AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Biogeosciences

RESEARCH ARTICLE
10.1002/2017JG003837

Key Points:
- Hydrogeophysical observations provide additional environmental covariates at high spatial scales
- Hydrogeophysical derived environmental covariates help predict soil properties and states
- Environmental covariates should be considered when designing future observation networks

Supporting Information:
- Supporting Information S1
- Data S1
- Data S2
- Data S3
- Data S4
- Data S5
- Data S6
- Data S7
- Data S8
- Movie S1
- Movie S2
- Movie S3

Correspondence to:
T. E. Franz,
tfranz2@unl.edu

Citation:
Franz, T. E., T. D. Loecke, A. J. Burgin, Y. Zhou, T. Le, and D. Moscicki (2017), Spatiotemporal predictions of soil properties and states in variably saturated landscapes, *J. Geophys. Res. Biogeosci.*, *122*, 1576–1596, doi:10.1002/2017JG003837.

Received 8 MAR 2017
Accepted 5 JUN 2017
Accepted article online 19 JUN 2017
Published online 5 JUL 2017

Spatiotemporal predictions of soil properties and states in variably saturated landscapes

Trenton E. Franz[1], Terrance D. Loecke[2], Amy J. Burgin[3], Yuzhen Zhou[4], Tri Le[4], and David Moscicki[1]

[1]School of Natural Resources, University of Nebraska–Lincoln, Lincoln, Nebraska, USA, [2]Kansas Biological Survey and Environmental Studies Program, University of Kansas, Lawrence, Kansas, USA, [3]Kansas Biological Survey, Environmental Studies Program and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas, USA, [4]Department of Statistics, University of Nebraska–Lincoln, Lincoln, Nebraska, USA

Abstract Understanding greenhouse gas (GHG) fluxes from landscapes with variably saturated soil conditions is challenging given the highly dynamic nature of GHG fluxes in both space and time, dubbed hot spots, and hot moments. On one hand, our ability to directly monitor these processes is limited by sparse in situ and surface chamber observational networks. On the other hand, remote sensing approaches provide spatial data sets but are limited by infrequent imaging over time. We use a robust statistical framework to merge sparse sensor network observations with reconnaissance style hydrogeophysical mapping at a well-characterized site in Ohio. We find that combining time-lapse electromagnetic induction surveys with empirical orthogonal functions provides additional environmental covariates related to soil properties and states at high spatial resolutions (~5 m). A cross-validation experiment using eight different spatial interpolation methods versus 120 in situ soil cores indicated an ~30% reduction in root-mean-square error for soil properties (clay weight percent and total soil carbon weight percent) using hydrogeophysical derived environmental covariates with regression kriging. In addition, the hydrogeophysical derived environmental covariates were found to be good predictors of soil states (soil temperature, soil water content, and soil oxygen). The presented framework allows for temporal gap filling of individual sensor data sets as well as provides flexible geometric interpolation to complex areas/volumes. We anticipate that the framework, with its flexible temporal and spatial monitoring options, will be useful in designing future monitoring networks as well as support the next generation of hyper-resolution hydrologic and biogeochemical models.

Plain Language Summary Understanding greenhouse gas emissions from landscapes with variable saturated soil conditions is challenging given the highly dynamic nature of greenhouse gases (GHGs) in both space and time, dubbed hot spots, and hot moments. On one hand, our ability to directly monitor these processes is limited by sparse in situ and surface observational networks. On the other hand, proximal and remote sensing approaches provide spatial data sets but are limited by indirect observations and infrequent mapping. In this work, we use a statistical framework to merge sparse sensor network observations with reconnaissance mapping. The framework addresses a critical gap that exists in many long-term monitoring networks, specifically how to combine sparse in situ monitoring networks with proximal sensing. We found that combining time-lapse electromagnetic induction surveys with landscape features was able to generate statistical models of (1) soil inventories (i.e. clay percent, total soil carbon, and soil nitrogen) across the study site and (2) spatiotemporal models of key abiotic factors (i.e. soil temperature, soil water content, and soil oxygen) that may affect GHGs. The developed framework allows for temporal gap filling of individual sensors as well provides flexible geometric interpolation making it well suited for the use in the next generation of hydrologic and biogeochemical models.

1. Introduction

Understanding greenhouse gas (GHG) fluxes from landscapes with variably saturated soil conditions (e.g., aquatic terrestrial interfaces (ATI)) is challenging given the highly variable nature of GHG fluxes in both space and time, dubbed hot spots, and hot moments [*McClain et al.*, 2003]. A fundamental challenge for monitoring, predicting, and mitigating landscape scale GHG fluxes is (1) where to place a finite number of soil sensors or surface flux chambers? and (2) how to upscale those observations to the landscape scale where stakeholder decisions are made? Recent work has addressed these two key questions with respect to total soil carbon

stocks across Alaska using existing soil sample locations and readily available environmental covariates [*Mishra and Riley*, 2015] as well as under various climate change scenarios [*Vitharana et al.*, 2017]. However, a robust statistical framework remains elusive for scales less than ~50 m due to the scale limitation of existing environmental covariates, missing critical environmental covariates on vegetation and hydrologic processes, and/or computationally resources [*Mishra et al.*, 2010; *Mishra and Riley*, 2015; and *Hengl et al.*, 2017]. In the absence of effective monitoring strategies, proper inventory calculations remain challenging and thus limits the effectiveness of establishing and regulating future carbon markets.

GHG fluxes involve complex abiotic and biotic processes and feedback operating at different scales [*Hillel*, 1998; *Blagodatsky and Smith*, 2012; *Rubol et al.*, 2013; *Mishra and Riley*, 2015]. With respect to abiotic processes, the flow of heat, vapor, and liquids are represented by several key state variables, soil temperature, soil water content, and soil oxygen [*Hillel*, 1998]. The spatiotemporal dynamics of each state variable within an ATI are controlled by both soil hydraulic/thermal properties and boundary conditions governing the fluxes [*Hillel*, 1998]. Often landscape position (i.e., elevation and surface flow convergence), soil textural class, and land use are used as criteria to select an experimental monitoring design as well as upscale point scale properties or states to the landscape. *Crow et al.* [2012] discuss the challenges of soil water content organization across scales being influenced by scale-dependent environmental covariates. With respect to biotic processes, soil inventories (i.e., total soil carbon and total nitrogen) are used as criteria to select representative sampling and monitoring locations [see, e.g., *Mishra et al.*, 2010; *Mishra and Riley*, 2012; and *Hengl et al.*, 2017].

As discussed above selecting appropriate monitoring locations requires an understanding of both surface and subsurface features and processes. Given the time, labor, and cost for collecting detailed soil cores to estimate landscape patterns of soil hydraulic/thermal properties and soil inventories, monitoring networks are often designed in the absence of such information or assessed after the installation of the monitoring network (as is the case with the current study). The challenge of effective GHG sampling design, monitoring state variables, upscaling state variables/fluxes, and understanding soil-atmosphere interactions is ubiquitous across national and global monitoring efforts (e.g., Ameriflux, OzFlux, Carboafrican, South African Ecological Observation Network, Fluxnet, National Ecological Observation Network, Long Term Ecological Research network, Long Term Agricultural Research network, and Critical Zone Observatory). Here we investigate if hydrogeophysical mapping can provide critical missing environmental information related to subsurface hydraulic/thermal properties and soil inventories at scales less than 50 m at an existing long-term ATI monitoring study site in Ohio, United States [*Jarecke et al.*, 2016].

While hydrogeophysical methods have a long history of use in geologic and hydrologic studies [*Jakosky*, 1950], their widespread use in disciplines such as soil biogeochemistry [*Binley et al.*, 2015] has been limited due to the "soft" data that is inherently generated. For example, in this study we used electromagnetic induction (EMI) mapping which measures the bulk electrical conductivity (ECa) of soil in the near surface [*McNeill*, 1980; *Abdu et al.*, 2008]. It is well known that ECa varies with soil temperature, soil water content, and soil cation-exchange capacity [*Friedman*, 2005], thus obfuscating which soil property or state is being measured [*Altdorff et al.*, 2017]. In order to better isolate soil properties and states from ECa, previous work has developed correction factors for changes in soil temperature [*Friedman*, 2005] and used time-lapse differencing to isolate time-varying components, such as subsurface water flow pathways [*Robinson et al.*, 2009; *Franz et al.*, 2011]. As an extension of time-lapse ECa differencing, here we will further deconstruct the variability in the images using the technique of empirical orthogonal functions (EOF) [*Perry and Niemann*, 2007] in order to better predict spatial soil properties [*Pedrera-Parrilla et al.*, 2016]. We will then investigate whether the deconstructed hydrogeophysical maps and landscape position (specifically elevation and topographic wetness index (TWI), as defined by *Sorensen et al.* [2006]) are good environmental covariates to predict soil properties and states across space and through time.

The primary objective of this work is to use a robust statistical framework to merge sparse sensor network observations with reconnaissance style hydrogeophysical mapping at a well-studied site in Ohio. The merged data sets will provide critical spatiotemporal monitoring of key abiotic state variables and properties that potentially influence surface GHG emissions (see *Graf et al.* [2012] and *Fóti et al.* [2016] for spatial soil respiration studies). Specifically, we will use time-lapse EMI imagery with other common environmental covariates to generate (1) spatial interpolation of clay percent, total soil carbon, and total nitrogen across the study site

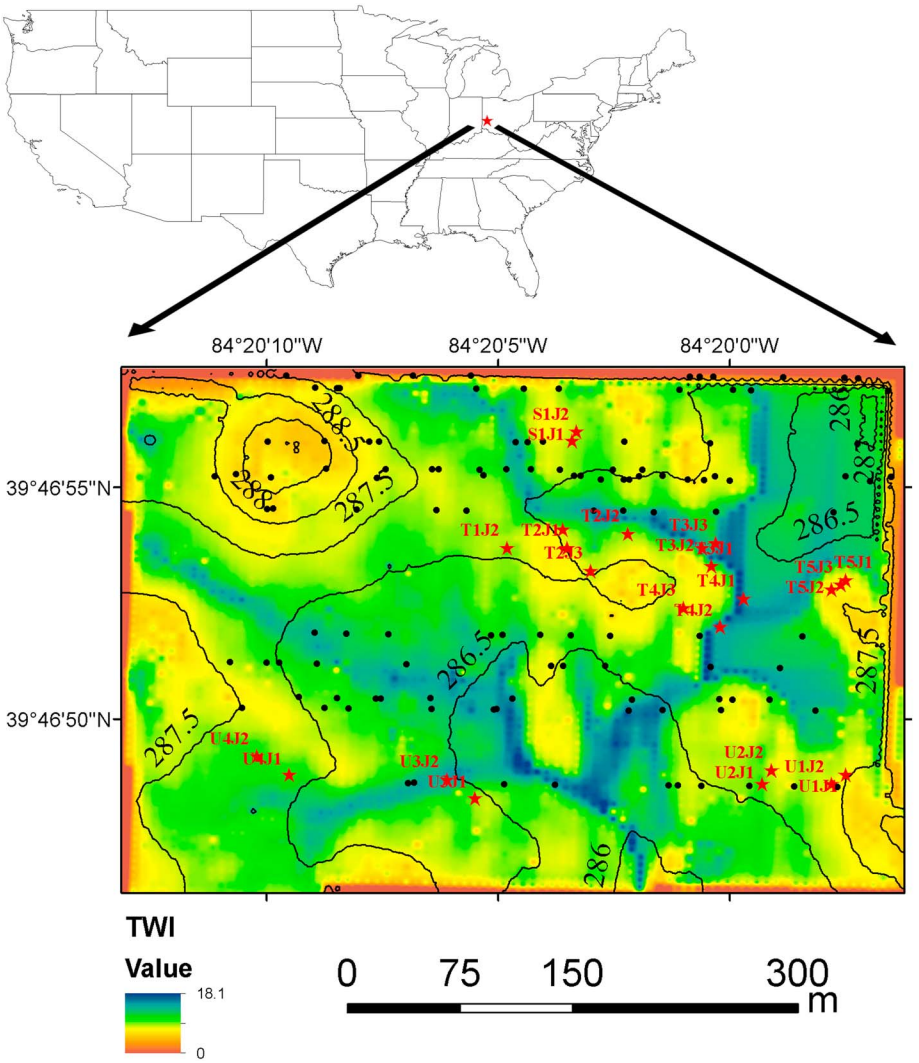

Figure 1. (top) Location of study site in western Ohio. (bottom) Layout of sensor network (red stars) and location of extracted soil cores (black dots). The site shows moderate changes in elevation (0.5 m contours included) and distinct surface convergence zones (shown by topographic wetness index and background surface). The site is bounded by a built up levee (north and east sides) that creates a permanent wetland feature. The site will then fill with water during periods of seasonal rainfall.

