#1275

Statistical Modeling and Prediction Methodologies for Large Scale Spatial Soil Salinity Characterization: A case study using calibrated electromagnetic measurements within the Broadview Water District.

Lesch, S.M., J.D. Rhoades, and D.L. Corwin

United States Salinity Laboratory

Technical Report # 13 1

4500 Glenwood Drive

Riverside, California 92501

INTRODUCTION

This report documents the statistical analysis and modeling procedures used to characterize the spatial soil salinity features within 5920 acres of the Broadview Water District. The soil salinity and electromagnetic induction (EM) data analyzed within this report are from a comprehensive salinity survey conducted during the months of April and May, 1991.

The general goal behind the Broadview survey project was to demonstrate how to rapidly and efficiently conduct a multiple depth soil salinity survey over a large area of agricultural landscape encompassing different types of cropped farmland. To achieve this goal the survey relied primarily on the collection of electromagnetic induction data (which can be gathered quite quickly), augmented with a much smaller subset of soil salinity samples. The idea was to then use the more abundant EM data to increase our knowledge of the spatial soil salinity levels by exploiting the correlation structure between the two data sets.

The details concerning the Broadview Water District survey area and EM/soil sampling design are presented in the SURVEY DESCRIPTION / METHODS section of this report. A series of modeling techniques used for predicting soil salinity levels from EM data are introduced in the STATISTICAL MODELING & PREDICTION METHODOLOGY section. Finally, the overall effectiveness of these modeling methodologies is examined within the RESULTS & DISCUSSION section.

All statistical programming for this project was carried out using SAS. Additionally, please note that the details concerning the analytical chemistry analysis; e.g., the actual laboratory techniques used for measuring soil salinity (ECe dS/m basis) are not included here. These techniques have been previously documented (Rhoades, et. all, 1989a,b).

SURVEY DESCRIPTION / METHODS

Survey Area

The total survey zone within the Broadview Water District consisted of 37 connected quarter sections, with each quarter section encompassing approximately 160 acres of land. Maps of the survey zone and sample site locations are shown in Figure 1. The identification codes associated with each of the quarter sections shown in Figure 1 were supplied by the Broadview Water District at the time of the survey.

The three dominant agricultural crops within the survey were cotton, tomatos, and alfalfa. Additionally, about 15% of the total survey area was fallow. Thirty three of the 37 quarter sections were supporting a single crop (or no crop) and were homogeneously managed at the time of the initial survey expedition. Irrigation methods consisted mostly of modified flood and sprinkler application techniques and application rates varied between quarter sections. Soil types ranged from loam to clay, with the clay content usually tending to increase with depth.

EM Survey and Soil Sampling Design

Electromagnetic induction measurements were acquired with an EM-38 meter on an 8 by 8 centric, systematic survey grid (100 meter spacing) at 64 sites within each quarter section. Four measurements were taken at each site; one pair of horizontal and vertical readings 10 cm above the soil surface and a second pair of readings 50 cm above the soil surface. Electromagnetic readings were taken directly above the center of the furrow in all fields with either 30 inch or 40 inch beds (supporting alfalfa and cotton, respectively), and directly above the center of the bed in the remaining tomato fields with 60 inch beds. All the electromagnetic readings were taken in line with (e.g., parallel to) the bed / furrow direction.

Spatial locations of all survey sites were acquired using portable GPS (global positioning system) equipment. Differentual correction of all GPS location data resulted in an absolute x,y coordinate accuracy of 5 to 10 meters.

Eight of the 64 survey sites were sampled for soil salinity; four sites were chosen which represented the observed range within the 10 cm electromagnetic readings and four additional sites were chosen randomly. Four of the 37 quarter sections had more than 8 sites sampled for soil salinity due to the presence of multiple fields within the quarter sections at the time of sampling (quarter sections 4-1, 4-3, 9-3, and 14-1). Each site was sampled in one foot increments down through four feet. In all, 1252 soil samples from 313 sample sites were analyzed in the laboratory for soil salinity (ECe, dS/m), gravimetric saturation percent (Sp), and percent water content (Pw). Additionally, soil samples from two sites within each quarter section were also analyzed for cation and anion levels, SAR, and boron toxicity concentration.

STATISTICAL MODELING & PREDICTION METHODOLOGY

Regression Model Development / Rationale

The goal of the survey process was to use the observed electromagnetic induction readings to estimate the soil salinity levels at all the non-sampled EM survey sites. This was accomplished by developing multiple linear regression equations which predicted the soil salinity levels using the EM readings. The usefulness of such an approach has been previously documented (Lesch et al, 1992).

The modeling approach used for this survey data required the development of individual, depth specific regression equations within each quarter section. To facilitate this modeling approach, the EM readings were transformed and decorrelated. The natural log transformation of each EM measurement was acquired and then principal component analysis was used to decorrelate these ln EM data within each quarter section. The first 2 principal component scores typically explained more than 99% of the total log EM variability and were used in place of the original four ln EM readings during all subsequent regression modeling. This was done in order the reduce the number of variables in the regression equations without degrading the models predictive capabilities.

In addition to electromagnetic induction data, knowledge of the soil texture has also been shown to increase the accuracy of salinity prediction equations (Rhoades, et. all, 1989c; Rhoades and Cot-win, 1990). The laboratory measured saturation percent (Sp) can be used as a measurement of soil texture; in general, the higher the Sp the greater the clay content. A preliminary analysis revealed that in most of the quarter sections the average Sp level within the first 4 feet of soil could be predicted reasonably well using the following 1st order trend surface model:

$$\text{predicted Sp} = pSp = \mathbf{B}_0 + \mathbf{B}_1 x + \mathbf{B}_2 y + \mathbf{B}_3 xy, \qquad (1)$$

where x and y represented the relative spatial coordinates within the field (quarter section) and $\mathbf{B}_0$ through $\mathbf{B}_3$ represented empirically determined parameter estimates. Hence, equation 1 was fit to each of the 37 quarter sections and then used to generate pSp texture values at every one of the original survey sites. This was carried out before the development of salinity models so that the pSp data could be used as an additional independent parameter within the soil salinity prediction equations.

Within a given quarter section, the soil salinity prediction model for a specific sample depth was defined to be:

$$\text{predicted } \ln(ECe) = lECe = B_0 + B_1 pcs1 + B_2 pcs2 + B_3 pcs12 + B_4 pSp, \qquad (2)$$

where *pcsl* represented the first principal component score, pcs2 represented the second score, pcsl2 represented an interaction term, e.g., $pcs12 = pcsl \times pcs2$, *pSp* represented the texture variable defined in equation 1, and B_0 through B, represented empirically determined parameter estimates. Note that since there were 37 quarter sections, each sampled at 4 distinct depths, a total of 148 individual salinity prediction equations had to be estimated.

Regression Model Hierarchy / Model Selection

During the initial model estimation process within some of the quarter sections, one or more of the parameter estimates associated with equation 2 often did not appear to be significantly different from 0 (as judged by the individual t-scores). Furthermore, occasionally the estimation of equation 2 resulted in a model which was clearly "over-fit"; e.g., the combination of independent regressor variables resulted in one or more extremely high leverage points (having HAT diagonal elements nearly equal to 1). To rectify these problems the following four "restricted" salinity prediction models were also estimated at each depth within every quarter section:

$$lECe = B_0 + B_1 pcs1 + B_2 pcs2 + B_3 pcs12, \qquad (3)$$

$$lECe = B_0 + B_1 pcs1 + B_2 pcs2 + B_4 pSp, \qquad (4)$$

$$1ECe = B_0 + B_1 pcs1 + B_2 pcs2, \qquad (5)$$

$$lECe = B_0 + B_1 pcs1, \qquad (6)$$

Equations 3 through 6 all represented sub-models of equation 2 in the sense that each of these restricted equations contained some, but not all, of the original four independent parameters contained in equation 2. Additionally, note that this modeling approach was essentially hierarchical; the second principal component score was never included in a model without the first score being present, and the interaction and/or spatial parameters were not included without the simultaneous inclusion of both principal component scores.

Since the ultimate purpose of the salinity prediction models was accurate salinity estimation, rather than parameter testing and/or interpretation per se, the PRESS (Prediction Sum of Squares) statistic was used as the criterion for final model selection (Myers, 1986; Weisberg, 1985). Selecting the model with the smallest PRESS achieved two desirable goals. First, it eliminated the initial model over fitting problem. Models with high leverage points produced high PRESS scores and hence these models, which typically contained 3 or 4 parameters, were rejected in favor of a more simple model (containing only 1 or 2 parameters). Second, all the parameter estimates associated with the model having the smallest PRESS typically appeared significantly different from 0. Thus, redundant and/or non-significant independent variables generally did not end up in the final prediction equation.

After the final regression equations had been chosen, a full analysis of the residuals was carried out. This analysis included checking the residuals for outliers, assessing the Normality and homogeneous variance assumptions, and constructing semi-variograms to detect the presence of spatial autocorrelation. Jack-knifed prediction statistics were also computed in order to determine the overall predictive capabilities of the regression models.

