Spatial Statistics 14 (2015) 70–90

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set

Caley K. Gasch [a,*], Tomislav Hengl [b], Benedikt Gräler [c], Hanna Meyer [d], Troy S. Magney [e], David J. Brown [a]

[a] Department of Crop and Soil Sciences, Washington State University, USA
[b] ISRIC — World Soil Information/Wageningen University and Research, The Netherlands
[c] Institute of Geoinformatics, University of Münster, Germany
[d] Department of Geography/Environmental Informatics, Philipps-Universität Marburg, Germany
[e] College of Natural Resources, University of Idaho, USA

ARTICLE INFO

Article history:
Received 1 November 2014
Accepted 1 April 2015
Available online 29 April 2015

Keywords:
Digital soil mapping
Random forests algorithm
Regression-kriging
Soil sensor network

ABSTRACT

The paper describes a framework for modeling dynamic soil properties in 3-dimensions and time (3D + T) using soil data collected with automated sensor networks as a case study. Two approaches to geostatistical modeling and spatio-temporal predictions are described: (1) 3D + T predictive modeling using random forests algorithms, and (2) 3D + T kriging model after detrending the observations for depth-dependent seasonal effects. All the analyses used data from the Cook Agronomy Farm (37 ha), which includes hourly measurements of soil volumetric water content, temperature, and bulk electrical conductivity at 42 stations and five depths (0.3, 0.6, 0.9, 1.2, and 1.5 m), collected over five years. This data set also includes 2- and 3-dimensional, temporal, and spatio-temporal covariates covering the same area. The results of (strict) leave-one-station-out cross-validation indicate that both models accurately predicted soil temperature, while predictive power was lower for water content, and lowest for electrical conductivity. The kriging model explained 37%, 96%, and 18% of the variability in water content, temperature, and electrical conductivity respectively versus 34%, 93%, and 5% explained by the random forests model. A less rigorous simple cross-validation of the random forests model

* Corresponding author.
E-mail address: caley.gasch@wsu.edu (C.K. Gasch).

http://dx.doi.org/10.1016/j.spasta.2015.04.001

indicated improved predictive power when at least some data were available for each station, explaining 86%, 97%, and 88% of the variability in water content, temperature, and electrical conductivity respectively. The high difference between the strict and simple cross-validation indicates high temporal auto-correlation of values at measurement stations. Temporal model components (i.e. day of the year and seasonal trends) explained most of the variability in observations in both models for all three variables. The seamless predictions of 3D + T data produced from this analysis can assist in understanding soil processes and how they change through a season, under different land management scenarios, and how they relate to other environmental processes.

1. Introduction

Comprehension of dynamic soil properties at the field scale requires measurements with high spatial and temporal resolution. Distributed sensor networks provide frequent *in situ* measurements of environmental properties at fixed locations, providing data in 2- or 3-dimensions and through time (Porter et al., 2005; Pierce and Elliott, 2008). While sensor networks produce ample data for observing dynamic soil properties, data processing for inference and visualization become increasingly difficult as data dimensionality increases. Ideally, the end product should consist of seamless interpolations that accurately represent the spatial and temporal variability in the property of interest. These products can then be used for predictions at unobserved locations, they can be integrated into process models, and they can simply aid in visualization of soil properties through space and time.

Multiple approaches have been developed for spatial interpolation of soil properties and digital soil mapping, including:

(1) multiple regression models based on the soil forming factors, terrain attributes, spatial coordinates, or derived principal components (McKenzie and Ryan, 1999);
(2) smoothing (splines) and neighborhood-based functions (Mitas and Mitasova, 1999);
(3) geostatistics, or kriging, and variations thereof (see overviews by McBratney et al. (2003) and Hengl (2009)).

Of these, regression-kriging (Odeh et al., 1995; Hengl et al., 2007), which combines a multiple regression model (a trend) with a spatial correlation model (a variogram) for the residuals, produces unbiased, continuous prediction surfaces. Regression-kriging has been adapted for soil mapping with great success, in part because of the flexibility in defining the trend model as a linear, non-linear, or tree-based relationship between the response and predictors. Furthermore, regression-kriging relies on the incorporation of auxiliary data, providing mechanistic support for the soil property predictions.

The widest application of regression-kriging in soil science has likely been for producing 2-dimensional (2D) maps (Hengl, 2009). However, soil data is often also collected at multiple depths, and geostatistical interpolation techniques can be expanded to represent soil predictions across both vertical and horizontal space (Malone et al., 2009; Veronesi et al., 2012). Global predictions of multiple soil properties obtained from 3-dimensional (3D) regression models were recently showcased by Hengl et al. (2014a). Here, spline functions define the vertical trend (depth) within the regression model, while horizontal trends are defined by covariate grids. These approaches are sufficient for understanding static soil properties across 2- and 3D space; however, modeling dynamic soil properties requires expansion of the geostatistical model to incorporate correlation in data through time (Heuvelink and Webster, 2001; Kyriakidis and Journel, 1999). Addition of temporal and/or spatio-temporal predictors can assist in explaining temporal variation in a response variable, but fitting a variogram model in 2D and time (2D + T) poses additional challenges (summarized by Heuvelink and Webster, 2001). Specifically, time exists in only one dimension and has a directional component, while

spatial properties might be correlated in vertical, horizontal, or 3D directions. The easiest solution for approximating dependence across both space and time is based on anisotropy scalings, which relate horizontal distances to distance in depth and temporal separation.

Modeling 2D + T data has successfully been implemented for predicting soil water from repeated field-wide measurements obtained with time-domain reflectometry, directly with ordinary kriging (Huisman et al., 2003), and with the incorporation of estimated daily evapotranspiration (Jost et al., 2005) and net precipitation (Snepvangers et al., 2003) as covariates. More recently, daily air temperature predictions have been produced from spatio-temporal interpolation models of weather station data at the regional (Hengl et al., 2012) and global (Kilibarda et al., 2014) scales. These models combine spatial covariates (terrain attributes) and spatio-temporal covariates (remotely sensed daily land surface temperature) in the trend model to explain greater than 70% of variation in weather station observations.

Previously, 3D + T data has been analyzed in a spatio-temporal context, wherein interpolations produce predictions in a slice-wise manner (i.e. by depth or by time point) (Bardossy and Lehmann, 1998; Wang et al., 2001; Wilson et al., 2003). To our knowledge geostatistical methods have not yet been expanded to produce predictions from data collected in 3D and time (3D + T). This may be, in part, due to the rarity of quality 3D + T data. With each added dimension, the number of observations required for accurate interpolation increases, as does the need for ancillary (covariate) data if a regression-kriging model is applied. Theoretically, adapting the existing regression-kriging framework for predicting in 3D + T can follow the same mathematical logic as the models that scale up from 2D to 3D or 2D to 2D + T (Heuvelink and Webster, 2001). In that context, existing geostatistical tools for interpolating spatio-temporal data can also assist in modeling 3D + T data (Pebesma, 2012; Pebesma and Gräler, 2013).

In this paper, we demonstrate two approaches for interpolating 3D + T soil water, temperature, and electrical conductivity data (collected from a distributed soil sensor network) at the field-scale: one that is based on using random forests algorithms, and one that is based on spatio-temporal kriging. The kriging model uses different dependence structures (i.e. variogram models) for horizontal, vertical, and temporal components, which are then combined using concepts from 2D + T geostatistics. The models were motivated by the existing geostatistical frameworks and incorporate spatial, temporal, and spatio-temporal covariates. We present the implementation of the models, accuracy assessments, visualization and applications of model output, and future directions for improvement with a long term objective to develop robust 3D + T models for mapping soil data that has been collected with high spatial and temporal resolution.

2. Materials and methods

2.1. The Cook Agronomy Farm data set

The R.J. Cook Agronomy Farm is a Long-Term Agroecosystem Research Site operated by Washington State University, located near Pullman, Washington, USA (46°47′N, 117°5′W; Fig. 1). The farm is 37 ha, stationed in the hilly Palouse region, which receives an annual average of 550 mm of precipitation (Western Regional Climate Center, 2013), primarily as rain and snow in November through May. Soils are deep silt loams formed on loess hills; clay silt loam horizons commonly occur at variable depths (NRCS, 2013). Farming practices at Cook Agronomy Farm are representative of regional dryland annual cropping systems (direct-seeded cereal grains and legume crops).

At 42 locations (stations), five 5TE sensors (Decagon Devices, Inc., Pullman, Washington) were installed at 0.3, 0.6, 0.9, 1.2, and 1.5 m depths. Locations were chosen from an existing non-aligned systematic grid and stratified across landscape units to represent the variability in terrain of Cook Agronomy Farm (Fig. 1). Every hour, the 5TE sensors measure:

(1) volumetric water content, (m^3/m^3),
(2) temperature, (°C),
(3) and bulk electrical conductivity, (dS/m).

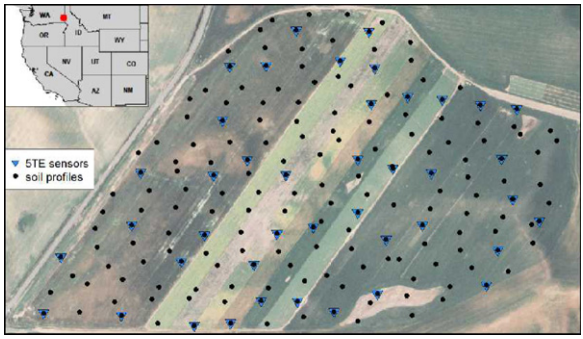

Fig. 1. Cook Agronomy Farm overview map with soil profile sampling points (dots) and instrumented locations (triangles). A total of 210 sensors (42 locations × 5 depths) have been collecting measurements of volumetric water content, temperature, and bulk electrical conductivity since 2009.

