a wide range of SWC conditions in order to utilize an empirical orthogonal function (EOF) with the purpose of identifying temporally stable sub-field (less than 1 km²) spatial patterns. Our study site consisted of three 65 ha agricultural fields located on the western fringe of the United States Corn Belt in the state of Nebraska. This selected study area is an ideal location for testing our hypothesis in a real-world setting for three reasons. 1) The fields are located in a river valley where soil units are often heterogeneous and create complex patterns due to fluvial formational processes. 2) The proximity to the river valley makes the fields highly utilized for commercial agriculture via sprinkler irrigation. 3) Given the aridity of the region and demand for water resources precision agriculture techniques are actively being tested and adopted for optimizing irrigation management. This natural resource dependent socioeconomic environment is a critical location for demonstrating the validity and utility of these approaches.

The primary objectives of this study are to: 1) identify temporally stable spatial patterns using hydrogeophysical methods and statistical techniques, 2) measure and compare water retention functions of soil cores extracted from the range of identified SWC regions that are relatively wet, average, and dry, and 3) quantify and benchmark the skill of using identified SWC patterns as a covariate to predict soil hydraulic parameters. The objectives were carried out on three fields with varying soil types and topography in the study region of western Nebraska. Results of this analysis are then compared to water retention functions determined from a standard and widely used benchmark, SSURGO and ROSETTA. Lastly, a framework for carrying out these objectives in novel environments is presented, specifying the likely number of hydrogeophysical maps and soil cores needed.

2. Materials and methods

2.1. Description of study sites

The study area is located in western Nebraska where the South Platte River enters the state (Fig. 1) (N 41.007°, W 102.192°). The three study sites are each approximately 65 ha and 10 km apart primarily under irrigated maize production. The study area is semi-arid where annual crop referenced (maize) evapotranspiration (ETc) is significantly higher than precipitation (P) (HPRCC, 2016). The 10-year average annual ETc is 440 mm/yr and average annual ETc is 820 (mm/yr), as measured by the High Plains Regional Climate Center weather station (HPRCC, 2016) located within the study area near Brule, Nebraska. Data obtained from SSURGO (Soil Survey Staff, 2016) indicates that soil texture in the area falls within two USDA textural classes: sandy loam and loam. Lidar elevation rasters at 1 m resolution for each field were obtained from the USGS (USGS, 2016). Using the elevation data, relative elevation was calculated by subtracting the lowest elevation in the field from all elevations in order to investigate the influence of local topography.

2.2. Near surface hydrogeophysics

2.2.1. Electromagnetic induction

Between March 2016 and May 2017, a minimum of three hydrogeophysical surveys were collected at each of the three study sites using an all-terrain vehicle (ATV). See Table 1 for exact dates of data collection. Bulk electrical conductivity (ECa) maps were collected using a Duaalem-215 electrical magnetic induction (EMI) sensor (DUALEM, Milton, Canada). The EMI sensor has dual-receiver with an exploration depth of 1 m. The EMI boom was towed behind an ATV on a plastic sled at speeds of 8–15 km hr⁻¹ with 8–10 row spacing (~7–9 m) taking about 90 min to complete each survey. A Hemisphere GPS XF101 DGPS (Juniper Systems, Inc., Logan, UT) unit recorded the location of each measurement. Following basic quality assurance and quality control of the raw ECa data (Franz et al., 2011), a spatial map with 5 by 5 m resolution was created using an inverse-distance weighting procedure. We
note here that temporal differences in ECa mapping stem from soil temperature, SWC, and soil solute concentration (Franz et al., 2015; Robinson et al., 2009). SWC has been shown to account for approximately 50% of this variability (Brevik et al., 2006). We take advantage of this fact here to use changes in ECa as an indicator of relative change in SWC spatial patterns.

