A simplified regional-scale electromagnetic induction — Salinity calibration model using ANOCOVA modeling techniques

Dennis L. Corwin a,⁎, Scott M. Lesch b,1

⁎ USDA-ARS, U.S. Salinity Laboratory, 450 West Big Springs Road, Riverside, CA 92507-4617, USA
b Riverside Public Utilities, 3435 14th St, Riverside, CA 92501, USA

A R T I C L E   I N F O

Article history:
Received 12 October 2012
Received in revised form 18 March 2014
Accepted 23 March 2014
Available online 13 April 2014

Keywords:
Soil spatial variability
Salinity mapping
Electromagnetic induction
Electrical resistivity
Proximal sensor
Response surface sampling

A B S T R A C T

Directed soil sampling based on geospatial measurements of apparent soil electrical conductivity (ECa) is a potential means of characterizing the spatial variability of any soil property that influences ECa, including soil salinity, water content, texture, bulk density, organic matter, and cation exchange capacity. Multi-field ECa survey data often exhibit abrupt changes in magnitude across field boundaries that complicate the calibration of ECa to soil salinity (i.e., ECe, electrical conductivity of the saturation extract) over large spatial extents. The primary objective of this study is to evaluate three regression techniques for calibrating ECa to ECe over spatial scales ranging from a few thousand to a hundred thousand hectares, where ECe was measured using electromagnetic induction equipment. The regression techniques include analysis of covariance (ANOCOVA), field specific regression (FSR), and common coefficient regression (CCR). An evaluation was made by comparing jack-knifed mean square prediction errors (MSPE) of ECe for two case studies: 2400 ha of the Broadview Water District in California’s San Joaquin Valley and roughly 100,000 ha of the west side of Kittson County in the Red River Valley of Minnesota. The ANOCOVA model outperformed the FSR and CCR regression models on a prediction accuracy basis with the smallest MSPE estimates for depth predictions of soil salinity. The implication of this evaluation is that once ANOCOVA models for each depth are established for a representative set of fields within a regional-scale study area, then the slope coefficients can be used at all future fields, thereby significantly reducing the need for ground-truth soil samples at future fields, which substantially reduces labor and cost. Land resource managers, producers, agriculture consultants, extension specialists, and Natural Resource Conservation Service field staff are the beneficiaries of regional-scale maps of soil salinity.

Published by Elsevier B.V.

1. Introduction

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 field- and landscape-scale processes related to agriculture and the environment, including solute transport, within-field variation in crop yield, and soil salinity accumulation, just to mention a few. Soil salinity accumulation is a major agricultural concern in arid and semi-arid soils throughout the world because it reduces crop yields due to osmotic and specific-ion toxicity effects and impairs soil permeability and tilth. Of the 13.2 × 10⁹ ha of land surface on the earth, only 1.5 × 10⁹ ha is cultivated and 23% of the cultivated land is estimated to be salt-affected, which comprises about 10% of the total arable land (Massoud, 1981). The influence of soil salinity on crop yield is well known in the plant salt tolerance literature (Maas, 1996). Maps of soil salinity assist producers in crop selection, irrigation management, and reclamation. However, spatial variation of dynamic soil properties, such as water content and salinity, are especially challenging to characterize spatially due to their temporal nature and their complex spatial nature.

The geospatial measurement of ECa is a sensor technology that has played, and continues to play, a major role in addressing the issue of field-scale spatial variability characterization, particularly in mapping soil salinity (Corwin and Lesch, 2005a). Geospatial measurements of ECa are spatially complex because they reflect the influence of several physical and chemical soil properties, including soil salinity, texture, water content, bulk density, organic matter, and cation exchange capacity. Subsequently, geospatial measurements of ECa are used to direct soil sampling as a means of characterizing spatial variability of those soil properties that correlate with ECa at

http://dx.doi.org/10.1016/j.geoderma.2014.03.019
0016-7061/Published by Elsevier B.V.
that particular study site. Characterizing spatial variability with EC$_i$-directed soil sampling is based on the notion that when EC$_i$ correlates with a soil property or properties, then spatial EC$_i$ information can be used to identify sites that reflect the range and spatial variability of the property or properties (Corwin and Lesch, 2005b).

