Validation of Sensor-Directed Spatial Simulated Annealing Soil Sampling Strategy

Elia Scudiero,* Scott M. Lesch, and Dennis L. Corwin

Abstract

Soil spatial variability has a profound influence on most agronomic and environmental processes at field and landscape scales, including site-specific management, vadose zone hydrology and transport, and soil quality. Mobile sensors are a practical means of mapping spatial variability because their measurements serve as a proxy for many soil properties, provided a sensor–soil calibration is conducted. A viable means of calibrating sensor measurements over soil properties is through linear regression modeling of sensor and target property data. In the present study, two sensor-directed, model-based, sampling scheme delineation methods were compared to validate recent applications of soil apparent electrical conductivity (EC) directed spatial simulated annealing against the more established EC–directed response surface sampling design (RSSD) approach. A 6.8-ha study area near San Jacinto, CA, was surveyed for EC and 30 soil sampling locations per sampling strategy were selected. Spatial simulated annealing and RSSD were compared for sensor calibration to a target soil property (i.e., salinity) and for evenness of spatial coverage of the study area, which is beneficial for mapping nontarget soil properties (i.e., those not correlated with EC). The results indicate that the linear modeling EC–salinity calibrations obtained from the two sampling schemes provided salinity maps characterized by similar errors. The maps of nontarget soil properties show similar errors across sampling strategies. The Spatial Simulated Annealing methodology is, therefore, validated, and its use in agronomic and environmental soil science applications is justified.

Core Ideas

- Sensor-directed sampling is valuable for mapping soil properties using few samples.
- Apparent electrical conductivity is a good proxy for soil salinity.
- Spatial simulated annealing and response surface design sampling are compared.
- EC-directed SSA is a valuable sampling approach for soil science applications.

The characterization of spatial variability of soil properties is a major area of interest in environmental and agronomic sciences because of its wide implications on most field- and landscape-scale processes (Corwin and Lesch, 2010). Indeed, soil spatial variability has great influence on the effectiveness of environmental practices, such as reclamation of saline-sodic soil by controlled leaching (Corwin et al., 2006) and field-scale solute transport (Corwin et al., 1999), and of agronomic practices such as seeding, fertilization, and irrigation (Zhang et al., 2002).

The most convenient way of representing spatial variability is through intensive geospatial surveys of on-the-go soil sensor data (Adamchuk et al., 2004). According to the sensor used, data can be used as a proxy for a variety of soil properties. In particular, measurements of apparent electrical conductivity (EC) are the most common geophysical measurement used for soil spatial characterization (Adamchuk et al., 2004; Corwin and Lesch, 2005). According to the geographical location where EC is measured, it can be a proxy for a wide range of soil properties, including soil salinity, water content, and texture (Corwin and Lesch, 2005). To map a specific (target) soil property, the EC readings require a calibration against actual soil analyses (Lesch and Corwin, 2008).

To minimize the number of soil samples needed for calibration, soil sampling locations are established from the spatial variability of the on-the-go sensor data. One of the best established sampling designs is the Response Surface Sampling Design (R SSD) (Lesch et al., 2000; Lesch et al., 1995; Lesch, 2005), which selects soil sample sites that best represent the frequency statistics of the soil sensor data and, at the same time, physically separates the sampling locations as far apart as possible to increase the chances of meeting the independent error assumption of ordinary linear regression modeling. In the United States, the RSSD is endorsed by the American Society of Civil Engineers (Corwin et al., 2012; Lesch, 2012) and by the Soil Science Society of America (Corwin and Scudiero, 2016). The
RSSD has been tested on a variety of sensors, including EC, (Lesch, 2005), remote sensing imagery (Fitzgerald et al., 2006), and radar data (Guo et al., 2016). The RSSD has previously been validated against classical, non–model-based designs, such as stratified random sampling (Corwin et al., 2010) and grid sampling (Guo et al., 2016), showing better performances for the calibration of soil sensor data with soil properties.