at high spatial resolutions (5 m) and (2) spatiotemporal models of soil temperature, soil water content, and soil oxygen. Given the universal challenge of scaling sparse monitoring networks to the landscape scale, we envision the identified hydrogeophysical mapping techniques, statistical framework, and upscaling procedure to be of widespread interest to both the hydrological and biogeochemical communities.

2. Materials and Methods

2.1. Site Description

The study site is located in the Great Miami Wetland Mitigation Bank, a 46 ha restored wetland in Trotwood, Montgomery County, Ohio, United States (39.7808°N, −84.3350°W; Figure 1) [*Jarecke et al.*, 2016]. The study site serves as a wetland mitigation bank to provide compensatory mitigation for impacted waters of the United States as authorized under the Clean Water Act. The project is funded and managed by the Five Rivers Metro Parks of Dayton, OH. The restoration involved extensive Earth moving during 2011 to create Earth dams around the study site boundary and planting native wetland and upland vegetation during 2012 (a mix of perennial and annual grasses and forbes). An earthen wall encompasses the northern and

eastern boundaries creating a seasonal wetland following rainy periods. Previously to 2010, the site was passively drained (via subterranean tile networks) for row crop production for over 100 years.

Average annual precipitation is 1005 mm, and average annual daily temperature is 10.8°C. The soils at the study site are characterized as poorly drained silty clay loam (Brookston, fine-loamy, mixed, active, and mesic Aeric Endoaqualfs). In November 2011, 120 soil cores uniformly distributed around the study site were collected and analyzed for soil texture, total soil carbon, and total nitrogen in four layers down to ~1 m depth (see Table S1 in the supporting information). The site contains moderate topographic relief (maximum slope of 2.75%; Figure 1) and defined drainage patterns (as reflected by the TWI; Figure 1).

2.2. In Situ Sensor Network

In April 2012, an in situ sensor network was installed at 24 locations (red stars in Figure 1) to record soil water content, soil temperature (SDI-12 hydra probes, Stevens Water, Portland, Oregon), and soil oxygen (SO-110 soil O_2 sensors, Apogee Instruments, Logan, Utah) at 30 min intervals and a depth of 10 cm. The soil O_2 sensor has internal temperature composition circuitry, and the data are expressed as relative concentration, i.e., the mixing ratio as a percentage with a maximum reading equal to atmospheric O_2 concentration of 20.9%. The O_2 sensors used the sensor mV output in open air at 100% humidity to calibrate individual O_2 sensors relative to the atmosphere [*Jarecke et al.*, 2016]. The soil O_2 sensors were installed horizontally at 10 cm of depth into the wall of a 20 cm deep pit by removing the shape and volume of soil to be occupied by the sensor and inserting the sensor into this void before back-filling the pit with soil. All sensor observations are transmitted through a cellular network to an offsite server. Due to sensor errors and power failures gaps exist in the measurement record for individual sensors and the entire network. For the work presented here we will analyze data over the first 19 months of the experiment, 1 May 2012 through 31 December 2013, and the 10 cm soil depth (see Table S2 for daily average values and Table S3 for location and corresponding environmental covariates).

2.3. Electromagnetic Induction (EMI) Surveys

Between 1 September 2015 and 22 November 2015 we collected seven EMI surveys at the study site using a Dualem-21S (Dualem Inc., Ontario, Canada) over a range of dry and wet soil conditions. Using an all-terrain vehicle, we pulled the EMI instrument with a plastic sled recording electrical conductivity measurements every second and integrating the GPS locations with an Archer Field computer (Juniper Systems, Logan, UT). The ~16 ha surveys took approximately 3 h to complete using 10 m transect spacing and driving at speeds of 8 kph or less in a boustrophendonic pattern (~11,000 data point per survey). The EMI works by transmitting a known magnetic field, which induces an electric field in the soil and thus a secondary magnetic response, which is then measured by the sensor [*McNeill*, 1980]. The ratio of the magnitude of the primary and secondary magnetic field is proportional to the apparent soil bulk electrical conductivity (ECa, mS m^{-1}) [*McNeill*, 1980]. Measurements of ECa have been used to quantify soil structure [*Hendrickx et al.*, 2002; *Hendrickx and Kachanoski*, 2002], clay content [*Triantafilis and Lesch*, 2005], and subsurface water flow paths [*Robinson et al.*, 2009; *Franz et al.*, 2011] by measuring the electromagnetic properties of the combined soil water matrix [*Friedman*, 2005]. We note that the EMI surveys (2015) were collected after the soil core extraction (2011) and period of in situ observations from the sensor network (2012 to 2013). Design of future sensor networks in combination with hydrogeophysical surveys will be discussed in section 4.1.

2.4. Empirical Orthogonal Function (EOF) Analysis

The purpose of EOF analysis is to decompose variability in observed data into a set of time-invariant orthogonal spatial patterns (EOFs) and a set of time series called expansion coefficients (ECs). Based on EOFs, the original coordinate system is rotated into a new system aligned along perpendicular axes. By retaining only significant EOFs, EOF analysis can effectively reduce the dimensionality of the data set while preserve most of the variability in the data. Detailed procedures for performing EOF analysis can be found elsewhere [*Perry and Niemann*, 2007; *Korres et al.*, 2010], and only a brief discussion is offered here. In this work we will perform EOF analysis on both time series sensor observations as well as repeated imagery from the EMI time-lapse mapping.

With respect to sensor time series EOF analysis, we take a data set with n locations and m observations at each location, where the spatial anomalies of the sensor observations can be computed as

$$a_i(t) = s_i(t) - \frac{1}{n}\sum_{j=1}^{n} s_j(t) \tag{1}$$

where $a_i(t)$ and $s_i(t)$ are the sensor observation spatial anomaly and sensor observation at location i and time t, respectively. A matrix of sensor observation spatial anomalies, $\boldsymbol{A}$ (capital letters in bold denote matrices), can be constructed as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_{11} & \cdots & \boldsymbol{a}_{1n} \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}_{m1} & \cdots & \boldsymbol{a}_{mn} \end{bmatrix} \tag{2}$$

Then, an empirical covariance matrix $\boldsymbol{V}$ can be calculated as

$$\boldsymbol{V} = \frac{1}{m-1}\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} \tag{3}$$

where the superscript $\boldsymbol{T}$ indicates the matrix transpose.

To perform EOF analysis, we find eigenvectors and eigenvalues for $\boldsymbol{V}$, which satisfy the following equation.

$$\boldsymbol{V}\times\boldsymbol{E} = \boldsymbol{E}\times\boldsymbol{L} \tag{4}$$

where $\boldsymbol{E}$ contains eigenvectors (i.e., EOFs) in columns

$$\boldsymbol{E} = \begin{bmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{n1} & \cdots & e_{nn} \end{bmatrix} \tag{5}$$

and $\boldsymbol{L}$ contains eigenvalues along the diagonal

$$\boldsymbol{L} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_{nn} \end{bmatrix} \tag{6}$$

The above procedure rotates the original coordinate axes with each axis indicating a sampling time into a new set of orthogonal coordinate axes with each eigenvector representing a new axis. The eigenvalues explain the variance in the data along the direction of each corresponding new axis, and the portion of the explained variance (EV_k) by the kth new axis in the total variance can be computed as

$$EV_k = \frac{l_{kk}}{\sum_{p=1}^{n} l_{pp}} \tag{7}$$

The eigenvectors are then arranged according to eigenvalues: the first axis explains the largest variance in the data, while each following axis explains the largest remaining variance and is orthogonal to other axes. The ECs are then found by projecting $\boldsymbol{A}$ onto $\boldsymbol{E}$.

$$\boldsymbol{F}=\boldsymbol{A}\times\boldsymbol{E} \tag{8}$$

where $\boldsymbol{F}$ contains each ECs in columns. Based on the explained variance only significant EOF/EC pairs are retained for the remaining analysis (here defined as a threshold of greater than ~8% explained variance; see *Peres-Neto et al.* [2005] for a complete discussion).

With respect to the EOF analysis of the EMI time-lapse imagery, the procedure is identical as above save the initial construction of spatial anomaly matrix $\boldsymbol{A}$. Here we computed the spatial anomalies for each EMI grid

location on the same sampling date. The resultant $\boldsymbol{V}$ is 7×7 resulting in seven EOF/EC pairs from the seven EMI images collected. With respect to the sensor observations they resulted in up to 24 EOF/EC pairs based on the number of sampling locations. Note that some sensor locations were not included in the analysis due to sparse records due to sensor failure. Following the EOF analysis the time series or spatial image can be reconstructed using only the retained EOF/EC pairs. We note that while a single set of EOFs are used the EC values are found at each time step using the sensor observations and equation (7). For missing individual sensor observations the spatial mean is used to gap fill the observation record. This procedure is a slight modification to other work [cf. *Perry and Niemann*, 2007], where previous work used only the spatial network mean and a seasonality function to estimate the time varying EC values. See R code in the supporting information for full methodological details.

2.5. Spatio-Temporal Statistical Analysis

The EOF/EC pairs from the time series analysis were combined with landscape environmental covariates (elevation, TWI, and ECa EOFs) to scale the observations across the study site for each day. We used multivariate linear regression to explore different statistical models for the analyses. Final model selection was based on the prediction sum of squares (PRESS), see attached R code in the supporting information. In addition, a cross-validation experiment was performed for all statistical models where appropriate. With respect to the spatial analysis we considered eight different interpolation methods: linear (V4 method in MATLAB in R2015b), inverse distance weighted with fixed radius (here 50 m), inverse distance weighted with fixed number of points (here 10), ordinary kriging, and regression kriging with elevation, TWI, ECa EOF1, and ECa EOF2. All spatial analyses were conducted in MATLAB R2015b using standard and user community libraries. A summary of all raw data and cross-validation experiments are provided in the supporting information. With respect to the time series analysis and cross-validation results, R code is provided in the supporting information. With respect to the spatial analyses and cross-validation results, MATLAB code can be requested from the corresponding author. Finally, a summary of each data set, statistical analyses, resulting products, and information is provided as a flowchart in Figure 2 to help orient the reader for the remainder of the manuscript. The flowchart also serves as a general framework for merging sparse observational networks with hydrogeophysical data and environmental covariates.

