Comparison of Sampling Strategies for Characterizing Spatial Variability with Apparent Soil Electrical Conductivity Directed Soil Sampling

Dennis L. Corwin[1], Scott M. Lesch[2], Eran Segal[3], Todd H. Skaggs[1] and Scott A. Bradford[1]
[1]USDA-ARS, U.S. Salinity Laboratory, 450 West Big Springs Road, Riverside, CA 92507-4617
Email: Dennis.Corwin@ars.usda.gov
[2]Riverside Public Utilities—Resource Division, 3435 14th St., Riverside, CA 92501
[3]Soil Physics and Irrigation, Gilat Research Center, Agricultural Research Organization,
Mobile Post Negev 2 85280, Israel

ABSTRACT

Spatial variability has a profound influence on a variety of landscape-scale agricultural issues including solute transport in the vadose zone, soil quality assessment, and site-specific crop management. Directed soil sampling based on geospatial measurements of apparent soil electrical conductivity (EC_a) is a potential means of characterizing the spatial variability of any soil property that influences EC_a including soil salinity, water content, texture, bulk density, organic matter, and cation exchange capacity. Arguably the most significant step in the protocols for characterizing spatial variability with EC_a-directed soil sampling is the statistical sampling design, which consists of two potential approaches: model- and design-based sampling strategies such as response surface sampling design (RSSD) and stratified random sampling design (SRSD), respectively. The primary objective of this study was to compare model- and design-based sampling strategies to evaluate if one sampling strategy outperformed the other or if both strategies were equal in performance. Using three different model validation tests, the regression equation estimated from the RSSD data produced accurate and unbiased predictions of the natural log salinity levels at the independently chosen SRSD sites. Design optimality scores (*i.e.*, D-, V-, and G-optimality criteria) indicate that the use of the RSSD design should facilitate the estimation of a more accurate regression model, *i.e.*, the RSSD approach should allow for better model discrimination, more precise parameter estimates, and smaller prediction variances. Even though a model-based sampling design, such as RSSD, has been less prevalent in the literature, it is concluded from the comparison that there is no reason to refrain from its use and in fact warrants equal consideration.

Introduction

Ever since the classic paper by Nielson *et al.* (1973) concerning the variability of field-measured soil water properties, the significance of within-field spatial variability of soil properties has been scientifically acknowledged and documented. Spatial variability of soil has been the focus of books (Bouma and Bregt, 1989; Mausbach and Wilding, 1991) and numerous comprehensive review articles (Warrick and Nielsen, 1980; Jury, 1985, 1986; White, 1988). The significance of soil spatial variability lies in the fact that it is a key component of any landscape-scale soil-related issue including solute transport in the vadose zone, site-specific crop management, and soil quality assessment.

The characterization of spatial variability is without question one of the most significant areas of concern in soil science because of its broad reaching influence on all field- and landscape-scale processes. There are a variety of methods for potentially characterizing soil spatial variability including ground penetrating radar, aerial photography, multi- and hyperspectral imagery, time domain reflectometry, and apparent soil electrical conductivity (EC_a). Although not commonly used, magnetometry is another method for potentially characterizing soil spatial variability (Rogers *et al.*, 2006). However, none of these approaches has been as extensively investigated for applications in agricultural geophysics as EC_a, which can be measured using either electrical resistivity (ER) or electromagnetic induction (EMI) (Corwin and Lesch, 2005a). The geospatial measurement of EC_a is a sensor technology that has played, and continues to play, a major role in addressing the issue of spatial variability characterization. Geospatial measurements of EC_a have been successfully used for (i) identifying the soil physical and chemical properties influencing crop yield patterns and soil condition, (ii) establishing the spatial variation of soil properties that influence the EC_a measurement, and (iii) characterizing the spatial distribution of soil properties influencing solute transport through the vadose zone (Corwin *et al.*, 1999, 2003a, 2003b, 2006; Kaffka *et al.*, 2005).

Since its early agricultural use for measuring soil salinity, the application of EC_a has evolved into a widely accepted means of establishing the spatial variability of several soil physical and chemical properties that influence the EC_a measurement (Corwin and Lesch, 2003, 2005a). Geospatial measurements of EC_a are well-suited for characterizing spatial variability for several reasons: (i) geospatial measurements of EC_a are reliable, quick, and easy to take; (ii) the mobilization of EC_a measurement equipment is easy and can be accomplished at a reasonable cost; and (iii) EC_a is influenced by a variety of soil properties for which the spatial variability of each could be potentially established. Corwin and Lesch (2005a) provide a compilation of literature pertaining to the soil physical and chemical properties that are either directly or indirectly measured by EC_a.

Because the geospatial measurement of EC_a is a complex spatially measured property of soil that reflects the influence of several soil physical and chemical properties (including soil salinity, texture, water content, bulk density, organic matter, and cation exchange capacity) it is rarely used to map a single property, but rather is used as a surrogate for general spatial variability of those soil physical and chemical properties that are spatially correlated with EC_a. As such, geospatial measurements of EC_a are used to direct soil sampling as a means of characterizing spatial variability of those soil properties that correlate with EC_a at that particular study site. Characterizing spatial variability with EC_a-directed soil sampling is based on the notion that when EC_a correlates with a soil property or properties, then spatial EC_a information can be used to identify sites that reflect the range and variability of the property or properties.

In instances where EC_a correlates with a particular soil property, an EC_a-directed soil sampling approach will establish the spatial distribution of that property with an optimum number of site locations to characterize the variability and keep labor costs minimal (Corwin *et al.*, 2003a). Details for conducting a field-scale EC_a survey for the purpose of characterizing the spatial variability of soil properties can be found in Corwin and Lesch (2005b). General guidelines appear in Corwin and Lesch (2003) and Corwin *et al.* (2003a, 2003b).

The basic elements of a field-scale EC_a survey for characterizing spatial variability include (i) EC_a survey design, (ii) geo-referenced EC_a data collection, (iii) soil sample design based on geo-referenced EC_a data, (iv) soil sample collection, (v) physico-chemical analysis of pertinent soil properties, (vi) spatial statistical analysis, (vii) determination of the dominant soil properties influencing the EC_a measurements at the study site, and (viii) GIS development (Corwin and Lesch, 2005b). Step (iii) is arguably the most critical step because it establishes the sample site locations based on the variation and magnitude of the geospatial EC_a measurements.

Currently, two EC_a-directed soil sampling design approaches are used: (i) design-based (probability) sampling and (ii) model-based (prediction) sampling. The former consists of the use of simple random, cluster, unsupervised classification, and stratified random sampling, whereas the latter typically relies on optimized spatial response surface sampling designs. Throughout the statistical literature model-based designs are less common, although some statistical research has been performed in this area (Valliant *et al.*, 2000). Nathan (1988) and Valliant *et al.* (2000) discuss the merits of design- and model-based sampling strategies in detail. Specific model-based sampling approaches, having direct application to agricultural and environmental survey work, are described by McBratney and Webster (1981), Lesch *et al.* (1995a, 1995b) Van Groenigen *et al.* (1999), and Lesch (2005). However, a comparison of the prediction results of model- and designed-based sampling has not been performed.

The objectives of this research are (i) to test statistically the validity of using a model-based sampling strategy in conjunction with ordinary regression modeling to quantify the spatial salinity (EC_e, dS m^{-1}) pattern in an agricultural field and (ii) to compare a model-based and design-based sampling strategy for purposes of estimating the ordinary linear calibration model between EC_a and EC_e.

Materials and Methods

Study Site Description

The on-farm research study site (lat. 33° 50′ 25.43″ N, long. 117° 00′ 14.93″ W) is located on Scott Brothers' Dairy Farm in San Jacinto in southern California's Riverside County (Fig. 1). The 32-ha field site provided an extensive range of spatial variability in EC_a needed to make a real-world evaluation of the sampling design comparison.