Among other sensor-directed sampling delineation methods that use soil sensor data (e.g., the Latin hypercube approach by Minasny and McBratney [2006]), recent attention has been given to the model-based optimization of sampling schemes using soil sensor data and Spatial Simulated Annealing (SSA). Sensor-directed SSA soil sampling can be used in a wide range of environmental studies from mineral exploration (Debba et al., 2005) to agronomic (Barca et al., 2015; Castrignanò et al., 2008; Scudiero et al., 2011) applications. The SSA algorithm for soil sampling delineation was originally presented by Van Groenigen and Stein (1998) and Van Groenigen et al. (1999, 2000). In practice, SSA can evenly fill an area of interest (similarly to RSSD) and, concurrently, can increase the coverage of the sampling scheme over areas of interest by means of a spatial weight matrix. The integration of EC (or other sensor) data is done by mapping local spatial variability with GIS tools such as a gradient map (i.e., a measure of the maximum difference between a cell and its contiguous neighbors), as shown by Barca et al. (2015), Castrignanò et al. (2008), and Scudiero et al. (2011).

The objective of this study was to validate the SSA approach against the more widely used RSSD. To do so, a 6.8-ha study area was surveyed for EC, and 30 soil sampling locations per sampling strategy were selected. The sampling strategies were compared for their ability to assure the best sensor calibration to a target soil property (i.e., salinity) and for their ability to provide even spatial coverage of the area, which is beneficial for properly mapping nontarget soil properties (i.e., those not correlated with soil sensor used).

Materials and Methods
Study Site

The study site (33°50’25.43” N, 117°00’14.93” W) is a 6.8-ha field located on Scott Brothers’ Dairy Farm in San Jacinto, CA (Fig. 1). The field is described in previous publications (Corwin et al., 2010; Huang et al., 2016). The field was selected because it is characterized by remarkable variation in soil properties over short distance ranges.

Soil Apparent Electrical Conductivity Survey

Geospatial EC measurements were obtained with the Geonics EM38 dual-dipole electrical conductivity meter (Geonics Ltd.) connected to a differential global positioning system. The EC survey followed the detailed survey protocols outlined by Corwin and Lesch (2005). The EC survey was conducted on 18 Apr. 2014. The survey consisted of geospatial EC measurements taken with mobile electromagnetic induction equipment where measurements were simultaneously taken both in the horizontal (EM, 0–0.75 m) and vertical coil configurations (EM, 0–1.5 m) every 2 s. Measurements were taken at 8995 locations on transects running in a north-south direction.

Fig. 1. Maps of study area in San Jacinto, CA. Spatial Simulated Annealing (square), Response Surface Sampling Design (triangle), and independent validation (star) sampling locations are shown.

Soil Sampling Optimization and Soil Laboratory Analyses

The SSA sampling strategy was compared with the RSSD. In the days after the EC survey, undisturbed soil cores were obtained at 60 locations (30 locations per design) at 0.3-m increments down to 1.8 m. Bulk density (g cm⁻³), gravimetric water content (WC; g g⁻¹), and gravel percentage (%) were determined. The soil was air-dried and ground to pass a 2-mm sieve.
Soil was wetted to saturation, and saturation percentage (SP; %) was recorded. The water extracted from the saturated soil paste was then analyzed for electrical conductivity (EC$_c$; dS·m$^{-1}$) and pH. The EC$_c$ is a measure of the amount of soluble ions in the soil (i.e., soil salinity). At the site, SP is a good measure of soil texture, showing $R^2$ = 0.83 for sand, $R^2$ = 0.90 for silt, and $R^2$ = 0.71 for clay (Corwin and Ahmad, 2015). The methods used for the soil analyses can be found in Dane and Topp (2002) and Sparks (1996).

Spatial Simulated Annealing

A key concept in SSA (Van Groenigen and Stein, 1998; Van Groenigen et al., 1999, 2000) is the minimization of a selected sampling optimization fitness function $\Phi(S)$ to its global minimum. Random perturbations of an initial set of sampling locations describing $\Phi(S)$ allow reducing its size (i.e., increasing the fitness). The Metropolis criterion (Metropolis et al., 1953) is used to determine a probability of acceptance for the perturbations not improving the fitness of $\Phi(S)$ to avoid the selection of local minima of the function as optimal sampling scheme. As the SSA optimization progresses, the $\Phi(S)$ gradually increases until the sampling scheme “freezes” at its optimal configuration. Details on the procedure can be found in Van Groenigen and Stein (1998).