3. Results

3.1. Spatial Models of Landscape Soil Properties

Spatially distributed estimates of soil properties over the landscape are often estimated from spatial interpolation of core samples collected from a grid or from available data sets like Soil Survey Geographic Database (SSURGO) [*Soil Survey Staff, Natural Resources Conservation Service*, 2016, http://websoilsurvey.nrcs.usda.gov/]. For fine-scale information and detailed analyses the SSURGO data sets are often too coarse. On the other hand, grid samples of soil cores are time and labor-intensive. Moreover, the spatial correlation length from grid sample soil properties is often short leading to problematic spatial interpolation from techniques like kriging without dense observations. As an alternative strategy, here we investigate the correlation between the EMI survey results with the 2011 soil core data set. While this strategy is not new [*Pedrera-Parrilla et al.*, 2016], here we argue that the EOF/EC pairs from the time-lapse imagery will be better environmental covariates with soil properties as individual EMI surveys are known to be affected by soil temperature, soil water content, pore water pH, etc. [*Friedman*, 2005].

Seven EMI surveys were conducted between September and November 2015 with two relatively dry surveys, two wet surveys, and three extremely wet surveys following seasonal rainfall (Figures 3a–3g). The ECa values were corrected for soil temperature changes, and erroneous readings were removed following the same procedures detailed in *Franz et al.* [2011]. Following quality assurance/quality control the temperature-corrected ECa values for each survey were interpolated to a 5 m grid using ordinary kriging (see Table S4 for processed data and corresponding environmental covariates). Next, the interpolated ECa values were analyzed with EOF to generate seven EOF/EC pairs (Figure 3h and Tables 1 and S4). EOF analysis indicated that the first EOF/EC pair explained 82.4% of the spatial variability followed by pair two which explained 6.7% of the variability. The other five EOF/EC pairs contributed less than 6% explained variance and were discarded. A leave one-leave-out cross-validation analysis of the seven images indicated that four surveys (ideally two wet and two dry) were sufficient to estimate the first two EOF/EC pairs, with less than

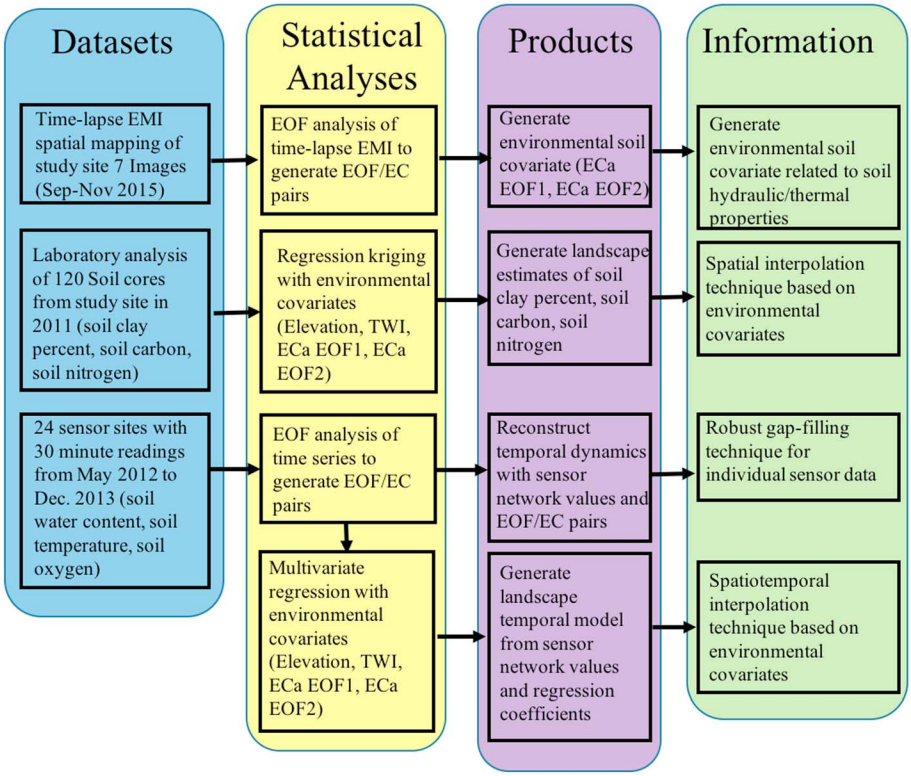

Figure 2. Flow diagram summarizing the available spatial and temporal data sets (blue box), statistical analyses (yellow box), generated products (purple box), and information (green box). The remaining figures, tables, videos, and supporting information summarize results of each box.

5% changes in EOF coefficients. The reduction in required EMI surveys will streamline this method for other applications in the future.

The two selected EMI EC/EOF pairs (ECa EOF1 and ECa EOF2) were then used as environmental covariates to investigate different statistical models using laboratory-derived properties obtained from the 120 soil cores extracted from the site in 2011 (Table S1). Each core was analyzed for sand, silt, clay weight percent, total soil carbon, and total nitrogen in four layers every 25 cm. The vertically averaged values from each core were then regressed against the corresponding 5 m grid site elevation, TWI, ECa EOF1, and ECa EOF2 (Table S1 and Figure 4). Figure 4 illustrates that clay weight percent, total soil carbon, and total nitrogen are strongly linear correlated to elevation, TWI, and ECa EOF1, albeit with significant scatter. Computing the semivariogram [*Goovaerts*, 1997] from all 120 cores resulted in correlation lengths of 120 m for clay, 35 m for total soil carbon, and 95 m for total nitrogen, thus requiring fairly dense grid sampling for spatial interpolation.

Next we performed a cross-validation experiment for eight different spatial interpolation methods for each of the three properties. Here we randomly selected 12, 30, 60, 90, and 108 soil cores to perform the spatial interpolation and compared the results against the remaining validation cores. We repeated this analysis 10 times and reported the average root-mean-square error (RMSE) for the validation cores (see Figure 5 for average results, Table S5 for all cross-validation simulation results, and Figure 6 for best fit results). With respect to clay weight percent we found the average cross-validation RMSE decreased from 4.57 to 3.18 (a 30.5% error reduction) using regression kriging with ECa EOF1 versus linear interpolation. Similarly, for total soil carbon we found an RMSE reduction from 0.776 to 0.555 (a 29.5% error reduction) using regression kriging with ECa EOF1 versus linear interpolation. With respect to total nitrogen we found an RMSE reduction from 0.1187 to 0.1165 (1.5% error reduction) using regression kriging with TWI versus linear interpolation. We note that several methods produced similar results for total nitrogen. Most importantly, the cross-validation results clearly illustrate greater RMSE error reduction in clay weight percent and total soil carbon weight percent

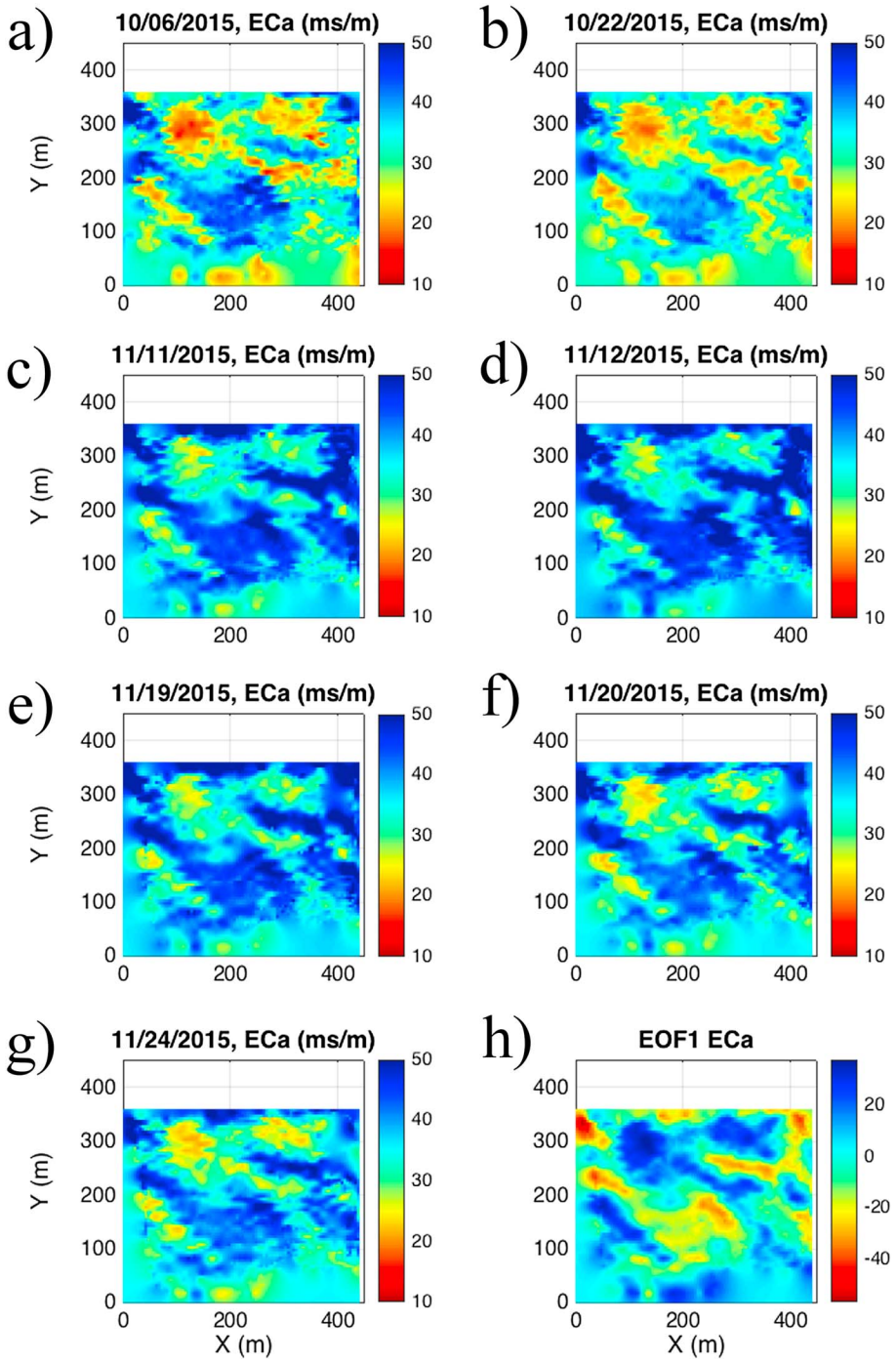

Figure 3. Maps of temperature corrected bulk electrical conductivity (ECa) from (a–g) seven survey dates and (h) first empirical orthogonal function (EOF) of ECa surveys.

using regression kriging with ECa EOF1 versus linear interpolation or other regression kriging with environmental covariates. Furthermore, we note that the use of regression kriging with ECa EOF1 will require fewer sampling locations for spatial interpolation as compared to linear interpolation or inverse distance weighted methods which are subject to "bulleyes" and/or smoothing local features. Consideration of soil environmental covariates, like ECa EOF1, may be beneficial for resampling efforts of existing sites or designing future sampling strategies to optimize the number and location of in situ samples.