Intensive EMI Survey

Geospatial EC_a measurements were obtained with the Geonics EM38 dual-dipole electrical conductivity

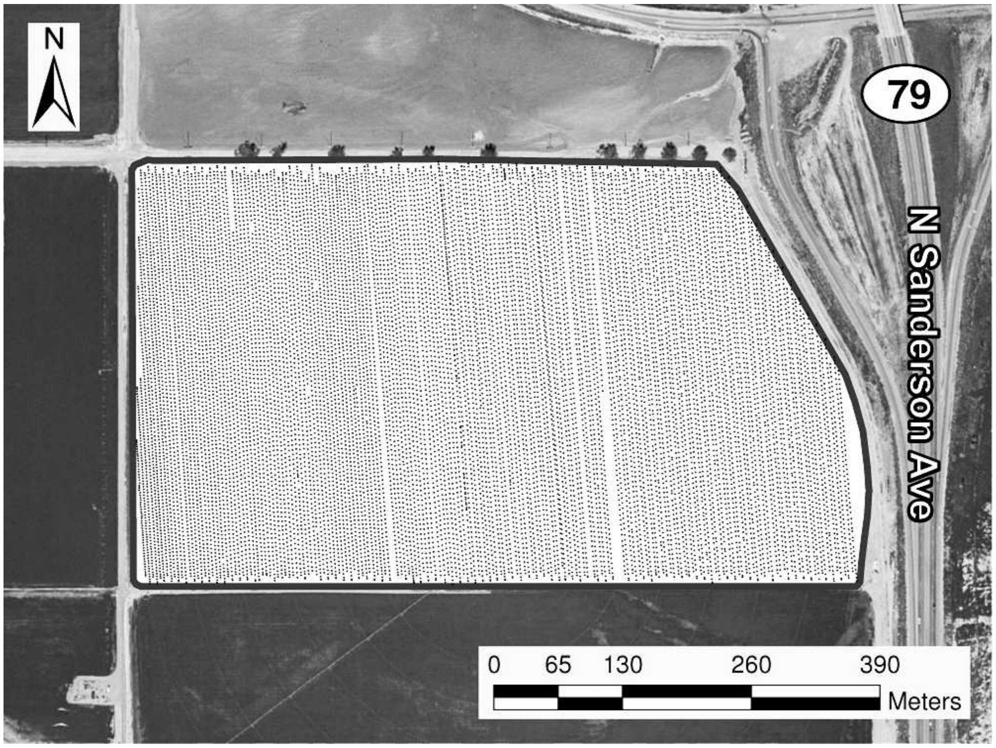

Figure 1. Scott Brothers' Dairy Farm study site located near San Jacinto, CA. Dots indicate 16,122 locations of electromagnetic induction measurements.

meter[1]. The EC_a survey followed the detailed survey protocols outlined by Corwin and Lesch (2005b). The EC_a survey was conducted 22–23 June 2006. The survey consisted of geospatial EC_a measurements taken with mobile EMI equipment where measurements were simultaneously taken both in the horizontal (EM_h) and vertical coil configurations (EM_v) every 5 m. Measurements were taken at 16,122 locations on transects running in a north-south direction as shown in Fig. 1. Table 1(a) indicates the EC_a summary statistics for all 16,122 sites.

Sampling Protocol Details

Apparent soil electrical conductivity serves as a surrogate to characterize the spatial variation of those soil properties that are found to influence EC_a within a field. Based on the variation in EC_a, soil sample sites were selected that reflect the range and variation in EC_a using a model- and design-based sampling strategy. Soil samples were collected for the following depth increments: 0–0.15, 0.15–0.30, 0.30–0.60, 0.60–0.90, 0.90–1.20, 1.20–1.50 m. Saturation extracts of the soil sample depth increments were prepared and the electrical conductivity of the saturation extracts (EC_e, dS m^{-1}) were measured using the method presented in Rhoades (1996). The depth-weighted average EC_e at each sample site was calculated over the 0–1.5 m depth using the 6 depth increments.

Both a 40-site model-based sampling plan (*i.e.*, response surface sampling design, RSSD) and a 30-site design-based site sampling plan (*i.e.*, stratified random sampling design, SRSD) were used to generate the full 70-site design. Table 1(b) indicates the summary statistics for the soil salinity (EC_e) for all 70 locations (*i.e.*, 40 RSSD sites and 30 SRSD sites) for composite 1.5-m core samples. The model-based sampling plan (*i.e.*, RSSD) was developed using the ESAP-RSSD software program, version 2.35 (Lesch *et al.*, 2000) and represented a composite of two 20-site RSSDs. The locations of these 40 sites are shown in Fig. 2. In principal, either of the two 20-site designs (Design A or B) can be used to estimate an ordinary linear calibration model, and two (or more) designs can be combined together in order to estimate a geostatistical mixed linear model.

The locations of the 30 SRSD sites are also shown in Fig. 2. The full SRSD is actually comprised of 20 primary sites and 10 secondary sites, where both sets of sites were selected by stratifying on blocks of sequentially acquired survey readings. Specifically, the SRSD sites were chosen by first randomly selecting one

[1] Geonics Ltd., Mississaugua, Ontario, Canada. Product identification is provided for the benefit of the reader and does not imply endorsement by USDA.

Table 1. Basic summary statistics of (a) apparent soil electrical conductivity (EC_a) survey data and (b) soil salinity (EC_e) samples.

(a) EC_a survey data (N = 16,122)			(b) EC_e ($dS\ m^{-1}$) samples (RSSD plans A & B: n = 40)	
	EM_h ($mS\ m^{-1}$)	EM_v ($mS\ m^{-1}$)		EC_e ($dS\ m^{-1}$)
Mean	41.81	47.40	Mean	2.06
Std. Dev	20.14	28.78	Std. Dev	1.44
Skewness	0.60	0.45	Skewness	1.29
Kurtosis	−0.35	−0.85	Min–Max	0.64–6.28
Quantiles:			(SRSD plan: n = 30)	
				EC_e (dS/m)
Minimum	10.8	7.2	Mean	1.78
1%	14.9	10.4	Std. Dev.	1.24
5%	17.3	12.5	Skewness	1.82
10%	18.9	14.0	Min–Max	0.66–5.97
25%	22.8	18.9		
Median	39.1	43.5		
75%	56.3	70.5		
90%	68.5	86.7		
95%	77.5	96.6		
99%	94.9	115.4		
Maximum	135.6	157.4		

sampling location from every 540 sequential survey locations and then randomly assigning this location into either the primary or secondary set (with 2/3's or 1/3's probability, respectively). In principal, either SRSD subdesign can also be used to estimate an ordinary linear model, although the larger 20-site (primary) design is obviously preferable because of the increased sample size.

With respect to the two research objectives, the aforementioned sampling designs were used as follows. For purposes of testing the model-based sampling strategy for accurately estimating a sensor calibration model (Objective 1), the full 40-site RSSD plan was treated as the calibration sampling plan and the full 30-site SRSD plan was used as an independent set of validation sites. In contrast, for purposes of comparing sampling strategies (Objective 2), it is necessary that all of the sampling designs contain the same number of sites. Thus, one of the two individual 20-site RSSD plans was compared and contrasted with the primary 20-site SRSD plan.

Statistical Methodology

Apparent soil electrical conductivity survey data represent just one type of ancillary sensor data that is commonly collected to help identify, quantify, and/or predict various soil or crop properties. Being spatial in nature (*i.e.*, referenced across a spatial domain), it is quite reasonable to consider some type of geostatistical modeling technique when attempting to calibrate such survey data to a specific soil or crop response variable. Numerous examples exist in the literature of geostatistical or spatial modeling approaches; the textbooks by Isaaks and Srivastava (1989), Wackernagel (1998), Webster and Oliver (2001), Schabenberger and Pierce (2002), and Schabenberger and Gotway (2005) are particularly relevant to the above mentioned calibration problem.

In addition to the commonly used geostatistical techniques like kriging with external drift or regression-kriging, ordinary linear regression models are also often employed when calibrating such data. In the mainstream statistical literature, it is well known that ordinary linear regression models represent a special case of a much more general class of models commonly known as linear regression models with spatially correlated errors (Schabenberger and Gotway, 2005), hierarchical spatial models (Banerjee *et al.*, 2004), or geostatistical mixed linear models (Haskard *et al.*, 2007). This broader class of models includes many of the geostatistical techniques familiar to soil scientists, such as universal kriging, kriging with external drift and/or regression-kriging, as well as standard statistical techniques like ordinary linear regression and analysis of covariance (ANOCOVA) models.