In instances where EC$_i$ correlates with a particular soil property, an EC$_i$-directed soil sampling approach will establish the spatial distribution of that property with an optimum number of site locations, which significantly reduces labor costs compared to grid sampling (Corwin et al., 2003a, 2003b). Details for conducting a field-scale EC$_i$ survey for the purpose of characterizing the soil spatial variability are in Corwin and Lesch (2005b). Protocols specifically for mapping soil salinity with EC$_i$-directed soil sampling are in Corwin and Lesch (2013). Corwin and Lesch (2005a) provide a compilation of literature pertaining to the soil physical and chemical properties that either directly or indirectly influence EC$_i$.

Regional-scale maps of soil salinity are needed by policy makers to establish the extent of the soil salinity problem and to monitor the impact of climate change and agriculture on soil salinization. The Red River Valley in the Midwestern USA is a perfect example of where this information is needed (Lobell et al., 2010). However, regional-scale mapping of soil salinity poses new challenges beyond those of field-scale salinity assessment due to the greater spatial extent.

A review paper by Metternicht and Zink (2003) and a recent special collection of papers in the Journal of Environmental Quality (2010, volume 39, issue 1) focusing on remote sensing of soil degradation provide several papers that present regional-scale salinity assessments using remote sensing (Caccetta et al., 2010; Furby et al., 2010; Lobell et al., 2010; Singh et al., 2010). In the study by Lobell et al. (2010) EC$_i$-directed sampling was used in combination with remote imagery (i.e., MODIS, Moderate Resolution Imaging Spectroradiometer) to map salinity over hundreds of thousands of hectares (Lobell et al., 2010). In essence, the EC$_i$-directed sampling provided ground-truth measurements of soil salinity to calibrate multi-year MODIS enhanced vegetation index (EVI) imagery, thereby providing a relationship between EVI and soil salinity.

An alternative and less complicated approach for assessing regional-scale soil salinity is to develop a relationship between EC$_i$ measured with electromagnetic induction (EMI) and soil salinity for an entire region. We hypothesize that regression techniques can calibrate EC$_i$ to soil salinity (where soil salinity is measured using the electrical conductivity of the saturation extract, EC$_e$, expressed in dS m$^{-1}$) for multiple fields extending over a regional scale, substantially reducing the need for future ground-truth soil sampling. The objective of this research is to evaluate three regression modeling techniques for calibrating EC$_i$ to EC$_e$ for multiple fields extending over a range of a few thousand to over a hundred thousand hectares and to establish the viability of using regression techniques for regional-scale salinity assessment.

2. Rationale for regional calibration with regression models

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 EC$_i$ survey data to a specific soil property such as salinity. Numerous examples exist in the literature of geostatistical or spatial modeling approaches. The textbooks by Schabenberger and Gotway (2005), Schabenberger and Pierce (2002), Webster and Oliver (2001), Wackernagel (1998), and Isaaks and Srivastava (1989) are particularly relevant to the calibration problem.

However, in addition to the commonly used geostatistical techniques, ordinary linear regression models are often used when calibrating 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.

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 calibrate EC$_i$ survey data to one or more target soil properties, such as salinity, i.e., EC$_e$ (Corwin and Lesch, 2005b). For example, field-scale soil salinity patterns are commonly mapped quite accurately using EC$_i$ survey data and ordinary linear regression models, since the residual error distribution typically exhibits only short-range spatial correlation (Corwin and Lesch, 2005b; Lesch and Corwin, 2008; Lesch et al., 2005). 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).

Multi-field EC$_i$ survey data often exhibit an abrupt change in magnitude across field boundaries, presenting a challenge to the conversion of EC$_i$ to EC$_e$ at large spatial extents of thousands to tens of thousands of hectares. The abrupt change is typically caused by a variety of reasons: (i) between-field variation in field average water content due to irrigation frequency and timing, (ii) between-field variation in soil texture, (iii) condition of the soil surface (e.g., till vs. no-till) due to management practices that effect soil compaction, (iv) surface geometry (i.e., presence or absence of beds and furrows), (v) temperature differences (i.e., EC$_i$ surveys conducted at different times of the year), and (vi) between-field spatial variation in salinity.

Calibration models are often used to adjust out an abrupt change. For instance, temperature corrections to EC$_i$ data are typically done using a multiplicative adjustment constant, i.e., EC$_{i25^\circ C} = f_T \cdot EC_i$, where $f_T$ is the soil temperature and $f_T$ is the temperature correction factor. Similarly, changes in bed–furrow geometry, surface conditions, soil texture, and water content are approximated in the same manner. Generally speaking, many secondary effects that influence EC$_i$-EC$_e$ calibration models are modeled as multiplicative in nature, at least to a first-order approximation (Corwin and Lesch, 2005b).