Table 1. Summary of Explained Variance for Ranked Empirical Orthogonal Function and Expansion Coefficient (EOF/EC) Pairs for Soil Temperature, Soil Water Content and Soil Oxygen Sensors, and Bulk Electrical Conductivity (ECa) Imagery[a]

	Explained Variance From Sensor/Chamber Time Series			Explained Variance Spatial Mapping
Ranked EOF/EC Pair	Soil Temperature	Soil Water Content	Soil Oxygen	ECa
1	***88.9%***	***38.7%***	***40.3%***	***82.4%***
2	4.5%	***25.5%***	***16.9%***	***6.7%***
3	2.3%	***12.4%***	***9.2%***	2.5%
4	1.3%	7.8%	***8.3%***	2.4%
5	1.1%	5.4%	4.8%	2.2%

[a]The bold italicized values represent the number of EOF/EC pairs used in subsequent statistical analyses.

3.2. Temporal Reconstruction and Gap Filling of In Situ Sensor Data Using EOF Analysis

Using up to 24 in situ sensor locations we used EOF to analyze daily soil water content, soil temperature, and soil oxygen at 10 cm depth between 1 May 2012 and 31 December 2013 (Figures 7a, 8a, and 9a and Tables S2 and S3). The EOF analysis indicated that the first EOF/EC pair (of the 24) were above the 8% threshold for soil temperature (88.9% explained variance), three EOF/EC pairs for soil water content (38.7%, 25.5%, and 12.4%), and four pairs for soil oxygen (40.3%, 16.9%, 9.2%, and 8.3%) (Table 1). Using the selected EOF/EC pairs, the reconstructed and gap-filled time series of the in situ data are illustrated in Figures 7b, 8b, and 9b. We note that ~200 observations were needed to estimate the time-invariant EOF coefficient for soil temperature and first two soil water content EOFs, while ~390 observations were needed to estimate the first two soil oxygen EOFs. All data were needed to estimate the third soil water content EOF and third and fourth soil oxygen EOFs. See R code in the supporting information for details on EOF estimation and gap filling procedure. Table 3 summarizes the individual sensor daily RMSE during calibration and reconstruction gap filling periods.

With respect to goodness of fit we find that 19 of the 24 soil temperature sensors have RMSE $<$1°C for the calibration period and 18 of 24 have RMSE $<$1 °C for the reconstruction period. Sensor T3J2 behaves the poorest (RMSE = 2.04°C). With respect to soil water content we find that 19 of the 20 sensors have RMSE $<$0.04 m^3/m^3 for the calibration period (threshold defined for validation of remote sensing products [*Entekhabi et al.*, 2010]) and 17 of the 20 have RMSE $<$0.04 m^3/m^3 for the reconstruction period. Sensor U2J2 behaves the poorest (RMSE = 0.120 m^3/m^3). With respect to soil oxygen we find that 10 of 22 sensors have RMSE $<$2% for the calibration period and 7 of 22 have RMSE $<$2% for the reconstruction period. Sensor U3J2 (RMSE = 6.67%) behaved the poorest. It is unclear if the larger reconstruction sensor errors are due to either failure of the individual sensor or the inability of EOF method to perform proper reconstruction (i.e., need higher-order EOF/EC pairs for reconstruction). In summary, soil temperature behaved the best and required only a single EOF/EC pair for reconstruction. Soil water content also behaved well (in comparison to remote sensing validation criterion) but required three EOF/EC pairs. Estimation of soil oxygen was the most challenging requiring four EOF/EC pairs and had the largest relative uncertainty compared to the expected range of measurement.

3.3. Spatial Estimates of In Situ Sensor Data Using EOF Analysis

Using the identified EOF/EC pairs, we investigated the association of the identified EOFs against the sensor's environmental covariates (elevation, TWI, ECa EOF1, and ECa EOF2; see Table S3 for raw data). Here we investigated a few simple statistical models ranging from linear, multivariate linear, and nonlinear. Models were selected based on PRESS values (see R code in the supporting information for full analysis and cross-validation results). While more robust model fitting is possible, the main objective of this work was to illustrate the correlation between environmental covariates and derived sensor time series EOFs, for spatiotemporal interpolation across the landscape.

Figure 10 illustrates the soil temperature EOF scatterplot versus environmental covariates, Figure 11 illustrates the soil water content EOFs scatterplot, and Figure 12 illustrates the soil oxygen EOFs. Table 3 summarizes the regression analysis, selected model, selected model coefficients, statistical fit, and cross-validation results. With respect to soil temperature, we find that the EOF is most correlated to TWI (R^2 of

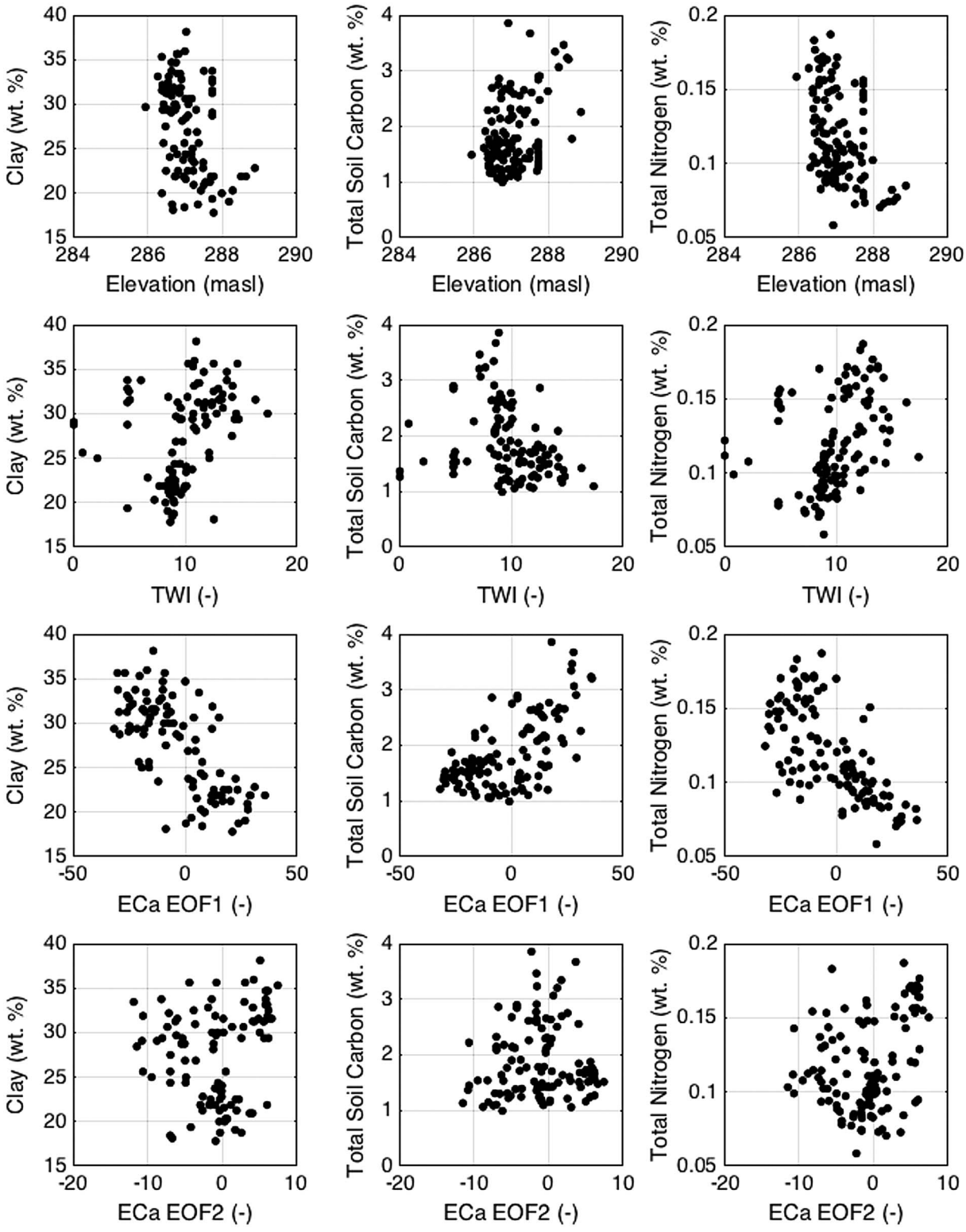

Figure 4. Scatterplots of clay weight percent, total soil carbon weight percent, and total nitrogen weight percent from 120 soil cores versus associated soil core elevation, topographic wetness index (TWI), bulk electrical conductivity first empirical orthogonal function (ECa EOF1), and bulk electrical conductivity second empirical orthogonal function EOF2 (5 m grid).

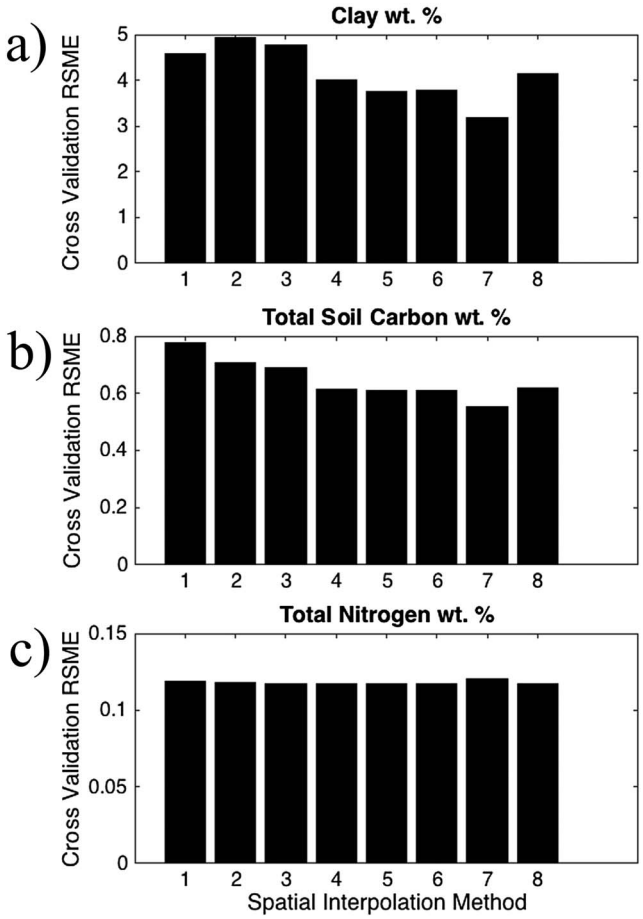

Figure 5. Cross validation results of (a) clay weight percent, (b) total soil carbon weight percent, and (c) total nitrogen weight percent for eight different spatial interpolation methods: (1) linear (V4 method in MATLAB in R2015b), (2) inverse distance weighted with fixed radius (here 50 m), (3) inverse distance weighted with fixed number of points (here 10), (4) ordinary kriging, (5) regression kriging with elevation, (6) regression kriging with topographic wetness index, (7) regression kriging with bulk electrical conductivity first empirical orthogonal function, and (8) regression kriging with bulk electrical conductivity second empirical orthogonal function. See Table S5.