2.2.2. Cosmic-ray neutron probe

The mobile CRNP has been used to quantify spatial patterns of SWC across a range of spatial scales, from transects across the state of Hawaii to mesoscale maps around Tucson Arizona and central Nebraska (Chrisman and Zreda, 2013; Desilets et al., 2010; Franz et al., 2015). Here we use the mobile CRNP to map the spatial variability of SWC within each 65 ha field over relatively short time periods (~1.5 h) using the same ATV and collection pattern as described above. We also note that minimal vegetation present at the time of sampling (<0.5 kg/m²) due to crop planting and harvest schedules on site. The mobile CRNP records epithermal neutron intensity integrated over one minute counting intervals. The change in epithermal neutron intensity is inversely correlated to the mass of hydrogen in the measurement volume (Zreda et al., 2012). The authors note that SWC changes are by far the largest change in hydrogen mass (Mclanett et al., 2014). Numerous validation studies across the globe (Bogena et al., 2013; Franz et al., 2016, 2011; Hawdon et al., 2014) have shown the CRNP to have area-average measurement accuracies of root mean square errors (RMSE) less than 0.03 cm³ cm⁻³ against a variety of industry standard SWC point scale probes. The measurement volume is roughly a disk, with a ~130–250 m radius and a penetration depth of 0.15–0.40 m (Köhli et al., 2015) depending on local conditions (e.g. elevation, water vapor, soil water content etc.). For simplicity, a constant penetration depth of 0.3 m was assumed for all surveys.

In order to provide a SWC map, first a spatial map of neutron intensity was estimated, then a calibration function was applied following details in Franz et al. (2015) for use in agricultural fields. We note that if spatial patterns are of only of interest, then the spatial neutron field could be used directly. However, the quantitative difference in SWC patterns may provide insight to the investigator to decide if differences between surveys are meaningful, whereas differences in neutron counts may be opaque. The neutron intensity map is created in two steps. First, a drop-in-the-bucket pre-processing step is applied (Chan et al., 2014), where a dense grid is generated (here 20 by 20 m) and all raw data points are found within a certain radius (here 50 m). Then, the average of all raw data found within the search radius is assigned to the grid center. This oversampling approach is necessary for sharpening the image

**Table 1** Summary of geophysical survey dates and explained variance of the 1st EOF.

<table>
<thead>
<tr>
<th>Field</th>
<th>CRNP Survey Dates</th>
<th>Explained Variance of 1st EOF (%)</th>
<th>ECa Survey Dates</th>
<th>Explained Variance of 1st EOF (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1S1</td>
<td>2016: 03/11, 03/22, 05/09 2017: 05/02, 05/03</td>
<td>73.3</td>
<td>2016: 03/11 2017: 05/02, 05/03</td>
<td>95.9</td>
</tr>
<tr>
<td>T1S3</td>
<td>2016: 03/11, 03/22, 06/07, 06/08</td>
<td>65.8</td>
<td>2016: 03/11 2017: 05/02, 05/03</td>
<td>91.4</td>
</tr>
<tr>
<td>T1S4</td>
<td>2016: 03/11, 03/22, 06/07, 06/08, 10/16, 10/17</td>
<td>67.1</td>
<td>2016: 03/11 2017: 05/02, 05/03</td>
<td>80.6</td>
</tr>
</tbody>
</table>
quality and is a common strategy used in remote sensing analyses (Chan et al., 2014) when overlapping area average observations are collected, as is the case with the CRNP in this study. Next, an inverse-distance-weighted approach is used on the resampled 20-m grid to provide the 5-m neutron intensity estimate. Finally, the neutron intensity gridded estimate is converted to SWC following Franz et al. (2015). The authors refer the reader to the rapidly growing CRNP literature (see Andreassen et al., 2017; Zreda et al., 2012) in lieu of providing full details of the methodology here.