Salinity Prediction / Estimation Techniques

Once the final regression equations had been validated, a range of salinity characteristics could be predicted. Define $\gamma_{i,j}$ to be the predicted natural log of the soil salinity for the ***jth*** site at the ***ith*** depth within a given quarter section, and let σ_i^2 represent the estimated MSE (mean square error) from the regression model used to generate this prediction. Furthermore, assume that all the individual site predictions are unbiased estimates of the true log salinity values and that the error associated with these predictions follows a Normal distribution with mean 0 and variance σ_i^2. For a given sample depth ***i,*** the following soil salinity estimates were constructed:

I. the predicted soil salinity at site j

$$\lambda_j = \exp(\gamma_{i,j})$$

II. the probability that the soil salinity at site j falls within the range [a,b]

$$\theta[a,b]_j = \int_{\omega1,\omega2} \Phi(x)dx,$$

where $\omega1 = (\ln(a)-\gamma_{i,j})/\sigma_i$
$\omega2 = (\ln(b)-\gamma_{i,j})/\sigma_i$
and $\Phi(x)$ represents a standard normal density

III. the predicted median soil salinity for the entire field

$$\Lambda = \exp(\textstyle\sum_{j=1,N} \upsilon_j \gamma_{i,j}),$$

where υ_j represents a weighting coefficient which sums to 1 (for centric systematic survey grids, υ_j = 1/N for all j)

IV. the % area within a field falling within the range [a,b]

$$\Theta[a,b] = 100\textstyle\sum_{j=1,N} \upsilon_j \theta[a,b]_j$$

Since site predictions were made at multiple depths, it was further assumed that the composite error structure followed a multivariate Normal distribution with mean vector 0 and covariance matrix Υ. Under these conditions it was possible to define an additional salinity characteristic, ζ, referred to as a “profile shape estimate”.

For surveys conducted at two depths, the profile shape estimate at site j is defined as:

$$\zeta_j = \exp(\gamma_{1,j}) \;/\; (\exp(\gamma_{1,j}) + \exp(\gamma_{2,j})), \qquad \text{where } 0 < \zeta_j < 1.$$

Dividing the numerator and denominator on the right hand side by $\exp(\gamma_{1,j})$ and then rearranging terms yields an estimate of the lower to upper salinity ratio:

$(1/\zeta_j)\text{-}1 = \exp(\gamma_{2,j})/\exp(\gamma_{1,j}).$

Taking the natural log on both sides of this equation then yields:

$\ln((1/\zeta_j)\text{-}1) = \gamma_{2,j} - \gamma_{1,j}.$

Hence, a nonlinear function of the profile shape estimate, or equivalently, the natural log of the salinity ratio estimate, can be expressed as a simple linear difference between the predicted log salinities at the two sample depths.

For this survey, which encompassed four sample depths, the following profile estimates were constructed:

V. the composite profile shape, ζ, and the geometric salinity ratio, δ, at site j

$$Cj = [\exp(\gamma_{1,j})\exp(\gamma_{2,j})]^{0.5} / ([\exp(\gamma_{1,j})\exp(\gamma_{2,j})]^{0.5} + [\exp(\gamma_{3,j})\exp(\gamma_{4,j})]^{0.5})$$

$$\delta_j = (1/\zeta_j)\text{-}1$$

VI. the probability that the geometric salinity ratio at site j falls within the range [a: 1,b: 1]

$\delta[a,b]_j = \int_{\alpha1,\alpha2} \Phi(x)dx,$ where al = $(\ln(a)\text{-}mu)/\sigma_p$
a2 = $(\ln(b)\text{-}mu)/\sigma_p$
mu = $0.5(\gamma_{4,j}+\gamma_{3,j}\text{-}\gamma_{2,j}\text{-}\gamma_{1,j})$
σ_p = appropriate error term (found from the covariance matrix Υ)
and $\Phi(x)$ again represents a standard normal density

VII. and finally, the % area within a field having profile ratios falling within the range [a: 1,b: 1]

$A[a,b] = \sum_{j=1,N} \upsilon_j \delta[a,b]_j.$

From the final regression equations and formulas I through VII, the following salinity characteristics were estimated within each individual quarter section:

1) the median salinity levels, by depth, which were used to describe the “typical” salinity level present within the field,

2) a salinity range matrix: e.g., the % area of the field, by depth, falling into one of the

four following salinity ranges, [0,2], [2,4], [4,8] and [> 8], which was used to describe the salinity variability occurring within the field,

3) the % area of the field having ratios exceeding 2:1, used to estimate the potential for large relative increases in upper root-zone salinity, and

4) the typical salinity ratio, constructed from the median profile shape estimate and used to categorize the efficiency of recent irrigation management practices.

Additionally, the predicted salinity levels at each depth were classified into one of the four ranges shown above within each quarter section, and then these interval data were used to make composite spatial salinity range and salinity ratio maps for the whole survey area.

Jack-knifed variance estimates which reflected the degree of salinity prediction accuracy were computed for both the median estimates and the salinity range matrix. Approximate variance estimates for each median field salinity level were computed as $\varrho_i^2 = \mathrm{PRESS}_i/n^2$, where PRESS, was the jack-knifed prediction sum of squares associated with the prediction equation at depth i and n was set equal to the original soil salinity calibration sample size. An approximate 95% confidence interval for each median estimate was then computed as $\exp(\sum_{j=1,N} \upsilon_j \gamma_{i,j} \pm 2\varrho_i)$.

To estimate the average error associated with the salinity range matrix predictions, class estimates within each cell were recomputed using the jack-knifed predicted salinity estimates associated with the calibration sample. The absolute differences between each jack-knifed and true cell estimate were then averaged to form an AAPE (average absolute percent error) value. Hence, if the AAPE associated with a range matrix was 5%, then on the average the percent area estimates based on the jack-knifed calibration sample set were found to be within 5% of the true values.

RESULTS AND DISCUSSION

Model Validation

Table 1 lists some summary statistics from the residual analysis performed after the model fitting process. Residuals associated with each of the four sampling depths appeared symmetric and approximately Normally distributed. Additionally, no “outliers” or unusual observations were found. The combined variance estimates across all 37 quarter sections were 0.0868, 0.1116, 0.1270, and 0.1441 for the four sampling depths, respectively.

Residual semivariograms were estimated within each quarter section and then averaged across all the quarter sections to produce an “average isotropic residual semivariogram plot” for each sample depth (Isaaks & Srivastava, 1988). These plots are shown in Figure 2; note that none of the semivariograms suggest that the residuals were spatially autocorrelated. This was encouraging, since spatially autocorrelated residuals would have implied that the regression equations were producing biased salinity predictions.

The jack-knifed residuals, computed by subtracting the jack-knifed predictions from the observed log salinity values, were also analyzed. Summary statistics for these jack-knifed residuals are shown in the bottom of Table 1. Like the standard residuals, the jack-knifed residuals appeared Normally distributed and spatially uncorrelated. Note that the combined variance estimates were typically about 50% larger than the variance estimates associated with the standard residuals, implying that the standard residual variance estimates under estimated the true prediction error of the regression models. This was probably a result of estimating the regression equations with such small sample sixes (usually n = 8). This also demonstrates why variance estimates from the jack-knifed residuals, rather than the standard residuals, are more appropriate for the construction of confidence intervals, etc.

Regression model coefficients of determination (R^2) are shown in Figure 3. Within the O-l sample depth, the predictive capabilities of the regression models tended to vary considerably. However, the predictive capabilities improved as the sampling depth increased. This was due to the increased salinity variability at the deeper depths, which tended to be the primary factor in determining the magnitude of the electromagnetic induction readings. The O-l ft salinity levels usually were consistently low within most of the quarter sections, and therefore the EM signals alone were often incapable of predicting the relatively minor changes in the spatial soil salinity levels. (More recent experimental research has indicated that the addition of a third electromagnetic measurement, such as a four-probe or wenner reading restricted to the near surface soil layer, can significantly increase the salinity prediction capabilities of an estimated regression equation.)

The overall correlations between the observed and predicted log salinity values for the four sample depths were 0.839, 0.905, 0.918, and 0.904. Plots of these data are shown in Figure 4. The percentage of correct classifications at each depth within the four salinity ranges [0,2], [2,4], [4,8], and [>8] were 75.4, 72.2, 74.1, and 68.7 percent, respectively. The predicted versus observed salinity classification results are shown in Table 2. Taken together, these correlation and classification values suggested reasonably good “prediction resolution”; e.g., the modeling approach appeared to successfully explain the majority of log salinity variability across the survey area.

The overall correlations between the observed and jack-knifed predicted log salinities for the four sample depths were 0.642, 0.796, 0.818, and 0.817, respectively. Plots of these data are shown in Figure 5. The percentage of correct jack-knifed classifications at each depth were 59.4, 57.2, 61.0, and 58.8 percent, respectively. This jack-knifed classification data is shown in Table 3.

While not excellent, these values did confirm that the “prediction reliability” was adequate.

Classification results pertaining to both the observed versus predicted and observed versus jack-knifed predicted salinity ratios are shown in Table 4. Using a dichotomous classification (ratio < 2:l and ratio > 2:l) the percentage of correctly classified ratios was 86.3 before jack-knifing and 77.6 percent after jack-knifing, respectively.

Overall, the analysis of the jack-knifed predictions suggested that the modeling approach could be successfully employed to combine the electromagnetic induction survey data with the sampled soil salinity data, and hence improve the final predictions of the spatial salinity characteristics.

Soil Salinity Characterization within Individual Quarter Sections

Summary statistics from the regression modeling approach applied to quarter section 11-4 are shown in Table 5. The following model attributes within each sampling depth are shown; the final non-zero independent parameter estimates contained within the regression equation, the model’s $\mathbf{R^2}$ estimate, the model’s mean square error estimate (MSE), the jack-knifed MSE estimate computed from the PRESS residuals, the correlation between the observed and predicted log salinity values, and the correlation between the observed and jack-knifed predicted log salinity values. Note that the four regression models associated with this quarter section appeared to have good prediction resolution and prediction reliability.

Table 6 compares the calibration sample log salinity mean and standard error estimates to the regression based estimates for this same quarter section. These data demonstrate the usefulness of the regression modeling approach. By exploiting the correlation between the EM survey data and the soil salinity samples, the final standard errors associated with the field mean log salinity estimates were reduced from 27 to 69 percent. Additionally, the regression equations effectively adjust the field mean estimates up or down to reflect the information contained in the more exhaustive EM survey data, and can assign predicted log salinity estimates at all the non-sampled survey sites (hence facilitating the production of more accurate spatial salinity maps).