More specifically, consider the case of surface geometry, i.e., presence and absence of beds and furrows in a field, where an EC$_i$ survey has been conducted. In the absence of any surface geometry, a simple power model describes the deterministic component of the EC$_i$–EC$_e$ relationship, i.e., EC$_i$ = $\beta$ * EC$_e^\gamma$ where $\beta$ is a coefficient and $i = 1, 2, 3, ... n$. To account for the surface geometry effect an additional dummy variable ($x$) and associated scaling parameter ($\theta$) are used, i.e., EC$_i$ = $\beta x^\theta$ * EC$_e^\gamma$ where $x_i = 1$ if there is a surface geometry effect and $x_i = 0$ otherwise. Under a log transformation, this multiplicative parameter becomes additive as shown in Eq. (1):

$$\ln\left(EC_i\right) \approx x_i \ln\left(\theta\right) + \ln\left(\beta\right) + \alpha \ln\left(EC_e\right) = \beta_0 + \beta_1 x_i + \alpha \ln\left(EC_e\right)$$

(1)

On a log–log scale, a simple linear regression model with an additional blocking (shift) parameter can adjust an abrupt change in any multiplicative EC$_e$ effect within a field. Note that Eq. (1) is a simple type of ANOCOVA model. In principle, this type of ANOCOVA modeling approach could be used to calibrate multiple-field EC$_i$ surveys to EC$_e$, provided that the assumptions in Eq. (1) are reasonable.

Consider a scenario where EC$_i$ survey data is acquired across multiple fields and assume that the number of soil sampling locations
collected in any given field is minimal (i.e., $n < 10$). In the absence of any useful spatial or geostatistical modeling approach under these conditions, basic regression modeling techniques are used. The regression techniques include field specific regression (FSR), common coefficient regression (CCR), and ANOCOVA. A FSR model is defined by Eq. (2):

$$\ln(\text{EC}_{i,j,k}) = \beta_{0,jk} + \beta_{1,jk} \ln(\text{EM}_{i,jk}) + \beta_{2,jk} \ln(\text{EM}_{h,ik}) + \epsilon_{ijk}. \quad (2)$$

Common coefficient regression model is defined by Eq. (3):

$$\ln(\text{EC}_{i,j,k}) = \beta_{0,j} + \beta_{1,j} \ln(\text{EM}_{v,ik}) + \beta_{2,j} \ln(\text{EM}_{h,ik}) + \epsilon_{ijk}. \quad (3)$$

Analysis of covariance model is defined by Eq. (4):

$$\ln(\text{EC}_{i,j,k}) = \beta_{0,jk} + \beta_{1,j} \ln(\text{EM}_{v,ik}) + \beta_{2,j} \ln(\text{EM}_{h,ik}) + \epsilon_{ijk}. \quad (4)$$

where $i$ refers to the soil sample site within a field ($i = 1, 2, 3, ..., n_i$), $j$ is the sample depth ($j = 1, 2, 3, ..., p$), $k$ is the field ($k = 1, 2, 3, ..., M$), $\text{EM}_h$ is the EC$_c$ measured with EMI in the vertical coil configuration (dS m$^{-1}$), and $\text{EM}_v$ is the EC$_c$ measured with EMI in the horizontal coil configuration (dS m$^{-1}$).

In the ANOCOVA model, the intercept parameter is uniquely estimated for each sampling depth and field, but the slope coefficients are only assumed to change across sampling depths (not across fields). For a specific field, the FSR model requires 3p parameter estimates (for p depths) and sufficient sample data must be acquired in each field to estimate these parameters: thus, for M fields a total of 3pM estimates are needed. In contrast, the CCR model requires just 3p estimates for an entire region. A set of calibration fields can be used to develop the CCR equations and no additional samples are needed. Unfortunately, the accuracy of a CCR model tends to be rather poor. With respect to parameter estimate requirements the ANOCOVA approach represents a compromise between the FSR and CCR models. Once a suitable set of calibration fields are identified to develop the ANOCOVA equations, then additional samples collected in new fields are only used to estimate the field-specific intercept values. The set of initial calibration fields require $pM + 2p$ initial parameter estimates. After developing these calibration estimates a new survey field requires just p parameter estimates (for p depths).