Table 1
Cook Agronomy Farm data set spatio-temporal covariates. DEM – Digital elevation model, TWI – SAGA wetness index, NDRE.M—Normalized Difference Red Edge Index (mean), NDRE.sd—Normalized Difference Red Edge Index (s.d.), Bt – Occurrence of Bt horizon, BLD – Bulk density of soil, PHI—Soil pH, Precip_cum—Cumulative precipitation in mm, MaxT_wrcc—Maximum measured temperature, MinT_wrcc—Minimum measured temperature, Crop—Crop type. Response variables include VW—soil volumetric water content in m^3/m^3, C—soil temperature in °C, and EC—soil bulk electrical conductivity in dS/m.

Code	2D	Depth	Time	Spatio-temporal support size		
	(x, y)	(d)	(t)	$\Delta x, y$	Δd(m)	Δt
DEM	✓			10 m	0	>10 yrs
TWI	✓			10 m	0	>10 yrs
NDRE.M	✓			10 m	0	3 yrs
NDRE.sd	✓			10 m	0	3 yrs
Bt	✓	✓		10 m	0.3	>10 yrs
BLD	✓	✓		10 m	0.3	>10 yrs
PHI	✓	✓		10 m	0.3	>10 yrs
Precip_cum			✓	Spatially constant	0	1 d
MaxT_wrcc			✓	Spatially constant	0	1 d
MinT_wrcc			✓	Spatially constant	0	1 d
Crop	✓		✓	10 m	0	1 year
VW, C, EC	✓	✓	✓	42 points on 0.37 km	0.3	1 d

Data are stored on Em50R data loggers (Decagon Devices, Inc., Pullman, Washington), which are buried to allow data collection regardless of farm operations (seeding, spraying, and harvest). The sensor network has been in operation since 2009. For the purpose of this article, hourly sensor data was aggregated to daily averages and all plots and statistical modeling refers to daily values. Sensor data collected for three years at one station and all five depths is illustrated in Fig. 2, and hexbin plots (Sarkar, 2008; Carr, 2014) illustrate the distribution of all observations of all three variables across depth in Fig. 3. Please note that absolute values of sensor readings require further correction for accurate interpretation. Thus, interpretation of the presented readings should focus on the observed relative changes.

In addition to the sensor readings, this data set contains spatial and temporal regression covariates either at 10 m resolution, or as a temporal measurement that is assigned to all possible locations in the area of interest at a given time step (hereafter, spatially constant). Dimensionality of the covariates differs: some covariates are available only in horizontal space (elevation, wetness index, vegetation images), some covariates are available as 3D images (soil properties) and some are available either in time (daily temperatures and rainfall from the nearest meteorological station) or space–time (cropping identity). The covariates used for modeling water content, temperature, and electrical conductivity are described in Table 1. Note that only the response variables (sensor readings) exist in 3D + T, while the predictor variables are a combination of 2D, 3D, 2D + T, and temporal covariates.

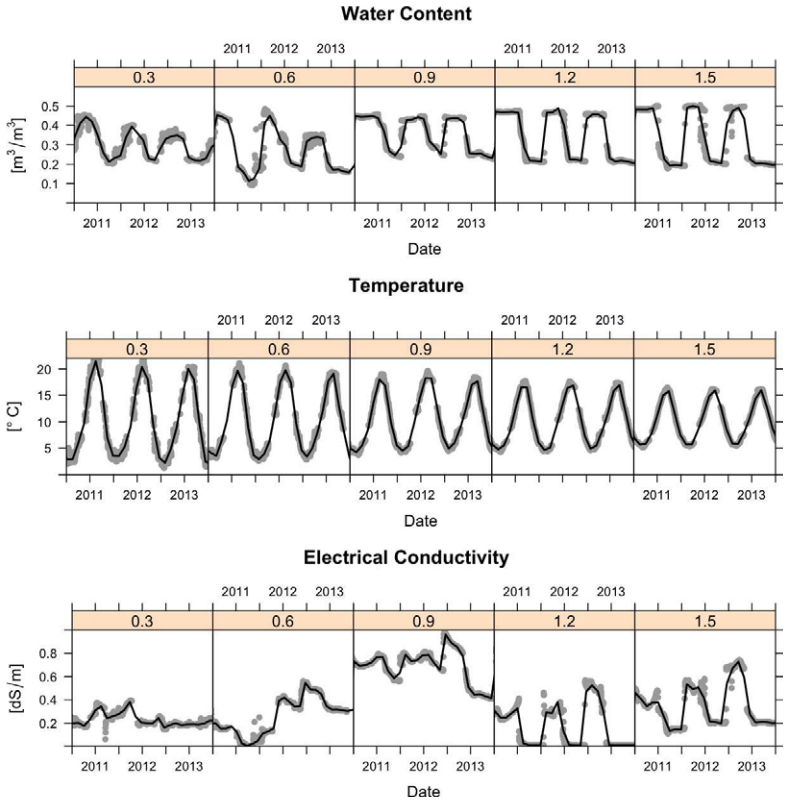

Fig. 2. Sensor values from five depths (0.3, 0.6, 0.9, 1.2, and 1.5 m) at one station at Cook Agronomy Farm from January 2011–January 2014. The black line indicates locally fitted splines (here used for visualization purposes only).

The SAGA wetness index, a modification of the topographic wetness index (Beven and Kirkby, 1979), was derived from the digital elevation model (DEM) using the RSAGA package (Brenning, 2013) for R (R Core Team, 2014). A total of 11 Level 3A RapidEye images satellite images acquired between 2011 and 2013 were used to incorporate vegetation patterns on Cook Agronomy Farm. Images were pre-processed exactly as in Eitel et al. (2011). Following image pre-processing, spectral bands (near infrared — NIR and red-edge — RE) were mathematically converted into the Normalized Difference Red-Edge Index (Barnes et al., 2000):

$$NDRE = \frac{NIR - RE}{NIR + RE}. \tag{1}$$

The RE region of the electromagnetic spectrum has been shown to be superior to red (as used in the Normalized Difference Vegetation Index, or NDVI, Tucker, 1979) for mapping variations in plant chlorophyll and nitrogen content (Carter and Knapp, 2001; Lichtenthaler and Wellburn, 1983; Eitel et al., 2007, 2008, 2009). The images were aggregated to produce one NDRE mean grid and one NDRE standard deviation grid, which were resampled from a 5 to a 10 m grid to align with other covariate grids.

The 3D maps for the occurrence of the Bt horizon, bulk density (g/cm^3), and soil pH were generated using 184 soil profiles distributed across Cook Agronomy Farm (Fig. 1) using the GSIF package for automated soil mapping (Hengl et al., 2014b). Soil profiles were described using the National Soil Survey Center NRCS USDA guidelines for soil profile description (National Soil Survey Center NRCS USDA, 2011). To make the maps, the presence or absence of a Bt horizon was interpolated using a logistic regression-kriging model and the DEM, apparent electrical conductivity grids, soil unit

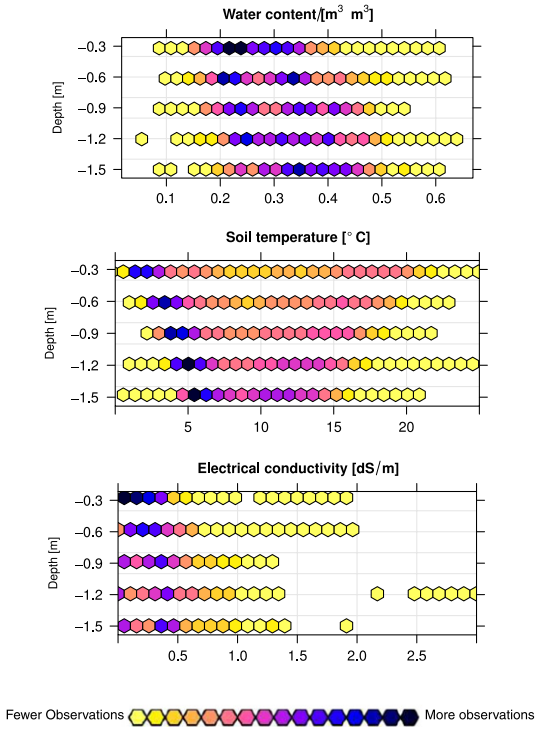

Fig. 3. Distribution of observations (based on all dates) for water content, temperature, and electrical conductivity across soil depth.

description map (NRCS, 2013) and depth as covariates. Bulk density and soil pH were predicted with regression-kriging models with the DEM, wetness index, soil mapping units, apparent electrical conductivity grids, and depth as covariates.

The daily meteorological data (precipitation, minimum and maximum temperature) were obtained from a weather station located 8 km from the farm in Pullman, WA (Western Regional Climate Center, 2013). Daily precipitation was transformed to cumulative precipitation, which reverts to zero after a period of precipitation. Meteorological covariates are only available in the time domain (i.e. they are assumed to be spatially constant).