Along with individual site and field average estimates, the regression equations were used to estimate the percent area in quarter section 11-4 contained within four specific salinity classes (e.g., a salinity range matrix). Table 7 shows how the AAPE (average absolute percent error) was derived from the jack-knifed predictions of the calibration salinity samples. In the column entitled "Obs", the observed percentage of calibration soil salinities falling within the four salinity ranges are shown. The next two columns show the final jack-knifed predicted percentages after applying Formula II to account for the regression model's MSE, as well as the resulting absolute percent errors (APE). The AAPE is simply the average of these APE values.

The final two columns in Table 7 show the jack-knifed predicted percentages and APE values, without correcting for the prediction error. Note that the resulting AAPE estimate increased when the prediction error was not taken into account. In general, the incorporation of the regression model MSE estimates into the salinity range matrix predictions through Formula II resulted in smaller AAPE estimates. The AAPE estimates were computed with and without accounting for the prediction errors during the salinity estimation phase of this study in all 37 quarter sections; 25 quarter sections had smaller AAPE values after adjusting for the prediction error. Furthermore, the mean AAPE decreased from 9.79% to 7.92%; representing about a 20% relative reduction in the overall percent error.

Table 8 shows an example "salinity appraisal sheet" for quarter section 11-4, which summarizes all of the relevant salinity characteristics associated with this quarter section. The first part of Table 8 lists the median soil salinity levels (dS/m), estimated from the mean log salinity values shown in Table 5, along with approximate 95% confidence intervals based on the standard error estimates from the jack-knifed residuals. The second part lists the salinity range matrix estimates (after incorporating all of the additional EM survey data). Finally, the third part list the "typical" salinity ratio and the percent area within the quarter section exceeding a 2:l ratio. The typical ratio was estimated using the median of the predicted profile shape values, which in this case was 0.249. The percent area estimate was derived using Formulas VI and VII.

Salinity appraisal sheets were constructed for each one of the 37 quarter sections during the estimation phase of this study. (In addition to the characteristics shown in Table 8, the final appraisal sheets also contained additional chemical analysis data, such as cation/anion, SAR, and boron levels, etc.) These sheets were then included in the descriptive report submitted to the Broadview Water District and subsequently distributed to the appropriate land owners.

Macro-Salinity Characterization across the Full Survey Area

Once the quarter section soil salinity estimates had been derived, it was possible to combine these data together to characterize the macro salinity patterns across the total survey area. The estimated median soil salinity levels by depth for each of the 37 surveyed quarter sections within the Broadview Water District are shown in Table 9. A graphical representation of these same data is shown in Figure 6 in bar chart format. The dominant trend in these data was not a spatial, but rather a depth effect; the median salinity levels consistently tended to increase with depth throughout the survey area. The spatial trends in these data were weak, at best. In fact, quarter sections with relatively high median salinity levels often occurred next to quarter sections having low median levels (for example, quarter sections 4-3 and 4-4). Overall, the quarter section salinity levels appeared to be primarily a function of management practices, rather than spatial location.

The cumulative salinity range matrix; e.g., the percent area of land, by depth, for the total survey zone of the Broadview Water District falling within the four different salinity ranges is shown in the first part of Table 10. Roughly 27.9% of the total survey area had soil salinity levels less than or equal to 2.0 dS/m within the O-l foot depth, only about 18.4% of the total survey area exceeded 4.0 dS/m, and less than 1.3% of the area exceeded 8.0 dS/m. Within the 3-4 foot depth, 5.1% of the survey area had salinity levels less than or equal to 2.0 dS/m, about 71.3% of the survey area exceeded 4.0 dS/m, and about 31.4% of the area exceeded 8.0 dS/m. There appeared to be consistent trends evident in these percentile estimates; as the depth increased the amount of land classified into the lowest two salinity ranges decreased, while the amount of land falling into the highest two ranges increased. As with the median salinity estimates, this trend reflected the increasing salinity with depth relationship evident throughout the survey area.

Figures 7, 8, 9, and 10 show the spatial salinity distributions across the total survey area for each of the four sample depths. The salinity range classifications were again defined to be [0,2], [2,4], [4,8], and [> 8] dS/m. The spatial relationship between the salinity levels at the deeper depths suggested, to a limited extent, large scale trends across the total survey area in the apparent salinity distribution. For example, at the 3-4 foot depth, a continuous sliver of land below 4.0 dS/m was apparent within the main area of the survey zone (beginning at quarter section 3-2 and slicing all the way down through quarter section 16-3). However, within the first two sample depths the spatial patterns primarily seemed to reflect the significant differences in management practices between fields.

The overall percent area within the total survey zone which exceeded a 2:l ratio is shown at the bottom of Table 10; e.g., 26.88 + 13.63 = 40.51%. Table 11 lists the typical profile ratios and the percent area of land with profile ratios exceeding 2:1 on a quarter section basis, and Figure 11 displays this same information graphically (in bar chart format). Finally, Figure 12 shows the spatial distribution of the dichotomously classified profile ratio data, using a 2:l ratio cutoff value.

The data shown in Table **11** and Figures 11 and 12 indicated that there were 11 quarter sections with pronounced profiles and excessive amounts of land area exceeded a 2:l ratio. These quarter sections were 4-1, 4-2, 4-3, 4-4, 8-4, 9-2, 10-2, 11-4, 14-3, 15-2, and 15-4. Not all of these quarter sections had unusually high salinity levels; for example, the median salinity levels in quarter sections 4-4 and 8-4 were rather low. None the less, these ratio survey data indicated that these quarter sections were most at risk for large relative increases in upper root-zone soil salinity levels, should a net upward flux of salts occur sometime in the future.

Concluding Remarks:

Advantages / Disadvantages of a Regression Model Salinity Estimation Approach

To better understand the strengths and limitations of the regression modeling estimation technique, it is useful to examine some other types of statistical modeling techniques which could have conceivably been used for the purposes of salinity estimation. Two obvious approaches would have been to perform either a Kriging or Cokriging analysis on the survey data. Stable and accurate estimations of the relevant semivariogram models should have been possible, given the relatively large number of salinity samples and EM survey sites within the entire survey area (N, =313 salinity samples per depth and $N_2=2368$ EM survey sites). A complete spatial statistical analysis and prediction model could have then been developed, allowing for the estimation of the same salinity characteristics previously discussed.

Both of these spatial modeling approaches were explored during the initial analysis phase of the survey data, but neither approach produced reliable prediction results. The generally poor performance of these models was understandable, given their underlying assumptions. In either a Kriging or Cokriging model a central model assumption is that of ***stationarity; e.g.,*** the sample data can be grouped together spatially under the assumption that the probability distribution of each random variable is the same. In reality, this assumption would have only been valid if each field (quarter section) within the survey area had supported the same type of crop, subjected to the same type of irrigation practices, and managed in the same manner. However, this was clearly not the case; crop types and cropping cycles were different, irrigation practices varied, and management techniques often changed considerably from one quarter section to the next. Furthermore, due to differences in soil texture and water content levels, two quarter sections with similar salinity levels often exhibited very different electromagnetic induction patterns.

Hence, the single greatest strength of the regression modeling approach was its cost-effectiveness. Individual regression equations which accounted for the field scale deterministic variations present within the survey area could be estimated using a sample size of $n = 8$ sites per quarter section. The estimation of reliable semivariogram models of the spatial salinity distributions within each quarter section would have required far more samples, and thus raised the overall survey cost and/or reduced the total potential survey area.

In general, another intrinsic advantage to a regression modeling approach is that it is easier to implement and relies on less subjective modeling techniques. Model parameter estimation and prediction criteria guidelines are based on standard ordinary least squares theory; furthermore, prediction jack-knifing is computationally simple.

Two disadvantages of regression modeling are 1) that the models do not incorporate spatial information from any additional surrounding survey and/or sample sites into the salinity predictions, and 2) it can not generate a salinity prediction at any non-survey site. Hence, regression models may tend to be less efficient since they ignore the surrounding spatial information, and their spatial point prediction patterns will be determined by the initial survey pattern.

The second restriction is not a serious disadvantage, as long as the initial EM survey grid follows an approximately centric systematic pattern and is dense enough to achieve the required spatial resolution for map production. Additionally, the final salinity predictions can always be interpolated into a denser grid using either Kriging or inverse distance squared algorithms. However, the loss in efficiency can become very serious if the regression residuals exhibit noticeable spatial autocorrelation. In this case the predictions will be spatially biased and the summary salinity estimates may be in serious error. On the other hand, when the residuals are relatively uncorrelated spatially the loss in efficiency will be minimal and is more than offset by the increased cost effectiveness achieved through the reduction in the sample size.

References

Isaaks, E.H. and R.M. Srivastava. 1989. Applied Geostatistics. Oxford University Press. New York, New York.

Lesch, SM., J.D. Rhoades, L.J. Lund and D.L. Corwin. 1992. Mapping soil salinity using Calibrated Electromagnetic Measurements. Soil Sci. Soc. AM. J. 56:540-548.

Myers, R.H. 1986. Classical and Modern Regression with Applications. Duxbury Press. Boston, Massachusetts.

Rhoades, J.D. and D.L. Corwin. 1990. Soil electrical conductivity: Effects of soil properties and application to soil salinity appraisal. Commun. Soil Sci. Plant Anal. 21:837-860.

Rhoades, J.D., N.A. Manteghi, P.J. Shouse and W.J. Alves. 1989a. Estimating soil salinity from saturated soil-paste electrical conductivity. Soil Sci. Soc. Am. J. 53:428-433.