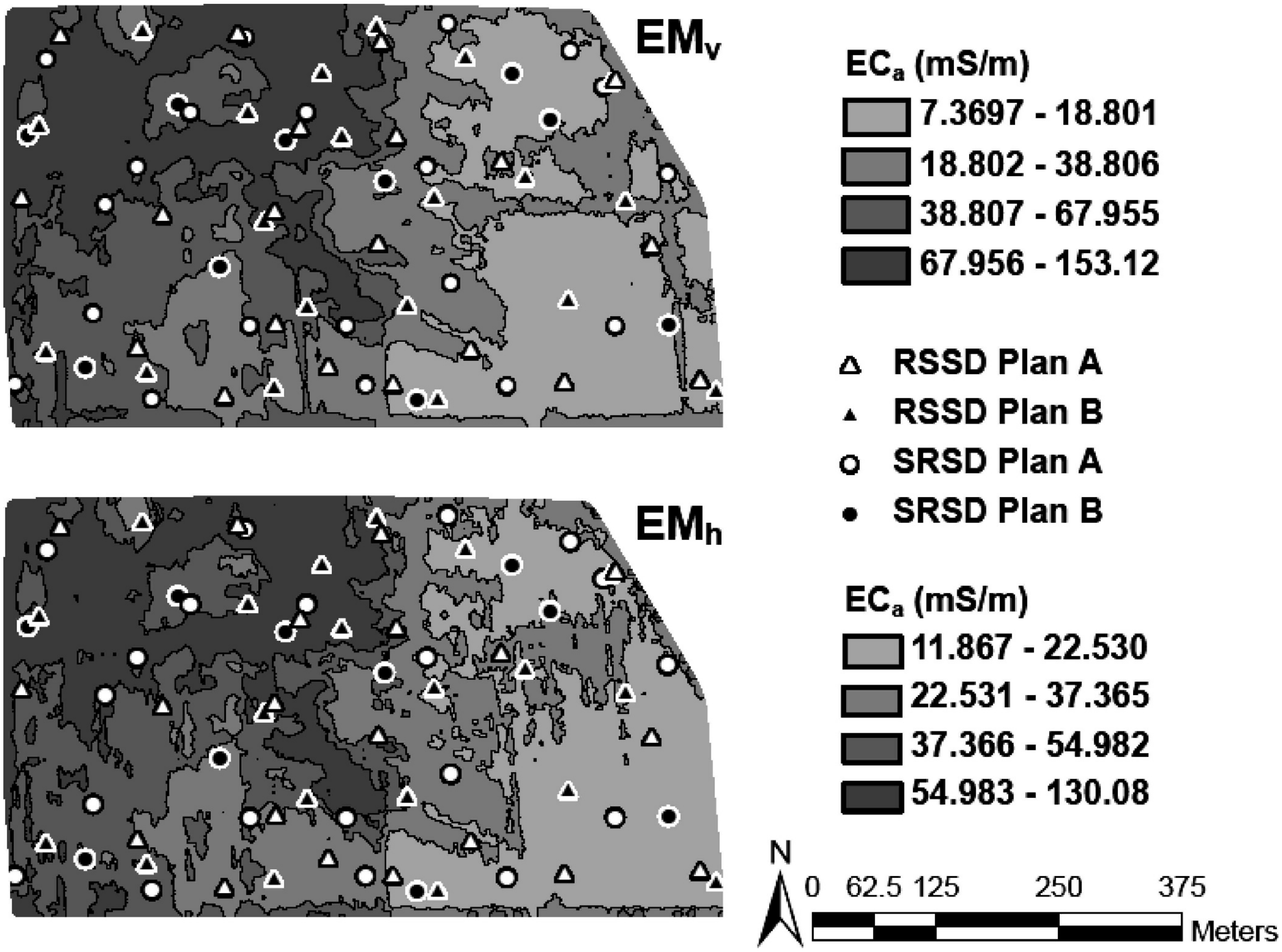

Figure 2. Model-based response surface sampling design (RSSD) plans (Designs A and B; triangles) and stratified random sampling design (SRSD) plans (Primary-A and Secondary-B; circles) overlaid upon the apparent soil electrical conductivity (EC_a) measurements taken with electromagnetic induction in the (a) vertical coil configuration (EM_v) and (b) horizontal coil configuration (EM_h).

Lesch and Corwin (2008) review the use of these different modeling techniques for calibrating remotely sensed survey data to soil properties. Lesch and Corwin (2008) also describe the necessary set of statistical assumptions for reducing a geostatistical mixed linear model to an ordinary linear model. Historically, ordinary linear models have often been used to accurately calibrate EC_a survey data to one or more target soil properties (Corwin and Lesch, 2005b). For example, field-scale soil salinity patterns are commonly mapped quite accurately using EC_a survey data and ordinary linear regression models, since the residual error distribution typically exhibits only short-range spatial correlation (Lesch and Corwin, 2008; Lesch *et al.*, 2005; Corwin and Lesch, 2005b). Therefore, a simpler linear regression model can be used in place of the full geostatistical model to generate a map with a high degree of prediction precision, provided that an appropriate sampling strategy is employed (Lesch, 2005).

To statistically address the two objectives of the study, we first discuss in Appendix A how an ordinary linear model can be derived from the more complicated geostatistical mixed linear model (see Appendix A—Geostatistical and Ordinary [Spatially Referenced] Linear Models). We then review both model-based and design-based sampling strategies (see Appendix A—Sampling Strategies for Spatially Referenced Linear Models) for estimating ordinary linear models and compare and contrast a model-based sampling design with a design-based, stratified random sampling strategy using three statistical design optimality criteria (*i.e.*, D-, V-, and G-optimality criteria) described in the Appendix A—*Sample design optimality criteria*. We also describe (see Appendix A—*Prediction validation tests for the ordinary linear model*) and employ three different model validation tests (*i.e.*, composite model F-test, joint-prediciton F-test, and mean-prediciton *t*-test) to verify that the regression equation estimated from the model-based sample data produces accurate and unbiased predictions of the natural log salinity levels at the independently chosen stratified random sample sites. We demonstrate both the model validation techniques and assessment of sampling strategies using data from a detailed soil salinity survey performed in San Jacinto, CA in 2006. All statistical analyses discussed were carried out using SAS/IML (SAS, 1999a) and SAS/STAT software (SAS, 1999b).

Model Specification

In most soil salinity surveys using EMI Geonics EM38 equipment, it is commonly observed that the natural log of the EM38 readings exhibit near linear relationships with the natural log EC_e levels, and that these natural log functions exhibit more homogeneous variance relationships (Lesch *et al.*, 2005). However, the $\ln(EM_h)$ and $\ln(EM_v)$ readings also tend to be highly correlated. To remove this multi-collinear signal effect, one can equivalently consider regressing on the first two standardized principal component scores computed from the natural log transformed EM38 readings (Lesch, 2005). Let z_1 and z_2 represent these calculated first and second principal component scores. The following four plausible natural log salinity/natural log sensor data relationships can then be specified for $y_i = \ln(EC_e)_i$:

$$y_i = \beta_0 + \beta_1(z_1)_i + \varepsilon(\mathbf{s})_i, \quad (1)$$

$$y_i = \beta_0 + \beta_1(z_1)_i + \beta_2(z_2)_i + \varepsilon(\mathbf{s})_i, \quad (2)$$

$$y_i = \beta_0 + \beta_1(z_1)_i + \beta_2(z_1^2)_i + \varepsilon(\mathbf{s})_i, \quad (3)$$

$$y_i = \beta_0 + \beta_1(z_1)_i + \beta_2(z_2)_i + \beta_3(z_1^2)_i + \varepsilon(\mathbf{s})_i. \quad (4)$$

Equation (1) relates the natural log salinity level to the collocated first principal component score, which is roughly proportional to the average value of the natural log EM38 readings. Similarly, Eq. (2) relates the natural log salinity level to both principal component scores, which is similar to regressing on the average and differenced EM38 values, respectively. Equation (3) extends Eq. (1) by allowing for a curvi-linear (quadratic) relationship between the natural log salinity and the principal component score. Likewise, Eq. (4) extends Eq. (2) in a similar manner.

In some EM38 surveys it is also not uncommon to observe a certain level of instrument drift and/or for the EM38 signal data to be simultaneously influenced by secondary soil properties that change slowly over the survey area (Robinson *et al.*, 2004). In either scenario, the accuracy of Eqs. (1)–(4) can be improved by adding first-order trend-surface parameters to the specified equations. Upon noting that $\mathbf{s}_i = \{u_{x,i}, u_{y,i}\}$ defines the coordinate locations of all the survey and sample locations, Eqs. (1)–(4) can be readily expanded to include additional coordinate location parameters; *i.e.*,

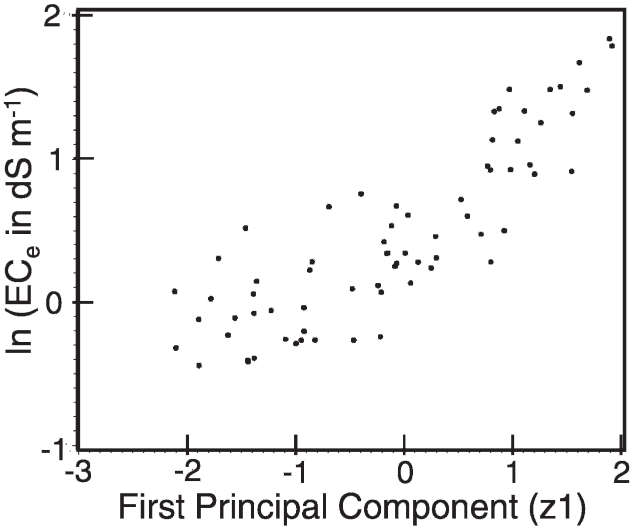

Figure 3. Relationship between the $\ln(EC_e)$ sample data and first principal component scores (as computed from the collocated EM38 signal data).

$$y_i = \beta_0 + \beta_1(z_1)_i + \beta_2(u_x)_i + \beta_3(u_y)_i + \varepsilon(\mathbf{s})_i, \quad (5)$$

$$y_i = \beta_0 + \beta_1(z_1)_i + \beta_2(z_2)_i + \beta_3(u_x)_i + \beta_4(u_y)_i + \varepsilon(\mathbf{s})_i, \quad (6)$$

$$y_i = \beta_0 + \beta_1(z_1)_i + \beta_2(z_1^2)_i + \beta_3(u_x)_i + \beta_4(u_y)_i + \varepsilon(\mathbf{s})_i, \quad (7)$$

$$y_i = \beta_0 + \beta_1(z_1)_i + \beta_2(z_2)_i + \beta_3(z_1^2)_i + \beta_4(u_x)_i + \beta_5(u_y)_i + \varepsilon(\mathbf{s})_i. \quad (8)$$

Equations (1)–(8) define eight plausible regression models that potentially can be used to describe the relationship between the natural log salinity and natural log sensor data.