2.3. Soil hydraulic property measurement

In each field, up to 18 soil cores were extracted at locations that encompassed the range of variability determined by the geophysical surveys and elevation (see supplementary data (DS02) for core locations). The sampling strategy was informed based on visual inspection of the maps (EOF and elevation), and sample locations were prioritized based on: 1) ensuring that the numerical scale of each data source had at least three locations sampled in the high, low, and mid values, 2) areas were avoided near known disturbances in soil (e.g., irrigation recirculation pits, center pivot roads) and 3) large areas with similar EOF values were prioritized over small areas with significant variability. These cores were undisturbed and extracted at a depth of 20 cm, inside a steel cylinder of volume 250 cm³ with a height of 5 cm (UMS, GmBH, Munich, Germany). Cores were placed in cold storage (4°C) until they were sampled in the laboratory. Water retention data was determined using two Decagon devices: a HYPROP and a WP4C to cover a wide range of soil tension values. The combination of both devices has been shown to produce reasonably continuous water retention data for a range of soil textures (Schelle et al., 2013). The HYPROP is a benchtop evaporation system that produces continuous measurement of both SWC and soil tension from saturation (pF ~ 0) to a pF of 3, where pF is the log10 of the absolute value of soil tension in units of cm. The WP4C utilizes the chilled mirror technique (Gee et al., 1992; Scanlon et al., 2002) and has a measurement range from pF 3 to pF 6, which was used to measure two points near a pF of 4.2 (typically one below and one above). Water retention data was fit using the constrained van Genuchten-Mualem model (Mualem, 1976; van Genuchten, 1980). Saturated hydraulic conductivity (Ksat) measurements were taken on the same core using a Decagon KSAT device under falling head. Soil bulk densities were taken after soil hydraulic parameters were measured, by dividing the dried mass (dried at 105°C for 24 h) by the known volume of the core. Saturated water contents (θs, cm³ cm⁻³) were calculated by:

\[ θ_s = 1 - \left( \frac{ρ_s}{ρ_f} \right) \]

(1)

where \( ρ_s \) was measured soil dry bulk density (g cm⁻³) and \( ρ_f \) is mineral grain density, assumed here as 2.65 g cm⁻³. Because \( ρ_s \) is a direct conversion of bulk density, only bulk density will be correlated with environmental covariates hereafter. Although bulk density can be a dynamic parameter (e.g., land management changes, compaction by traffic, erosion) we note here that conditions were fairly consistent over the approximately 1 year the surveys were conducted over. This combined with the extracted depth of 20 cm we do not expect significant changes to have occurred.

The remainder of the van Genuchten-Mualem model (Mualem, 1976; van Genuchten, 1980) soil hydraulic model is:

\[ θ(h) = \begin{cases} \theta_r + \frac{θ_s - θ_r}{1+10^{a(h+b)}}, & h < 0 \\ θ_s, & h ≥ 0 \end{cases} \]

(2)

\[ K(Se) = K_{sat} \times S_e^{1/2} \times \left[1 - \left(1 - S_e/m\right)^n\right]^2 \]

(3)

where \( θ \) is SWC (cm³ cm⁻³); \( θ_r \) (cm³ cm⁻³) and \( θ_s \) (cm³ cm⁻³) are residual and saturated SWC, respectively; \( h \) (cm) is pressure head; \( K \) (cm day⁻¹) and \( K_{sat} \) (cm day⁻¹) are unsaturated and saturated hydraulic conductivity, respectively; and \( S_e \) is saturation degree (-) calculated as:

\[ S_e = (\theta - θ_r)/(θ_s - θ_r) \]

(4)

With respect to the fitting factors, \( a \) (1/cm) is inversely related to air entry pressure, \( n \) (-) measures the pore size distribution of a soil with \( m = 1-1/n \), and \( l \) (-) is a parameter accounting for pore space tortuosity and connectivity, assumed to be equal to 0.5 here.

2.4. Statistical methods

2.4.1. Empirical orthogonal function (EOF)

To identify the spatial variability of ECa from EMI measurements and SWC from CRNP measurements, an EOF analysis was used on both the EMI ECa and CRNP SWC geophysical property maps. Full details on the multivariate statistical EOF analysis are provided in previous literature (Korres et al., 2010; Perry and Niemann, 2007) and only a summary is provided here. The EOF analysis decomposes the observed SWC and ECa variability measured by the hydrogeophysical surveys into a set of orthogonal spatial patterns (EOFs), which are invariant in time, and a set of time series called expansion coefficients (ECs), which are invariant in space (Perry and Niemann, 2007). Multiplication of the EOFs and ECs will exactly reconstruct the original data. Often the number of necessary coefficients (i.e., eigenvectors) to reconstruct most of the data is less than the original dataset (i.e., determined by the ranked eigenvalues), thus the procedure can be used to reduce the dimensionality of the dataset while preserving the key information, here dominant geophysical property spatial patterns. The authors note that EOF is nearly identical to Principal Component Analysis (PCA) save the splitting of axes of variation into spatial and temporal coefficients instead of arbitrary linear combinations.