Rhoades, J.D., B.L. Waggoner, P.J. Shouse and W.J. Alves. 1989b. Determining soil salinity from soil and soil-paste electrical conductivities: Sensitivity analysis of models. Soil Sci. Soc. Am. J. 53:1368-1374.

Rhoades, J.D., S.M. Lesch, P.J. Shouse and W.J. Alves. 1989c. New calibrations for determining soil electrical conductivity: Depth relations from electromagnetic measurements. Soil Sci. Soc. Am. J. 53:74-79.

Weisberg, S. 1985. Applied Linear Regression. Second Edition. John Wiley & Sons. New York, New York.

Tables & Figures

Table 1. Residual and jack-knifed residual mean, variance, and quantile estimates, by depth (N=313).

Residuals

Residual Statistics	Depth o-1 ft	1-2 ft	2-3 ft	3-4 ft
mean:	0	0	0	0
variance:	0.0868	0.1116	0.1270	0.1441
quantiles				
95:	0.377	0.480	0.474	0.482
90:	0.279	0.292	0.342	0.321
75:	0.138	0.131	0.136	0.150
50:	0.009	0.004	-0.00 1	0.016
25:	-0.132	-0.126	-0.144	-0.141
10:	-0.3 14	-0.301	-0.323	-0.345
05:	-0.391	-0.443	-0.454	-0.587

Jack-Knifed Residuals

Residual Statistics	Depth o-1 ft	l-2 ft	2-3 ft	3-4 ft
mean:	0.0013	0.0159	0.0015	0.0082
variance:	0.1276	0.1513	0.1764	0.1755
quantiles				
95:	0.538	0.642	0.716	0.661
90:	0.428	0.520	0.548	0.523
75:	0.217	0.245	0.277	0.253
50:	0.014	0.015	-0.005	0.029
25:	-0.234	-0.216	-0.242	-0.229
10:	-0.466	-0.446	-0.482	-0.519
05:	-0.593	-0.626	-0.702	-0.713

Table 2. Classification results, observed versus predicted salinity range data.

Depth o-1 ft		Predicted ECe range							
		[0,2]		[2,4]		[4,8]		[>8]	
Obs	[0,2]	63.2	(55)	36.8	(32)	0.0	(0)	0.0	(0)
ECe	[2,4]	7.7	(13)	86.3	(145)	6.0	(10)	0.0	(0)
range	[4,8]	0.0	(0)	36.4	(20)	63.6	(35)	0.0	(0)
	[>8]	0.0	(0)	33.3	(1)	33.3	(1)	333	(1)

Overall correct classification: (55 + 145 + 35 + 1)/3 13 = 75.4 %

Depth 1-2 ft		Predicted ECe range							
		[0,2]		[2,4]		[4,8]		[>8]	
Obs	[0,2]	71.4	(45)	28.6	(18)	0.0	(0)	0.0	(0)
ECe	[2,4]	10.1	(13)	77.5	(100)	12.4	(16)	0.0	(0)
range	[4,8]	0.0	(0)	13.8	(11)	77.5	(62)	8.7	(7)
	[>8]	0.0	(0)	2.4	(1)	51.2	(21)	46.4	(19)

Overall correct classification: (45 + 100 + 62 + 19)/3 13 = 72.2 %

Depth 2-3 ft		Predicted ECe range							
		[0,2]		[2,4]		[4,8]		[>8]	
Obs	[0,2]	65.6	(21)	31.3	(10)	3.1	(1)	0.0	(0)
ECe	[2,4]	6.5	(7)	75.7	(81)	15.9	(17)	1.9	(2)
range	[4,8]	0.0	(0)	24.0	(18)	693	(52)	6.7	(5)
	[>8]	0.0	(0)	1.0	(1)	20.2	(20)	78.8	(78)

Overall correct classification: (21+ 81+ 52 + 78)/3 13 = 74.1 %

Depth 3-4 ft		Predicted ECe range							
		[0,2]		[2,4]		[4,8]		[>8]	
Obs	[0,2]	43.5	(10)	47.8	(11)	8.7	(2)	0.0	(0)
ECe	[2,4]	8.4	(7)	59.1	(49)	32.5	(27)	0.0	(0)
range	[4,8]	0.0	(0)	24.2	(22)	67.0	(61)	8.8	(8)
	[>8]	0.0	(0)	0.0	(0)	18.1	(21)	81.9	(95)

Overall correct classification: (10 + 49 + 6 1 + 95)/3 13 = 68.7 %

Table 3. Classification results, observed versus jack-knifed predicted salinity range data.

Depth 0-1 ft — Jack-Knifed Predicted ECe range

		[0,2]		[2,4]		[4,8]		[>8]	
Obs	[0,2]	**48.3**	**(42)**	51.7	(45)	0.0	(0)	0.0	(0)
ECe	[2,4]	16.7	(28)	**72.6**	**(122)**	10.1	(17)	0.6	(1)
range	[4,8]	1.8	(1)	54.6	(30)	**40.0**	**(22)**	3.6	(2)
	[>8]	0.0	(0)	33.3	(1)	66.7	(2)	**0.0**	**(0)**

Overall correct classification: (42+122+22+0)/313 = **59.4 %**

Depth 1-2 ft — Jack-Knifed Predicted ECe range

		[0,2]		[2,4]		[4,8]		[>8]	
Obs	[0,2]	**57.1**	**(36)**	42.9	(27)	0.0	(0)	0.0	(0)
ECe	[2,4]	15.5	(20)	**65.9**	**(85)**	18.6	(24)	0.0	(0)
range	[4,8]	0.0	(0)	31.3	(25)	**55.0**	**(44)**	13.7	(11)
	[>8]	0.0	(0)	4.9	(2)	61.0	(25)	**34.1**	**(14)**

Overall correct classification: (36+85+44+14)/313 = **57.2 %**

Depth 2-3 ft — Jack-Knifed Predicted ECe range

		[0,2]		[2,4]		[4,8]		[>8]	
Obs	[0,2]	**46.9**	**(15)**	46.9	(15)	6.2	(2)	0.0	(0)
ECe	[2,4]	13.1	(14)	**58.9**	**(63)**	24.3	(26)	3.7	(4)
range	[4,8]	0.0	(0)	33.3	(25)	**57.3**	**(43)**	9.4	(7)
	[>8]	0.0	(0)	1.0	(1)	28.3	(28)	**70.7**	**(70)**

Overall correct classification: (15+63+43+70)/313 = **61.0 %**

Depth 3-4 ft — Jack-Knifed Predicted ECe range

		[0,2]		[2,4]		[4,8]		[>8]	
Obs	[0,2]	**26.1**	**(6)**	56.5	(13)	17.4	(4)	0.0	(0)
ECe	[2,4]	13.3	(11)	**45.8**	**(38)**	39.7	(33)	1.2	(1)
range	[4,8]	2.2	(2)	27.5	(25)	**57.1**	**(52)**	13.2	(12)
	[>8]	0.0	(0)	0.9	(1)	23.3	(27)	**75.8**	**(88)**

Overall correct classification: (6+38+52+88)/313 = **58.8 %**

Table 4. Classification results, observed versus predicted and observed versus jack-knifed predicted salinity profile data.

		Predicted Ratio range			
		[below 2: l]		[above 2: l]	
Obs Ratio range	[below 2:1]	88.2	(172)	11.8	(23)
	[above 2:1]	16.9	(20)	83.1	(98)

Overall correct classification: (172 + 98)/3 13 = 863 %

		Jack-Knifed Predicted Ratio range			
		[below 2:1]		[above 2:1]	
Obs Ratio range	[below 2:1]	82.1	(160)	17.9	(35)
	[above 2:1]	29.7	(20)	703	(83)

Overall correct classification: (160+ 83)/313 = 77.6 %

Table 5. Summary statistics for regression equations fit to the salinity calibration data from quarter section 11-4.

Sample Depth	Regression Equation non-zero parameters	R^2	MSE	PSSMSE†	Corr (O,P)†2	Corr (O,J)†3
o-1 ft	pcsl, pcs2, pcsl2, psp	89.6%	0.0276	0.0598	0.946	0.822
l-2ft	pcs1,pcs2, pcs12, pSp	96.8%	0.0345	0.0708	0.984	0.930
2-3 ft	pcs1, pcs2, pSp	88.8%	0.0958	0.1406	0.942	0.805
3-4 ft	pcs1, pcs2, pSp	97.0%	0.0124	0.0230	0.985	0.946

†: calculated as PSSMSE = $PRESS_i/8$, where PRESS, represents the jack-knifed prediction sum of squares associated with the *ith* sample depth

†2: correlation between the observed and predicted log soil salinities

†3: correlation between the observed and jack-knifed predicted log soil salinities

Table 6. Comparison between the calibration sample log salinity mean and standard error (based on 8 sample sites) and the final regression estimated log salinity mean and standard error (based on all 64 survey sites) for quarter section 11-4.

Sample Depth	Calibration Sample mean	Calibration Sample std.err.	Regression Equation mean	Regression Equation std.err†	% reduction in std.err estimate†2
o-1 ft	0.860	0.1191	0.999	0.0865	27.4
l-2 ft	1.401	0.2418	1.414	0.0941	61.1
2-3 ft	1.956	0.2473	1.922	0.1406	43.1
3-4 ft	2.417	0.1727	2.476	0.0536	69.0

†: stderr[regression] = $(PSSMSE/8)^{0.5}$

†2: calculated using the following formula
100 - l00(stderr[regression]/stderr[calibration])

Table 7. Estimation of AAPE using jack-knifed predictions of the calibration salinity samples from quarter section 11-4 (with and with out applying formula II).