0.607, validation R^2 or 0.515). This indicates high confidence in our spatiotemporal predictions of soil temperature across the landscape. With respect to soil water content we find that the EOFs are best correlated with elevation, TWI, and ECa EOF2. The R^2 for SWC EOF1 is high (R^2 of 0.650, validation R^2 or 0.510), while SWC EOF2 (R^2 of 0.249, validation R^2 or 0.061) and SWC EOF3 (R^2 of 0.234, validation R^2 or 0.043) behave much poorer. While we are confident in the most important first SWC EOF spatial predictor, the lower fits for SWC EOF2 and EOF3 make our landscape predictions less certain. With respect to soil oxygen we find that elevation, TWI, ECa EOF1, and ECa EOF2 are all correlated to the four soil oxygen EOFs. The R^2 for soil oxygen EOF1 (R^2 of 0.636, validation R^2 or 0.524), EOF2 (R^2 of 0.256, validation R^2 or 0.105), and EOF4 (R^2 of 0.469, validation R^2 or 0.246) are high, while soil oxygen EOF3 (R^2 of 0.082, validation R^2 or 0.000) is poor. Our confidence in the landscape prediction of soil oxygen will be less certain than soil temperature and soil water content. However, the strong correlation with soil oxygen EOF1 and EOF2 should provide reasonable predictions across the landscape.

Using the sensor network values and gap filling missing data with the spatial mean we are able to construct a continuous time series of EC values for soil temperature, soil water content, and soil oxygen. In addition, we created a 5 by 5 m grid for the study site environmental covariates including elevation, TWI, ECa EOF1, and ECa EOF2 (Table S4). Using the model fits from Table 3 we are able to construct all the EOFs for the landscape. Multiplying the ECs and EOFs we are able to construct a continuous 5 by 5 m resolution daily soil temperature, soil water content, and soil oxygen product for the study site (Movies S1–S3 in the supporting information). We note that while daily data were used to estimate the ECs, the temporal resolution can be used to the finest grain of the sensor network (here 30 min observations). Lastly, we note that the geometry of the selected grid is also somewhat arbitrary and will be discussed more in section 4.2.

4. Discussion

4.1. Environmental Controls of Soil Properties

The design of landscape-scale GHG monitoring networks is very challenging in ATI with variably saturated conditions and influences from both abiotic and biotic factors. The variable GHG fluxes are notoriously difficult to monitor in both time and space. Furthermore, the selection of representative sampling locations and

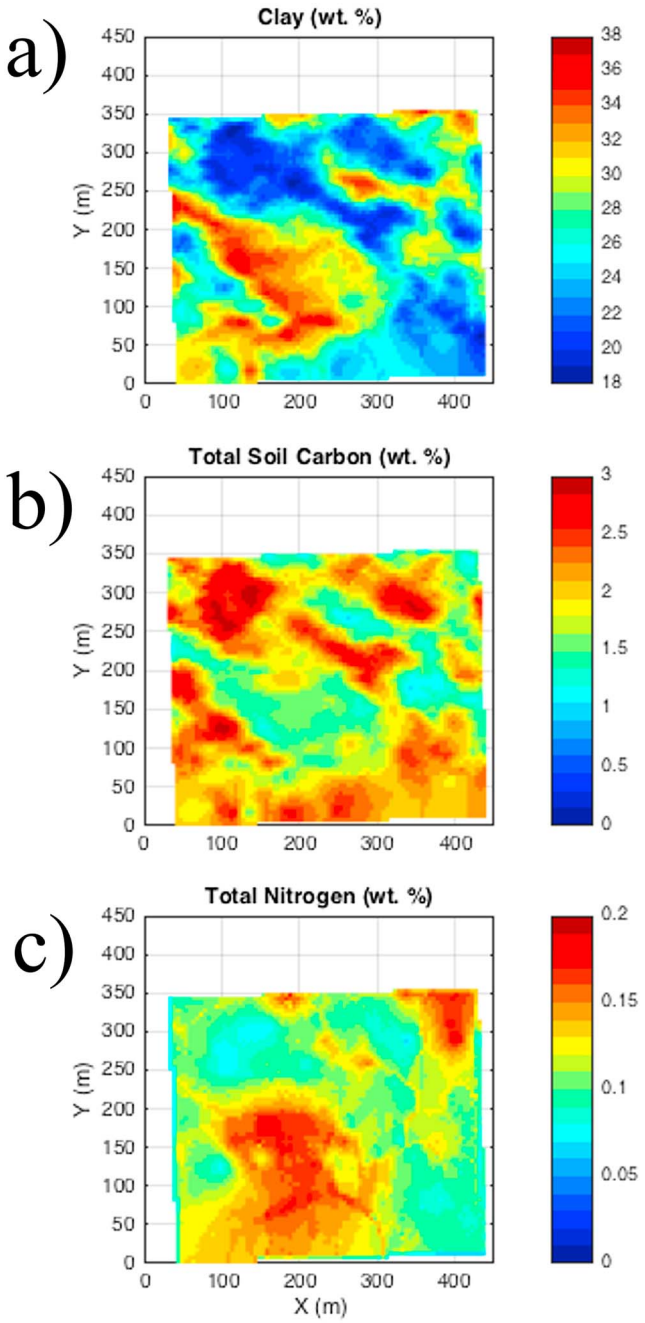

Figure 6. Spatial interpolation of (a) clay weight percent, (b) total soil carbon weight percent, and (c) total nitrogen weight percent using regression kriging and environmental covariates. See Table S4.

number of sensors and soil cores are often based on readily available information and economic considerations. We suspect that those decisions are made in the absence of high-quality information about spatial soil properties in the near subsurface. In this work we have shown that EOF analysis of time-lapse EMI provides a key and often missing soil covariate [*Pedrera-Parrilla et al.*, 2016]. We found ECa EOF1 to be a good predictor of spatial soil inventories, reducing cross-validation RMSE by 30.5% in clay weight percent, and 29.5% in total soil carbon weight percent (Figure 5 and Table S5) using regression kriging versus linear interpolation. Only marginal benefits were found for total nitrogen using regression kriging, perhaps related to the small range of observed values.

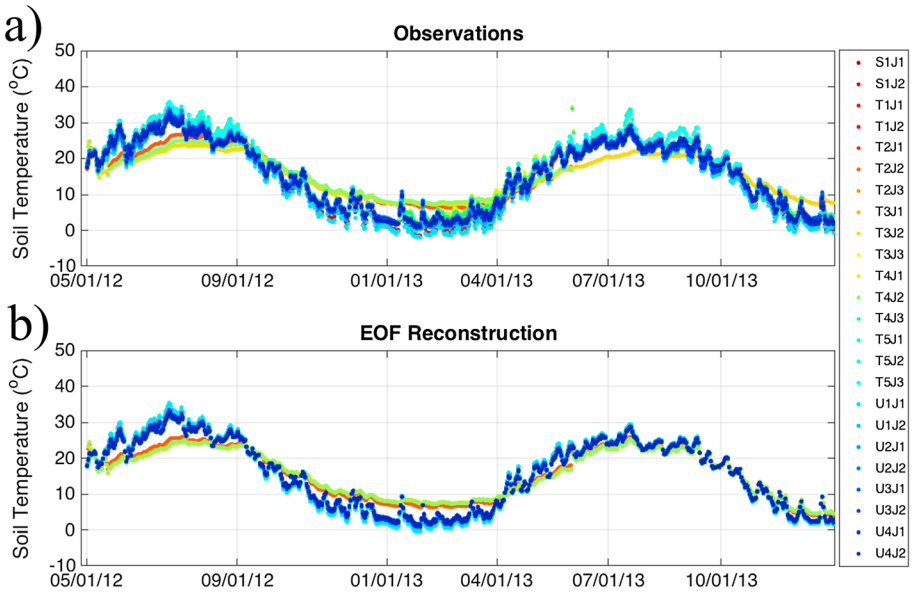

Figure 7. (a) Time series of observed soil temperature from 24 sensor locations and (b) time series of EOF reconstructed values. See Table 2 for statistical metrics.

Table 2. Statistical Summary of Individual Sensor Locations for Empirical Orthogonal Function (EOF) Analysis During Calibration and Reconstruction Periods[a]

	Time Series Analysis					
	Soil Temperature (°C)		Soil Water Content (cm^3/cm^3)		Soil Oxygen (%)	
Sensor/Chamber Location	Calibration RMSE	Reconstruction RMSE	Calibration RMSE	Reconstruction RMSE	Calibration RMSE	Reconstruction RMSE
S1J1	0.58	0.69	NA	NA	1.34	1.49
S1J2	1.18	0.81	0.031	0.033	NA	NA
T1J1	0.39	0.35	0.020	0.017	1.52	2.04
T1J2	0.41	0.39	0.020	0.017	2.52	2.71
T2J1	0.47	0.47	0.009	0.011	1.94	2.47
T2J2	0.70	0.83	0.009	0.011	2.35	5.43
T2J3	0.39	0.55	0.009	0.011	1.21	1.97
T3J1	0.48	0.49	0.007	0.005	1.82	1.93
T3J2	0.54	2.04	0.007	0.005	1.87	1.81
T3J3	0.83	0.69	0.007	0.005	2.03	1.87
T4J1	1.58	1.22	NA	NA	1.92	3.41
T4J2	0.36	0.64	NA	NA	1.82	4.20
T4J3	1.11	1.06	NA	NA	2.33	4.48
T5J1	1.14	1.56	0.007	0.024	2.44	2.52
T5J2	0.64	0.74	0.007	0.024	2.14	2.64
T5J3	0.92	1.42	0.007	0.024	2.33	4.25
U1J1	1.74	1.61	0.014	0.110	NA	NA
U1J2	0.80	0.40	0.019	0.023	1.45	2.49
U2J1	0.46	0.35	0.016	0.021	2.02	3.67
U2J2	0.38	0.47	0.022	0.120	2.68	1.82
U3J1	0.63	0.63	0.057	0.074	2.89	1.55
U3J2	0.62	0.87	0.034	0.037	2.75	6.67
U4J1	0.39	0.36	0.024	0.020	1.82	2.26
U4J2	0.58	0.42	0.023	0.018	2.30	2.12

[a]Reconstruction periods contain incomplete records due to erroneous sensor readings of failures. NA stands for not available as not enough data were available during this time period.

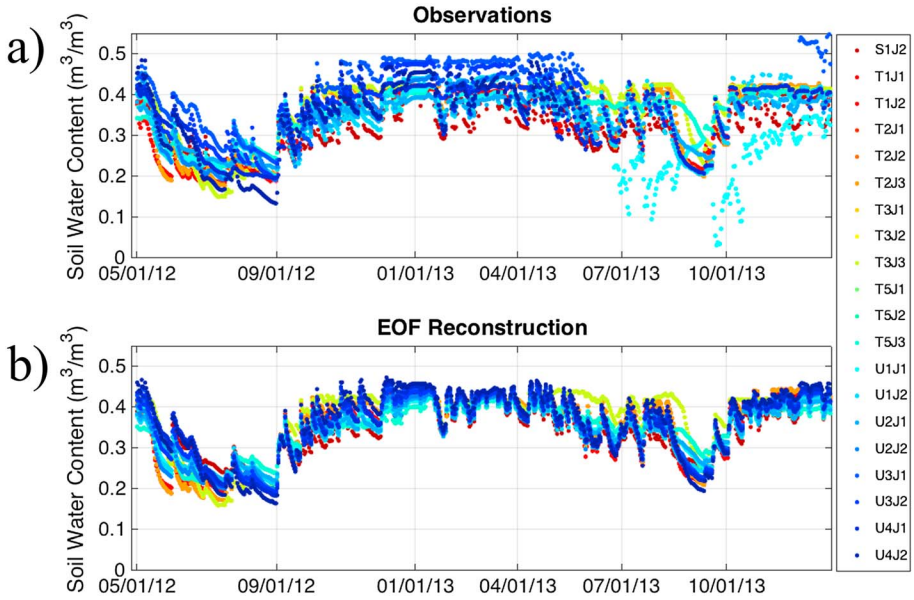

Figure 8. (a) Time series of observed soil water content from 20 sensor locations and (b) time series of EOF reconstructed values. See Table 2 for statistical metrics.