In this paper, the selected sampling optimization fitness function accounts for field shape and uneven distribution of soil properties (Van Groenigen et al., 2000). This was done by defining $\Phi(S)$ with the Minimization of the Weighted Means of the Shortest Distances (MWMSD) criterion (see Van Groenigen et al. [2000] for detailed formulation of the MWMSD criterion). The MWMSD criterion aims to allocate sampling points in areas of expected maximum variations (Barca et al., 2015) while spreading such points as far apart as possible (i.e., minimizing the Euclidean distance of any arbitrary unsampled location to the nearest sampled location). The spatial variations of EC$_c$ data can be chosen as a weight for the MWMSD criterion (Barca et al., 2015; Castrignanò et al., 2008; Scudiero et al., 2011). Specifically, the gradient map of EC$_c$ was used as the spatial weight. The gradient map measures the maximum differences (in degrees) between a target cell and its contiguous neighbors (i.e., a high degree of change between neighboring cells indicates high local spatial variability). The EC$_c$ gradient map was obtained from the principal component analysis (PCA) of the EM$_h$ and EM$_m$ maps, with ArcMap 10.2. According to the analysis of Eigen values, the first component of the PCA expressed 94.1% of the variance of the EM$_h$ and EM$_m$ maps. The first component had $R^2$ = 0.88 for EM$_h$ and $R^2$ = 0.97 for EM$_m$. The PCA (first component only) and PCA-gradient maps are provided in Supplemental Fig. S2 and S3a, respectively.

The freeware SANOS (Van Groenigen and Stein, 1998) was used to optimize the SSA sampling scheme. To assure that the best fitness for the annealing optimization was reached, the procedure was repeated seven times, with a 24-h maximum computing time, using different initial seeds.

Response Surface Sampling Design

The RSSD is briefly described (for more detail, refer to Lesch [2005]). The RSSD is tailored to optimize the calibration of intensively surveyed soil sensor data to actual soil analyses of target properties using ordinary least square (OLS) linear regression modeling. The RSSD aims to select a limited number of soil samples ($m$) to optimize the estimation of the regression parameter and to eliminate or minimize the effects of the spatially dependent error structure on this estimation process. The RSSD optimization is a multistep procedure. In Step 1, the sensor data are transformed into a centered, scaled, and decorrelated matrix via PCA; in Step 2, a traditional response surface approach (Myers et al., 2009) is selected to, theoretically, facilitate the optimal parameterization of the model associated with the PCA matrix; in Step 3, an initial set of $q$ “candidate” locations is selected ($q > m$) from the PCA matrix that best matches the $m$ design levels specified by the response surface design from Step 2; and in Step 4, an interactive algorithm, aiming to maximize some function of the minimum separation (Euclidean) distance between adjacent site locations, is used to select a set of $m$ sampling locations from the $q$ candidate sites. Step 4 is necessary to minimize (or eliminate) possible short-range residual correlation effects and to support the use of OLS modeling.

The freeware ESAP (Lesch et al., 2000) was used to optimize the RSSD sampling scheme.

Data Analysis

Features of the two sampling scheme designs were compared, including their ability of assuring the best sensor calibration to a target property and their ability to provide even spatial cover of the study area for properly mapping nontarget soil properties correlated with EC$_c$.

ArcMap 10.2 nearest neighbor statistics were performed, calculating the average nearest neighbor distance and the average nearest neighbor ratio (ANNR) (see Mitchell [2005] for details on the calculation and significance-testing of the ANNR). Practically speaking, a statistically significant ANNR <1 indicates clustering, whereas when ANNR is >1 the spatial dataset tends toward dispersion.

The frequency statistics of the EM$_h$ and EM$_m$ data across the entire site were analyzed and compared with those at the sampling sites. Additionally, the Anselin Local Moran’s I statistics (Anselin, 1995) tool in ArcMap 10.2 was used to identify clusters of low and high values in the EM$_h$ and EM$_m$ data. A 5-m threshold distance for local neighborhood selection was used. See Mitchell (2005) for details about the calculation and significance-testing of the Anselin Local Moran’s I.