ECe Range	Sample Depth	Obs†	formula II applied Prd†2	APE	formula II not applied Prd†3	APE
[0,2]	1	37.5	45.2	7.7	50.0	12.5
	2	50.0	41.7	8.3	37.5	12.5
	3	12.5	12.7	0.2	12.5	0.0
	4	0.0	0.4	0.4	0.0	0.0
[2,4]	1	25.0	25.2	0.2	25.0	0.0
	2	25.0	24.6	0.4	25.0	0.0
	3	25.0	27.7	2.7	25.0	0.0
	4	25.0	22.5	2.5	25.0	0.0
[4,8]	1	12.5	0.7	11.8	0.0	12.5
	2	12.5	21.3	8.8	25.0	12.5
	3	12.5	34.8	22.3	37.5	25.0
	4	62.5	43.2	19.3	37.5	25.0
[>8]	1	0.0	0.0	0.0	0.0	0.0
	2	0.0	0.1	0.1	0.0	0.0
	3	25.0	26.3	1.3	12.5	12.5
	4	75.0	73.6	1.4	87.5	12.5
Average Absolute Percent Error (AAPE):				5.46		7.81

†: observed % of calibration soil salinities, by depth, falling within the given salinity range

†2: jack-knifed predicted % of calibration soil salinities, by depth, estimated to be within the given salinity range after accounting for the prediction error through the application of formula II (shown in the text)

†3: jack-knifed predicted % of calibration soil salinities, by depth, failing within the given salinity range (without correcting for the prediction error)

Table 8. Example of salinity "apprasial sheet" for quarter section 11-4. The salinity characteristics shown below include the median field ECe estimates, with confidence intervals, salinity range matrix data, and profile ratio data.

Quarter Section Identification Code: **11-4**

I. Median Salinity Estimates (dS/m)

Depth	Median ECe	Aproximate 95% CI
o-1 ft	2.72	[2.3 , 3.2]
l-2 ft	4.11	[3.4 , 5.0]
2-3 ft	6.84	[5.2, 9.1]
3-4 ft	11.89	[10.7, 13.2]

II. Rank Estimates: % area of field falling within specific salinity classes.

	ECe range (dS/m)			
Depth	[0,2]	[2,4]	[4,8]	[above 8]
o-1 ft	27.9	50.9	**16.7**	**4.4**
l-2 ft	15.3	24.5	**47.6**	**12.6**
2-3 ft	9.6	13.0	**27.2**	**50.1**
3-4 ft	0.0	2.7	**19.0**	**78.3**

Average Absolute Percent Error (AAPE): **5.45**

III. Profile Shape / Ratio Estimates

⇒ Typical 2-4 to O-2 ft ECe ratio: 3.01:1

% area of field with ECe ratio greater than 2 to 1: 78.9

Table 9. **Median soil salinity estimates (dS/m) by depth for each of the 37 quarter sections within the survey zone of the Broadview Water District.**

Quarter Section Identification Code	Salinity Estimates (dS/m) Depth o-1 ft	1-2 ft	2-3 ft	3-4 ft
3-l	2.94	4.17	4.63	4.44
3-2	2.60	5.01	4.03	3.62
3-3	1.96	1.97	2.74	4.33
3-4	2.11	2.42	3.50	5.58
4-l	4.50	8.98	16.6	19.1
4-2	3.05	3.55	7.84	11.8
4-3	3.35	4.83	8.11	11.3
4-4	1.93	1.69	3.48	5.39
8-2	2.86	4.34	4.61	7.40
8-4	2.18	3.22	4.77	6.57
9-1	2.93	3.79	5.43	6.35
9-2	1.61	2.05	3.62	5.51
9-3	3.16	4.18	5.96	6.04
9-4	2.22	2.75	4.60	5.22
10-l	2.49	3.66	4.17	4.51
10-2	2.09	3.71	5.74	6.07
10-3	2.20	2.41	4.52	4.77
10-4	4.27	4.15	4.40	4.45
11-3	2.45	1.67	2.98	4.17
11-4	2.72	4.11	6.84	11.9
13-1	1.95	3.24	4.88	5.55
13-2	2.20	3.14	4.10	4.75
13-4	1.70	1.97	2.75	2.98
14-1	1.92	2.42	3.59	5.21
14-2	4.06	6.37	6.10	5.34
14-3	2.96	3.55	6.52	9.94
14-4	2.22	3.86	4.59	6.34
15-l	2.86	4.26	5.98	6.45
15-2	2.16	2.59	5.81	6.21
15-3	3.40	2.89	3.96	4.57
15-4	2.31	2.68	5.27	6.14
16-1	3.15	3.63	4.66	5.20
16-2	2.66	3.68	4.24	3.97
16-3	2.25	2.82	3.28	4.40
16-4	4.63	4.22	4.26	4.38
18-3	3.36	5.64	6.41	6.53
18-4	3.73	5.44	7.38	8.48

Table 10. Salinity Range & Ratio Matrix data: percent area of land, by depth, within the survey zone of the Broadview Water District falling within specific salinity classes and/or exhibiting specific profile shapes.

	Salinity Range (dS/m)			
Depth	[0,2]	[2,4]	[4,8]	[above 8]
o-1 ft	27.87	53.69	17.15	1.28
1-2 ft	20.08	39.23	32.12	8.57
2-3 ft	8.56	30.12	39.01	22.3 1
3-4 ft	5.09	23.64	39.85	31.42

Average Absolute Percent Error (AAPE): 2.84

Profile Shape	
2-4 to O-2 ft Salinity Ratio	% Area of land
less than 1.5:l	35.75
1.5:1 to 2:l	23.74
2:l to 3:l	26.88
greater than 3:l	13.63

Average Absolute Percent Error (AAPE): 1.37

Table 11. Typical profile ratios and percent area of land with ratios exceeding 2:l for each of the 37 quarter sections within the survey zone of the Broadview Water District.

Quarter Section Identification Code	[X: 1] Median Profile Ratio	% Area w/ratio > 2:l
3-l	1.33	9.2
3-2	1.03	7.8
3-3	1.95	41.4
3-4	2.07	49.7
4-l	2.83	80.6
4-2	3.06	81.0
4-3	2.47	73.6
4-4	2.58	68.2
8-2	1.81	35.6
8-4	2.30	61.0
9-1	1.68	42.2
9-2	2.45	83.3
9-3	1.52	28.7
9-4	2.10	52.4
10-l	1.46	5.3
10-2	2.26	67.8
10-3	1.90	49.0
10-4	1.06	1.1
11-3	1.78	31.8
11-4	3.01	78.9
13-1	2.02	53.4
13-2	1.65	27.4
13-4	1.66	26.1
14-1	2.00	49.8
14-2	1.14	7.8
14-3	2.54	73.9
14-4	1.87	39.1
15-1	1.73	37.4
15-2	2.49	74.8
15-3	1.32	13.7
15-4	2.42	65.7
16-1	1.45	9.7
16-2	1.32	4.4
16-3	1.42	23.6
16-4	0.93	0.9
18-3	1.59	22.2
18-4	1.66	20.6

Figure Captions

Fig 1. Quarter section identification codes and locations of all sample sites within the Broadview Water District survey area.

Fig 2. Average isotropic residual semivariograms, by depth, showing the lack of spatial autocorrelation within the residual distributions.

Fig 3. Coefficients of determination (R^2 values) from the regression models used to transform the electromagnetic induction data into predicted log soil salinity levels.

Fig 4. Observed versus predicted log soil salinity levels, by depth, throughout the entire survey area.

Fig 5. Observed versus jack-knifed predicted log soil salinities, by depth, throughout the entire survey area.

Fig 6. Estimated median quarter section soil salinity levels (dS/m).

Fig 7. Spatial soil salinity distribution across the entire survey area, O-l foot depth.

Fig 8. Spatial soil salinity distribution across the entire survey area, l-2 foot depth.

Fig 9. Spatial soil salinity distribution across the entire survey area, 2-3 foot depth.

Fig 10. Spatial soil salinity distribution across the entire survey area, 3-4 foot depth.

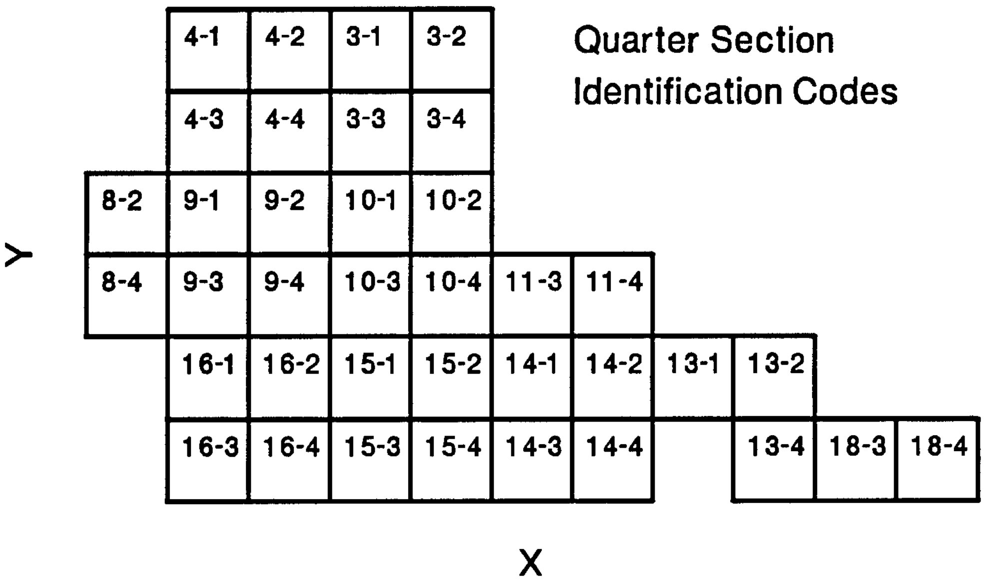
Fig 11. Typical salinity profile ratios and percent area of land exceed a 2:l ratio within each quarter section.