Elevation and TWI were, not surprisingly, the primary controlling factors of soil temperature, soil water content [*Western and Bloschl*, 1999], and soil oxygen. The ECa EOFs were also found to be good secondary covariates for soil water content and soil oxygen (Table 3). Given the rapid collection time of EMI imagery at this scale (~3 h to collect a 16 ha survey at a 5 m spatial resolution), noncontact data collection method, and required number of surveys to estimate the EOF coefficients (here four surveys, two wet and two dry), we anticipate that hydrogeophysical surveys [cf. *Binley et al.*, 2015] could provide critical a priori information used in future network design. Moreover, the surveys could be made after sensor placement to assess

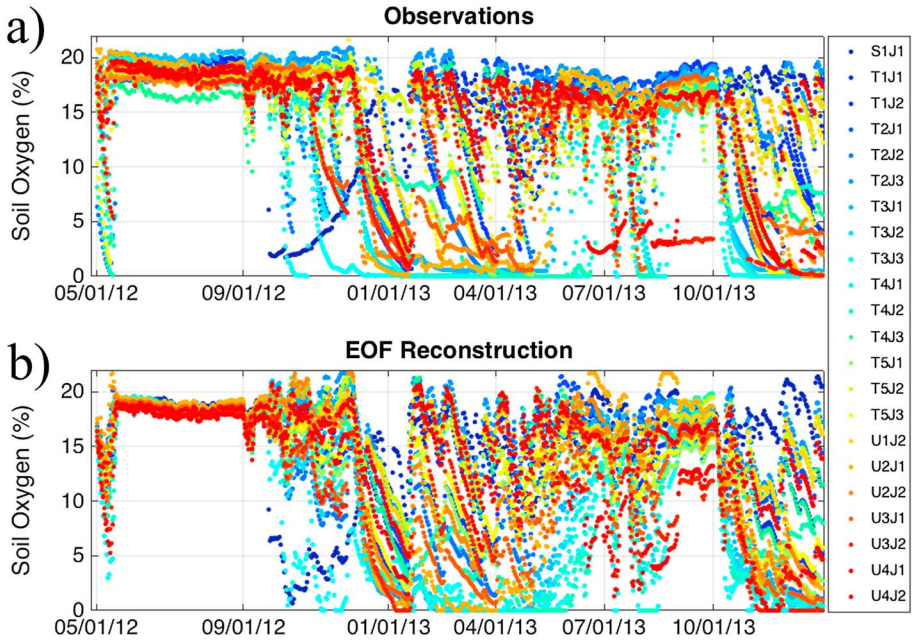

Figure 9. (a) Time series of observed soil oxygen from 22 sensor locations and (b) time series of EOF reconstructed values. See Table 2 for statistical metrics.

Table 3. Summary of Statistical Models Selected for Spatial Predictors of Selected Empirical Orthogonal Function and Expansion Coefficient (EOF/EC) Pairs[a]

	Soil Temperature (°C)			Soil Water Content (cm^3/cm^3)			Soil Oxygen (%)		
Selected Model	$z = a + bx + cx^2$	$z = a + bx + cy + dx^2 + ey^2$	$z = a + bx$	$z = a + bx + cy^2$	$z = a + bx + cy$	$z = a + bx + cy + dw$	$z = a + bx^2$	$z = a + bx + cy^2$
x	TWI	Elevation	ECa EOF2	ECa EOF2	Elevation	Elevation	ECa EOF2	TWI
y	–	TWI	–	TWI	TWI	TWI	–	ECa EOF1
w	–	–	–	–	–	ECa EOF1	–	–
z	Soil Temp EC/ EOF1	SWC EC/EOF1	SWC EC/ EOF2	SWC EC/EOF3	Soil Oxygen EC/ EOF1	Soil Oxygen EC/ EOF2	Soil Oxygen EC/ EOF3	Soil Oxygen EC/ EOF4
Coefficient								
a	7.626E − 01	−4.052E + 05	−1.000E − 04	−1.223E − 01	−2.040E + 02	−8.262E + 01	−3.309E − 02	1.274E − 01
b	−1.820E − 01	2.825E + 03	3.000E − 02	−2.460E − 02	7.119E − 01	2.864E − 01	1.619E − 03	−2.577E − 02
c	1.000E − 02	4.350E − 01	–	1.400E − 03	−2.016E − 02	4.720E − 02	–	6.08E − 04
d	–	−4.924E + 00	–	–	–	8.285E − 04	–	–
e	–	−1.606E − 02	–	–	–	–	–	–
Statistical Fit								
R^2	0.607	0.650	0.249	0.234	0.636	0.256	0.082	0.469
Validation R^2	0.515	0.510	0.061	0.043	0.524	0.105	0.000	0.246
RMSE	0.128	0.132	0.194	0.196	0.129	0.129	0.204	0.155
validation RMSE	0.156	0.220	0.217	0.226	0.152	0.213	0.221	0.187

[a]Models were selected based off minimizing RMSE. Resulting daily spatiotemporal products for soil temperature, soil water content, and soil oxygen are provided in Movies S1 to S3.

the spatial representativity of existing networks (cf. *Werbylo and Niemann* [2014] for statistical design information). With respect to this work, Figure 13 illustrates scatterplots of elevation, TWI, and ECa EOF1 versus existing sensor placement. While the two interquartiles of each environmental covariate are well represented the outer two quartiles are poorly represented by the sensor placement. This limited sampling range might help explain the one identified soil temperature outlier (Figure 10) and poor spatial models for higher-order soil water content and soil oxygen EOFs (Figures 11 and 12). In addition, our spatiotemporal predictions of soil temperature, soil water content, and soil oxygen are likely highly uncertain in these under sampled quartiles.

4.2. Spatio-Temporal Interpolation of Soil Properties and States

Figure 2 presents a detailed flowchart summarizing the data sets, statistical analysis, products, and information. The framework outlines a robust procedure for spatial interpolation of soil inventories, temporal gap filling of individual sensor data, and spatiotemporal interpolation of sensor data. The ECa EOFs provided critical and often missing information about spatial soil properties (clay and total soil carbon weight percent). The ECa EOFs were also shown to be correlated to spatiotemporal patterns of soil water content and likely to be related to soil hydraulic parameters, both of which are needed to define hydraulic diffusivity [*Hillel*, 1998; *Moldrup et al.*, 2001]. We speculate that combining spatial diffusivity maps with soil inventories will be critical to properly identify and characterize GHG hot spots and hot moments in future work. Furthermore, the composition of the GHG fluxes will be partly controlled by soil temperature and other biotic factors [*Jarecke et al.*, 2016;

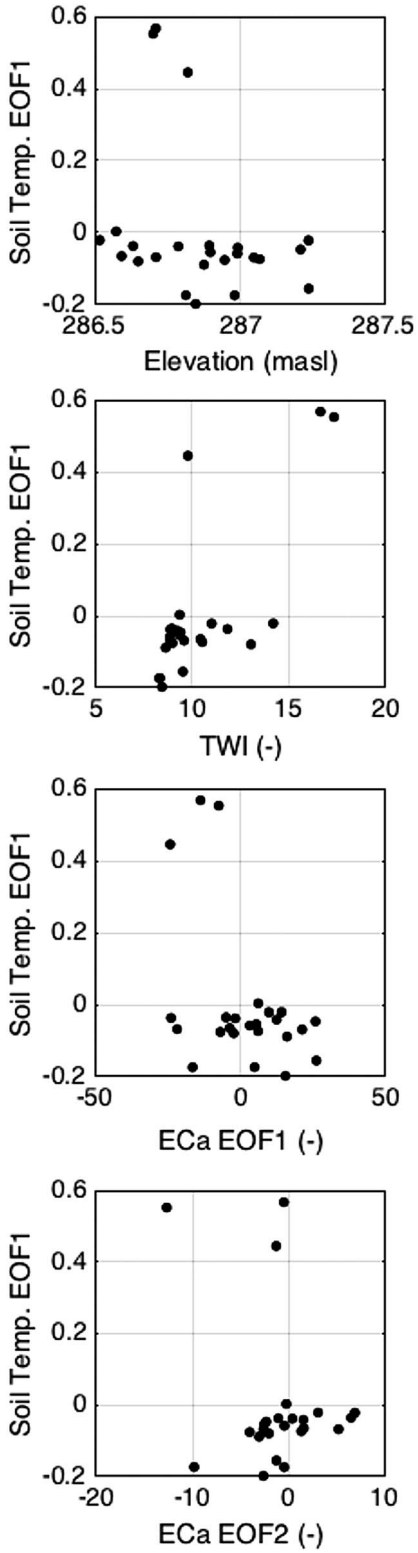

Figure 10. Scatterplots of soil temperature first empirical orthogonal function (soil temp. EOF1) versus associated sensor location's elevation, topographic wetness index (TWI), bulk electrical conductivity first empirical orthogonal function (ECa EOF1), and bulk electrical conductivity second empirical orthogonal function (ECa EOF2). See Table 3 for selected model and summary statistics.