As the only 2D + T covariate we used the cropping system classification maps, which are available each year from 2006 through 2013. The crop identities include: barley, canola, garbanzo, lentil, pea, or wheat, each with either a spring or winter rotation.

All sensor observations and covariates were assembled into a spatio-temporal regression matrix, using the overlay functionality of the spacetime package in R (Pebesma, 2012). The resulting spatio-temporal regression matrix was very large—even though we only included measurements from 42 stations, the matrix contained close to a quarter million records (about four years of daily measurements at 42 locations and five depths i.e. $4 \times 365 \times 42 \times 5 = 306,600 - \text{missing data} = 219,240$ water content observations, 222,614 temperature observations, and 222,065 conductivity observations).

2.2. Conceptual foundation for 3D + T modeling

We model water content, soil temperature, and electrical conductivity as a spatio-temporal process of a continuous variable Z, where Z varies over space and time. The statistical model of such a process

is typically composed of the sum of a trend and a stochastic residual (Burrough, 1998; Heuvelink et al., 2012; Kilibarda et al., 2014). In this case we begin with a 3D + T model of the form:

$$Z(x, y, d, t) = m(x, y, d, t) + \varepsilon'(x, y, d, t) + \varepsilon''(x, y, d, t) \quad (2)$$

where x, y, d, t are the space–time coordinates, d is depth from the land surface, m is the trend, $\varepsilon'(x, y, d, t)$ is the spatio-temporally correlated stochastic component and $\varepsilon''(x, y, d, t)$ is the uncorrelated noise. We model the trend (m) as a function of spatial (2D or 3D), temporal, or spatio-temporal explanatory variables (covariates, such as in Table 1) available over the entire spatio-temporal domain of interest.

2.3. 3D + T random forests model

The trend model, m in Eq. (2), can be fitted using linear regression or some kind of Generalized Linear Model depending on the distribution of the target variable (Pinheiro and Bates, 2009). Our focus here is on fitting the trend model using random forests algorithms (Breiman, 2001) for two main reasons. First, with random forests algorithms, the target variable does not need to assume specific distributions or adhere to linear relationships (Ahmad et al., 2010; Kuhn and Johnson, 2013). Second, random forests is advantageous for fitting a predictive model for a multivariate data set with high dimensionality. A disadvantage of random forests models, on the other hand, is that model fitting can be computationally intensive, which may become a limitation as data set complexity increases. The second disadvantage is that random forests typically tends to over-fit data sets that are particularly noisy (Statnikov et al., 2008).

We model, for example changes in soil water content, in the form:

```
R> fm = VW ~ DEM + TWI + NDRE.M
   + NDRE.Sd + Bt + BLD + PHI + Precip_cum
   + MaxT_wrcc + MinT_wrcc + cdayt + Crop
```

where `DEM + TWI +... + Crop` are the covariates (see also Table 1) both measured at the same x, y, d, t locations, `VW` is the volumetric water content, and `cdayt` is the transformed cumulative day, computed as:

$$cdayt = \cos\left([t_D - \phi] \cdot \frac{2\pi}{365}\right) \quad (3)$$

where t_D is the linear date (cumulative days), ϕ is the time delay from the coldest day and a trigonometric function is assumed to model seasonal fluctuation of daily temperature. The predictive model, based on the spatio-temporal regression matrix (`regm.VW`) is:

```
R> rfm.VW <- randomForest(fm, data = regm.VW)
```

The random forests prediction model from above can be used to generate predictions for any position in space and time, provided that all covariates are available at that location, but it does not provide inference on the mean trend and spatio-temporal correlation structure as in a regression-kriging model that has interpretable parameters.

In theory, 3D + T residuals of this model could be further analyzed for spatio-temporal autocorrelation and used for kriging. However, in this specific study, examination of residuals obtained from the random forests models for all three variables revealed the absence of any correlation structure over horizontal space (x, y). Since the random forests models explained a high amount of the variability in the data (>90% for all three response variables), all residual variation was considered to be uncorrelated noise ($\varepsilon''(x, y, d, t)$ in Eq. (2)).

2.4. 3D + T kriging model

To explore an alternative approach to spatio-temporal random forests modeling, we developed a 3D + T regression-kriging model based on existing geostatistical methods. In this case, we use

the same model as in Eq. (2), except we do not use any gridded or meteorological covariates to explain the trend model (m). Instead, to model the observed water content, temperature, and electrical conductivity, we only use simple seasonal detrending. Because annual patterns of weather conditions influence these soil properties in a systematic way (see Fig. 2), detrending is necessary before we can apply any kriging. Moreover, because strength of seasonality decreases with depth and shows some delay in time, separate seasonal models were fit at each depth. Daily soil temperatures throughout the year nicely follow a sine curve with intercept c, amplitude a and shift b for the day of the year t_D^* (1–365) given by:

$$s_C(t_D^*) = c + a \cdot \sin\left(\frac{b + t_D^*}{365} \cdot 2 \cdot \pi\right). \tag{4}$$

The other two variables, water content and electrical conductivity, require a somewhat more complex function because values are fairly stable during periods of crop inactivity. These correspond to sustained minima during the dry season (late summer to autumn) and sustained maxima after winter recharge (late winter to spring). The seasonal function for these variables is:

$$s_V(t_D^*) = c + a \cdot \cos\left(breaks(t_D^*) \cdot \pi\right) \tag{5}$$

with:

$$breaks(t_D^*) := \begin{cases} 1 + \dfrac{t_D^* + 365 - b_4}{b_1 + 365 - b_4}, & t_D^* \le b_1 \\ 0, & b_1 < t_D^* \le b_2 \\ \dfrac{t_D^* - b_2}{b_3 - b_2}, & b_2 < t_D^* \le b_3 \\ 1, & b_3 < t_D^* \le b_4 \\ 1 + \dfrac{t_D^* - b_4}{b_1 + 365 - b_4}, & t_D^* \le b_4 \end{cases}$$

where $1 \le b_1 < \cdots < b_4 \le 365$ are four consecutive break points during one year, which resemble the on- and offset of sustained minima and maxima. Hence, the function $\cos(breaks(t_D^*) \cdot \pi)$ connects two plateaus at 1 (from b_1 to b_2) and -1 (from b_3 to b_4) with smooth transitions along a stretched cosine curve. The parameters c and a in Eq. (5) correspond to an intercept and amplitude respectively, V indicates the variable (water content or electrical conductivity).

The models in Eqs. (4) and (5) use purely mathematical functions that can be used to describe the seasonality of this data set. An alternative approach would be to use the daily mean value of sensor readings as the trend. We were interested in using these parameters to learn about how the seasonal trends of the measured soil properties change across depths. In analyses where such interpretation is unnecessary, the simpler approach may be adequate.

Assuming that the remaining residual is normally distributed and has zero mean, only its variance–covariance remains to be specified. To tackle the 3D + T data set, we assume a metric covariance model over horizontal and vertical distances after an isotropy scaling has been applied. The more general set-up would yield a 3D variogram surface in 4-dimensional space (gamma $\sim$ horizontal distance + depth + time) and can thus be reduced to the simpler 2D surface (gamma $\sim$ 3D distance + time).

In order to obtain an objective estimate of the anisotropy ratio between horizontal and vertical distances, we calculated 2D empirical variograms where each day is used as a repetition of the process (i.e. distances are only calculated within each day and not across time). Based on this variogram surface, a pure metric model can be estimated and its anisotropy scaling can then be used to construct pseudo 3D data where the depth value has been rescaled by the anisotropy ratio.

The sum-metric variogram structure for the spatial, temporal, and spatio-temporal (*'joint'*) components, treated as mutually independent, is defined as (Heuvelink et al., 2012):

$$\gamma(h, u) = \gamma_S(h) + \gamma_T(u) + \gamma_{ST}(\sqrt{h^2 + (\alpha \cdot u)^2}), \tag{6}$$

where $\gamma(h, u)$ is the semivariance of variable Z for 3D distances in space (h) and in time (u), γ_S, γ_T are spatial and temporal components respectively, each with a sill, range, and nugget. The joint space–time component, γ_{ST}, also includes a parameter for the conversion of temporal separation (u) to spatial distance (h), denoted α. Variogram parameters are estimated from the observations and then fit with a metric semivariance function, used in kriging to predict Z at unobserved space–time points. For example, kriging predictions are produced from water content observations as:

```
R> svgmVW3DT <- variogramST(resid~1, VW.st)
R> fvgmVW3DT <- fit.StVariogram(svgmVW3DT,
                vgmST("sumMetric",
                space=vgm(sill, model, range, nugget),
                time=vgm(sill, model, range, nugget),
                joint=vgm(sill, model, range, nugget),
                stAni=ratio)
R> predVW.resid <- krigeST(resid~1, VW.st, Pred.st,
    fvgmVW3DT)
```

where `resid` $\sim$ `1` defines the sample variogram for the water content residuals after detrending, which are stored in the spacetime object, `VW.st`. The sample variogram `svgmVW3DT` is used to fit a 3D + T sum-metric model, `vgmST`, wherein the variogram for each component is defined with user inputs for initial model parameters (partial sill, model type, range, nugget, and anisotropy ratio α), based on inspection of the sample variogram. The fitted variogram `fvgmVW3DT` is then used to make predictions at unobserved locations, stored in a spacetime object `Pred.st`. The residual predictions `predVW.resid` are added to the seasonal trend to obtain predicted water content at any space–time point. The formulas of kriging in the spatio-temporal domain do not differ fundamentally in a mathematical or statistical sense from those of spatial kriging (Heuvelink et al., 2012).