Results and Discussion

Model Identification, Estimation, and Validation

Figure 3 shows the observed relationship between the natural log salinity measurements at the 70 sampling locations and the collocated first principal component scores (derived from the natural log transformed EM38 signal data). The pattern is clearly curvi-linear, suggesting that Eqs. (3), (4), (7), or (8) might represent plausible

Table 2. Regression model summary statistics and parameter estimates; Eq. (4). RSSD represents responses surface sample design and SRSD represents stratified random sample design.

Model	Sample size (n)	R^2	Root mean square error (RMSE)	F-score	Pr > F
Combined	70	0.814	0.272	96.25	<0.001
RSSD	40	0.805	0.293	49.38	<0.001
SRSD	30	0.843	0.248	46.57	<0.001
			Parameter estimates and test statistics		
	Parameter	**Estimate**	**Std. error**	**t-score**	**Pr > |t|**
Combined Sample Set (n = 70)	β_0	0.372	0.049	7.67	<0.001
	β_1	0.493	0.031	16.14	<0.001
	β_2	−0.060	0.030	−2.01	0.048
	β_3	0.116	0.029	3.98	<0.001
RSSD Samples (n = 40)	β_0	0.421	0.067	6.39	<0.001
	β_1	0.504	0.043	11.60	<0.001
	β_2	−0.068	0.039	−1.72	0.093
	β_3	0.108	0.038	2.86	0.007
SRSD Samples (n = 30)	β_0	0.285	0.076	3.75	<0.001
	β_1	0.477	0.043	11.17	<0.001
	β_2	−0.035	0.049	−0.72	0.481
	β_3	0.142	0.049	2.88	0.008

models for describing the natural log EC_e/natural log EC_a relationship. After estimation of Eq. (8), the *t*-test associated with the z_2 parameter estimate was found to be nearly significant at the 0.05 significance level ($p = 0.058$). However, the u_x and u_y coefficients appeared to be only marginally important ($p = 0.157$ and $p = 0.085$, respectively). Additionally, Eq. (4) produced a slightly smaller jack-knifed mean square error estimate than Eq. (8), although this difference was fairly trivial (<1%). Thus, the most parsimonious model was Eq. (4).

Table 2 lists the model summary statistics and parameter estimates for Eq. (4), after estimating this model using all 70 sampling locations. This model produced an R^2 value of 0.814 and a root mean square error (RMSE) estimate of 0.272. All four parameter estimates were statistically significant; three of the four estimates were highly significant ($p < 0.001$). Additionally, the residual errors associated with Eq. (4) appeared to be Normally distributed, devoid of any outliers, and spatially uncorrelated; the empirical residuals passed both the Shapiro-Wilk test for Normality ($SW = 0.9854$, $p = 0.593$) and the Moran test for spatial correlation ($z_M = 0.778$, $p = 0.218$). These results confirm that an ordinary linear model can be used in place of a more elaborate geostatistical model for purposes of predicting the natural log salinity levels from the associated (natural log transformed and de-correlated) sensor readings.

After identifying and estimating a suitable ordinary linear regression calibration model, we performed the tests described in the Appendix (see Appendix A—*Prediction validation tests for the ordinary linear model*) to validate the suitability of the 40-site model-based sampling plan. Table 2 shows the corresponding summary statistics and parameter estimates for Eq. (4) using the 40-site model-based (RSSD) and 30-site design-based sampling plans (SRSD), respectively. The results shown in Table 2 suggest that both sampling plans yield very similar model summary statistics and parameter estimates.

A formal test of the hypothesis of equivalent parameter estimates was carried out using a composite model F-test. Likewise, a joint-prediction F-test and mean-prediction *t*-test were used to formally test the accuracy of the Eq. (4) natural log salinity predictions derived from the 40-site RSSD plan (*i.e.*, to test if the predictions associated with the 30 randomly chosen SRSD locations were sufficiently accurate and globally unbiased). Table 3 shows the model validation test results for the above mentioned tests. The composite model F-test produced an F-score of 0.63 ($p = 0.640$), suggesting that the Eq. (4) parameter estimates associated with the RSSD and SRSD plans were statistically equivalent. Likewise, the joint-prediction F-test produced an F-score of 0.69 ($p = 0.846$), suggesting that the predictions associated with the 30 randomly chosen

Table 3. Model validation test results for Eq. (4).

Model Eq. (4)	F-score	Pr > F
Composite model F-test (ndf = 4, ddf = 62)[1]	0.63	0.640
Joint-prediction F-test (ndf = 30, ddf = 36)[2]	0.69	0.846
	t-score	**Pr > |t|**
Mean-prediction t-test (ndf = 1, ddf = 36)[2]	−1.23	0.226

[1] Eq. (4) with unique parameters for each sample set.
[2] Eq. (4) estimated using response surface sampling design (RSSD) data.

SRSD locations were unbiased and within the specified precision of the regression model. Figure 4 shows these 30 observed and predicted ln(EC_e) measurements (r = 0.916), where Eq. (4) was in turn estimated using only the data from the 40-site RSSD plan. The mean-prediction *t*-test score was also non-significant ($t = -1.23$, p = 0.226), suggesting that the average predicted natural log salinity value across these 30 SRSD locations was also statistically equivalent to the average measured value. Thus, when Eq. (4) was estimated using only the 40 RSSD plan sites, the resulting equation passed all three model validation tests.

Figure 5 shows the predicted spatial salinity pattern for the entire field, after using the 40-site RSSD plan to estimate Eq. (4). Likewise, Fig. 6 shows the predicted salinity pattern generated by the same equation, but estimated using the 30-site SRSD plan. Clearly, both sampling designs result in nearly equivalent spatial salinity maps.

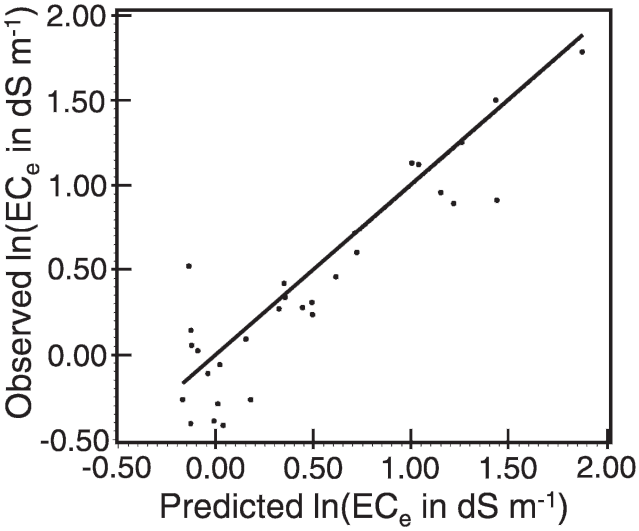

Figure 4. Observed versus predicted ln(EC_e) sample data at 30 SRSD locations, where SRSD represents the stratified random sampling design.

Sampling Design Optimality Scores

The 40-site model-based sampling plan was generated by combining two 20-site RSSD plans together (*e.g.*, Designs A and B). We used this compositing approach in order to develop a larger (n = 40) sampling plan, specifically in case a geostatistical mixed linear model needed to be estimated. Additionally, although both 20-site designs were generated with the same ESAP-RSSD software program, the response surface sampling designs used to generate Design A were purposely degraded (in order to keep the sampling locations selected by both 20-site designs from being located too close together). Hence, from a statistical perspective, Design B would represent the preferable model-based sampling plan for estimating an ordinary linear model, if data from only one RSSD were available.

To assess the suitability of using a 20-site, model-based design to calibrate a regression equation, we computed and compared the design optimality scores for the RSSD plan B design with the 20-site primary SRSD plan. The design optimality scores for Eqs. (1)–(8) are shown in Table 4 for both sampling designs. The optimality criteria computed in Table 4 include the D-, V-, and two G-optimality scores for each of the eight hypothesized regression models. Note that the RSSD always outperforms the equivalent sized SRSD with respect to all three optimality criteria. The score differences are especially pronounced for the D_{opt} and G_{max} criteria, as well as for the more complex regression functions (Eqs. (4) and (8), respectively). For Eq. (4) specifically, the use of the RSSD results in about a 4% reduction in the expected average prediction, a 7.5% reduction in the 90th percentile variance, and about a 36% reduction in the expected maximum prediction variance.