Using this approach, the EOF analysis is able to contextualize the behavior of geophysical property at any given point in the field relative to the mean geophysical property of the field as a whole. For example, points that are persistently dry relative to the mean will be represented with a negative reprojected coefficient. Similarly, points that are persistently wet relative to the mean will be represented with a positive coefficient. The magnitude of each coefficient is assigned based on the difference between the mean behavior of the field and the mean behavior of each respective point. Each point is then spatially reprojected and a continuous surface is created. EOF surfaces from the ECa and CRNP mapping along with the LIDAR elevation data will serve as the three environmental covariates utilized in this study following (Franz et al., 2017).

2.4.2. Regression of environmental covariates and soil hydraulic parameters

Following the EOF analysis, measured soil hydraulic parameters were regressed with the environmental covariates using a simple linear model to determine correlation. This provides the ability to spatially estimate soil hydraulic parameters using the exhaustive spatial datasets. Similar approaches have been carried out in other studies (Pedrera-Parrilla et al., 2016) referring to this as PCA instead of EOF. We note here that because the EOF analysis provides results that are both invariant and incommensurate, regressions from one study field will not be comparable to another.

2.4.3. Bootstrap validation

In order to determine 1) the accuracy of the regressed parameter relative to measured parameter and 2) how many
Fig. 2. Repeat CRNP surveys taken at T154 (approximately 65 ha in size). Black lines indicate SSURGO soil unit boundaries. Black circles indicate locations where soil samples were extracted in the field. Surveys are presented in chronological order. See Table 1 for survey dates. See supplementary data 1 for survey data.
samples are necessary for a RMSE to converge, we utilized a bootstrap validation analysis using the statistical package R (Version 3.3.3 2017). This was carried out by randomly selecting n-1 samples (where n is the number of samples extracted at each study site), building a simple linear model, and then determining the RMSE of the remaining validation samples relative to the model predicted value. This process was repeated 1000 times, and then repeated again with n-2 for the training set and so on until only 3 samples were used as a training set, with the rest used as a validation set. Results are also contextualized with a comparison of using SSURGO soil texture (sand, silt, and clay percentages) and bulk density data as inputs to the ROSETTA pedotransfer function model to estimate soil hydraulic parameters. We assumed this framework is a reasonable benchmark given the widespread use of ROSETTA (Schaap et al., 2001) with the hydrological and agricultural communities.

Fig. 3. The underlying spatial patterns identified by the 1st EOF (T1S4 - approximately 65 ha in size). Initially, as more surveys are included into the EOF analysis, new features emerge in the spatial pattern. This is followed by the spatial pattern converging with only minor changes in the spatial boundaries. Black lines are SSURGO soil unit boundaries and black circles are locations where soil cores were extracted. See supplementary data 1 for survey data.
3. Results and discussion

3.1. Near surface hydrogeophysical surveys

Fig. 2 illustrates seven CRNP rover surveys along with the accompanying calculated 1st EOF within field T1S4, serving as an example case to present the EOF result along with the underlying spatial data. Table 1 presents geophysical survey summary data collected in each field, along with the associated statistical information from the EOF analyses. In the case of T1S4, the southwest edge of the field tends to be relatively dry (SWC 0.15–0.20 cm³ cm⁻³). The north central part of the field tends to be relatively wet (SWC 0.35–0.40 cm³ cm⁻³) when compared to the southwest edge of the field. Both of these patterns are highlighted in the 1st EOF result demonstrating the efficacy of the method. We note that similar results were found in the other two field sites so only the EOFs of the geophysical properties will be presented. A supplementary table (DS01) is provided with the 5 m processed data for all surveys and study sites.