2.5. Cross-validation

We run cross-validation for the two spatio-temporal prediction approaches (3D + T random forests model and 3D + T kriging after detrending) separately. Moreover, we run two versions of cross-validation for the random forests model:

(1) 3D + T random forests prediction (RF):
- `RF-loc`: strict cross-validation, using leave-onestation-out iterations of model fitting and validation, and
- `RF-rnd`: simple cross-validation, by randomly subsetting space–time points, and using 5-fold sets of model fitting and validation,

(2) 3D + T regression kriging (kriging):
- `kriging-loc`: leave-one-station-out using the fitted variogram model, then validation.

Specific details of the cross-validation methods appear below, but first, it is important to emphasize that fundamental differences between the two modeling approaches do not allow the predictions for cross-validation to be obtained in the exact same way. In particular, the RF model is informed directly by the observations rather than a parametric model. So, if observations are removed, a new model is developed, driven by the included observations. Conversely, the kriging model quantifies the variability in the data and how it changes with distance. The inherent replication of point pairs within each lag distance buffers the resulting variogram model from the removal of an observation. These differences materialize in the cross-validation steps as follows: once a RF model has been fit with all data, the same model cannot be used on a subset of the observations (a training set) to make predictions, so a leave-one-out approach for n observations requires n training models, each unique, for n predictions. This differs from the kriging cross-validation in that the same theoretical variogram model – developed from all observations – is applied to each of the n training sets to make n predictions because automatically re-fitting the variogram for each training set would be cumbersome and is unlikely to produce considerably different variogram models.

For strict cross-validation of the RF model (`RF-loc`), 42 models were iteratively trained, each using the data of 41 stations (a 'station' includes all five depths and all time points, a 5-variate time series). Each model was then applied to predict on the respective withheld 5-variate time series. The results of the strict cross-validation indicate predictive performance at new, unsampled locations. For simple cross-validation of the RF model (`RF-rnd`), 10% of observations were randomly subset from the full set of space–time points, and subject to 5-fold cross-validation. This less-rigorous approach provides information on predictive performance when at least some observations exist at all locations, and is useful for understanding the accuracy of interpolating missing data at an existing sample location.

To validate the kriging model (`kriging-loc`), we assumed the variogram model to be known and used the fitted model for all predictions in the cross-validation. Each of the 42 stations (including all five depths) was removed from the data set in turn. This withheld 5-variate time series was then predicted using the remaining data. For computational reasons, the prediction was limited to the closest 500 spatio-temporal neighbors (using anisotropy scalings for the 3D + T distances) from a temporal window of ± 10 days for prediction.

For each variable and each model approach, we calculated standard model performance measures: root mean square error (RMSE), mean absolute error (MAE), mean error (ME), and coefficient of determination (R^2) for observations and predictions obtained in cross-validation. As a baseline comparison, spatially constant predictions were made from the seasonal models alone (Eqs. (4) and (5)) for each variable and each depth. The same four cross-validation statistics were computed for these predictions. Although we do not apply exactly the same cross-validation procedures to the two methodological approaches, we assume that the cross-validation results will reveal useful information about each model's performance.

2.6. Software implementation

All analysis was conducted in R (R Core Team, 2014) unless otherwise noted in the text. Preparation of sensor network data and covariate data was assisted by the following packages: aqp (Beaudette and Roudier, 2013), gdata (Warnes et al., 2014), GSIF (Hengl et al., 2014b), gstat (Pebesma and Gräler, 2013), plyr (Wickham, 2014), raster (Hijmans et al., 2014), rgdal (Bivand et al., 2014), RSAGA (Brenning, 2013), and spacetime (Pebesma, 2012). The randomForest package (Liaw and Wiener, 2002) was used for the RF modeling. The kriging approach was mainly based on the gstat package (Pebesma, 2004) in combination with the spacetime package (Pebesma, 2012). The lattice (Sarkar, 2014) and plotKML (Hengl et al., 2015) packages were used for data visualization.

A subset of this data set (for the period Jan. 1, 2011–Dec. 31, 2012) and example code for the main processing steps has been added to the GSIF package (Hengl et al., 2014b) for demonstration and can be obtained by calling `?cookfarm` after loading the package.

3. Results

3.1. 3D + T random forests model

The importance plots for predicting water content, temperature, and electrical conductivity with the RF models are shown in Fig. 4. The covariates with higher importance will influence the prediction more if randomly permuted within the model. The mean decrease in accuracy metric (%IncMSE) indicated that the cumulative date was the most important predictor for all three variables, followed by crop identity for water content and electrical conductivity, and soil pH and crop identity for soil temperature. The decrease in mean squared error (IncNodePurity) also indicated that cumulative day was important for modeling water content, and all three weather covariates were important for soil temperature. Soil properties (pH, Bt presence, and bulk density) were most important for modeling bulk electrical conductivity by the same metric.

The randomForest package reported that the RF models, based only on covariate data, explain 93% of the variance in water content, 98% in temperature, and 93% in conductivity observations. As described in Section 2.3, we did not fit space–time variograms to the residuals because residual

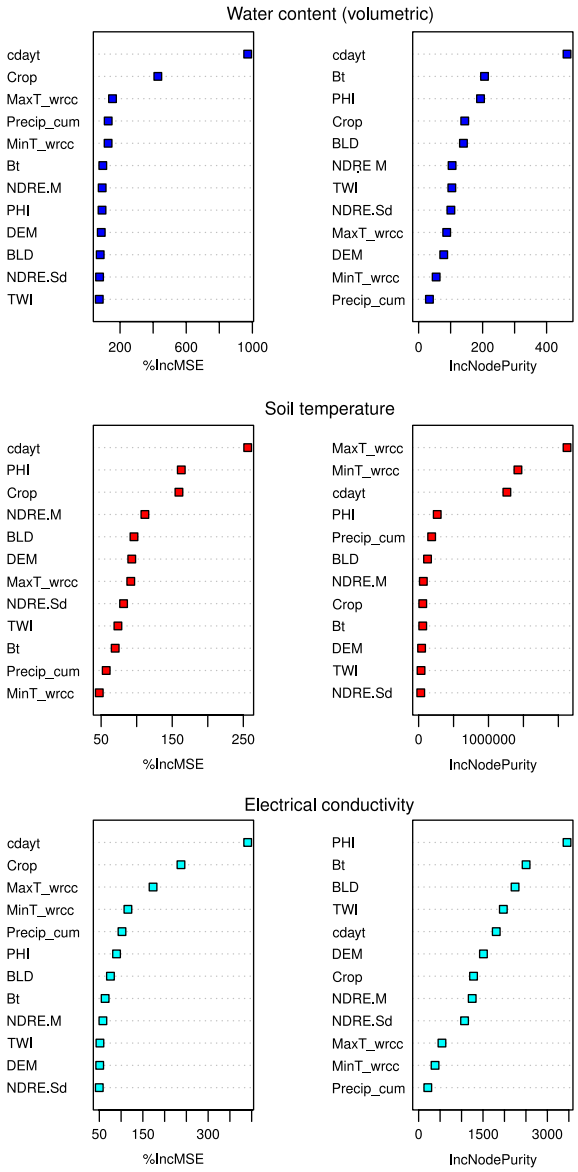

Fig. 4. Importance plots (covariates sorted by importance) derived using the randomForest package (Liaw and Wiener, 2002). %IncMSE is the mean decrease in accuracy; IncNodePurity is the decrease in mean squared error.

variation did not display any strong spatio-temporal correlation. Further processing would produce pure-nugget variograms (of uncorrelated noise), which do not impart any additional explanatory power.

Prediction surfaces for water content (for the first day of five months in 2012) produced directly from the RF model are shown in Fig. 5. This period of time represents the growing season, when large changes in water content occur as crops develop and rapidly extract soil water. Prediction maps for water content, soil temperature, and electrical conductivity for the whole period of observation (space–time prediction stacks) can be obtained by contacting the authors.

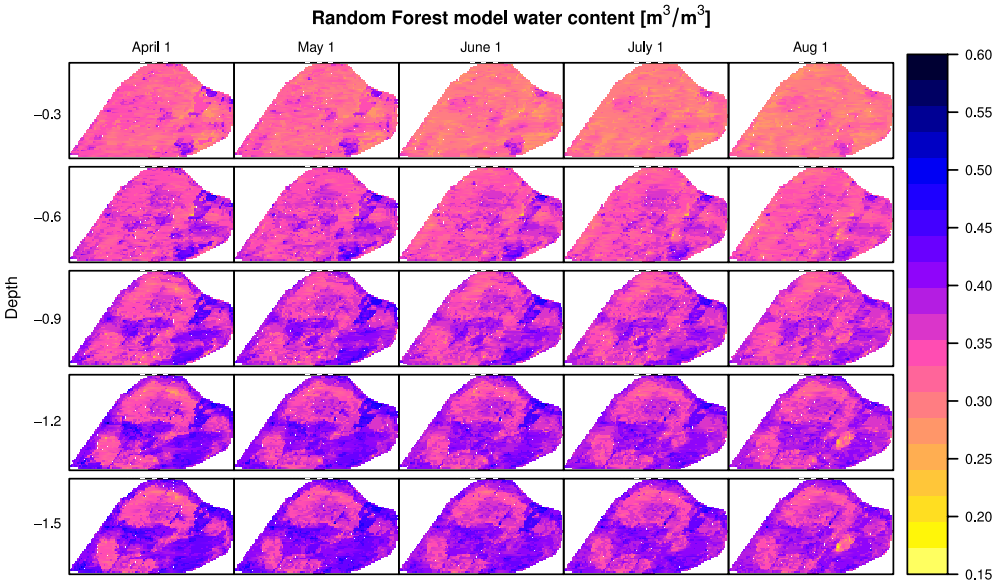

Fig. 5. Spatio-temporal predictions of soil water content at Cook Agronomy Farm for the growing season in 2012 using the random forests (RF) model. Note that relative changes in water content are accurate, but absolute sensor readings require correction.