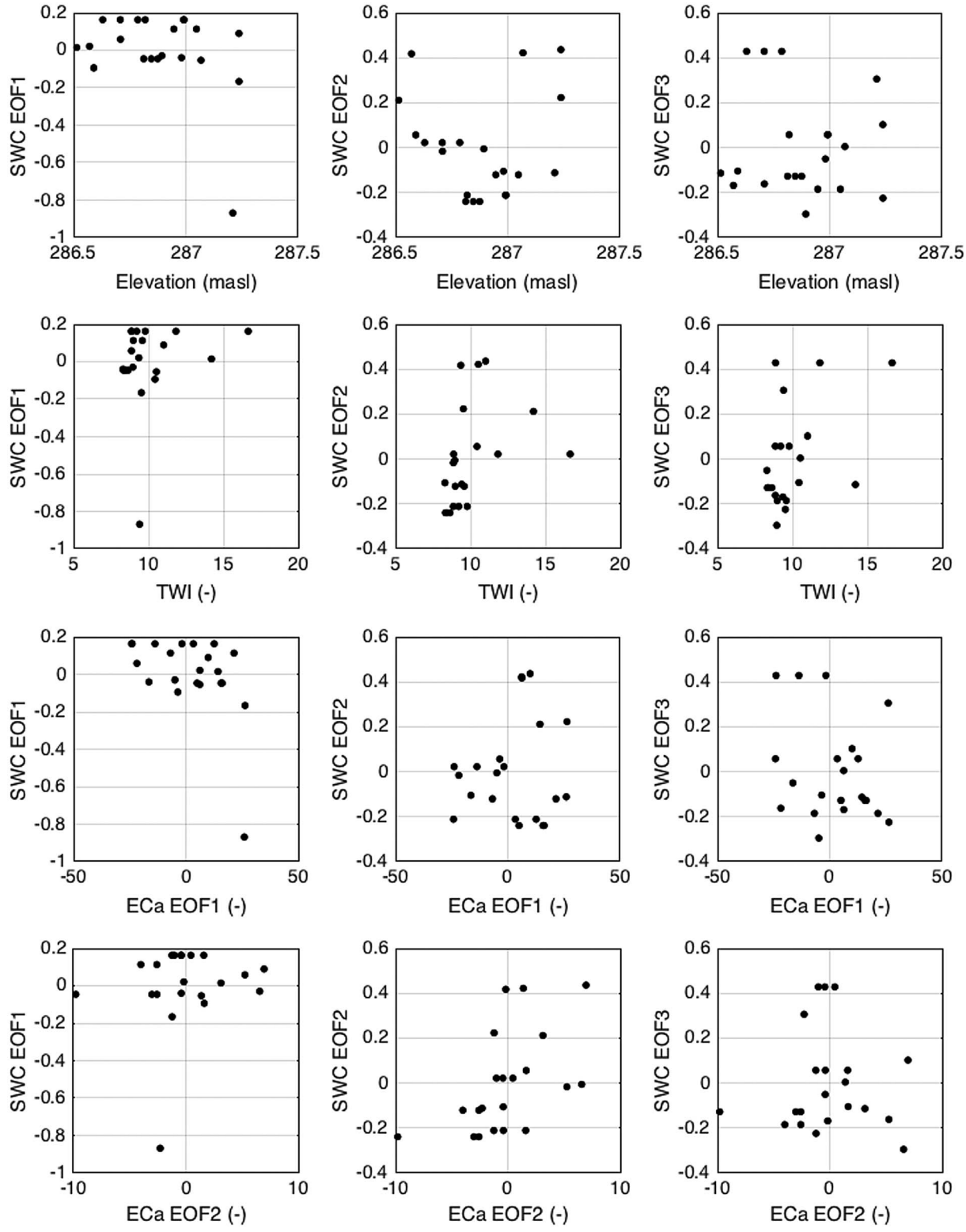

Figure 11. Scatterplots of soil water content first, second, and third empirical orthogonal functions (SWC EOF1, SWC EOF2, and SWC EOF3) versus associated sensor location's elevation, topographic wetness index (TWI), bulk electrical conductivity first empirical orthogonal function (ECa EOF1), and bulk electrical conductivity second empirical orthogonal function (ECa EOF2). See Table 3 for selected model and summary statistics.

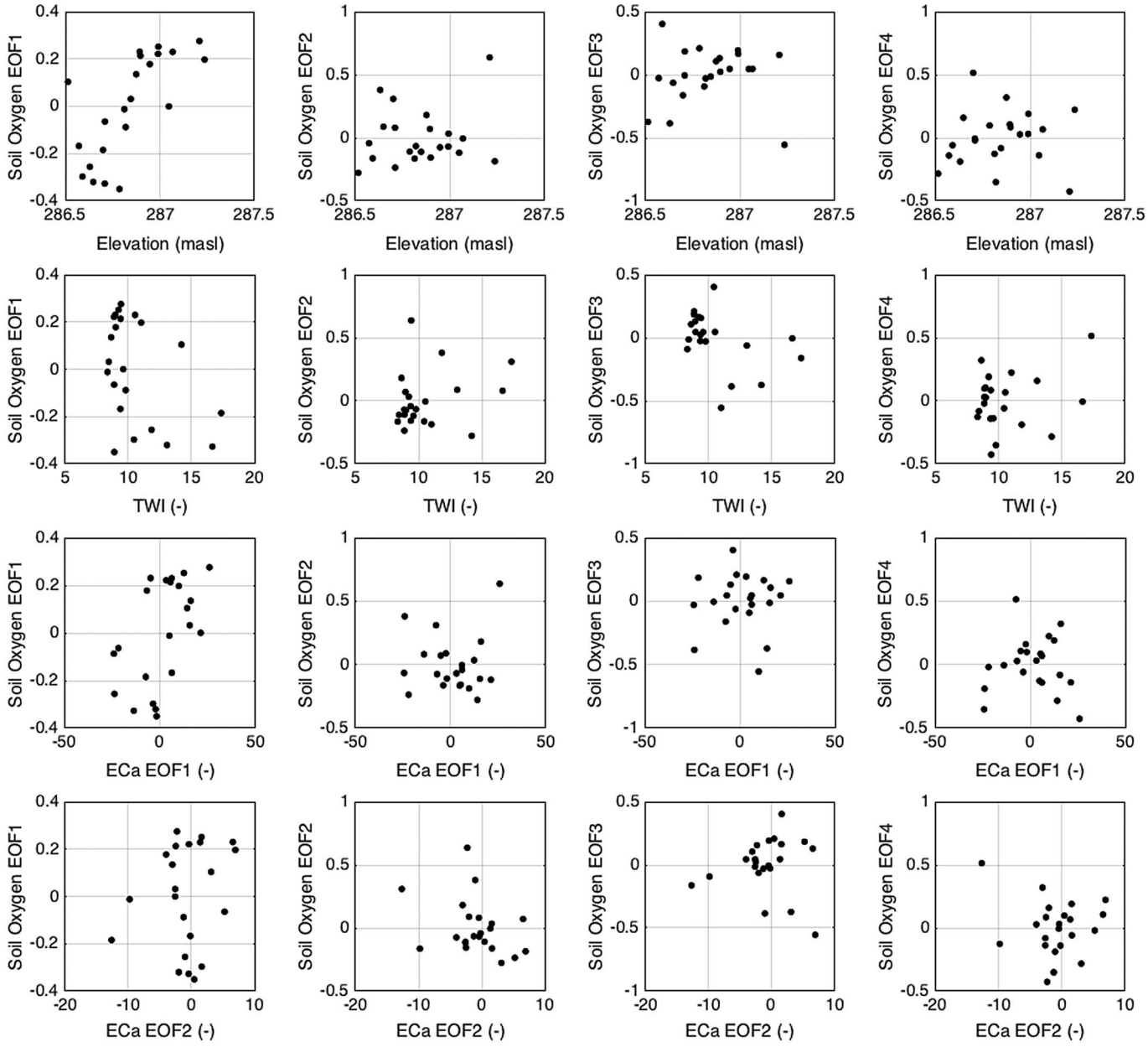

Figure 12. Scatterplots of soil oxygen first, second, third, and fourth empirical orthogonal functions (soil oxygen EOF1, SWC EOF2, and SWC EOF3) versus associated sensor location's elevation, topographic wetness index (TWI), bulk electrical conductivity first empirical orthogonal function (ECa EOF1), and bulk electrical conductivity second empirical orthogonal function (ECa EOF2). See Table 3 for selected model and summary statistics.

Rubol et al., 2013]. A critical first step toward fully understanding hot spots, hot moments, and GHG composition is providing the underlying abiotic state variables that affect the flow of heat, vapor, and liquids in the subsurface. The framework presented here provides a path forward that will be useful for many long-term monitoring networks that face similar challenges of merging spatial and temporal data.

In the era of big data and ever increasing computation resources, the next generation of hydrological and biogeochemical models will require better observational data [*Peters-Lidard et al.*, 2017]. Observational data are essential in calibrating, validating, and evaluating model predictions, thus justifying model predictions and recommendations of resource allocation. Currently, continental-scale hydrologic models have reached

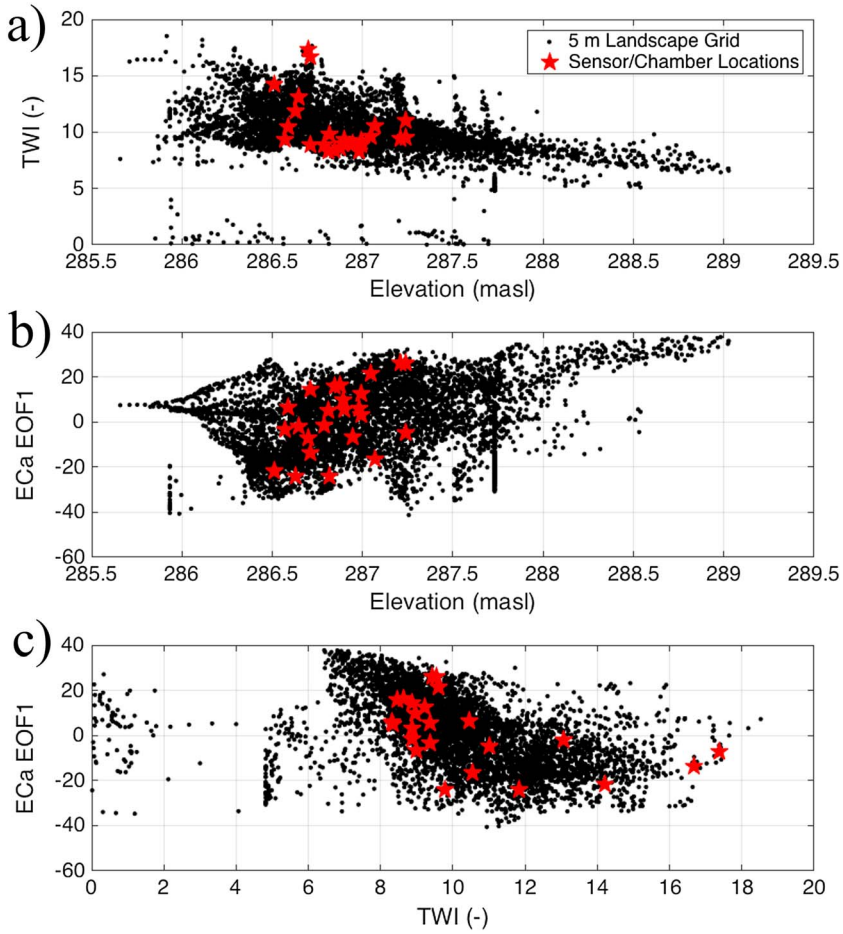

Figure 13. Location of in situ sensors/surface chamber measurements versus elevation, topographic wetness index (TWI), and bulk electrical conductivity first empirical orthogonal function (ECa EOF1).

spatial resolutions of ~30 m [*Chaney et al.*, 2016]. Unfortunately, remote sensing data sets at this spatial resolution are often limited to infrequent overpasses (cf. Landsat with 16 day overpass and 30–50 m spatial resolution). As the research community thinks about coupling of hyper-resolution hydrological models with biogeochemical models [e.g., *Rubol et al.*, 2013], robust monitoring networks are needed to compliment and properly benchmark the model output to understand model behavior and information content [*Best et al.*, 2015; *Nearing et al.*, 2016]. In addition, flexibility in both temporal and spatial monitoring is critical given the various physical and biotic processes cooccurring in surface GHG fluxes. We note that while we picked a 5 m grid for convenience in this work, the selected geometry is somewhat arbitrary and can be used to match physically based numerical model's meshes/volumes, further maximizing the data sets use in evaluating model performance.