While there are similarities amongst all the CRNP surveys, the location of the wettest areas varies from survey to survey. This observation underscores the need for repeat geophysical mapping. To highlight this further, Fig. 3 illustrates how the EOF analysis evolves as more maps are added into the analysis. Of particular note we find that between the two and three survey analysis, new wet features emerge. As more surveys are added in, the boundaries of EOF features tend to converge with the five and seven survey EOF analyses being fairly similar. This underscores the need for multiple surveys in the attempt to link hydrogeophysical techniques with soil properties, particularly where fine scale information is desired for agricultural management decisions.

3.2. Soil hydraulic parameter measurements

Raw observations from the soil water retention function (WRF) measurements are presented in Fig. 4. While data obtained from SSURGO indicated 2 textural classes for each field, we note the wide range of water retention functions. The nature of an evaporation experiment provides significantly denser data in the relatively wet portion of the WRF which is critical for constraining its shape. A wide range of soils were collected during the sampling effort as reflected by the spread of WRFs and sample hydraulic property results can be found in the supplementary data. Measurement of saturated hydraulic conductivity produced a wide range of values. This is consistent with similar studies often finding a range of at least one order of magnitude (Gwenzi et al., 2011; Papanicolaou et al., 2015). Residual water content only ranged from 0 to 0.05 cm³ cm⁻³ with most samples set to 0 in the fitting process. Due to the lack of variability in residual water content, this variable was not regressed against the environmental covariates.

3.3. Separation of WRFs using hydrogeophysics

Fig. 5 presents the fitted WRFs of cores extracted from each field. In both T1S1 and T1S4, both hydrogeophysical methods were able to separate the range of WRFs. For example, in the upper right plot, the WRFs with a low CRNP EOF value (represented with a red color) group together and those with a high CRNP EOF value (represented with a blue color) group together. The coarser textured samples have WRFs that group lower and finer samples group higher, which was consistent with the EOF values. T1S3 had little spatial variability in the WRFs sampled across the field (except for hₐ), and as a result proved difficult for any method to describe the variability. This speaks to the limitation of the method – in fields with no to minimal soil property variability, the hydrogeophysical methods may not provide a robust correlation to predict small variations. We also note that in this same field, WRFs of coarse textured soil samples were observed in higher relative elevations and vice versa for finer textured soils. However, this is contrasted with the opposite trend observed in the other two fields and this highlights the challenge of predicting small variation in soil properties from elevation alone. We also note that the performance of relative elevation was likely enhanced by the hydrogeophysical surveys informing ideal sampling locations. If relative elevation was used as the only covariate, performance may have been reduced.

Given that the environmental covariates were able to separate fitted WRFs, we further investigated the correlation between the environmental covariates, the laboratory estimated WRF parameters, and bulk density. However, we note that correlations may be somewhat limited or weak due to equifinality associated with the fitting process and nature of soil water flow (Beven and Freer, 2001; Binley et al., 1989) as well as the scale mismatch between the geophysics measurement volumes (5 by 5 m) and the extracted soil cores. Correlations between all three covariates are presented in Table 3. The three environmental covariates were all correlated amongst themselves (Pearson’s r ranging from 0.63 to 0.95). Given the tendency for topographically low areas to typically be relatively wet and often the most clay rich in the field, the results were not unexpected. Correlations between the environmental covariates and both hₐ and Kₚₐ were low ranging from 0.1 to 0.45. Both of these parameters (hₐ and Kₚₐ) are defined in the wet range of SWC and previous work has shown that both parameters drive fluxes under wet conditions (Jiménez-Martínez et al., 2009; Wang et al., 2015; Wang and Franz, 2015). Because high SWC within a field limits the practical feasibility of mapping (too wet for a vehicle to travel), we are unable to capture geophysical patterns on the very wet end of the curve. The lack of our ability to map in wet conditions may limit our ability to predict these parameters (hₐ and Kₚₐ) spatially and deserves more attention in future studies. We speculate the spatial pattern for the very wet end is likely different (follows topography more closely) and persists for a much shorter period of time. Future work should focus on collecting spatial datasets during these wet short time periods using unmanned aerial systems and multispectral data as summarized by (Minasny and McBratney, 2016).
3.4. Bootstrapping validation