Linear regressions were used to calibrate the EC$_c$ data to the measurements of soil salinity (EC$_s$) (Lesch and Corwin, 2008). In particular, because EC$_c$ data are a multiplicative function of salinity, soil tortuosity (depending on soil texture, density and particle geometry, particle pore distribution, and organic matter content), and water content (Archie, 1942; Corwin and Lesch, 2014), the following regression modeling was used:

$$EC_c = \beta \times EM^\alpha \times \varepsilon^*$$

where $\alpha$ and $\beta$ are coefficients that take into account the effects of soil tortuosity and water content, and $\varepsilon^*$ is a random (multiplicative) error component (in dS·m$^{-1}$). In Eq. [2], the error component is a multiplicative factor defined as the ratio between EC$_c$ and the explanatory term of the equation (Tian et al., 2013). Once a natural logarithm transformation is performed on
Eq. [2], the multiplicative nature of the EC_a–EC_b relationship becomes additive:

$$\ln(EC_a) = \ln(\beta) + \alpha \times \ln(EM) + \varepsilon \tag{3}$$

where $\varepsilon$ is a random (additive) error component, equal to $\ln(\varepsilon^*)$.

Equation [3] can be parameterized using an OLS approach, provided the assumption that the residuals are normally distributed and spatial independence is respected (Lesch and Corwin, 2008). Equation [3] was parameterized with a backward stepwise parameter selection procedure with Statistica 12 (StatSoft Inc.), using EM_a and EM_b as independent variables. The regression residuals were examined for spatial autocorrelation (Moran’s I Test for Residual Spatial Autocorrelation) (Cliff and Ord, 1981). The regression models showing a significant residual spatial autocorrelation were then recalculated with the maximum likelihood approach (Lesch and Corwin, 2008) to avoid biased parameter estimation (Cressie, 1993) using the spdep library (Bivand et al., 2011) in R 2.15.1 (R Development Core Team, 2012).

As pointed out by Lesch (2005), sensor-directed soil sampling scheme selection is useful for mapping soil properties that sensibly influence the measurements of the selected sensor. Maps of other soil variables are likely to be biased by the sampling scheme because their variance is unlikely to be properly represented across the study site. The two sampling designs were tested for their use in mapping soil properties not correlated with EC_a. For each sampling scheme, soil maps (for selected soil properties) were created with inverse distance weighting (IDW) of the 30 soil samples.

Maps Quality Assessment

For an independent quality assessment of the maps produced using the datasets from the two sampling schemes, 16 additional soil samples were used (Fig. 1). The 16 validation locations were sampled in August 2014. Soil sampling is discussed by Huang et al. (2016). The validation locations were sampled at 0.3-m intervals down to 1.5 m. The soil samples were analyzed for EC_a, pH, WC, and SP. These sampling locations were used to independently test the quality of the EC_a maps obtained from Eq. [3] and the inverse distance weighting interpolations of SP and pH. The independent validation was possible because the selected properties (i.e., EC_a, SP, and pH) were expected to show little to no change in the 4 mo separating the two soil sampling campaigns because no changes had occurred at the site to cause perturbations to the system.

Results and Discussion

Soil Apparent Electrical Conductivity Data

The EC_a measurements in the two EM38 dipole orientations showed very similar spatial variability (Pearson $r = 0.86$). The two variables showed a variance inflation factor (i.e., $1 \times (1 - R^2)^{-1}$) of 3.8 (4, when ln transformed); consequently, they were not considered collinear (O’Brien, 2007). The EM_a values were slightly smaller than those of EM_b, with an average of 0.35 and 0.43, a minimum of 0.16 and 0.24, and a maximum of 0.70 and 0.89 dS m$^{-1}$, respectively. The EM_a and EM_b data had similar standard deviations (0.08 and 0.10, respectively) and were slightly skewed (skewness of 0.74 and 0.79, respectively) (Fig. 2).

Overview on the Soil Sampling Schemes

The two sampling schemes are reported in Fig. 1 and 2. According to the average nearest neighbor statistics, both the SSA and RSSD sampling schemes were significantly ($p << 0.001$) dispersed, with ANNR of 1.95 and 1.45, respectively. This indicates that both designs performed well in maximizing the distance between sampling points. When using OLS linear regression modeling, the spacing between sampling points should be maximized to ensure the best chance that the independent error assumption is adequately met (Lesch, 2005). In particular, the SSA performed better than the RSSD under this aspect, with an average nearest neighbor distance of 37.3 m (30.8 m for RSSD).