5. Summary and Conclusions

In this work we present a framework for merging sparse sensor observations with near-surface hydrogeophysical mapping to understand complex spatiotemporal abiotic properties and states in variably saturated landscapes. The framework addresses a critical gap that exists in many long-term monitoring networks, specifically how to combine sparse in situ monitoring networks with proximal sensing at high spatial resolutions (~5 m). By using EOF analysis and regression kriging we are able to predict spatial soil inventory assessments of total soil carbon and total nitrogen. Moreover, the environmental covariates were combined with in situ sensor EOF/EC pairs to make spatiotemporal predictions of key abiotic state variables, namely, soil temperature, soil water content, and soil oxygen at high spatial (5 m) and

temporal resolutions (daily). The framework provides robust temporal gap filling of individual sensor data as well as flexible geometric interpolation to complex shapes that are needed by physically based numerical models. We anticipate that the framework, with its flexible temporal and spatial monitoring options, will provide monitoring data that are well suited for the next generation of hyper-resolution hydrologic and biogeochemical models.

Acknowledgments
This work is supported by the National Science Foundation Award 1613325 and the NASA/USDA joint program on Carbon Science (awards 2011-03007 and 2011-00829). We would also like to thank the Five Rivers Metro Park for access to the study site and Karla Jarecke for preliminary hydrogeophyscial surveys. T.E.F. would also like to acknowledge the financial support of the USDA National Institute of Food and Agriculture, Hatch project 1009760. Supporting data are included as five data sets, three text documents, and three videos in the supporting information.

References

Abdu, H., D. A. Robinson, M. Seyfried, and S. B. Jones (2008), Geophysical imaging of watershed subsurface patterns and prediction of soil texture and water holding capacity, *Water Resour. Res.*, *44*, W00D18, doi:10.1029/2008WR007043.

Altdorff, D., C. von Hebel, N. Borchard, J. van der Kruk, H. R. Bogena, H. Vereecken, and J. A. Huisman (2017), Potential of catchment-wide soil water content prediction using electromagnetic induction in a forest ecosystem, *Environ. Earth Sci.*, *76*(3), 11, doi:10.1007/s12665-016-6361-3.

Best, M. J., et al. (2015), The plumbing of land surface models: Benchmarking model performance, *J. Hydrometeorol.*, *16*(3), 1425–1442, doi:10.1175/jhm-d-14-0158.1.

Binley, A., S. S. Hubbard, J. A. Huisman, A. Revil, D. A. Robinson, K. Singha, and L. D. Slater (2015), The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales, *Water Resour. Res.*, *51*, 3837–3866, doi:10.1002/2015WR017016.

Blagodatsky, S., and P. Smith (2012), Soil physics meets soil biology: Towards better mechanistic prediction of greenhouse gas emissions from soil, *Soil Biol. Biochem.*, *47*, 78–92, doi:10.1016/j.soilbio.2011.12.015.

Chaney, N. W., P. Metcalfe, and E. F. Wood (2016), HydroBlocks: A field-scale resolving land surface model for application over continental extents, *Hydrol. Processes*, *30*(20), 3543–3559, doi:10.1002/hyp.10891.

Crow, W. T., A. A. Berg, M. H. Cosh, A. Loew, B. P. Mohanty, R. Panciera, P. de Rosnay, D. Ryu, and J. P. Walker (2012), Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products, *Rev. Geophys.*, *50*, RG2002, doi:10.1029/2011RG000372.

Entekhabi, D., et al. (2010), The Soil Moisture Active Passive (SMAP) mission, *Proc. IEEE*, *98*(5), 704–716, doi:10.1109/jproc.2010.2043918.

Fóti, S., et al. (2016), Meta-analysis of field scale spatial variability of grassland soil CO_2 efflux: Interaction of biotic and abiotic drivers, *Catena*, *143*, 78–89, doi:10.1016/j.catena.2016.03.034.

Franz, T. E., E. G. King, K. K. Caylor, and D. A. Robinson (2011), Coupling vegetation organization patterns to soil resource heterogeneity in a central Kenyan dryland using geophysical imagery, *Water Resour. Res.*, *47*, W07531, doi:10.1029/2010WR010127.

Friedman, S. P. (2005), Soil properties influencing apparent electrical conductivity: A review, *Comput. Electron. Agric.*, *46*(1–3), 45–70, doi:10.1016/j.compag.2004.11.001.

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, pp. 9–56, Oxford Univ. Press, New York.

Graf, A., M. Herbst, L. Weihermüller, J. A. Huisman, N. Prolingheuer, L. Bornemann, and H. Vereecken (2012), Analyzing spatiotemporal variability of heterotrophic soil respiration at the field scale using orthogonal functions, *Geoderma*, *181*, 91–101, doi:10.1016/j.geoderma.2012.02.016.

Hendrickx, J. M. H., B. Borchers, D. L. Corwin, S. M. Lesch, A. C. Hilgendorf, and J. Schlue (2002), Inversion of soil conductivity profiles from electromagnetic induction measurements: Theory and experimental verification, *Soil Sci. Soc. Am. J.*, *66*(3), 673–685.

Hendrickx, J. M. H., and R. G. Kachanoski (2002), Nonintrusive electromagnetic induction, in *Methods of Soil Analysis. Part 1*, edited by J. Dane and C. Topp, pp. 1297–1306, Soil Sci. Soc. of Am., Madison, Wis.

Hengl, T., et al. (2017), SoilGrids250m: Global gridded soil information based on machine learning, *PLoS One*, *12*(2), 40, doi:10.1371/journal.pone.0169748.

Hillel, D. (1998), *Environmental Soil Physics*, pp. 291–304, Academic Press, San Diego, Calif.

Jakosky, J. J. (1950), *Exploration Geophysics*, 2nd ed., pp. 1–23, Trija Publ., Los Angeles, Calif.

Jarecke, K. M., T. D. Loecke, and A. J. Burgin (2016), Coupled soil oxygen and greenhouse gas dynamics under variable hydrology, *Soil Biol. Biochem.*, *95*, 164–172, doi:10.1016/j.soilbio.2015.12.018.

Korres, W., C. N. Koyama, P. Fiener, and K. Schneider (2010), Analysis of surface soil moisture patterns in agricultural landscapes using empirical orthogonal functions, *Hydrol. Earth Syst. Sci.*, *14*(5), 751–764, doi:10.5194/hess-14-751-2010.

McClain, M. E., et al. (2003), Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems, *Ecosystems*, *6*(4), 301–312, doi:10.1007/s10021-003-0161-9.

McNeill, J. D. (1980), Electromagnetic terrain conductivity measurement at low induction numbers, Technical Note TN-6.

Mishra, U., and W. J. Riley (2012), Alaskan soil carbon stocks: Spatial variability and dependence on environmental factors, *Biogeosciences*, *9*(9), 3637–3645, doi:10.5194/bg-9-3637-2012.

Mishra, U., and W. J. Riley (2015), Scaling impacts on environmental controls and spatial heterogeneity of soil organic carbon stocks, *Biogeosciences*, *12*(13), 3993–4004, doi:10.5194/bg-12-3993-2015.

Mishra, U., R. Lal, D. S. Liu, and M. Van Meirvenne (2010), Predicting the spatial variation of the soil organic carbon pool at a regional scale, *Soil Sci. Soc. Am. J.*, *74*(3), 906–914, doi:10.2136/sssaj2009.0158.

Moldrup, P., T. Olesen, T. Komatsu, P. Schjonning, and D. E. Rolston (2001), Tortuosity, diffusivity, and permeability in the soil liquid and gaseous phases, *Soil Sci. Soc. Am. J.*, *65*(3), 613–623.

Nearing, G. S., D. M. Mocko, C. D. Peters-Lidard, S. V. Kumar, and Y. L. Xia (2016), Benchmarking NLDAS-2 soil moisture and evapotranspiration to separate uncertainty contributions, *J. Hydrometeorol.*, *17*(3), 745–759, doi:10.1175/jhm-d-15-0063.1.

Pedrera-Parrilla, A., E. Van de Vijver, M. Van Meirvenne, A. J. Espejo-Perez, J. V. Giraldez, and K. Vanderlinden (2016), Apparent electrical conductivity measurements in an olive orchard under wet and dry soil conditions: Significance for clay and soil water content mapping, *Precis. Agric.*, *17*(5), 531–545, doi:10.1007/s11119-016-9435-z.

Peres-Neto, P. R., D. A. Jackson, and K. M. Somers (2005), How many principal components? Stopping rules for determining the number of non-trivial axes revisited, *Comput. Stat. Data Anal.*, *49*(4), 974–997, doi:10.1016/j.csda.2004.06.015.

Perry, M. A., and J. D. Niemann (2007), Analysis and estimation of soil moisture at the catchment scale using EOFs, *J. Hydrol.*, *334*(3–4), 388–404, doi:10.1016/j.jhydrol.2006.10.014.

Peters-Lidard, C., M. Clark, L. Samaniego, N. E. C. Verhoest, T. van Emmerik, R. Uijlenhoet, K. Achieng, and T. E. Franz (2017), Scaling, similarity, and the fourth paradigm for hydrology, *Hydrol. Earth Syst. Sci. Discuss.*, doi:10.5194/hess-2016-695.

Robinson, D. A., I. Lebron, B. Kocar, K. Phan, M. Sampson, N. Crook, and S. Fendorf (2009), Time-lapse geophysical imaging of soil moisture dynamics in tropical deltaic soils: An aid to interpreting hydrological and geochemical processes, *Water Resour. Res.*, *45*, W00D32, doi:10.1029/2008WR006984.

Rubol, S., S. Manzoni, A. Bellin, and A. Porporato (2013), Modeling soil moisture and oxygen effects on soil biogeochemical cycles including dissimilatory nitrate reduction to ammonium (DNRA), *Adv. Water Resour.*, *62*, 106–124, doi:10.1016/j.advwatres.2013.09.016.

Soil Survey Staff, Natural Resources Conservation Service (2016), United States Department of Agriculture, Web soil survey. [Available at http://websoilsurvey.nrcs.usda.gov/, accessed in July, 2016].

Sorensen, R., U. Zinko, and J. Seibert (2006), On the calculation of the topographic wetness index: Evaluation of different methods based on field observations, *Hydrol. Earth Syst. Sci.*, *10*(1), 101–112.

Triantafilis, J., and S. M. Lesch (2005), Mapping clay content variation using electromagnetic induction techniques, *Comput. Electron. Agric.*, *46*(1–3), 203–237, doi:10.1016/j.compag.2004.11.006.

Vitharana, U. W. A., U. Mishra, J. D. Jastrow, R. Matamala, and Z. Fan (2017), Observational needs for estimating Alaskan soil carbon stocks under current and future climate, *J. Geophys. Res. Biogeosci.*, *122*, 415–429, doi:10.1002/2016JG003421.

Werbylo, K. L., and J. D. Niemann (2014), Evaluation of sampling techniques to characterize topographically-dependent variability for soil moisture downscaling, *J. Hydrol.*, *516*, 304–316, doi:10.1016/j.jhydrol.2014.01.030.

Western, A. W., and G. Bloschl (1999), On the spatial scaling of soil moisture, *J. Hydrol.*, *217*(3–4), 203–224, doi:10.1016/s0022-1694(98)00232-7.