Results of the bootstrapping validation are illustrated in Fig. 6. In general, most cross validation RMSE reduction values converged after 5 samples selected within the training set of up to 18 samples per site. This is a significant finding considering each site was approximately 65 ha. In most agricultural soil sampling, ~1 ha grid sampling is recommended (http://cropwatch.unl.edu/ssm/soil-sampling), requiring 65 samples to cover this area. This highlights the potential savings in cost, time, and labor of a priori hydrogeophysical mapping being able to reduce the sampling effort by up to 90%. Often, densely gridded strategies are carried out in order to ensure the variability in a field is captured, given that the underlying spatial variability is unknown. However, by using the proposed environmental covariates, the range of the variability can be rapidly identified within a field, and then sampled strategically. We note that additional research is needed to validate this finding of 5 sample locations per 65 ha, particularly where underlying soil heterogeneities and correlation length scales of soil texture vary.

Additional summary statistics are presented in Table 2. To serve as a reference benchmark to compare RMSE values, average parameters were calculated from all samples in each field. RMSE reduction relative to SSURGO was calculated as:

$$\text{RMSE Reduction} = \left( 1 - \frac{\text{RMSE}_{\text{covariate}}}{\text{RMSE}_{\text{SSURGO}}} \right) \times 100$$

Where RMSE_{covariate} is the RMSE using the covariate prediction obtained by bootstrapping (with a training set of 17 samples), and RMSE_{SSURGO} using the SSURGO based PTF prediction. Across all parameters, RMSE values were reduced on average by 64% relative to predictions from SSURGO (and ROSETTA where applicable). Even in fields with low correlations between the parameters and environmental covariates, low RMSEs were also obtained. In these cases,
while the environmental covariates may not have served as a better estimate relative to using the mean value of measured parameters, they likely reduced the number of samples necessary to obtain a representative mean value compared to a gridded sampling strategy. Therefore, even in fields with relatively little soil variability, these methods are still useful to ensure the range of soil variability is sampled.

3.5. Prediction of $K_{sat}$ and $\alpha$ parameters

Given the practical restrictions of driving a vehicle over very wet soil, the geophysical methods were unable to capture the extreme wet end of the range of SWC. We speculate that capturing the very wet spatial pattern may be key to spatial prediction of both $K_{sat}$ and $\alpha$ as these parameters control the magnitude of fluxes at the wet end of soils. With this in mind, future work may test this hypothesis by placing a CRNP on a rotating center pivot lateral that is able to more efficiently move through the field under wet conditions. In non-irrigated areas, a dense SWC sensor grid may be able to capture spatial patterns. Using the combination of these environmental covariates to inform placement of sensors shows some promise to aid in experimental design (Barker et al., 2017).

3.6. Informing management decisions in agriculture

Current agricultural practices are shifting to finer and finer scale management given the advent of Real Time Kinematic GPS. Soil hydrology is often a key underlying factor in yield differences...
within a field. However, most producers lack soil data that is of the same resolution that their planters, sprayers, and yield monitors provide. While commercial products such as VERIS (Tualatin, OR) exist to help bridge this gap, such technologies currently only map each field once, and are therefore more susceptible to temperature and soil solute differences impacting ECa and soil property correlation.

Numerous commercial modeling efforts (The Climate Corporation, Encirca, ClearAg etc.) are currently attempting to inform both SWC and nitrogen management. As these models and others (Encirca, ClearAg etc.) are currently attempting to inform both