As depicted in the histograms in Fig. 2, the frequency distributions of the sampling designs were very similar for EM_a and EM_b. Indeed, the (two-sample) Kolmogorov–Smirnov did not indicate significant differences between the two datasets. However, the RSSD dataset was slightly more skewed than the SSA dataset for EM_a (skewness = 0.35 for RSSD and 0.16 for SSA) and EM_b (skewness = 0.32 for RSSD and 0.24 for SSA). Previous work by Johnson et al. (2005) and Guo et al. (2016) reported that the RSSD favors the identification of sampling locations at extremes of the used sensor measurements, but this was not the case in this study, as shown by the histograms in Fig. 2.

As depicted in the Anselin Local Moran’s I maps (Fig. 2), most of the SSA samples were located at the edge of clusters of high and low EC_a values. As discussed by Scudiero et al. (2011), the use of a EC_a gradient map as spatial weight in the SSA optimization allows for intensification of sampling locations in areas of high local heterogeneity, which should ease spatial modeling of the target soil variable (Castrignanò et al., 2008). The gradient map is, in essence, similar to the local variance map (Woodcock and Strahler, 1987), which is often used in remote sensing as a measure of map spatial variability (in terms of variance). As shown by Woodcock and Strahler (1987), local heterogeneity is a function of the selected map support: significant changes in the best fitness for the MWMSD criterion should, therefore, be expected when gradient maps were to be created at different supports. The effect of block support on the EC_a gradient maps is shown in Supplemental Fig. S3. In contrast to SSA, the RSSD sampling allocation is based on the original sensor data matrix and therefore is not biased by the selected map support.

Mapping Soil Salinity

At the study site, the average vertical profile of soil salinity (Table 1) is a common salinity profile for irrigated agriculture where salinity has been leached from the upper profile, resulting in lower salinity in the top soil and increasing salinity with depth in the lower portion of the root zone where the greater portion of plant roots are located (at the study site, between 0.3 and 1.2 m), a peak in salinity just below the root zone, and lower salinity below the root zone. The bulge in salinity, which occurs at the lower portion of the root zone and just below the root zone, reflects salts that have been leached from the root zone and have increased in concentration due to the removal of water by the plant roots.

The two sampling schemes rendered salinity (EC_a) datasets with similar frequency statistics across the six sampled depth
The Kolmogorov–Smirnov test did not highlight any significant difference in the distribution of EC values between the two sampling designs at any of the considered depth increments. The similarity in frequency distribution between SSA and RSSD datasets allows for a direct comparison of the goodness-of-fit scores for the ECa–ECe calibration equations (Achen, 1982).

Table 2 presents the regression model coefficients and intercepts used to calibrate the ECa readings to observed soil salinity values, at all depths, for both sampling designs. For both sampling schemes, the OLS residual independence assumption was respected in most cases, except at 0.6 to 0.9 m and 1.5 to 1.8 m for SSA and at 0.3 to 0.6 m and 0.9 to 1.2 m for RSSD. These four regressions were corrected with the ML approach as suggested by Lesch and Corwin (2008) and returned spatially independent residuals. The regressions presented in Table 2 are significant and have normally distributed residuals.

All depth increments, except 1.2 to 1.5 m, showed the same model formulation for the two sampling schemes.

### Table 1. Frequency statistics for the measured soil salinity, down to 1.8 m depth, for the Spatial Simulated Annealing and the Response Surface Sampling Design soil sampling schemes.

<table>
<thead>
<tr>
<th>Soil sampling schemes†</th>
<th>Frequency statistics</th>
<th>EC‡</th>
<th>dS m⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0–0.3 m</td>
<td>0.3–0.6 m</td>
<td>0.6–0.9 m</td>
</tr>
<tr>
<td>SSA</td>
<td>average</td>
<td>2.76</td>
<td>5.03</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.18</td>
<td>2.70</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>6.28</td>
<td>10.77</td>
</tr>
<tr>
<td></td>
<td>minimum</td>
<td>1.46</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>skewness</td>
<td>1.51</td>
<td>0.43</td>
</tr>
<tr>
<td>RSSD</td>
<td>average</td>
<td>2.71</td>
<td>4.78</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.28</td>
<td>2.64</td>
</tr>
<tr>
<td></td>
<td>maximum</td>
<td>6.60</td>
<td>9.74</td>
</tr>
<tr>
<td></td>
<td>minimum</td>
<td>1.35</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>skewness</td>
<td>1.19</td>
<td>0.04</td>
</tr>
</tbody>
</table>