Table 2
Parameters of the seasonality functions (Eqs. (4) and (5)) for water content (VW), soil temperature (C) and electrical conductivity (EC) at each depth. The parameters represent the intercept (c), amplitude (a) and shift (b) in seasonal effects at each depth.

Var.	Depth (m)	c	a	b	b_1	b_2	b_3	b_4
	0.3	0.26	0.06		45	128	223	257
	0.6	0.29	0.06		66	153	228	280
VW	0.9	0.31	0.06		71	152	258	282
	1.2	0.32	0.05		71	169	266	289
	1.5	0.35	0.03		117	186	205	245
	0.3	9.7	−8.9	63				
	0.6	9.6	−7.4	55				
C	0.9	9.5	−6.3	46				
	1.2	9.5	−5.4	38				
	1.5	9.4	−4.6	30				
	0.3	0.20	0.05		55	133	219	241
	0.6	0.31	0.08		93	160	225	248
EC	0.9	0.37	0.11		77	148	254	290
	1.2	0.41	0.09		84	184	225	274
	1.5	0.44	0.07		62	110	281	332

3.2. 3D + T kriging model

Fitted parameters for the seasonality functions (Eqs. (4) and (5)) are listed in Table 2. The seasonal effects varied by depth: as depth increases, the change in soil properties was delayed, and the amplitude of the change, on average, increased for water content and decreased for soil temperature. For electrical conductivity, the amplitude was highest at 0.9 m. High temperatures corresponded with low water content and associated conductivity.

Table 3 lists the h/v ratios for horizontal–vertical distance scaling, as well as the variogram parameters for each variable. We set the h/v ratios so that 1 m in depth horizontally corresponded to 21 m for water content, 516 m for soil temperature, and 53 m for electrical conductivity.

Table 3
Variogram parameters for each variable. VW is water content, C is soil temperature, EC is electrical conductivity, $\frac{h}{v}$ is the anisotropy ratio for horizontal–vertical distances (m); st-vgm is the sum-metric component of the spatio-temporal variogram; $\frac{sp}{t}$ is the anisotropy ratio between spatial and temporal (m/days) distances (α); sill, range, and nugget are variogram parameters; and the semivariance function of each model is either Exponential (Exp) or Spherical (Sph). Sill and nugget units are the same as the measured variable.

Var.	$\frac{h}{v}$	st-vgm	$\alpha = \frac{sp}{t}$	Sill	Model	Range	Nugget
VW	21	Joint	0.20	0.005	Exp	32 m	0
C	516	Space		0.26	Exp	97 m	0.39
		Time		4.69	Exp	147 days	0
		Joint	0.48	0.27	Sph	20 m	0
EC	53	Joint	0.06	0.06	Exp	21 m	0

Table 4
Global cross-validation statistics including the spatially constant predictions based on the fitted seasonality functions, the kriging model (kriging-loc), and two sets of statistics for the RF model (RF-loc and RF-rnd). VW is water content, C is soil temperature, and EC is electrical conductivity, RMSE is root mean squared error, MAE is mean absolute error, ME is mean error, and R^2 is coefficient of determination. The R^2 for EC was calculated on the log scale, due to a skewed distribution.

Var.	Approach	RMSE	MAE	ME	R^2
VW	Season	0.08	0.06	0.00	0.31
	Kriging-loc	0.07	0.06	0.00	0.37
	RF-loc	0.07	0.06	0.00	0.34
	RF-rnd	0.03	0.02	0.00	0.86
C	Season	1.37	1.03	0.00	0.93
	Kriging-loc	0.98	0.70	0.01	0.96
	RF-loc	1.30	0.96	0.06	0.93
	RF-rnd	0.94	0.67	0.00	0.97
EC	Season	0.27	0.20	0.00	0.13
	Kriging-loc	0.27	0.19	−0.01	0.18
	RF-loc	0.31	0.21	0.00	0.05
	RF-rnd	0.10	0.05	0.00	0.88

The water content and electrical conductivity variogram models only contained the metric component (γ_{st}), each with four parameters (sill, range, nugget, and the anisotropy parameter α), while soil temperature used a sum-metric model with spatial, temporal, and joint components as in Eq. (6). The lack of pure spatial and temporal components in water content and conductivity indicated that these correlation structures appeared to be sufficiently modeled through a metric model. In all three cases, the correlation in time was stronger over larger separation distances, indicated by anisotropy ratios (sp/t) that were less than one. For example, correlation at 1 m was equal to correlation at 5 days for water content, 2 days for temperature, and 17 days for electrical conductivity. This translates to the inclusion of more temporal neighbors than spatial neighbors when making kriging predictions. Sample and 3D + T fitted variograms are depicted in Fig. 6, along with isolated 3D spatial and temporal components. Please note that the optimization of the spatial and temporal components of each 3D + T variogram is done based on the full spatio-temporal model. Hence, the fit represents the entire variogram surface. As a result, the individual space and time components may not intersect the sample data and appear as a poor fit compared with the overall surface. Prediction surfaces for water content during the 2012 growing season were also created from the 3D + T kriging model, shown in Fig. 7.

3.3. Model accuracy

For all three variables, Fig. 8 shows hexbin plots of observed versus predicted values with the full RF model, strict cross-validation of the RF model (`RF-loc`), and cross-validation of the kriging model (`kriging-loc`). Table 4 lists the global cross-validation statistics for the two models in addition to the spatially constant seasonal models used for detrending.

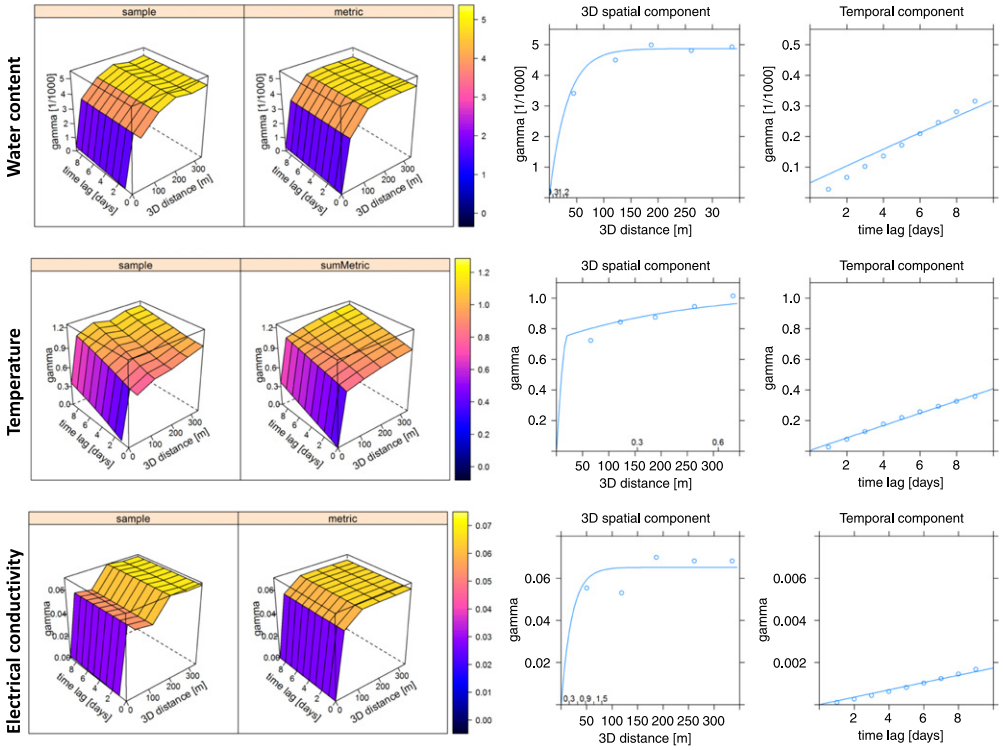

Fig. 6. Spatio-temporal sample variogram, metric variogram, and isolated 3D spatial and temporal components for water content, temperature, and electrical conductivity. The double axis on the 3D variogram illustrates the relationship between vertical and horizontal depths.

The goodness of fit between observations and predictions using the full RF model was >90% for all three variables. However, under strict cross-validation (`RF-loc`), the predictive power of the RF model decreased, especially for water content (34%) and conductivity (5%). The R^2 values for soil temperature remained high in cross-validation. The less rigorous cross-validation procedure (`RF-rnd`) demonstrated stronger predictive power and lower error for all three variables, with 86%, 97%, and 88% of variability explained for water content, temperature, and conductivity, respectively.