Fig 12. Spatial salinity profile ratio distribution across the entire survey area.

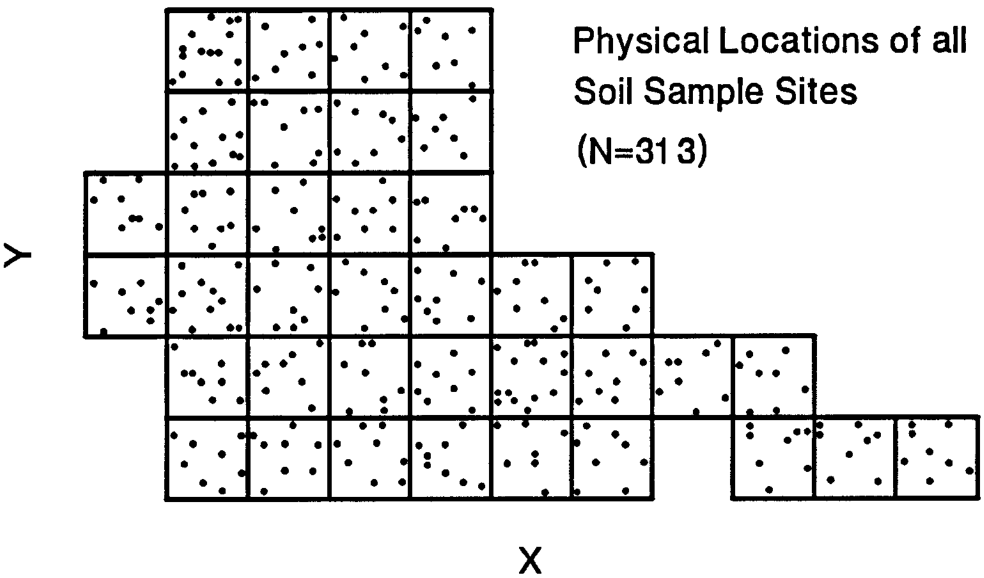
FIGURE 1

Broadview Water District Area

U.S.S.L. Salinity Survey: May 1991



Quarter Section Identification Codes



Physical Locations of all Soil Sample Sites (N=313)

FIGURE 2

Isotropic Residual Semivariograms: by Depth

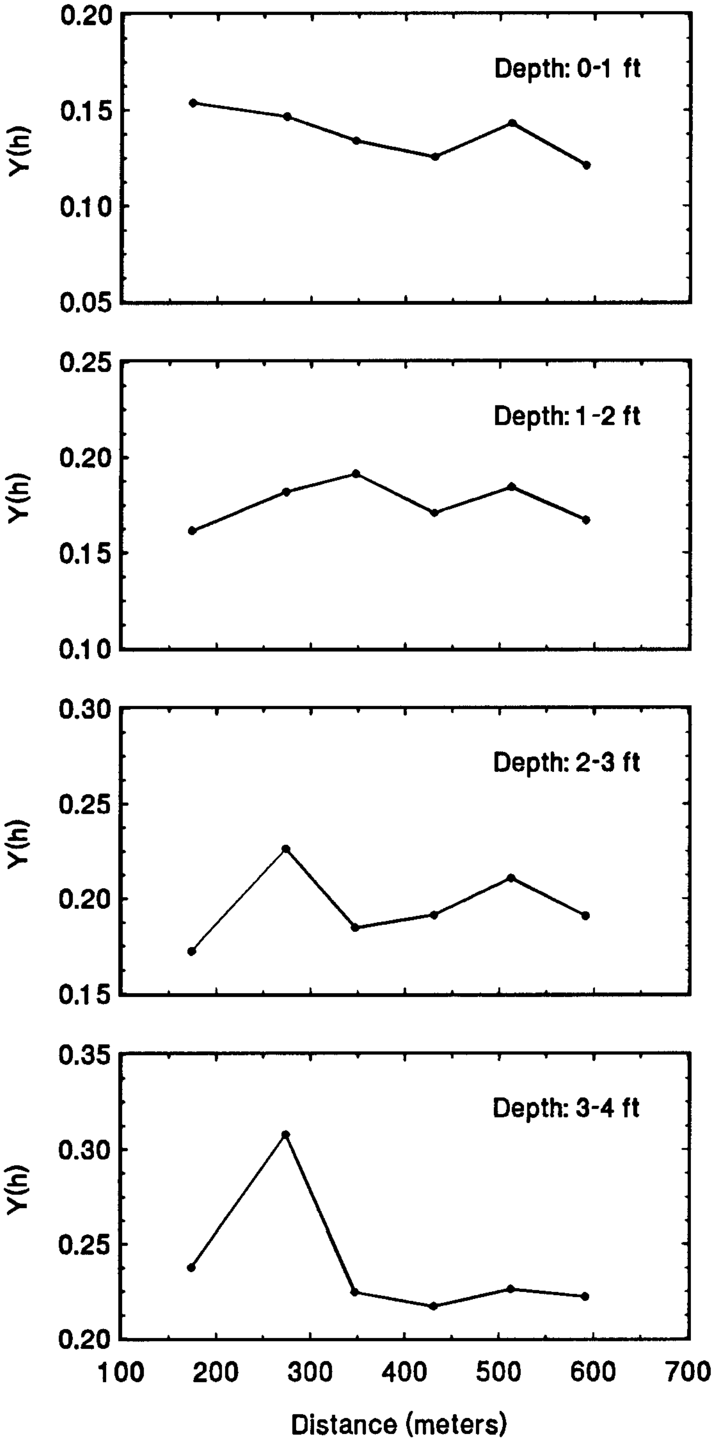
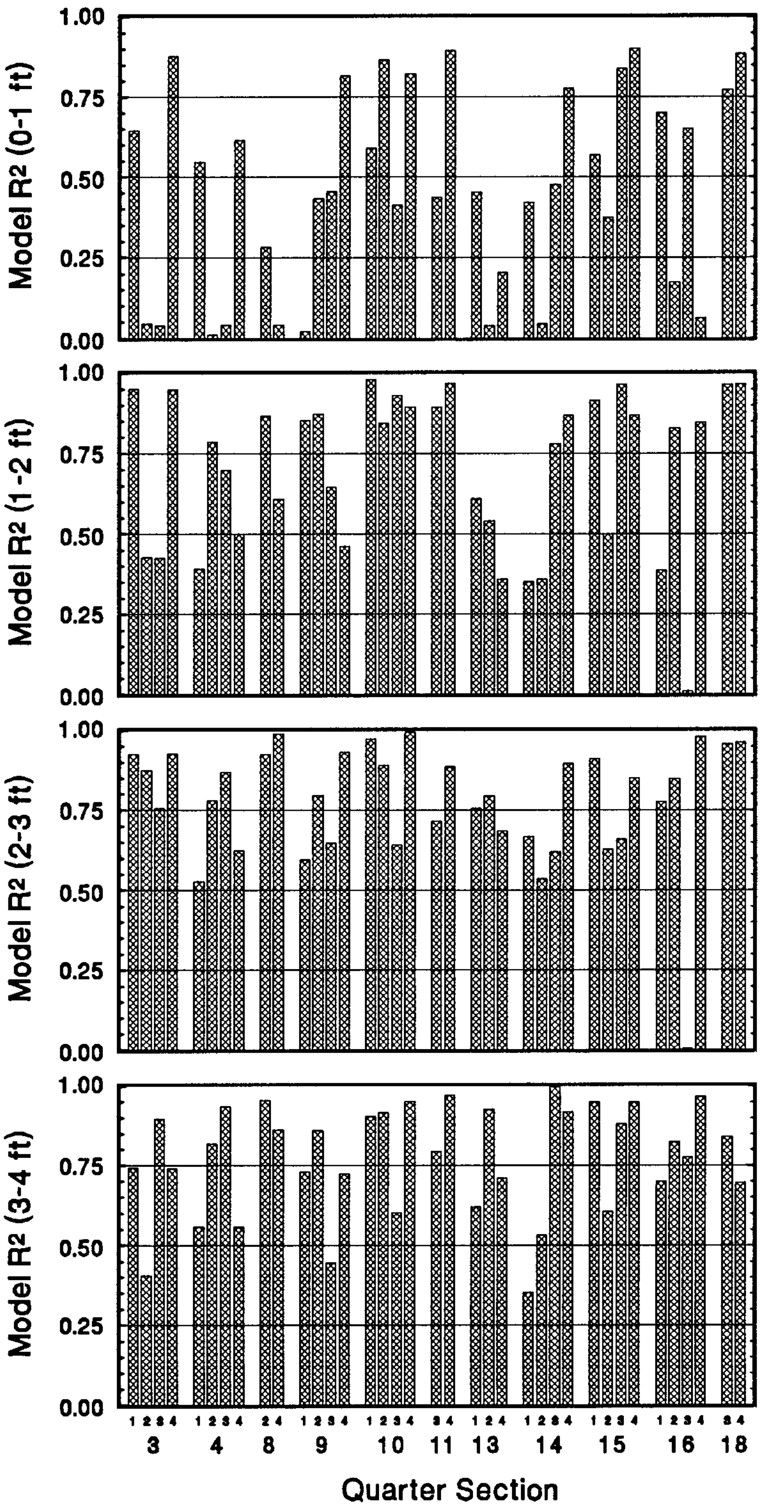
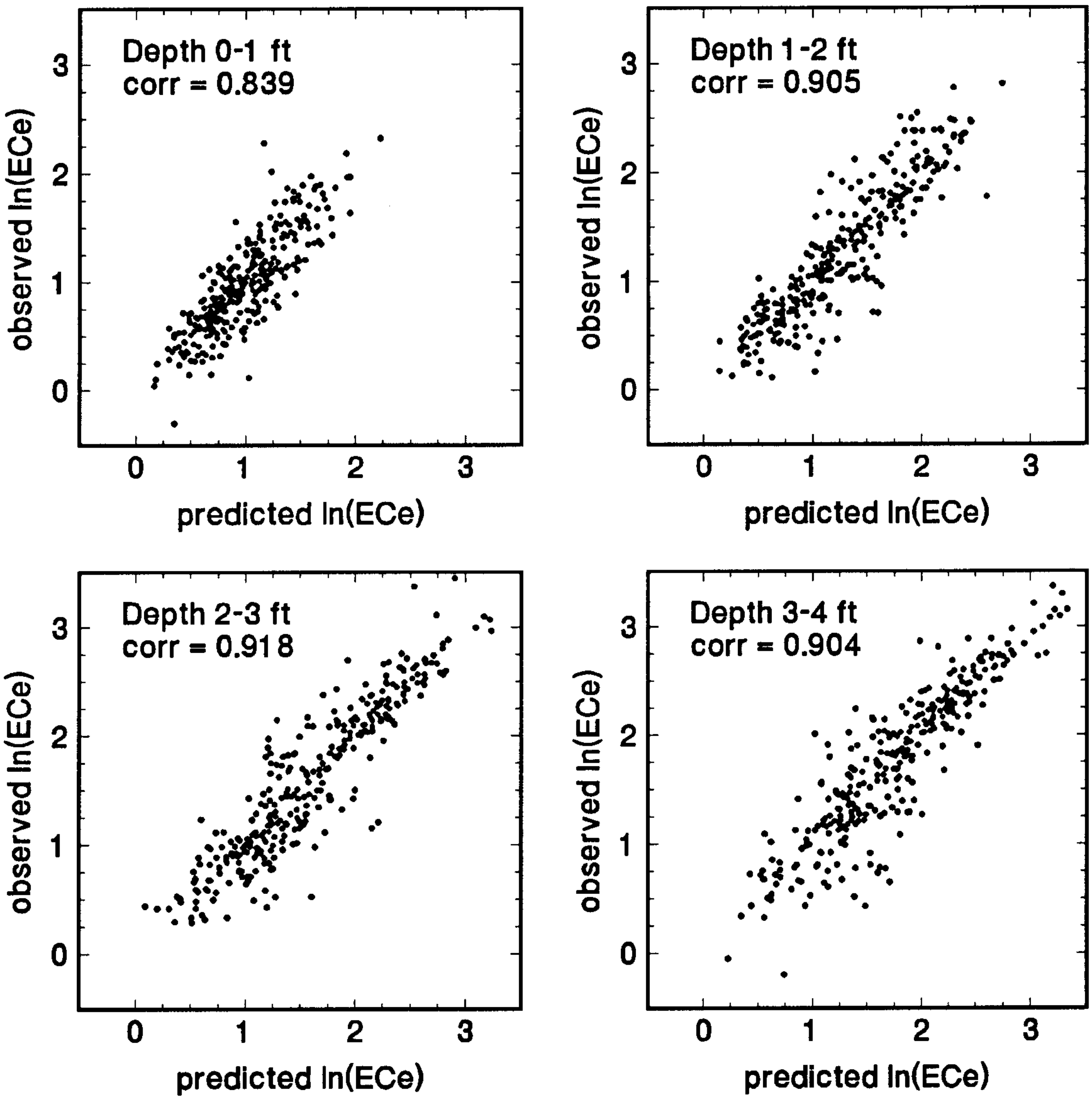