Although these optimality scores quantify just one sampling event, the results shown in Table 4 are expected. The RSSD tends to select sampling locations that exhibit a greater range in the observed EMI signal levels and hence the target soil property, provided that the soil property and signal data are strongly correlated (Myers and Montgomery, 2002). Therefore, RSSD yields many desirable statistical properties, such as high regression

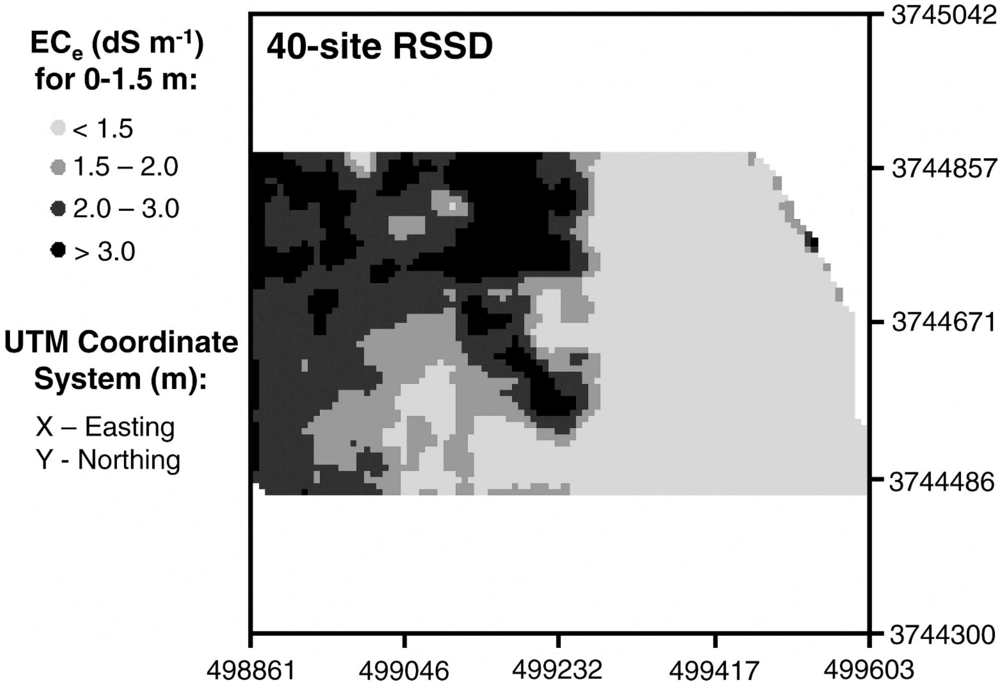

Figure 5. Predicted spatial EC$_e$ pattern from Eq. (4), using the n = 40 RSSD locations, where RSSD represents the response surface sampling design.

coefficient precision and low average prediction variance. A more detailed discussion on the statistical performance characteristics of this prediction-based sampling approach can be found in Lesch (2005).

A practical example of this effect is shown in Table 5, which summarizes the mean square prediction error (MSPE) estimates for Eqs. (3), (4), (7), and (8) for each design. In Table 5, these MSPE estimates have

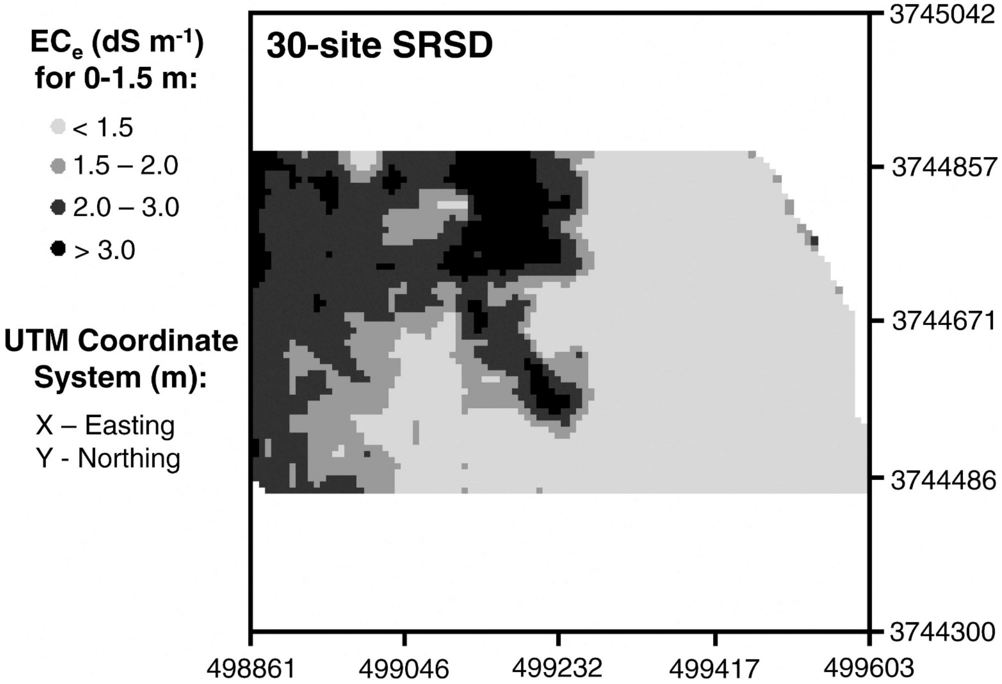

Figure 6. Predicted spatial EC$_e$ pattern from Eq. (4), using the n = 30 SRSD locations, where SRSD represents the stratified random sampling design.

Table 4. Response suface sampling design (Design B) and stratified random sampling design (Primary Design) D-, V-, and G-optimality scores for Eqs. (1)–(8).

		Design optimality scores			
Sample design	Regression equation	D_{opt}	V_{opt}	G_{max}	G_{90}
RSSD (Design B)	(1)	1.43	1.09	1.25	1.13
	(2)	2.31	1.12	1.81	1.19
	(3)	2.72	1.11	1.88	1.14
	(4)	4.02	1.14	1.94	1.23
	(5)	3.30	1.18	1.73	1.27
	(6)	4.96	1.21	2.15	1.32
	(7)	5.87	1.20	1.92	1.30
	(8)	8.05	1.24	2.16	1.36
SRSD (Primary Design)	(1)	1.12	1.10	1.30	1.15
	(2)	1.46	1.14	2.03	1.24
	(3)	0.78	1.15	2.82	1.22
	(4)	0.85	1.19	3.03	1.33
	(5)	1.87	1.19	2.22	1.31
	(6)	2.30	1.24	2.38	1.40
	(7)	1.02	1.25	3.96	1.41
	(8)	0.83	1.33	5.55	1.58

been calculated as:

$$\mathrm{MSPE}_{\mathrm{Eq,\ Design}} = (1/30)\sum_{i=1}^{30}(y_i - \hat{y}_i)^2, \qquad (9)$$

using the 20 sites associated with RSSD Design A and the 10 secondary SRSD sites (*i.e.*, 30 independent samples not considered in Table 4). For the regression equations without any trend surface parameters, these MSPE estimates are essentially equivalent. However, upon including the first-order trend surface parameters, the RSSD Design B produces MSPE estimates that are about 10% to 16% lower than the primary SRSD. Additionally, Fig. 7 shows the predicted EC_e map generated by Eq. (4) when this equation was calibrated using just the 20 RSSD Design B sites. Note that Fig. 7 is essentially identical to Fig. 5 (and Fig. 6), confirming that this 20-site, model-based sampling plan (*i.e.*, RSSD) can in fact be used to accurately estimate the specified calibration model.

Table 5. Mean square prediction error (MSPE) estimates for Eqs. (3), (4), (7), and (8). RSSD represents responses surface sample design and SRSD represents stratified random sample design.

	MSPE estimates	
Regression equation	RSSD (design B)	SRSD (primary design)
Eq. (3)	0.129	0.125
Eq. (4)	0.120	0.119
Eq. (7)	0.125	0.150
Eq. (8)	0.131	0.147

Conclusions

The model validation tests presented in Table 3 show that the model-based sampling design can be reliably used to estimate the $\ln(EC_e)$ calibration function and to provide an accurate and unbiased prediction of the validation sample sites chosen by the SRSD. The regression models estimated using the model- and design-based sampling plans can not be judged to be significantly different. Additionally, the field average $\ln(EC_e)$ predictions and salinity prediction maps produced by the two sampling approaches are quite similar.

Even though both RSSD and SRSD approaches provided similar validation results, it is apparent from the design optimality scores shown in Table 4 that the use of the RSSD should in principle facilitate the estimation of a more accurate regression model; *i.e.*, this sampling approach should allow for better model discrimination, more precise parameter estimates, and smaller prediction variances. These issues are obviously important in practice, since a sampling plan needs to allow for both effective model selection and accurate parameter estimation. Although only one property that influences EC_a was used in the comparison of model-

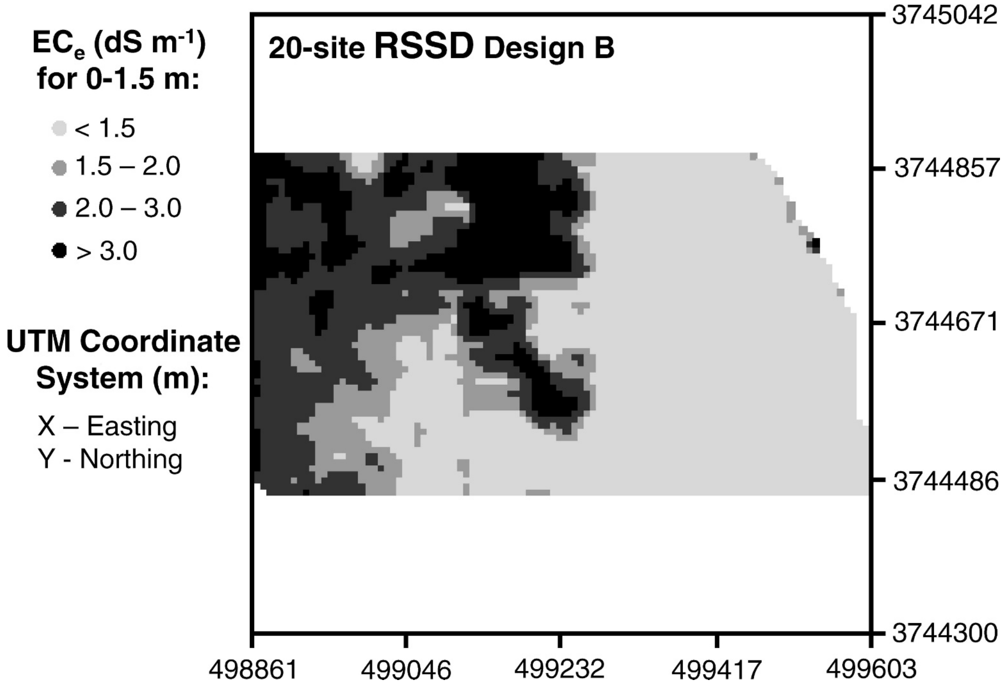

Figure 7. Predicted spatial EC_e pattern from Eq. (4), using the n = 20 RSSD Design B sampling locations, where RSSD represents the response surface sampling design.

and design-based sampling approaches, past experience has shown that any property that strongly correlates with EC_a at a given site will render similar results.