† RSSD, Response Surface Sampling Design; SSA, Spatial Simulated Annealing.
‡ Soil salinity measured as the electrical conductivity of the soil saturation extract.
all depth increments, except for the slope of EM, at 0.9 to 1.2 m, the intercept and slope coefficients for the SSA and RSSD regressions showed overlapping confidence intervals, suggesting that the EC, – EC, regressions were not significantly different across sampling designs. In general, in terms of back-transformed regression $R^2$ and RMSE and of independent validation RMSE, there was not a sampling scheme that clearly performed better than the other (Table 2). However, the RSSD regressions were characterized by lower parameter standard errors and OLS regression standard errors, which indicate better regression performance in terms of regression prediction interval size. In practice, the unsampled EC, population is expected to be slightly closer to the RSSD regression line than that of SSA.

The EC, – EC, relationships were not extremely strong ($R^2 < 0.6$). With the salinity values observed in the selected portion of the field, it is expected that the EC, readings from the EM38 would be affected not only by salinity but also by other soil properties (Corwin and Lesch, 2005). For reference, a global (i.e., comprising data from both sampling designs) natural logarithm transformed regression for the composite 0- to 1.8-m soil profile would explain 72% of the variance of EM, using EC, WC, bulk density, and pH as explanatory variables, whereas the use of EC, alone would explain only 49%.

### Mapping Soil Properties Unrelated to Apparent Electrical Conductivity

Systematic (i.e., grid-like) sampling is expected to perform better than clustered sampling for spatial interpolation when no information on the spatial variability of the target variable is available (Roberts et al., 2004; Zimmerman et al., 1999). Because of its greater average nearest neighbor distance (i.e., 37.3 m for SSA and 30.8 m for RSSD), the SSA sampling scheme was expected to render better spatial interpolations for those soil variables not correlated with EM, and EM.

For both sampling schemes, the soil properties not showing a significant relationship with EM, or with EM, were SP at 0.9 to 1.2 m and 1.2 to 1.5 m and pH at 0.3 to 0.6 m and 0.6 to 0.9 m. These soil properties were mapped using IDW. The distributions of SP and pH at the considered depth intervals did not show significant differences between the SSA and RSSD datasets (i.e., the data from the two sampling schemes were considered as samples of the same populations). The IDW interpolations were evaluated for their RMSE score during a “leave-one-out” cross-validation and for their RMSE with the 16 independent validation locations. The SP and pH interpolations obtained with the SSA data points were of similar quality to those obtained with the RSSD sampling scheme (Table 3). This evidence suggests that the bigger average-nearest-neighbor distance observed between SSA locations was not significantly more beneficial for the IDW interpolations because both of the sampling optimization designs performed well in covering the study site evenly.

### Conclusions

This study supports the use of EC,–directed sampling using a SSA sampling scheme approach. The SSA approach proved to be a powerful tool to select soil samples for linear modeling calibrations of soil sensor measurements to observed soil properties. Additional benefits of the SSA approach are the possibility of considering previously sampled locations to assure dispersed samples (Scudiero et al., 2011; Van Groenigen

---

**Table 2. Linear model calibration of soil apparent electrical conductivity readings at 0–0.75 m and 0–1.5 m to measured soil salinity fitting statistics and goodness-of-fit metrics.**