3. Materials and methods

Two study sites were selected to evaluate the three regression techniques as potential regional-scale calibration models of EC$_c$ to EC$_a$. The chosen sites provided a challenging and rigorous evaluation of the regression techniques by testing sites with wide ranging differences in areal extent, soil properties, bed–furrow surface geometry, management, and geographic location, each of which can in influence the EC$_c$ to EC$_a$ calibration. To evaluate the prediction accuracy of the three regression techniques the jack-knifed mean square prediction error (MSPE) was calculated for each model at each study site.

3.1. Description of case study sites

Two study sites of disparate size and property characteristics were selected to evaluate the regression techniques: (1) 2396 ha of the Broadview Water District (latitude: 36°50′08″ N; longitude: 120°33′55″ W) located on the west side of the San Joaquin Valley (WSJ) near Firebaugh, CA, and (2) approximately 100,000 ha on the west side of Kittson County, MN (latitude: 48°46′ N; longitude: 96°56′ W) in the Red River Valley (RRV).

The 2396-ha study area within the Broadview Water District consisted of 37 contiguous quarter sections, with each quarter section encompassing approximately 64 ha of land. The dominant agricultural crops within the survey area were alfalfa (Medicago sativa L.), cotton (Gossypium hirsutum), and tomato (Lycopersicon Lycopersicum). Additionally, about 15% of the total survey area was fallow. Thirty-three of the 37 quarter-sections supported a single crop (or no crop). The bed–furrow surface geometry consisted of 90, 100, 110, and 160 cm bed–furrow designs. All fields had been laser leveled. Four quarter-sections contained multiple fields. The remaining 33 quarter-sections were each single fields of 64 ha. Irrigation methods consisted of flood irrigation down furrows and sprinkler irrigation. Irrigation application rates varied between quarter sections. Soil textures ranged from sandy loam to clay, with the clay content usually increasing with depth.

Kittson County resides in the northwestern corner of Minnesota, consisting of dryland farms growing spring wheat (Triticum aestivum L.), soybeans (Glycine max (L) Merr.), sugar beets (Beta vulgaris L.), and alfalfa (M. sativa L.). Kittson County is part of the Red River Valley (RRV), which mostly consists of a glacial lake plain of nearly level topography with occasional slight depressions bordered by outwash plains and sloping slightly (about 10 cm for each km) toward the Red River. Soils in the RRV represent some of the more saline soils found in the USA, with the negative economic impact of salinity estimated at $50 million annually. Fields varied in size. Unlike Broadview Water District, the soils in Kittson County are relatively uniform, consisting of heavy textures soils with clay contents ranging from 60–80%.

3.2. Electromagnetic induction (EMI) EC$_a$ surveys

Geospatial EC$_a$ measurements were obtained with the Geonics EM38 electrical conductivity meter. The EC$_a$ surveys followed the survey procedures that ultimately led to the protocols and guidelines outlined by Corwin and Lesch (2003, 2005b) and Corwin and Lesch (2013). The EC$_a$ survey was conducted for Broadview Water District in May of 1991 and for Kittson County in the spring of 2007. The surveys consisted of geospatial EC$_a$ measurements taken with mobile EMI equipment. Electromagnetic induction measurements were taken in the horizontal (EM$_h$) and vertical coil configurations (EM$_v$). The inclusion of both EM$_h$ and EM$_v$ measurements provides a shallow (i.e., EM$_h$) and deep (i.e., EM$_v$) measurement of electrical conductance (EC), which provides information regarding the salinity profile with depth. The inclusion of both measurements in soil sample site selection often enhances the characterization of the spatial variability of EC.

The EMI survey for Broadview Water District was one of the first landscape-scale EC$_a$ surveys using first-generation mobile EMI equipment (i.e., Geonics single-dipole EM38) that did not take a steady stream of simultaneous EM$_h$ and EM$_v$ EC$_a$ measurements. Rather, the mobile EMI equipment went from one point to the next, stopping at each point to take an EC$_a$ measurement in the vertical coil configuration and then in the horizontal coil configuration. Measurements were taken in a 10-by-10 grid within each quarter section (i.e., 64 ha), totaling 3800 locations for the entire Broadview Water District.