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Metric</th>
<th>T1S1</th>
<th>T1S3</th>
<th>T1S4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SSURGO</td>
<td>Relative Elevation</td>
<td>CRNP EOF</td>
<td>Eca EOF</td>
</tr>
<tr>
<td>Bulk Density</td>
<td>Average (g/cm³)</td>
<td>1.60</td>
<td>1.60</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>RMSE (g/cm³)</td>
<td>0.25</td>
<td>0.25</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>RMSE/Average (%)</td>
<td>16</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>RMSE Reduction (%)</td>
<td>–</td>
<td>70</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Average (-)</td>
<td>1.422</td>
<td>1.422</td>
<td>1.422</td>
</tr>
<tr>
<td></td>
<td>RMSE (-)</td>
<td>0.265</td>
<td>0.144</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td>RMSE/Average (%)</td>
<td>19</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>RMSE Reduction (%)</td>
<td>–</td>
<td>46</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Average (1/cm)</td>
<td>0.016</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>RMSE (1/cm)</td>
<td>0.008</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>RMSE/Average (%)</td>
<td>53</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>RMSE Reduction (%)</td>
<td>–</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Average (cm/day)</td>
<td>51</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>RMSE (cm/day)</td>
<td>141</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>RMSE/Average (%)</td>
<td>276</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>RMSE Reduction (%)</td>
<td>–</td>
<td>91</td>
<td>91</td>
</tr>
</tbody>
</table>

Table 2
Correlation matrices of environmental covariates and soil hydraulic parameters. Correlations greater than 0.6 are marked in bold.

Table 3
Summary statistics for the cross-validation analysis.

3.7. Framework for use at novel sites

As best practice for use in novel settings we recommend the following procedure. Conduct a minimum of four hydrogeophysical surveys over a range of wet and dry field conditions.
(2017) found that 4–5 maps at varying water contents established 1st EOF coefficients within 5%. Fig. 3 also illustrated minimal changes in EOF values and zone locations following 3–4 SWC maps. Following completion of these surveys, we recommend extracting 5–7 cores (i.e. ~1 core per 10 ha) spanning the range of observed 1st EOF values and elevation data. Here we found that this number of local samples reduces RMSE by approximately 50%. We note that greater number of samples did not significantly reduce the cross validation RMSE and illustrate a diminishing return on information gained. We note that additional research is needed to validate these recommendations, particularly in study sites where the underlying soil heterogeneities and correlation length scales of soil texture may vary.

3.8. Environmental covariate selection

Within this work three environmental covariates were correlated with soil hydraulic parameters. The two geophysical methods have uncertainties inherent to measurement with changing state variables (e.g. temperature, SWC, etc.). While the EOF analysis helps reduce the impact of these time varying factors, a portion of the spatial variance remains unexplained. With regards to the error in the geophysical data, we believe that the first EOF axis of explained variance serves as a proxy (Table 1). While relative elevation is often correlated with soil textures, in areas with lower relief this correlation may not be as dependable (as was the case in T1S3). For these reasons selecting one data source to predict soil hydraulic parameters can be challenging, and best-case use will likely incorporate a portfolio of environmental covariates.

4. Conclusions

In this work, we tested different environmental covariates to help constrain the spatial variability of WRFs and to predict soil hydraulic parameters where no measurement information was used in a cross-validation experiment. We note that using hydrogeophysics to inform a more strategic sampling approach would drastically reduce the number of extracted samples, cutting the number by up to ~90% compared to current soil sampling strategies presented by agricultural extension. Using these approaches, we were able to reduce the RMSE of soil hydraulic parameters described in SSURGO (and using ROSETTA where applicable) by 64% on average. We anticipate that such datasets will provide a key missing piece of information to better evaluate the next generation of watershed and crop models to aid in real-time management decisions. Future work will focus on collecting geophysical data over very wet SWC in order to help better predict α and Ksat. Furthermore, future modeling work will evaluate the impact of these different soil hydraulic parameters on both water fluxes and the fate of fertilizers.

Acknowledgements

This research is supported financially by the Daugherty Water for Food Global Institute at the University of Nebraska and The Nature Conservancy. TEF would also like to acknowledge the financial support of the USDA National Institute of Food and Agriculture, Hatch project #1009760. Access to field sites and datasets is provided by land owners, The Nature Conservancy, Western Nebraska Irrigation Project, and South Platte Natural Resources District. A special thank you to the local land owners, Jacob Fritton, John Gates, Catherine Finkenbiner, William Avery, and Matthew Russell for various assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jhydrol.2018.03.046.

References