<table>
<thead>
<tr>
<th>Depth</th>
<th>Sampling design</th>
<th>Regression method</th>
<th>Intercept</th>
<th>$\hat{E}_n$</th>
<th>$\hat{E}_b$</th>
<th>$R^2$</th>
<th>Regression RMSE</th>
<th>Independent validation RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–0.3 m</td>
<td>SSA</td>
<td>OLS</td>
<td>1.76 (0.27)**</td>
<td>2.12 (0.67)**</td>
<td>-1.69 (0.65)*</td>
<td>0.30</td>
<td>0.99</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>RSSD</td>
<td>OLS</td>
<td>1.90 (0.23)**</td>
<td>2.20 (0.42)**</td>
<td>-1.65 (0.43)**</td>
<td>0.50</td>
<td>0.91</td>
<td>1.06</td>
</tr>
<tr>
<td>0.3–0.6 m</td>
<td>SSA</td>
<td>OLS</td>
<td>3.19 (0.42)**</td>
<td>3.78 (1.02)**</td>
<td>-2.63 (0.99)*</td>
<td>0.34</td>
<td>2.21</td>
<td>2.67</td>
</tr>
<tr>
<td></td>
<td>RSSD</td>
<td>ML</td>
<td>3.26 (0.34)**</td>
<td>3.19 (0.63)**</td>
<td>-1.89 (0.66)**</td>
<td>0.56</td>
<td>4.06</td>
<td>2.46</td>
</tr>
<tr>
<td>0.6–0.9 m</td>
<td>SSA</td>
<td>ML</td>
<td>2.67 (0.43)**</td>
<td>1.49 (0.40)**</td>
<td>NS#</td>
<td>0.26</td>
<td>3.53</td>
<td>2.82</td>
</tr>
<tr>
<td></td>
<td>RSSD</td>
<td>OLS</td>
<td>2.76 (0.35)**</td>
<td>1.42 (0.31)**</td>
<td>NS</td>
<td>0.26</td>
<td>1.86</td>
<td>2.46</td>
</tr>
<tr>
<td>0.9–1.2 m</td>
<td>SSA</td>
<td>OLS</td>
<td>3.02 (0.37)**</td>
<td>1.83 (0.36)**</td>
<td>NS</td>
<td>0.44</td>
<td>1.77</td>
<td>1.82</td>
</tr>
<tr>
<td></td>
<td>RSSD</td>
<td>ML</td>
<td>2.47 (0.30)**</td>
<td>1.08 (0.21)**</td>
<td>NS</td>
<td>0.34</td>
<td>3.28</td>
<td>1.59</td>
</tr>
<tr>
<td>1.2–1.5 m</td>
<td>SSA</td>
<td>OLS</td>
<td>2.71 (0.37)**</td>
<td>1.55 (0.36)**</td>
<td>NS</td>
<td>0.34</td>
<td>1.58</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>RSSD</td>
<td>OLS</td>
<td>2.14 (0.24)**</td>
<td>NS</td>
<td>1.20 (0.27)**</td>
<td>0.27</td>
<td>1.54</td>
<td>1.83</td>
</tr>
<tr>
<td>1.5–1.8 m</td>
<td>SSA</td>
<td>ML</td>
<td>2.07 (0.40)**</td>
<td>NS</td>
<td>1.42 (0.47)**</td>
<td>0.19</td>
<td>3.20</td>
<td>n.a.††</td>
</tr>
<tr>
<td></td>
<td>RSSD</td>
<td>OLS</td>
<td>1.93 (0.28)**</td>
<td>NS</td>
<td>1.12 (0.31)**</td>
<td>0.15</td>
<td>1.40</td>
<td>n.a.††</td>
</tr>
</tbody>
</table>

* Significant at the 0.05 probability level.
** Significant at the 0.01 probability level.
*** Significant at the 0.001 probability level.
† RSSD, Response Surface Sampling Design; SSA, Spatial Simulated Annealing.
‡ ML, maximum-likelihood; OLS, ordinary least square.
§ Regression $R^2$ and RMSE and of independent validation RMSE, there was not a sampling scheme that clearly performed better than the other (Table 2). However, the RSSD regressions were characterized by lower parameter standard errors and OLS regression standard errors, which indicate better regression performance in terms of regression prediction interval size. In practice, the unsampled EC, population is expected to be slightly closer to the RSSD regression line than that of SSA.

The EC, – EC, relationships were not extremely strong ($R^2 < 0.6$). With the salinity values observed in the selected portion of the field, it is expected that the EC, readings from the EM38 would be affected not only by salinity but also by other soil properties (Corwin and Lesch, 2005). For reference, a global (i.e., comprising data from both sampling designs) natural logarithm transformed regression for the composite 0- to 1.8-m soil profile would explain 72% of the variance of EM, using EC, WC, bulk density, and pH as explanatory variables, whereas the use of EC, alone would explain only 49%.