FIGURE 3

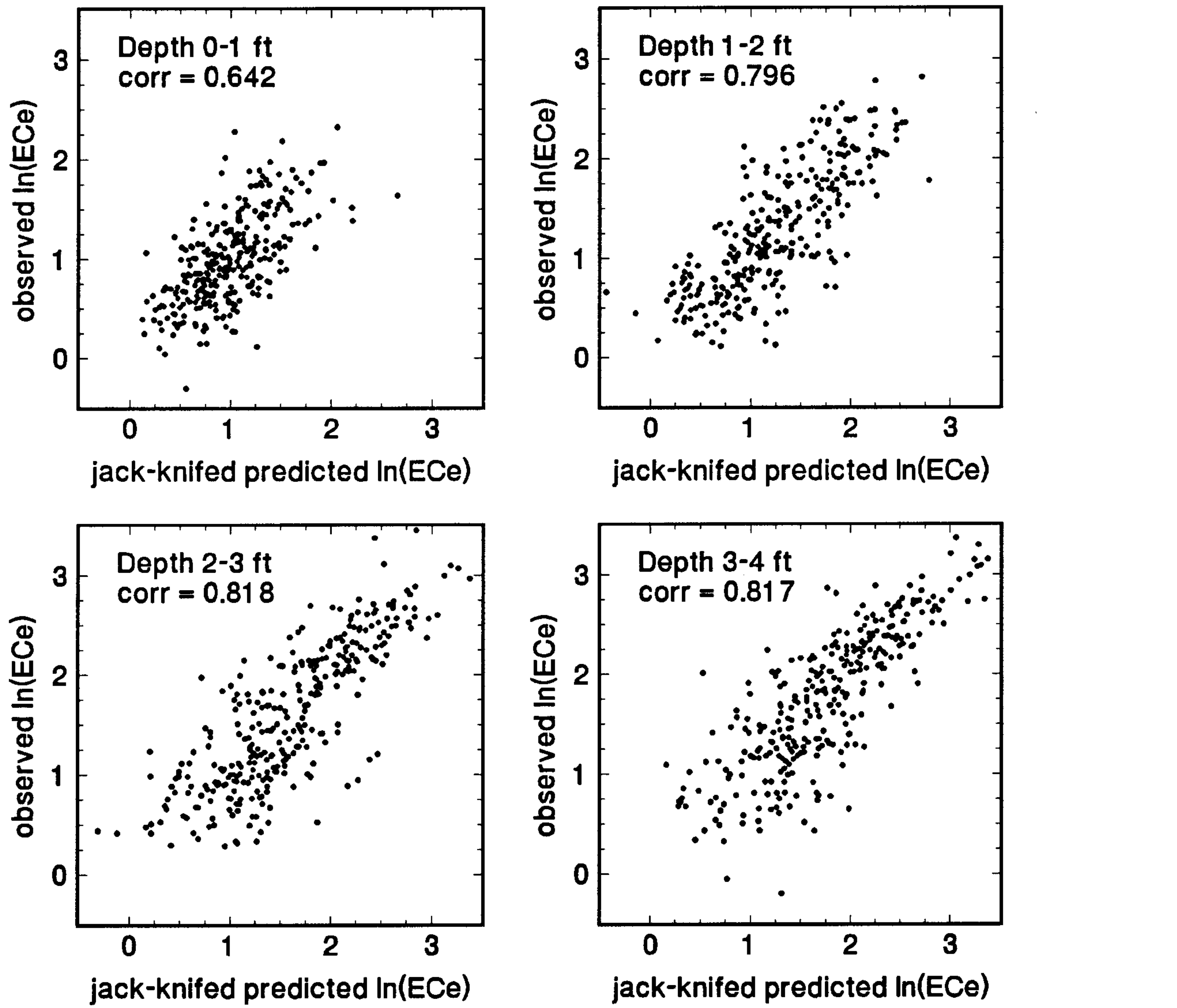
Regression Model Coefficients of Determination



Prediction Resolution: observed v.s. predicted ln(ECe)



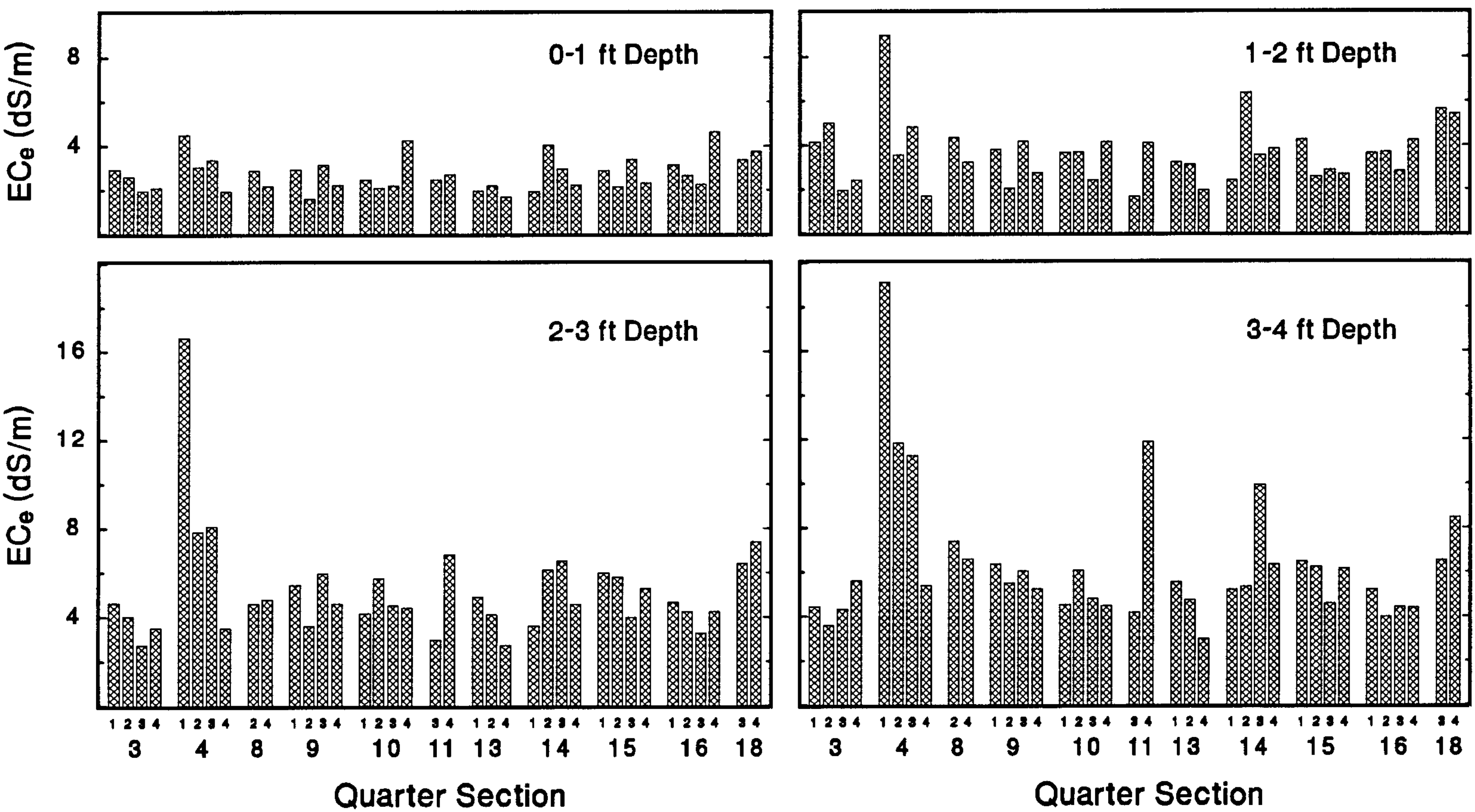
Prediction Reliability: observed v.s. jack-knifed predicted ln(ECe)