The significance of this sampling design comparison is that an alternative sampling approach (*i.e.*, RSSD) has been shown to be viable and that RSSD in principle provides an increased level of assurance of the spatial characterization of soil properties with EC_a-directed soil sampling. The level of technical knowledge needed for RSSD is greater than other sampling designs but software is available (*i.e.*, ESAP; Lesch *et al.*, 2000) that significantly reduces the statistical expertise necessary to create a RSSD plan from geo-referenced EC_a data.

Acknowledgments

The authors wish to acknowledge the numerous hours of diligent technical work that was performed in the field and in the laboratory by several technicians whose efforts and conscientiousness were crucial to the success of the projects including Jack Jobes, Kevin Yemoto, Clay Wilkinson, Harry Forster, and JoAn Fargerlund.

References

Banerjee, S., Carlin, B.P., and Gelfand, A.E., 2004. Hierarchical modeling and analysis for spatial data: CRC Press, Boca Raton, FL, 454 pp.

Bouma, J., and Bregt, A.K. (eds.), 1989, Land qualities in space and time: Pudoc, Wageningen, The Netherlands, 352 pp.

Brandsma, A.S., and Ketellapper, R.H., 1979, Further evidence on alternative procedures for testing of spatial autocorrelation amonst regression disturbances: *in* Exploratory and explanatory statistical analysis of spatial data, Bartels, C.P.A., and Ketellapper, R.H. (eds.), Martinus Nijhoff, Boston, MA, 113–136.

Brus, D.J., and de Gruijter, J.J., 1993, Design-based versus model-based estimates of spatial means: Theory and application in environmental soil science: Environmetrics, **4**, 123–152.

Brus, D.J., and Heuvelink, G.B.M., 2007, Optimization of sample patterns for universal kriging of environmental variables: Geoderma, **138**, 86–95.

Cook, R.D., and Weisberg, S., 1999. Applied regression including computing and graphics: John Wiley, New York, NY, 632 pp.

Corwin, D.L., and Lesch, S.M., 2005a, Apparent soil electrical conductivity measurements in agriculture: Comput. Electron. Agric., **46**(1–3) 11–43.

Corwin, D.L., and Lesch, S.M., 2003, Application of soil electrical conductivity to precision agriculture: Theory, principles, and guidelines: Agron. J., **95**, 455–471.

Corwin, D.L., and Lesch, S.M., 2005b, Characterizing soil spatial variability with apparent soil electrical conductivity: I. Survey protocols: Comput. Electron. Agric., **46**(1–3) 103–133.

Corwin, D.L., Carrillo, M.L.K., Vaughan, P.J., Rhoades, J.D., and Cone, D.G., 1999, Evaluation of GIS-linked model of salt loading to groundwater: J. Environ. Qual., **28**, 471–480.

Corwin, D.L., Lesch, S.M., Oster, J.D., and Kaffka, S.R., 2006, Monitoring management-induced spatio-temporal changes in soil quality through soil sampling directed by apparent electrical conductivity: Geoderma, **131**, 369–387.

Corwin, D.L., Lesch, S.M., Shouse, P.J., Soppe, R., and Ayars, J.E., 2003b, Identifying soil properties that influence cotton yield using soil sampling directed by apparent soil electrical conductivity: Agron. J., **95**(2) 352–364.

Corwin, D.L., Kaffka, S.R., Hopmans, J.W., Mori, Y., Lesch, S.M., and Oster, J.D., 2003a, Assessment and field-scale mapping of soil quality properties of a saline-sodic soil: Geoderma, **114**(3–4) 231–259.

Haining, R., 1990. Spatial data analysis in the social and environmental sciences: Cambridge University Press, Cambridge, UK, 411 pp.

Haskard, K.A., Cullis, B.R., and Verbyla, A.P., 2007, Anisotropic Matèrn correlation and spatial prediction using REML: J. Agric. Biol. Environ. Statistics, **12**, 147–160.

Isaaks, E.H., and Srivastava, R.M., 1989. An introduction to applied geostatistics: Oxford University Press, New York, NY, 561 pp.

Jury, W.A., 1985, Spatial variability of soil physical parameters in solute migration: A critical literature review *in* Electrical Power Research Institute (EPRI) Report EA-4228, EPRI, Palo Alto, CA, 73 pp.

Jury, W.A., 1986, Spatial variability of soil properties *in* Vadose Zone Modeling of Organic Pollutants, Hern, S.C., and Melancon, S.M. (eds.), Lewis Publishers, Chelsea, MI, 245–269.

Kaffka, S.R., Lesch, S.M., Bali, K.M., and Corwin, D.L., 2005, Relationship of electromagnetic induction measurements, soil properties, and sugar beet yield in salt-affected fields for site-specific management: Comput. Electron. Agric., **46**(1–3) 329–350.

Lesch, S.M., 2005, Sensor-directed response surface sampling designs for characterizing spatial variation in soil properties: Comp. Electron. Agric., **46**(1–3) 153–180.

Lesch, S.M., and Corwin, D.L., 2008, Prediction of spatial soil property information from ancillary sensor data using ordinary linear regression: Model derivations, residual assumptions and model validation tests: Geoderma, **148**, 130–140, 2008.

Lesch, S.M., Corwin, D.L., and Robinson, D.A., 2005, Apparent soil electrical conductivity mapping as an agricultural management tool in arid zone soils: Comp. Electron. Agric., **46**, 351–378.

Lesch, S.M., Rhoades, J.D., and Corwin, D.L., 2000, ESAP-95 version 2.10R: User manual and tutorial guide, Research Rpt. 146: USDA-ARS, U.S. Salinity Laboratory, Riverside, CA, USA, 169 pp.

Lesch, S.M., Strauss, D.J., and Rhoades, J.D., 1995a, Spatial prediction of soil salinity using electromagnetic induction techniques: 1. Statistical prediction models: A comparison of multiple linear regression and cokriging: Water Resour. Res., **31**, 373–386.

Lesch, S.M., Strauss, D.J., and Rhoades, J.D., 1995b, Spatial prediction of soil salinity using electromagnetic induction techniques: 2. An efficient spatial sampling algorithm suitable for multiple linear regression model identification and estimation: Water Resour. Res., **31**, 387–398.

Lieberman, G.J., 1961, Prediction regions for several predictions from a single regression line: Technometrics, **3**, 21–27.

Mausbach, M.J., and Wilding, L.P. (eds.), 1991, Spatial variabilities of soils and landforms, SSSA Special Publication 28: Soil Sci. Soc. Am., Madison, WI, 270 pp.

McBratney, A.B., and Webster, R., 1981, The design of optimal sampling schemes for local estimation and mapping of regionalized variables: II. Program and examples: Comput. Geosci., **7**, 335–365.

Minasny, B., McBratney, A.B., and Walvoort, D.J.J., 2007, The variance quadtree algorithm: use for spatial sampling design: Comput. Geosci., **33**, 383–392.

Müller, W.G., 2001, Collecting spatial data: Optimum design of experiments for random fields, 2nd ed.: Physica-Verlag, Heidelberg, Germany, 242 pp.

Müller, W.G., and Zimmerman, D.L., 1999, Optimal designs for variogram estimation: Environmetrics, **10**, 23–37.

Myers, R.H., 1986, Classical and modern rtegression with applications: Duxbury Press, Boston, MA, 359 pp.

Myers, R.H., and Montgomery, D.C., 2002. Response surface methodology: Process and product optimization using designed experiments, 2nd ed.: John Wiley, New York, N.Y, 690 pp.

Nathan, G., 1988, Inference based on data from complex sample designs, *in* Handbook of statistics, Krishnaiah, P.R., and Rao, C.R. (eds.), Vol. 6: Elsevier, Amsterdam, The Netherlands, 247–266.

Nielsen, D.R., Biggar, J.W., and Erh, K.T., 1973, Spatial variability of field-measured soil-water properties: Hilgardia, **42**(7) 215–259.

Rao, C.R., and Toutenburg, H., 1995. Linear models: Least squares and alternatives, Springer-Verlag, New York, NY, 435 pp.