### Mapping Soil Properties Unrelated to Apparent Electrical Conductivity

Systematic (i.e., grid-like) sampling is expected to perform better than clustered sampling for spatial interpolation when no information on the spatial variability of the target variable is available (Roberts et al., 2004; Zimmerman et al., 1999). Because of its greater average nearest neighbor distance (i.e., 37.3 m for SSA and 30.8 m for RSSD), the SSA sampling scheme was expected to render better spatial interpolations for those soil variables not correlated with EM, and EM.

For both sampling schemes, the soil properties not showing a significant relationship with EM, or with EM, were SP at 0.9 to 1.2 m and 1.2 to 1.5 m and pH at 0.3 to 0.6 m and 0.6 to 0.9 m. These soil properties were mapped using IDW. The distributions of SP and pH at the considered depth intervals did not show significant differences between the SSA and RSSD datasets (i.e., the data from the two sampling schemes were considered as samples of the same populations). The IDW interpolations were evaluated for their RMSE score during a “leave-one-out” cross-validation and for their RMSE with the 16 independent validation locations. The SP and pH interpolations obtained with the SSA data points were of similar quality to those obtained with the RSSD sampling scheme (Table 3). This evidence suggests that the bigger average-nearest-neighbor distance observed between SSA locations was not significantly more beneficial for the IDW interpolations because both of the sampling optimization designs performed well in covering the study site evenly.

### Conclusions

This study supports the use of EC,–directed sampling using a SSA sampling scheme approach. The SSA approach proved to be a powerful tool to select soil samples for linear modeling calibrations of soil sensor measurements to observed soil properties. Additional benefits of the SSA approach are the possibility of considering previously sampled locations to assure dispersed samples (Scudiero et al., 2011; Van Groenigen
et al., 2000) and to minimize kriging interpolation variance (Montanari et al., 2012; Van Groenigen et al., 1999). These options are not available for the RSSD in ESAP. A drawback of the use of sensor-directed SSA is that the gradient (or local variance) maps are a function of the selected mapping support.

Future research will focus on comparing the two sampling strategies for soil property mapping with other spatial statistics methods (e.g., co-kriging, regression kriging) other than the linear regression method used in this study and for which the RSSD was optimized and on the comparison of these two model-based, sensor-directed sampling schemes optimization methods versus other model-based (e.g., Latin hypercube methods) and sample-based (e.g., stratified random sampling) methods.

Acknowledgments

The authors thank Michael Bagtang, Wes Clary, and Kevin Yemoto for numerous hours of diligent technical work performed in the field and the laboratory and Bruce Scott of Scott Brother’s Dairy Farm for allowing unrestricted access to the study site.

References


Table 3. Quality assessment for the inverse distance weighting interpolations for soil saturation percentage (SP) and pH.

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Depth interval</th>
<th>Sampling design†</th>
<th>IDW exponential weight</th>
<th>l o o cv.$ RMSE</th>
<th>Independent validation RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP</td>
<td>0.9–1.2</td>
<td>SSA</td>
<td>4.0</td>
<td>6.02</td>
<td>4.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RSSD</td>
<td>1.1</td>
<td>5.85</td>
<td>4.19</td>
</tr>
<tr>
<td></td>
<td>1.2–1.5</td>
<td>SSA</td>
<td>1.0</td>
<td>5.41</td>
<td>3.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RSSD</td>
<td>1.9</td>
<td>5.14</td>
<td>4.33</td>
</tr>
<tr>
<td>pH</td>
<td>0.3–0.6</td>
<td>SSA</td>
<td>1.0</td>
<td>0.23</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RSSD</td>
<td>5.4</td>
<td>0.32</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>0.6–0.9</td>
<td>SSA</td>
<td>2.4</td>
<td>0.31</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RSSD</td>
<td>1.0</td>
<td>0.23</td>
<td>0.45</td>
</tr>
</tbody>
</table>

† RSSD, Response Surface Sampling Design; SSA, Spatial Simulated Annealing.
‡ Inverse distance weighting.
§ Leave-one-out cross-validation.