The seasonal models alone predicted all variables well, with the kriging models only capturing a bit more variability. As with the RF model, the kriging model was most successful at predicting soil temperature. The R^2 of the kriging model for the highly variable electrical conductivity was low at 18%. Both the RF and kriging models had difficulty predicting the infrequently high conductivity values.

4. Discussion

4.1. Model performance

In this paper we examined two approaches to producing continuous predictions from 3D + T point observations of three dynamic soil variables, measured daily at the field scale, by a 3D sensor network, for multiple years, and on complex terrain that hosts rotating cropping systems. First, we assembled a highly dimensional spatio-temporal regression matrix, and when fit with random forests algorithm, covariates successfully explained the variability in observations. All of the measured variables displayed seasonal patterns (Fig. 2), so temporal covariates explained much of the variability in the observations. Cumulative day was an important covariate for all three soil variables, as was crop identity. At Cook Agronomy Farm, the field is divided into multiple strips, which are the basis for

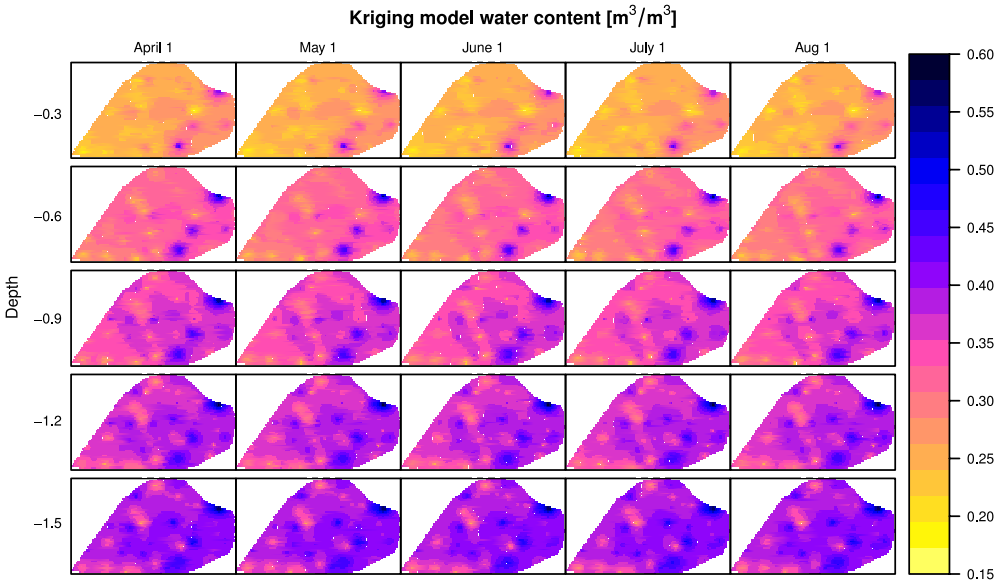

Fig. 7. Spatio-temporal predictions of soil water content at Cook Agronomy Farm for the growing season in 2012 using the kriging model. Note that relative changes in water content are accurate, but absolute sensor readings require correction.

crop rotations. Different cropping systems have different patterns of water use, biomass production, rooting depth, and influences on the soil surface e.g. shading, residue production, and interception of precipitation (Al-Mulla et al., 2009; Qiu et al., 2011). These characteristics are likely responsible for the differences in dynamic soil properties between the strips, and from year to year—thus, they can explain both spatial and temporal variability.

We expected precipitation to be an important predictor of soil water content; however, weather covariates, were only deemed important according to the decrease in mean squared error metric. While precipitation is the only source of soil water in this dryland agricultural system, evapotranspiration also plays an important role in controlling soil water content, along with terrain and soil properties (Cantón et al., 2004; Hébrard et al., 2006). Perhaps inclusion of estimated evapotranspiration as a covariate, as in Jost et al. (2005), would complement our covariate set in predicting soil water. The confounding and interacting effects of weather, terrain, and soil properties that influence soil water content were likely not recognized by the random forests model, as covariates are assessed individually.

Similarly, we expected air temperatures to be important in explaining variability in soil temperature. Daily minimum and daily maximum temperatures indeed had high importance, according to one of the rankings; however, air temperatures may not be representative of heat fluxes at the soil surface, due to crop influences mentioned above.

Soil bulk electrical conductivity is correlated with soil moisture, organic matter, soil salinity, and soil texture (Friedman, 2005). Accordingly, we expected covariates that are important in predicting soil water content to also predict conductivity, in addition to soil properties related to soil texture (bulk density and Bt horizon presence). These covariates were ranked with high importance in the RF model.

According to the strict cross-validation, the predictive success of the RF model decreased as the variability of the target variable increased. This suggests that the model was sensitive to the micro-scale variation in the data, rather than capturing the general spatio-temporal trend of the data. While the random forests algorithm generally tries to resist overfitting (Breiman, 2001), instances of overfitting have been documented (Statnikov et al., 2008). Conversely, under the simple cross-validation, the predictive power was strong. Clearly, the inclusion of at least some space–time points at a location were crucial for making predictions at each location using the random forests algorithm. The 42

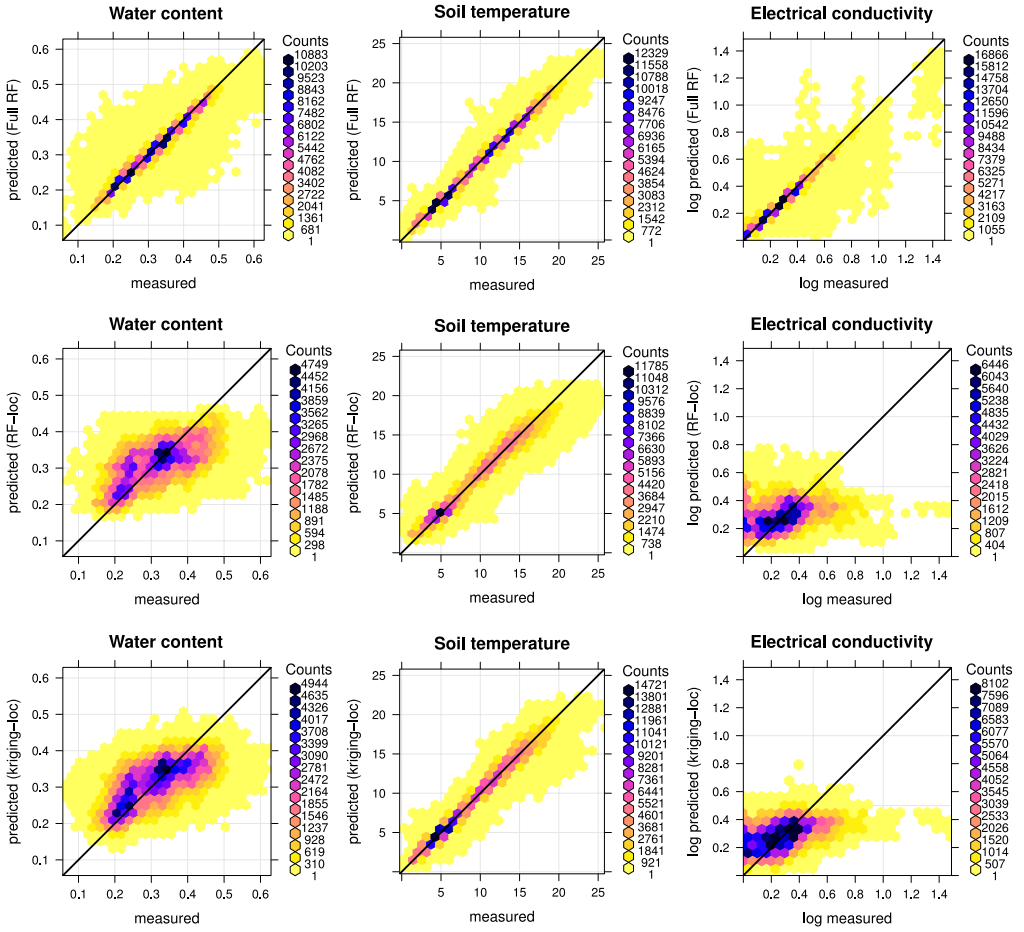

Fig. 8. Hexbin plots for observed and predicted values for the full RF model showing goodness of fit (top), strict cross-validation of the RF model (center), and of the kriging model (bottom).

instrumented stations are intentionally stratified across the terrain and soil feature space, and no two locations are the same. We suspect that the stations are sparse enough across the complex landscape of Cook Agronomy Farm that predicting new, unique locations occurs with higher error. It would be interesting to see if additional sensor stations would improve predictive power, and/or if model performance was improved in a more uniform study area. Identifying the optimal sample size for high predictive accuracy in a complex study area is a question that still needs to be addressed. Through this analysis, we have also realized that there are multiple ways of dividing the data set for cross-validation of these models—each providing different information about dependence across space, time, or both. Here, we applied validation methods familiar to spatial analysis, but we suspect that these methods are limited for handling complex 3D + T data. In the future, we hope to explore cross-validation methods that better assess predictive power through space, time, and their interaction.

We also expanded the kriging framework to accommodate the 3D + T data. These models first required that we de-trend the data with depth-dependent seasonality functions. The parameters of the seasonality functions that we fit demonstrate that all three variables experienced a temporal delay as soil depth increases. These results reflect the infiltration process during soil water recharge, and later in the season, water draw-down by crop roots at increasing depths. Similarly, seasonal soil temperature changes experienced a lag as soil insulation increases with depth. Soil electrical

conductivity followed a similar seasonal pattern as water content, but with the largest minima and maxima at depths where clay horizons occur. These depth-dependent temporal patterns explained most of the variability in all three variables, akin to in the RF model.