For Kittson County, EC$_a$ surveys were conducted on 20 fields using mobile EMI equipment that takes a continuous stream of EC$_a$ measurements every 5 s (i.e., Geonics dual-dipole EM38 or EM38DD). The 20 fields were identified using a stratified random sampling of enhanced vegetative index (EVI) data from 7 years of Moderate Resolution Imaging Spectroradiometer (MODIS) imagery as described in Lobell et al. (2010). Geospatial EC$_a$ readings were taken at 2000–5000 locations within each field. Specific details describing the EC$_a$ survey and soil sampling of Kittson County are in Lobell et al. (2010). The EC$_a$-directed
sampling survey followed the protocols outlined by Corwin and Lesch (2005b).

3.3. Soil sampling

Apparent soil electrical conductivity serves as a surrogate to characterize the spatial variation of those soil properties that are found to influence ECa within a field. In both studies, soil sample sites that reflect the range and variation in ECa were selected. To achieve this, the model-based sampling strategy in the ESAP software package was used to select the calibration sampling locations (Lesch et al., 1995a, 1995b, 2000). The model-based sampling strategy is a response surface sampling design, which in essence creates a 3-D surface of the ECa measurements and based on the range and variation selects locations that characterize the ECa variation while maximizing the distances between adjacent sampling locations. For Broadview Water District, 4–8 sample site locations were identified in each field using response surface sampling designs. However, only fields with six or more sample site locations were used to develop FSR, CCR, and ANOCOVA calibration models. For Kittson County, six sample site locations were identified in each field using response surface sampling designs. Soils samples were collected at 30-cm depth increments for both Broadview Water District and Kittson County. Samples were collected to a depth of 1.2 m in Broadview Water District and 1.5 m in Kittson County. A saturation extract of each soil sample was prepared and the electrical conductivity in Broadview Water District and Kittson County. Samples were collected at 30-cm depth increments for both Broadview Water District and Kittson County. A saturation extract of each soil sample was prepared and the electrical conductivity of the saturation extract (ECe, dS m−1) was measured using the method presented in Rhoades (1996). Saturation percentage and water content were determined gravimetrically.

Fig. 1 shows the 43 contiguous fields of Broadview Water District, with 315 sample site locations. Only the 38 fields with six or more sample site locations were used for calibration. Kittson County contained 120 sample site locations, i.e., six locations in each of 20 non-contiguous fields located throughout the county. Additional details concerning the ECa-directed soil sampling are presented in Corwin et al. (1999) for Broadview Water District and Lobell et al. (2010) for Kittson County.

3.4. Statistical evaluation of prediction error

The prediction accuracy of the three regression techniques was evaluated by a comparison of their jack-knifed mean square prediction error (MSPE) shown in Eq. (5):

\[
MSPE = \frac{1}{N_{jk}} \sum_{i,j,k} (y_{ijk} - \hat{y}_{ijk(-i)})^2
\]

Table 1 Comparison of FSR (field specific regression), ANOCOVA (analysis of covariance), and CCR (common coefficient regression) model calibration statistics for 4 depth increments of 38 contiguous fields in the Broadview Water District study site with 6–8 soil sample calibration sites per field.

<table>
<thead>
<tr>
<th>Model</th>
<th>Statistic</th>
<th>Soil depth increment (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0–30</td>
</tr>
<tr>
<td>FSR</td>
<td>R²</td>
<td>0.666</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.104</td>
</tr>
<tr>
<td>ANOCOVA</td>
<td>R²</td>
<td>0.491</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.112</td>
</tr>
<tr>
<td>CCR</td>
<td>R²</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.188</td>
</tr>
</tbody>
</table>

\[ R² = \text{coefficient of determination; MSE = mean square error.} \]

where \( y_{ijk} = \ln(\text{ECa}_{ijk}) \) and \( \hat{y}_{ijk(-i)} \) represents the model predicted \( \ln(\text{ECa}_{ijk}) \) where the \( i^{th} \) observed natural log salinity measurement has not been used to calibrate the model. This ‘leave-one-out’ approach is commonly used as a quantitative diagnostic tool to test the robustness and accuracy of regression models used for prediction (Myers, 1986).