Rhoades, J.D., 1996, Salinity: Electrical conductivity and total dissolved solids, *in* Methods of soil analysis: Part 3—Chemcial methods, Sparks, D.L. (ed.), SSSA Book Series No. 5: Soil Sci. Soc. Am., Madison, WI, USA, 417–435.

Robinson, D.A., Lebron, I., Lesch, S.M., and Shouse, P., 2004, Minimizing drift in electrical conductivity measurements in high temperature environments using the EM38: Soil Sci. Soc. Am. J., **68**, 339–345.

Rogers, M.B., Baham, J.E., and Draglia, M.I., 2006, Soil iron content effects on the ability of magnetometer surveying to locate buried agricultural drainage pipes: Appl. Eng. Agric., **22**(5) 701–704.

Russo, D., 1984, Design of an optimal sampling network for estimating the variogram: Soil Sci. Soc. Am. J., **48**, 708–716.

SAS Institute Inc., 1999a, SAS/IML User's Guide, Version 8: SAS Institute Inc., Cary, NC, 846 pp.

SAS Institute Inc., 1999b, SAS/STAT User's Guide, Version 8: SAS Institute Inc., Cary, NC.

Schabenberger, O., and Gotway, C.A., 2005, Statistical methods for spatial data analysis: CRC Press, Boca Raton, FL, 497 pp.

Schabenberger, O., and Pierce, F.J., 2002, Contemporary statistical models for the plant and soil sciences: CRC Press, Boca Raton, FL, 757 pp.

Shapiro, S.S., and Wilk, M.B., 1965, An analysis of variance test for Normality (complete samples): Biometrika, **52**, 591–611.

Thompson, S.K., 1992, Sampling: John Wiley, New York, NY, 367 pp.

Tiefelsdorf, M., 2000, Modeling spatial processes: The identification and analysis of spatial relationships in regression residuals by means of Moran's I: Springer-Verlag, New York, NY, 167 pp.

Upton, G., and Fingleton, B., 1985, Spatial data analysis by example, John Wiley, New York, NY, 410 pp.

Valliant, R., Dorfman, A.H., and Royall, R.M., 2000, Finite population sampling and inference: A prediction approach, John Wiley, New York, NY, USA, 504 pp.

Van Groenigen, J.W., Siderius, W., and Stein, A., 1999, Constrained optimisation of soil sampling for minimisation of the kriging variance: Geoderma, **87**, 239–259.

Wackernagel, H., 1998. Multivariate geostatistics, 2nd ed.: Springer-Verlag, Berlin, Germany, 387 pp.

Warrick, A.W., and Nielsen, D.R., 1980, Spatial variability of soil physical properties in the field, *in* Applications of soil physics, Hillel, D. (ed.), Academic Press, New York, NY, 319–344.

Warrick, A.W., and Myers, D.E., 1987, Optimization of sampling locations for variogram calculations: Water Resour. Res., **23**, 496–500.

Webster, R., and Oliver, M.A., 2001, Geostatistics for environmental scientists: John Wiley, New York, NY, 315 pp.

Weisberg, S., 1985, Applied linear regression, 2nd ed.: John Wiley, New York, NY, 324 pp.

White, I., 1988, Measurement of soil physical properties in the field, *in* Flow and transport in the natural environment: Advances and applications, Steffen, W.L., and Denmead, O.T. (eds.), Springer-Verlag, Berlin, Germany, 59–85.

Zhu, Z., and Stein, M.L., 2006, Spatial sampling design for prediction with estimated parameters: J. Agric. Biol. Environ. Statistics, **11**, 24–44.

APPENDIX A
STATISTICAL CONSIDERATIONS

Geostatistical and Ordinary (Spatially Referenced) Linear Models

In a field survey where remote sensor readings are collected, the sensor data is generally used to predict a specific, unobserved soil property. For example, assume that a dense grid of EC_a data has been collected across a particular field and soil samples have been collected at some of these survey locations for purposes of calibrating the EC_a data, with the intention to use these sensor and calibration sample readings to estimate a model that in turn will be used to predict the detailed spatial pattern of the soil property. Assume that the relationship between the remote sensing data and target soil property (*e.g.*, soil salinity) can be adequately approximated using the following geostatistical mixed linear model (Haskard *et al.*, 2007):

$$\mathbf{y} = \mathbf{X}\beta + \eta(\mathbf{s}) + \xi(\mathbf{s}), \qquad \text{(A-1)}$$

where $\mathbf{y}$ represents an $(n \times 1)$ vector of observed soil property data, $\mathbf{s}$ represents the corresponding vector of paired (s_x, s_y) survey location coordinates, $\mathbf{X}$ represents an $(n \times p)$ fixed data matrix that includes observed functions of sensor readings and possibly also the survey location coordinates, β represents a $(p \times 1)$ vector of unknown parameter estimates, $\eta(\mathbf{s})$ represents a 0-mean, second order stationary spatial Gaussian error process, and $\xi(\mathbf{s})$ represents a vector of jointly independent Normal(0, σ_n^2) random variables. Lesch and Corwin (2008) discuss Eq. (A-1) in detail and describe when and under what conditions this equation reduces to a spatially referenced, ordinary linear model.

Generally speaking, the most common justification for using an ordinary linear model is when the physical locations of the calibration sample sites are spread sufficiently far apart, such that the residual errors do not exhibit any spatially correlated structure. Conceptually, this condition occurs when one samples beyond the effective range of the residual covariance structure; Lesch and Corwin (2008) refer to such residuals as "effectively uncorrelated". More formally, under this assumption the model calibration errors in Eq. (A-1) can be treated as being statistically independent. Thus, from a statistical estimation viewpoint, Eq. (A-1) is no different from an ordinary linear regression model and can therefore be fit using ordinary least squares estimation techniques.

In the remainder of this section we will assume that the residual errors in Eq. (A-1) are spatially uncorrelated, thus allowing Eq. (A-1) to be re-expressed as an ordinary, spatially referenced linear model:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon(\mathbf{s}). \qquad \text{(A-2)}$$

Under this residual error assumption, one can readily verify that the best linear unbiased estimate for β becomes:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \qquad \text{(A-3)}$$

with a corresponding variance of

$$Var(\hat{\beta}) = \tau^2(\mathbf{X}^T\mathbf{X})^{-1}, \tag{A-4}$$

for $Var(\varepsilon(\mathbf{s})) = \tau^2$. Thus, when the residual spatial independence assumption holds, one can in turn use standard results from (non-spatial) linear modeling theory to assess the statistical optimality of competing sampling designs and/or the statistical validity of a fitted model, etc.

Sampling Strategies for Spatially Referenced Linear Models

Both design- and model-based sampling strategies can be employed to estimate spatially referenced linear models. Design-based sampling strategies have a well developed underlying theory and can be useful in many spatial applications (Thompson, 1992; Brus and de Gruijter, 1993). Likewise, model-based sampling strategies have been applied to the optimal collection of spatial data by Müller (2001); the specification of optimal designs for variogram estimation by Russo (1984), Warrick and Myers (1987), and Müller and Zimmerman (1999); the estimation of spatially referenced leaching requirement models by Lesch *et al.*, (1995b) and Lesch (2005), and the estimation of geostatistical linear models by Zhu and Stein (2006), Brus and Heuvelink (2007), and Minasny *et al.*, (2007).

The goals of this study were (1) to relate EC_a survey data to soil salinity measurements using a spatially referenced linear model, (2) to assess the suitability of using a model-based sampling strategy for eliciting and estimating a suitably specified linear model to produce these salinity predictions, and (3) to statistically compare and contrast a model-based and design-based sampling strategy. Thus, two different sampling designs were employed to estimate the regression model relating salinity to EC_a: a RSSD plan and a SRSD plan.

Sample design optimality criteria. For a hypothesized ordinary linear model, various statistical criteria have been proposed in the response surface design literature for assessing the "optimality" of competing sampling designs (Myers and Montgomery, 2002). Most of these criteria measure either the expected precision of the regression model parameter estimates (*e.g.*, D- and A-optimality) or quantify some measure of precision in the model predictions (*i.e.*, G-, V-, and Q-optimality). In this study we chose to compare and contrast the RSSD and SRSD plans using the D-, V-, and G-optimality criteria.

Let **X** represent the design matrix associated with a specific regression model, $\mathbf{x}_i$ represent the regression vector associated with the ith survey location, and *p* represent the number of parameters in the regression model (including the intercept). Additionally, let *n* and *N* represent the number of calibration soil sample sites and EC_a survey sites, respectively. The D-, V- and G-optimality scores can then be defined as follows:

$$D_{opt} = |\mathbf{X}^T\mathbf{X}|/n^p, \tag{A-5}$$

$$V_{opt} = (1/N)\sum_{i=1}^{N}\left(1 + \mathbf{x}_i{}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i\right), \text{ and} \tag{A-6}$$

$$G_{max} = \max\left(1 + \mathbf{x}_i{}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i\right)_{i=1,..,N}, \tag{A-7}$$

where the function $|\cdot|$ represents the determinant of a matrix. Intuitively, the D_{opt} score measures the expected precision in the regression model parameter estimates; larger scores imply higher precision and a sampling design that maximizes this score is said to be D-optimal. The V_{opt} score measures the expected average prediction error associated with the linear model predictions, assuming that the prediction errors are spatially independent. A smaller score implies a smaller average prediction error and a sampling design that minimizes this score is said to be V-optimal. Likewise, the G_{max} score measures the expected maximum prediction error in the regression model predictions, again assuming that the prediction errors are spatially independent. A sampling design that minimizes this score is said to be G-optimal.