3D + T variograms parameters indicated that spatial heterogeneity was high, while temporal correlation was stronger over longer separation distances (spatial range parameters were shorter than temporal range parameters). Soil temperature was correlated over shorter time periods, but was more constant over vertical space (as indicated by the h/v ratio). Water content was correlated over longer time periods, but over shorter vertical space. This translates to temperature changes in the soil occurring at a faster rate than changes in water content, but water content was more variable across 3D space. Electrical conductivity was the least dynamic of all, because it is partially dependent on static soil properties e.g. clay content (Corwin and Lesch, 2005). The presented 3D + T kriging approach only uses *day of the year* as a covariate. Including some of the many covariates used in the random forests approach to define the regression trend might also improve the performance of regression-kriging for this data set.

For both modeling approaches, temporal patterns explained most of the variability in the observations, while spatial components were secondary. Spatial heterogeneity is high at Cook Agronomy Farm, with hilly terrain, variable soil horizonation, and multiple crop rotations. Our ability to predict this spatial complexity with high precision was limited with only 42 stations. Thus, the high temporal sampling density within this data set seems to be more important to our modeling efforts.

4.2. Interpretation of model predictions

All three soil variables show interesting patterns through the soil profile, across horizontal space and time. The range of water content was higher in the shallower soils, which are exposed to extremely wet and extremely dry conditions. Additionally, on average, soil water was retained in deeper soil, relative to shallower depths. This was similar to soil temperature, where deeper soils are insulated from extreme air temperatures, in both cold and warm seasons. Electrical conductivity was variable through the profile, with some higher values occurring in shallow soils, possibly due to fertilizer application (De Neve et al., 2000; Eigenberg et al., 2002). High values also occurred at the 1.2 m depth, which may be an indication of accumulated carbonates or other materials. It is important to note that the electrical conductivity readings represent the conductivity of the bulk soil (including solid and liquid states). These values may be converted to conductivity of the soil solution, which would be of interest for assessing soil salinity specifically related to land and vegetation management. Soil solution salinity is calculated from the bulk conductivity using the dielectric permittivity, soil temperature, and water content measured by the sensors (Decagon Devices, Inc., 2014; Hilhorst, 2000). Depending on the research question, either bulk or soil solution conductivity could be interpolated with the methods described here. It is possible that soil solution electrical conductivity may display less variability and be easier to predict in space and time.

The prediction surfaces produced from the RF model showed more fine-scale variability, compared to the kriging predictions. This was a result of the inclusion of crop, terrain, and soil covariates in the predictive model. Within the prediction surfaces, spatial patterns of covariate features are visible; particularly for the covariates that ranked with high importance in the models (e.g. cropping strips and Bt presence in Fig. 5). The only spatial information provided by the kriging model was the spatio-temporal correlation around each sample point—causing the speckled appearance of the map. Nevertheless, in both cases, we can see that deeper soil retained water when shallow soil was dry late in the growing season. The seasonality and draw-down of soil water was more apparent in the RF model predictions, than in the kriging predictions, particularly in deeper soil. Certainly, the kriging predictions provide a more spatio-temporally smoothed representation of the response variables, compared with the RF model.

4.3. Final conclusions and future directions

We have demonstrated two approaches for interpolating dynamic 3D + T soil data. We observed that both models were highly successful in predicting soil temperature and that the predictive power

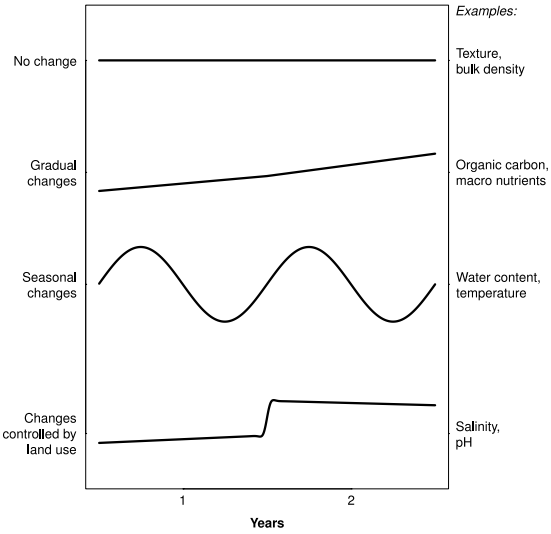

Fig. 9. Types of soil variables in terms of temporal stability or change.

decreased as property variability increased—particularly when data from a station was entirely absent. The temporal components in each model contributed most to explaining all three soil variables across depth, emphasizing the importance of the seasonal changes in this data set. Modeling changes in soil properties through time is, perhaps most interesting for variables where such change can be observed at temporal scales of a few days to a few years (Fig. 9). Certainly, dynamic properties that irregularly or erratically change will require innovative modeling approaches for explaining such temporal behavior.

It should be noted that these methods are experimental and invite modification and improvement. The results presented here are specific to the Cook Agronomy Farm data set; the work serves as a case study for exploring 3D + T interpolation approaches, and a basis upon which we can build. We observed that temporal autocorrelation and time (day of the year) largely contribute to the portion of variation that we can explain. A future direction could include combining a random forests model with residual kriging. Given such a large data set, we can experiment with thinning the regression matrix to remove spatial and/or temporal correlation from the random forests model, and integrate those predictions with spatio-temporal kriging.

Development of 3D + T models to create continuous predictions from point data will allow dynamic soil properties to be incorporated into spatially-explicit process and biophysical models. These spatio-temporal predictions of soil water content, temperature, and electrical conductivity, as well as the 3D maps of basic soil properties such as pH and bulk density, can inform precision agricultural practices. All these soil variables can assist in understanding site specific characteristics of Cook Agronomy Farm, such as crop performance, or risk of fertilizer loss to the groundwater or the atmosphere. The fitted initial spatio-temporal models can also be used to optimize soil monitoring networks (Heuvelink et al., 2012) and/or recommend sampling and modeling strategies for properties that might co-vary through space and time.

3D + T predictions of key soil properties also assist in visualizing dynamic below-ground properties, which, unlike above-ground properties, cannot be observed with photography or remote sensing. Time-lapse animations of 3D soil properties provide information that is difficult to access through static, piece-wise, representations. As a supplement (see Appendix A) to this paper, we have included KML (Keyhole Markup Language) files to illustrate how 3D + T predictions can be visualized in an interactive browser such as Google Earth.

Modeling data in 3D + T is not limited to soil or agricultural applications. Any point data collected in 3D and through time could benefit from 3D + T interpolations. In short, 3D + T models allow us

to visualize and access knowledge about dynamic properties that are difficult to directly observe. As technologies for monitoring ecosystem properties improve and high resolution spatial data collection becomes cheaper and easier, the majority of soil maps could become 3D + T.

Acknowledgments

The authors wish to thank David Huggins, Dave Uberuaga, Erin Brooks, Colin Campbell, Doug Cobos, Maninder Chahal, and Matteo Poggio for developing and maintaining the sensor network and collecting covariate data at Cook Agronomy Farm. This project was funded by the Site-Specific Climate Friendly-Farming project, provided by USDA-NIFA award #2011-67003-30341. This work was possible thanks to the software packages for organizing, visualizing, and analyzing soil data (Beaudette and Roudier, 2013) and spatio-temporal data (Pebesma and Bivand, 2013; Pebesma, 2012; Pebesma and Gräler, 2013). The authors are grateful to the R open source software community (R Core Team, 2014) for providing and maintaining numerous spatial and spatio-temporal analysis packages used in this work.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.spasta.2015.04.001.