3.5. Validation of the ANOCOVA approach for Kittson County, MN

Once the ANOCOVA calibration models have been established for each sampling depth from the form of the fitted equations defined by Eq. (6),

\[
\ln(\hat{y}_{ijk}) = \mu_{ijk} + \beta_{v(i)} \ln\left(EM_{v(i)}\right) + \beta_{h(i)} \ln\left(EM_{h(i)}\right)
\]

where \( \hat{y}_{ijk} \) represents the predicted \( \ln(\text{ECa}_{ijk}) \) for the \( i^{th} \) field site, \( j^{th} \) depth, and the \( k^{th} \) field and \( \mu_{ijk} \) represents the estimated intercept for the \( j^{th} \) depth and the \( k^{th} \) field, then the EMv and EMh slope coefficients, \( \beta_{v} \) and \( \beta_{h(j)} \), respectively, can be extracted and used for all future field surveys conducted across the region. In any new field surveyed in the region, these EMv and EMh slope coefficients can be used to estimate the field specific intercepts and associated model mean square error (MSE). This can be done using as few as two sampling locations in the field by following the four steps below:

Step 1: For each \( i^{th} \) sample location and \( j^{th} \) depth at \( n \) total locations in a field compute the shifted intercept residual \( \omega_{ij} \) using Eq. (7),

\[
\omega_{ij} = \ln(\text{ECa}_{ij}) + \beta_{1(i)} \ln\left(EM_{v(i)}\right) + \beta_{2(i)} \ln\left(EM_{h(i)}\right)
\]

Fig. 1. Physical locations of all 315 soil sample sites for 37 quarter sections of Broadview Water District. Soil samples were taken in May of 1991. Boundary lines delineate 43 fields. Thirty-eight of the 43 fields contained six or more sample sites, which were used for calibration.
4. Results and discussion

4.1. Broadview Water District

Table 1 displays a summary of the model statistics (i.e., $R^2$ and mean square error statistics, MSE) for the FSR, ANOCOVA, and CCR calibration models of Broadview Water District at each 30-cm depth increment down to 1.2 m. The data used was restricted to the 38 contiguous fields having 6–8 soil sample calibration sites because the FSR models could not be accurately calibrated using less than six sites. The $R^2$ and MSE statistics essentially measure how well each model “fits” the sample data. Without question, the CCR models fit the sample data the worst. The FSR models for all four depth increments produce the best-fit model statistics with the highest $R^2$ and lowest MSE. However, the ANOCOVA models outperform the FSR models on a prediction accuracy basis (i.e., MSE) as shown in Table 2, with the ANOCOVA models for each depth increment having the lowest MSE. On a per field basis, the ANOCOVA models produce smaller MSE estimates in 30 out of 38 fields (79%).

In the Broadview Water District survey, the predictions can be improved using advanced spatial modeling techniques because the fields are spatially contiguous. However, the main point is that the simple ANOCOVA approach works very well, particularly with respect to prediction accuracy, even though there is significant between-field textural variation, diverse irrigation management practices, and a wide range of crops and surface geometry conditions.

4.2. Kittson County, MN

Table 3 shows the model statistics for the FSR, ANOCOVA, and CCR calibration models of Kittson County, MN, at each 30-cm depth increment down to 1.5 m for the 20 non-contiguous fields analyzed in this study. As was the case for Broadview Water District, that the FSR calibration models have the best-fit statistics for all five depths with the ANOCOVA model a close second and CCR model trailing. However, Table 4 shows that the ANOCOVA calibration models produce the smallest MSE estimates for all five depths. Additionally, 15 of the 20 fields exhibit smaller MSE estimates when compared to the FSR calibration models.

Further data that clearly shows the viability of the ANOCOVA approach as a means of calibrating EC$_e$ measurements to salinity (i.e., EC$_s$) is presented in Fig. 2, which shows the measured EC$_s$ versus jack-knifed predicted EC$_s$, averaged over 0–150 cm for the 20 fields in Kittson County. The correlation coefficient between measured and predicted EC$_s$ over 0–150 cm at 120 locations is 0.951.