When N is large, it can also be useful to compute G scores for certain quantiles associated with the prediction error distribution, such as the 90th or 95th quantile, etc. The interpretation of such a "G-quantile" score (for example, the 90th quantile score) would be that 90% of the regression model predictions exhibit an expected relative prediction error that is less than this value, etc. Finally, note that each of the previously defined optimality scores can be computed for any plausible (hypothesized) ordinary linear model.

If an ordinary linear model is to be successfully used in place of the geostatistical mixed linear model, then more restrictive modeling assumptions need to be met. In addition to the Gaussian error process, the residual errors associated with the calibration sample site locations must be at least approximately uncorrelated. Thus, some type of test for residual spatial correlation should always be performed before deciding to adopt the ordinary linear modeling approach.

Residual diagnostic tests. A formal test for spatial correlation in the residual pattern can be carried out using either a nested likelihood ratio test or via the Moran residual test statistic (Upton and Fingleton, 1985; Haining, 1990; Tiefelsdorf, 2000; Schabenberger and Gotway, 2005). The likelihood ratio test can only be performed after first estimating a suitable geostatistical mixed linear model (Schabenberger and Gotway, 2005). In contrast, the Moran test can be carried out directly on the ordinary linear model residuals.

The Moran residual test statistic (δ_M) is defined as:

$$\delta_M = \frac{\mathbf{r}^T\mathbf{W}\mathbf{r}}{\mathbf{r}^T\mathbf{r}}, \qquad \text{(A-8)}$$

where $\mathbf{r}=\mathbf{y}-\mathbf{X}\hat{\beta}$ (*e.g.*, the vector of observed model residuals), $\mathbf{W}$ represents a suitably specified proximity matrix, and $\hat{\beta}$ is calculated using Eq. (A-3). While the specification of $\mathbf{W}$ can be application-specific, in most soil survey applications it is generally reasonable to specify $\mathbf{W}$ as a scaled inverse distance squared matrix. Under such a specification, where d_{ij} represents the computed distance between the i^th^ and j^th^ sample locations, the $\{w_{ij}\}$ elements associated with the i^th^ row of the $\mathbf{W}$ matrix are defined as:

$$w_{ii}=0 \ \text{ and } \ w_{ij}=d_{ij}^{-2} \Big/ \sum_{j=1}^{n} d_{ij}^{-2}. \qquad \text{(A-9)}$$

Brandsma and Ketellapper (1979) provide the formulas for computing both the mean and variance of the Moran test statistic; see also Lesch (2005) and Lesch and Corwin (2008). The corresponding Moran test score can then be computed as:

$$z_M = (\delta_M - E(\delta_M)) \Big/ \sqrt{Var(\delta_M)}, \qquad \text{(A-10)}$$

and compared to the upper (one-sided) cumulative standard Normal probability density function.

A test score in excess of 1.65 ($\alpha \approx 0.05$) is normally interpreted as being statistically significant. Provided that the fixed effects in the regression model have been correctly specified, such a test score implies that the ordinary linear regression model residuals exhibit significant spatial correlation. In this situation, the linear regression parameter estimates and survey predictions may be highly inefficient and the mean square error estimate and parameter test statistics may be substantially biased. If sufficient data is available (or additional data can be collected), then a suitable spatial or geostatistical linear modeling approach should instead be employed.

In addition to the spatially independent residual error assumption, one must also verify that the model residuals satisfy the usual standard Gaussian error assumption and that the hypothesized model is correctly specified. Fortunately, most well known residual analysis techniques used in an ordinary regression analysis are just as useful when applied to a spatially referenced linear model. These include assessing the assumption of residual Normality using quantile-quantile plots and the Shapiro-Wilk test (Shapiro and Wilk, 1965), detecting outliers and/or high leverage points using plots of internally or externally studentized residuals, and detecting model specification bias using residual versus prediction plots, partial regression leverage plots, and added variable plots (Myers, 1986).

Prediction validation tests for the ordinary linear model. Suppose that a plausible linear model has been specified that describes some type of soil property / survey data relationship. Suppose also that a Moran or likelihood ratio test has been used to verify that the residual errors are spatially uncorrelated and that the other usual residual assumptions hold. Hence, our spatially referenced linear model can be expressed in matrix notation as $\mathbf{y}=\mathbf{X}\beta+\varepsilon(\mathbf{s})$, where $\varepsilon(\mathbf{s}) \sim N(0, \tau^2\mathbf{I}_n)$ and $\mathbf{y}$ and $\mathbf{X}$ are defined as in Eq. (A-2).

In most surveys, the ultimate goal will be to use the fitted equation for prediction purposes, but assume first that we wish to assess the "validity" of our fitted linear model. There are three types of statistical tests that can be readily employed to assess the validity of the spatially referenced linear model. These can all be expressed as F-tests (and/or *t*-tests), and are based on the idea of data partitioning. Thus, assume that we can partition the $n=n_1+n_2$ sample sites into a primary calibration set and a secondary validation set. Given this data partition of sample sites, assume further that we wish to fit the model using the primary calibration data and then test its prediction adequacy using the secondary validation data.

First, with respect to the pooled (calibration + validation) data set, note that the pooled $\mathbf{y}$ vector and $\mathbf{X}$ matrix can be partitioned as:

$$\mathbf{y}=\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} \ \text{ and } \ \mathbf{X}=\begin{bmatrix}\mathbf{X}_1 & 0\\ 0 & \mathbf{X}_2\end{bmatrix}, \qquad \text{(A-11)}$$

where the subscripts index the calibration and validation data sub-sets and the dimension of the partitioned design matrix is $(n_1+n_2) \times 2p$. Given this partition, a "composite model" F-test can be performed by fitting the partitioned equation:

$$\mathbf{y}=\mathbf{X}_1\beta_1+\mathbf{X}_2\beta_2+\varepsilon(\mathbf{s}), \qquad \text{(A-12)}$$

and then testing if $\beta_1=\beta_2$ (Cook and Weisburg, 1999). This is one of the better known, standard model validation testing techniques suggested in the statistical literature; additional details can be found in most regression textbooks (Weisburg, 1985; Myers, 1986; Cook and Weisburg, 1999).

Two other useful model validation tests are the "joint-prediction" F-test and "mean-prediction" *t*-test.

The joint-prediction F-test can be performed by first estimating the linear regression model using just the calibration data, next calculating the joint set of prediction errors across the validation sites as:

$$\mathbf{r}_2 = \mathbf{y}_2 - \mathbf{X}_2\hat{\beta}_1, \qquad \text{(A-13)}$$

and then by computing the statistic:

$$F_1 = \mathbf{r}_2^T\mathbf{V}^{-1}\,\mathbf{r}_2/s_1^2 \quad \text{where } \mathbf{V} = (\mathbf{I} + \mathbf{X}_2(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_2^T). \qquad \text{(A-14)}$$

This test statistic, originally suggested by Lieberman (1961), essentially defines the joint (simultaneous) prediction region for multiple predictions from a single regression model. Given the EU residual assumption and under the null hypothesis (*i.e.*, that the fitted calibration model is correct), F_1 follows a central $F(n_2, n_1 - p)$ distribution where n_2 and $n_1 - p$ represent the number of validation sites and the (calibration) model degrees of freedom, respectively, and s_1^2 represents the estimated calibration model mean square error (MSE) estimate (Lieberman, 1961; Rao and Toutenburg, 1999). In a similar manner, the mean-prediction *t*-test can be performed by first calculating the average prediction error as:

$$\bar{r} = \mathbf{q}^T\mathbf{r}_2 \quad \text{where } \mathbf{q}^T = [\,1/n_2, \quad \ldots \quad 1/n_2\,], \qquad \text{(A-15)}$$

and then computing the statistic:

$$t_1 = \bar{r}/(s_1\sqrt{h}) \quad \text{where } h = \left[(1/n_2) + (\mathbf{q}^T\mathbf{X}_2(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_2^T\mathbf{q})\right]. \qquad \text{(A-16)}$$

Note that t_1 follows a central *t* distribution (with $n_1 - p$ degrees of freedom) under the null hypothesis, where s_1 represents the square root of the calibration model MSE estimate (Rao and Toutenburg, 1999).

Intuitively, the composite model F-test represents a test for non-equivalent parameter estimates across the partitioned calibration and prediction (validation) sample sites. In contrast, the joint-prediction F-test assesses the ability of the regression model (fit using the calibration data only) to make unbiased predictions at all new validation sites, and simultaneously tests if these predictions are within the specified tolerance (precision) of the model. The mean-prediction *t*-test follows from the joint-prediction F-test, and hence assesses the ability of the regression model to make an unbiased prediction of the average value across the new n_2 validation sites.