References

Ahmad, S., Kalra, A., Stephen, H., 2010. Estimating soil moisture using remote sensing data: A machine learning approach. Adv. Water Resour. 33 (1), 69–80.
Al-Mulla, Y., Wu, J., Singh, P., Flury, M., Schillinger, W., Huggins, D., Stöckle, C., 2009. Soil water and temperature in chemical versus reduced-tillage fallow in a Mediterranean climate. Appl. Eng. Agric. 25, 45–54.
Bardossy, A., Lehmann, W., 1998. Spatial distribution of soil moisture in a small catchment. Part 1: geostatistical analysis. J. Hydrol. 206, 1–15.
Barnes, E., Clarke, T., Richards, S., Colaizzi, P., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R., Li, H., Moran, M., 2000. Proceedings of the International Conference of Precision Agriculture, 5th, Bloomington, MN, 16–19 July.
Beaudette, D., Roudier, P., 2013. aqp: Algorithms for Quantitative Pedology. R package version 1.4. URL: http://CRAN.R-project.org/package=aqp.
Beven, K., Kirkby, M., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24, 43–69.
Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., Rouault, E., 2014. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.9–1. URL: http://CRAN.R-project.org/package=rgdal.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Brenning, A., 2013. RSAGA: SAGA Geoprocessing and Terrain Analysis in R. R package version 0.93-6. URL: http://CRAN.R-project.org/package=RSAGA.
Burrough, P. (Ed.), 1998. Principles of Geographical Information Systems, second ed. Oxford University Press, Oxford.
Cantón, Y., Solé-Benet, A., Domingo, F., 2004. Temporal and spatial patterns of soil moisture in semiarid badlands of SE Spain. J. Hydrol. 285, 199–214.
Carr, D., 2014. hexbin: Hexagonal Binning Routines. R package version 1.26-2. URL: http://CRAN.R-project.org/package=hexbin.
Carter, G., Knapp, A., 2001. Leaf optical properties in higher plants: linking spectral charactersitics to stress and chlorophyll concentration. Am. J. Bot. 88, 677–684.
Corwin, D., Lesch, S., 2005. Apparent soil electrical conductivity measurements in agriculture. Comput. Electron. Agric. 46, 11–43.
Decagon Devices, Inc., 2014. 5TE Water Content, EC and Temperature Sensor. Pullman, WA. URL: http://manuals.decagon.com/Manuals/13509_5TE_Web.pdf.
De Neve, S., Van de Steene, J., Hartmann, R., Hofman, G., 2000. Using time domain reflectometry for monitoring mineralization of nitrogen from soil organic matter. Eur. J. Soil Sci. 51, 295–304.
Eigenberg, R., Doran, J., Nienaber, J., Ferguson, R., Woodbury, B., 2002. Electrical conductivity monitoring of soil condition and available N with animal manure and a cover crop. Agric. Ecosyst. Environ. 88, 183–193.
Eitel, J., Long, D., Gessler, P., Hunt, E., 2008. Combined spectral index to improve ground-based estimates of nitrogen status in dryland wheat. Agron. J. 100, 1694–1702.
Eitel, J., Long, D., Gessler, P., Hunt, E., Brown, D., 2009. Sensitivity of ground-based remote sensing estimates of wheat chlorophyll content to variation in soil reflectance. Soil Sci. Soc. Am. J. 73, 1715–1723.
Eitel, J., Long, D., Gessler, P., Smith, A., 2007. Using in-situ measurements to evaluate the new rapideye satellite series for prediction of wheat nitrogen status. Int. J. Remote Sens. 28, 4183–4190.
Eitel, J., Vierling, L., Litvak, M., Long, D., Schulthess, U., Ager, A., Krofcheck, D., Stoscheck, L., 2011. Broadband, red-edge information from satellites improves early stress detection in a new mexico conifer woodland. Remote Sens. Environ. 115, 3640–3646.

Friedman, S.P., 2005. Soil properties influencing apparent electrical conductivity: a review. Comput. Electron. Agric. 46 (1), 45–70.
Hébrard, O., Voltz, M., Andrieux, P., Moussa, R., 2006. Spatio-temporal distribution of soil surface moisture in a heterogeneously farmed Mediterranean catchment. J. Hydrol. 329, 110–121.
Hengl, T., 2009. A Practical Guide to Geostatistical Mapping. Lulu.com, Amsterdam, Netherlands.
Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014a. SoilGrids1km—global soil information based on automated mapping. PLOS ONE 9 (8).
Hengl, T., Heuvelink, G.B.M., Perčec Tadić, M., Pebesma, E., 2012. Spatio-Temporal Prediction of Daily Temperatures Using Time-Series of MODIS LST Images—Springer. Theoretical and Applied Climatology.
Hengl, T., Heuvelink, G., Rossiter, D.G., 2007. About regression-kriging: from equations to case studies. Comput. Geosci. 33 (10), 1301–1315.
Hengl, T., Kempen, B., Heuvelink, G., Malone, B., Hannes, R., 2014b. GSIF: Global Soil Information Facilities. R package version 0.4-1. URL: http://CRAN.R-project.org/package=GSIF.
Hengl, T., Roudier, P., Beaudette, D., Pebesma, E., 2015. plotKML: Scientific visualization of spatio-temporal data. J. Stat. Softw. 63 (5), 1–25. URL: http://www.jstatsoft.org/v63/i05/.
Heuvelink, G.B.M., Griffith, D.A., Hengl, T., Melles, S.J., 2012. Sampling design optimization for space–time kriging. In: Mateau, J., Müller, W.G. (Eds.), Spatio-Temporal Design. John Wiley & Sons, Ltd., pp. 207–230.
Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: past, present, and future. Geoderma 100, 269–301.
Hijmans, R., van Etten, J., Mattiuzzi, M., Sumner, M., Greenberg, J., Lamigueiro, O., Bevan, A., Racine, E., Shortridge, A., 2014. raster: Geographic data analysis and modeling. R package version 2.3-12. URL: http://CRAN.R-project.org/package=raster.
Hilhorst, M., 2000. A pore water conductivity sensor. Soil Sci. Soc. Am. J. 64, 1922–1925.
Huisman, J.A., Snepvangers, J.J.J.C., Bouten, W., Heuvelink, G.B.M., 2003. Monitoring temporal development of spatial soil water content variation: comparison of ground penetrating radar and time domain reflectometry. Vadose Zone J. 2, 519–529.
Jost, G., Heuvelink, G., Papritz, A., 2005. Analysing the space–time distribution of soil water storage of a forest ecosystem using spatio-temporal kriging. Geoderma 128 (3), 258–273.
Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E., Perčec Tadić, M., Bajat, B., 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J. Geophys. Res.: Atmos. 119 (5), 2294–2313.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer.
Kyriakidis, P.C., Journel, A.G., 1999. Geostatistical space—time models: A review. Math. Geol. 31 (6), 651–684.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2 (3), 18–22. URL: http://CRAN.R-project.org/doc/Rnews/.
Lichtenthaler, H., Wellburn, A., 1983. Determination of total carotenoids and chlorophylls a and b of leaf extracts in different solvents. Biochem. Soc. Trans. 11, 591–592.
Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152.
McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89 (1–2), 67–94.
Mitas, L., Mitasova, H., 1999. Spatial interpolation. In: Longley, P., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (Eds.), Geographical Information Systems: Principles, Techniques, Management and Applications. Vol. 1. Wiley, pp. 481–492.
National Soil Survey Center NRCS USDA, 2011. Field Book for Describing and Sampling Soils, third ed. US Department of Agriculture, Lincoln, Nebraska.
Natural Resource Conservation Service (NRCS), 2013. Whitman county, WA soil survey. URL: http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm.
Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67, 215–226.
Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691.
Pebesma, E., 2012. Spacetime: Spatio-temporal data in R. J. Stat. Softw. 51 (7), 1–30. URL: http://www.jstatsoft.org/v51/i07/.
Pebesma, E., Bivand, R., 2013. sp: classes and methods for spatial data. R package version 1.0-5. URL: http://CRAN.R-project.org/package=sp.
Pebesma, E., Gräler, B., 2013. gstat: spatial and spatio-temporal geostatistical modelling, prediction and simulation. R package version 1.0-16. URL: http://CRAN.R-project.org/package=gstat.
Pierce, F.J., Elliott, T.V., 2008. Regional and on-farm wireless sensor networks for agricultural systems in eastern Washington. Comput. Electron. Agric. 61, 32–43.
Pinheiro, J., Bates, D., 2009. Mixed-Effects Models in S and S-PLUS. In: Statistics and Computing, Springer.
Porter, J., Arzberger, P., Braun, H.-W., Bryant, P., Gage, S., Hansen, T., Hanson, P., Lin, C.-C., Lin, F.-P., Kratz, T., Michener, W., Shapiro, S., Williams, T., 2005. Wireless sensor networks for ecology. BioScience 55 (7), 561–572.
Qiu, H., Huggins, D., Wu, J., Barber, M., Mccool, D., Dun, S., 2011. Residue management impacts on field-scale snow distribution and soil water storage. Trans. ASABE 54, 1639–1647.
R Core Team, 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.
Sarkar, D., 2008. Lattice: Multivariate Data Visualization with R. Springer.
Sarkar, D., 2014. lattice: Lattice graphics. R package version 0.20-29. URL: http://CRAN.R-project.org/package=lattice.
Snepvangers, J.J.J.C., Heuvelink, G.B.M., Huisman, J.A., 2003. Soil water content interpolation using spatio-temporal kriging with external drift. Geoderma 112, 253–271.
Statnikov, A., Wang, L., Aliferis, C., 2008. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics 9 (1), 319.
Tucker, C., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150.
Veronesi, F., Corstanje, R., Mayr, T., 2012. Mapping soil compaction in 3D with depth functions. Soil Tillage Res. 124, 111–118.
Wang, J., Fu, B., Qiu, Y., Chen, L., Wang, Z., 2001. Geostatistical analysis of soil moisture variability on Da Nangou catchment of the loess plateau China. Environ. Geol. 41, 113–120.

Warnes, G., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., MacQueen, D., Magnusson, A., Rogers, J., et al. 2014. gdata: Various R programming tools for data manipulation. R package version 2.13.3. URL: http://CRAN.R-project.org/package=gdata.

Western Regional Climate Center, 2013. Climate summary, Pullman, WA. URL: http://www.wrcc.dri.edu.

Wickham, H., 2014. plyr: Tools for splitting, applying and combining data. R package version 1.8.1. URL: http://CRAN.R-project.org/package=plyr.

Wilson, D., Western, A., Grayson, R., Berg, A., Lear, M., Rodell, M., Famiglietti, J., Woods, R., McMahon, T., 2003. Spatial distribution of soil moisture over 6 and 30 cm depth, mahurangi, river catchment, New Zealand. J. Hydrol. 276, 254–274.