![Fig. 2. Measured EC$_s$ versus jack-knifed predicted EC$_s$, averaged over 0–150 cm for 20 fields in Kittson County, MN, using the ANOCOVA calibration model. Solid line represents the 1:1 line.](image)
Another way to evaluate the prediction reliability of the ANOCOVA approach is to look at the predictions on a field basis for the MSPE range as shown in Table 5. Table 5 summarizes the distribution of the ANOCOVA model jack-knifed MSPE estimates into four classes; these classes “grade” the reliability of the salinity predictions. Overall, 75% (15 out of 20 fields) exhibit excellent to fair (Grade A, B, or C) prediction reliability. Estimates of MSPE greater than 0.4 suggest that the salinity levels in a particular field are not well described, or more specifically are not strongly correlated with associated ECα survey data measured with EMI equipment; only 5 of the 20 fields fall into this “unacceptable” class (Grade U). Fig. 3a–d shows the distribution of measured versus jack-knifed predicted salinity (i.e., ECe) for the groups of fields exhibiting A, B, C, and U prediction accuracy grades, respectively, based on the criteria in Table 5. The scatter of points around the 1:1 line in Fig. 3d shows the lack of one-to-one correspondence between measured and predicted ECe in the five fields graded as unacceptable. The Kittson County electromagnetic induction ECα slope coefficients of EMh and EMv estimated for ANOCOVA calibration Eq. (6) are shown in Table 6 for the top three depth increments (0–30, 30–60, and 60–90 cm) and estimating the intercepts for each of these depth increments from Eqs. (7)–(9), ANOCOVA predicted salinity levels (i.e., ECe) were estimated from the Kittson County validation data set. The ANOCOVA predicted ECes are compared to observed ECes from the validation data set for the three depth increments in Fig. 4. The correlation coefficient for this data is 0.845, indicating high prediction accuracy well within acceptable limits. On a field-by-field basis, the distribution of MSE into the various grades based on the criteria shown in Table 5 indicates that only Field 4 with three sample locations had an unacceptable grade (Table 7).

4.3. Validation of the ANOCOVA approach for Kittson County, MN

Validation of the traditional FSR calibration technique was not possible because of insufficient data in the Kittson County validation data set. In contrast, the previously developed ANOCOVA models can be easily validated with the limited validation data set (i.e., EMh and EMv, measurements and soil samples taken at 30-cm increments to a depth of 90 cm at only 3–4 locations). Using the ECe slope coefficients of EMh and EMv estimated for ANOCOVA calibration Eq. (6) shown in Table 6 for the top three depth increments (0–30, 30–60, and 60–90 cm) and estimating the intercepts for each of these depth increments from Eqs. (7)–(9), ANOCOVA predicted salinity levels (i.e., ECe) were estimated from the Kittson County validation data set. The ANOCOVA predicted ECes are compared to observed ECes from the validation data set for the three depth increments in Fig. 4. The correlation coefficient for this data is 0.845, indicating high prediction accuracy well within acceptable limits. On a field-by-field basis, the distribution of MSE into the various grades based on the criteria shown in Table 5 indicates that only Field 4 with three sample locations had an unacceptable grade (Table 7).

5. Conclusion

A comparison of ANOCOVA versus FSR shows that one parameter estimate for each field is needed for ANOCOVA, rather than three for FSR, resulting in greater estimation precision and increased model stability as long as the ANOCOVA assumption holds. In addition, the ANOCOVA approach requires less sample data per field once the basic ANOCOVA models have been developed. The ANOCOVA modeling approach completely outperformed the CCR equations. Admittedly, the ANOCOVA approach requires at least two sampling locations in each new field, but the ANOCOVA approach can be used to assess prediction accuracy (i.e., model reliability) based on a limited sample set. The field-specific intercept estimates in the ANOCOVA modeling approach successfully adjust for the majority
of field specific effects that tend to shift ECa data patterns multiplica-
tively from one field to the next under a log transformation.

The ANOVA modeling approach performs well and represents
a viable regional-scale calibration technique. It represents a practical
compromise between individually calibrating a regression model to
each field and forcing a common (static) model across all fields in
the survey area. Its simplicity is particularly appealing. Once the
slope coefficients have been determined, a practitioner does not
need any advanced statistical concepts to implement the technique.
All calculations are easily performed in a spreadsheet.

Fields can be calibrated with as few as two sampling locations once
the ANOVA equations have been developed. The ramifications of
this finding are significant for irrigation districts in the southwestern
USA, the National Resource Conservation Service (NRCS), or any agency
or consulting firm that conducts regular field-scale ECe-directed
sampling surveys using EMI within a basin or region. By pooling the ECa
and ECe data over a region and using the ANOVA approach, the cali-
b ration of ECa to ECe for any new

\[ y_{ijk} = \mu + \beta_i (EMc) + \gamma_j (EMh) + \epsilon_{ijk} \]

where \( y_{ijk} \) is the predicted EC in the \( i \)-th field, \( j \)-th depth, and the \( k \)-th field and \( \beta_i \) is estimated intercept for the \( i \)-th depth and the \( k \)-th field.

\[ 1 \text{ ANOVA calibration Eq. (6):} \quad y_{ijk} = \mu + \beta_i (EMc) + \gamma_j (EMh) + \epsilon_{ijk} \]

Table 6

<table>
<thead>
<tr>
<th>Sample depth (j) in cm</th>
<th>Kittson County electromagnetic induction ECa slope coefficients for ANOVA calibration Eq. (6)†</th>
<th>( \beta_1 (EMa) )</th>
<th>( B_0 (EMa) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–30</td>
<td>-0.747</td>
<td>1.282</td>
<td></td>
</tr>
<tr>
<td>30–60</td>
<td>1.296</td>
<td>-0.196</td>
<td></td>
</tr>
<tr>
<td>60–90</td>
<td>3.965</td>
<td>-1.740</td>
<td></td>
</tr>
<tr>
<td>90–120</td>
<td>3.238</td>
<td>-1.607</td>
<td></td>
</tr>
<tr>
<td>120–150</td>
<td>3.033</td>
<td>-1.414</td>
<td></td>
</tr>
</tbody>
</table>

† ANOVA calibration Eq. (6):

<table>
<thead>
<tr>
<th>Field</th>
<th># of Sample sites</th>
<th>Computed MSE</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>0.176</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.133</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.134</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.593</td>
<td>U</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.254</td>
<td>C</td>
</tr>
</tbody>
</table>

Fig. 4. Measured EC versus ANOVA predicted EC for Kittson County validation data set consisting of five fields and three depth increments of 0–30, 30–60, and 60–90 cm. Solid line represents the 1:1 line.

References


Caccetta, P., Dunne, R., George, R., McFarlane, D., 2010. A methodology to estimate the future

extremal oil of dryland salinity in the southeast of Western Australia. J. Environ. Qual. 39, 26–34.


Corwin, D.L., Lesch, S.M., 2005a. Apparent soil electrical conductivity measurements in ag-


versity Press, New York, NY.

Lesch, S.M., 2005. Sensor-directed response surface sampling designs for characterizing

Lesch, S.M., Corwin, D.L., 2008. Prediction of spatial soil property information from an-
cillary sensor data using ordinary linear regression: model derivations, residual as-
sumptions and model validation tests. Geoderma 148, 130–150.

Lesch, S.M., Strauss, D.J., Rhoades, J.D., 1995a. Spatial prediction of soil salinity using elec-
 tromagnetic induction techniques: I. Statistical prediction models: a comparison of

Lesch, S.M., Strauss, D.J., Rhoades, J.D., 1995b. Spatial prediction of soil salinity using elec-
tromagnetic induction techniques: 2. An efficient spatial sampling algorithm suitable for

Lesch, S.M., Rhoades, J.D., Corwin, D.L., 2000. ESAP-95 version 2.10R: user manual and tu-
torial guide. Research Rpt. 146, USDA-ARS, U.S. Salinity Laboratory, Riverside, CA.

Lesch, S.M., Corwin, D.L., Robinson, D.A., 2005. Apparent soil electrical conductivity map-

Massoud, F.I., 1981. Salt affected soils at a global scale and concepts for control. FAO Land
and Water Development Div., Technical Paper: Food and Agriculture Organization of
the United Nations, Rome, Italy (21 pp.).

Acknowledgments

The authors wish to acknowledge the numerous hours of diligent
technical work performed in the laboratory by two technicians, Lena
Ting and Nahid Vistein, whose efforts and conscientiousness were cru-
cial to the successful analysis of Broadview Water District soil samples.
The authors also acknowledge NRCS scientists and field staff includ-
ing Dr. James Doolittle, who conducted the EMI ECa field surveys in
Kittson County, and Mike Ulmer, Keith Andersen, Dave Potts, Manuel Matos, and Matt Baltes, who obtained ECc-directed soil samples in
Kittson County.

Appendix A. Practical examples of regional-scale calibration procedures.

Table 7

Kittson County ANOVA calibration model validation results for each of the five fields
based on the mean square error (MSE) classified by the criteria in Table 5.

<table>
<thead>
<tr>
<th>Field</th>
<th># of Sample sites</th>
<th>Computed MSE</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>0.176</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.133</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.134</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.593</td>
<td>U</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.254</td>
<td>C</td>
</tr>
</tbody>
</table>