Broadview Water District Area

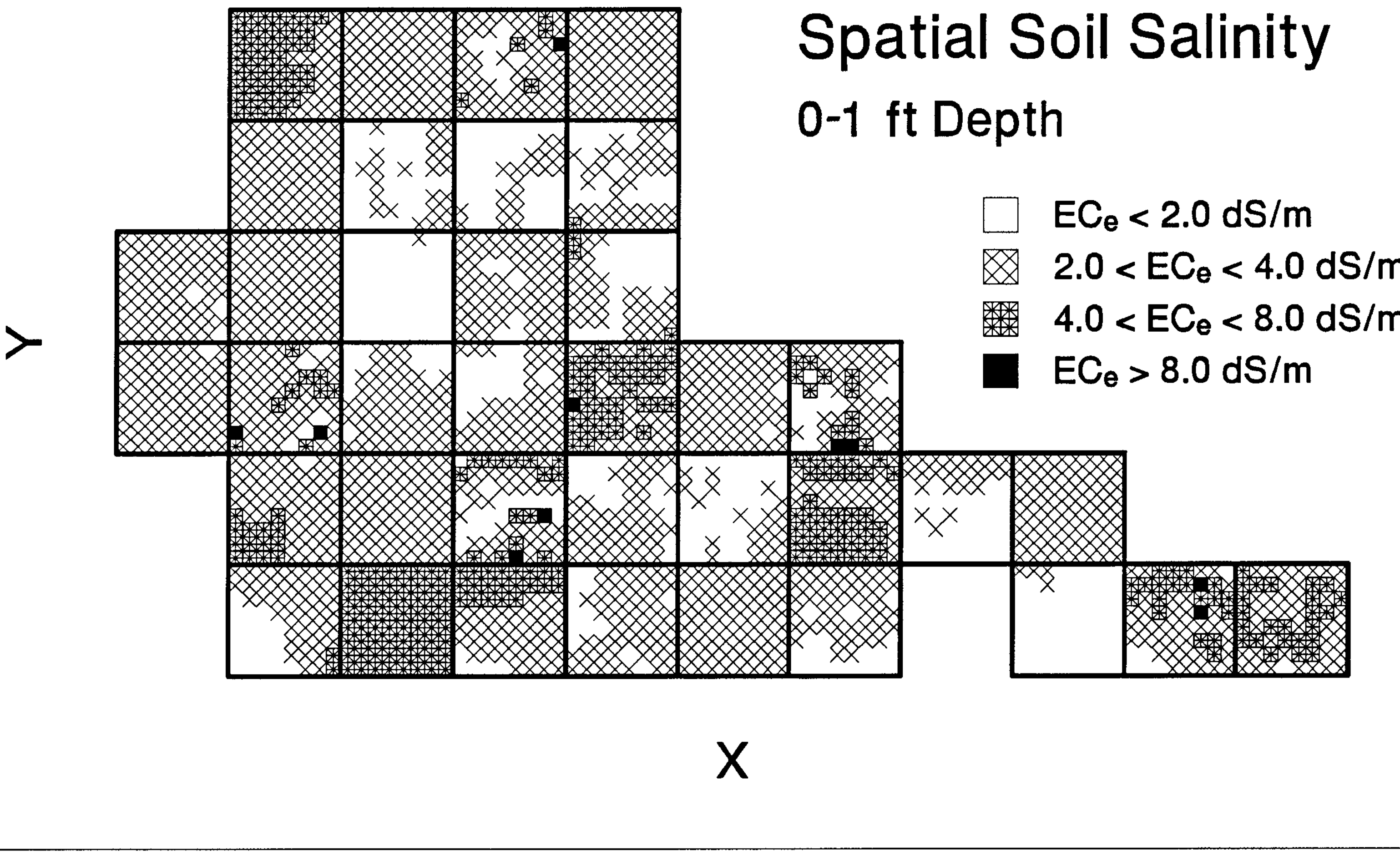
U.S.S.L Salinity Survey: May 1991

Median Quarter Section Soil Salinity Levels



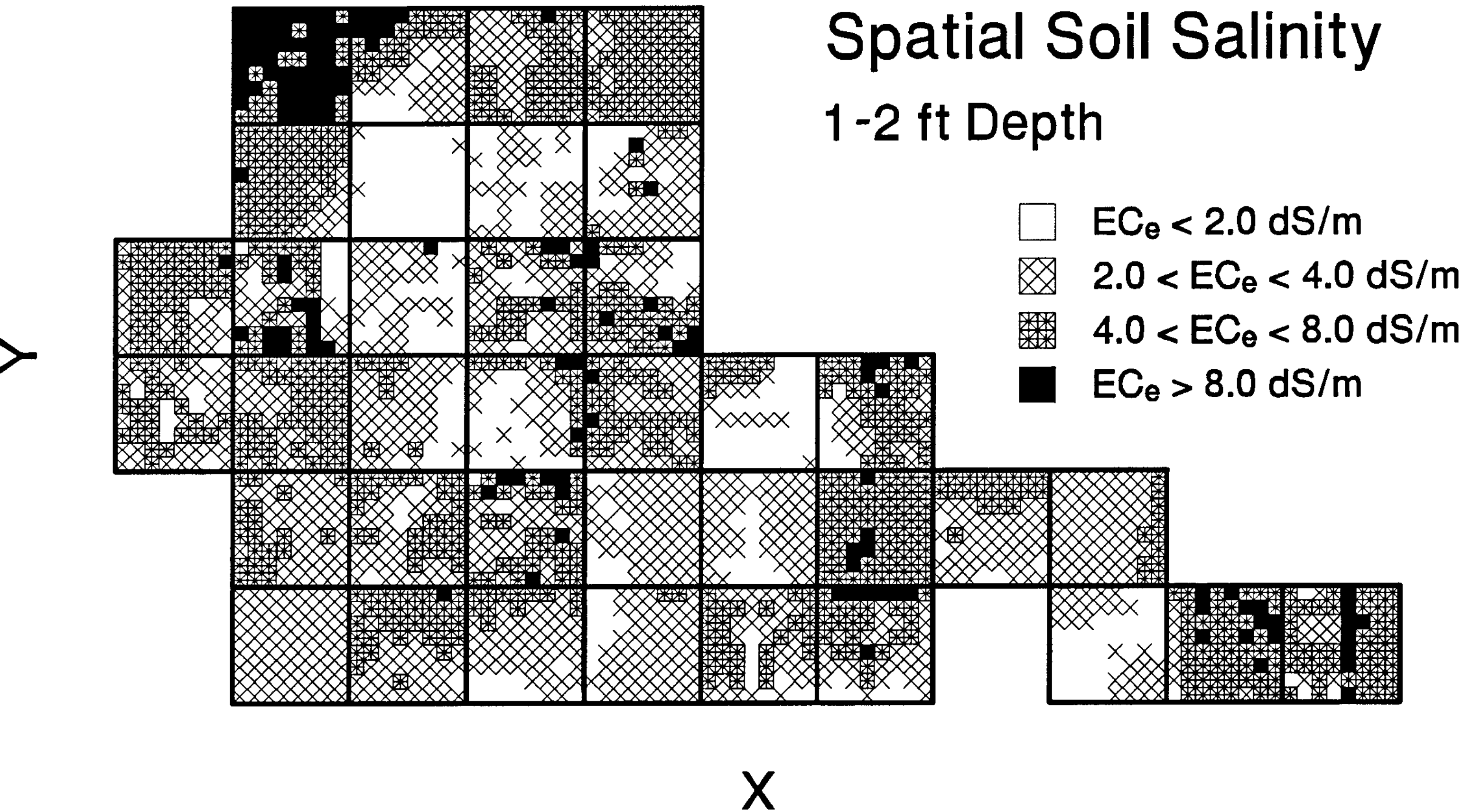
Broadview Water District Area

U.S.S.L. Salinity Survey: May 1991



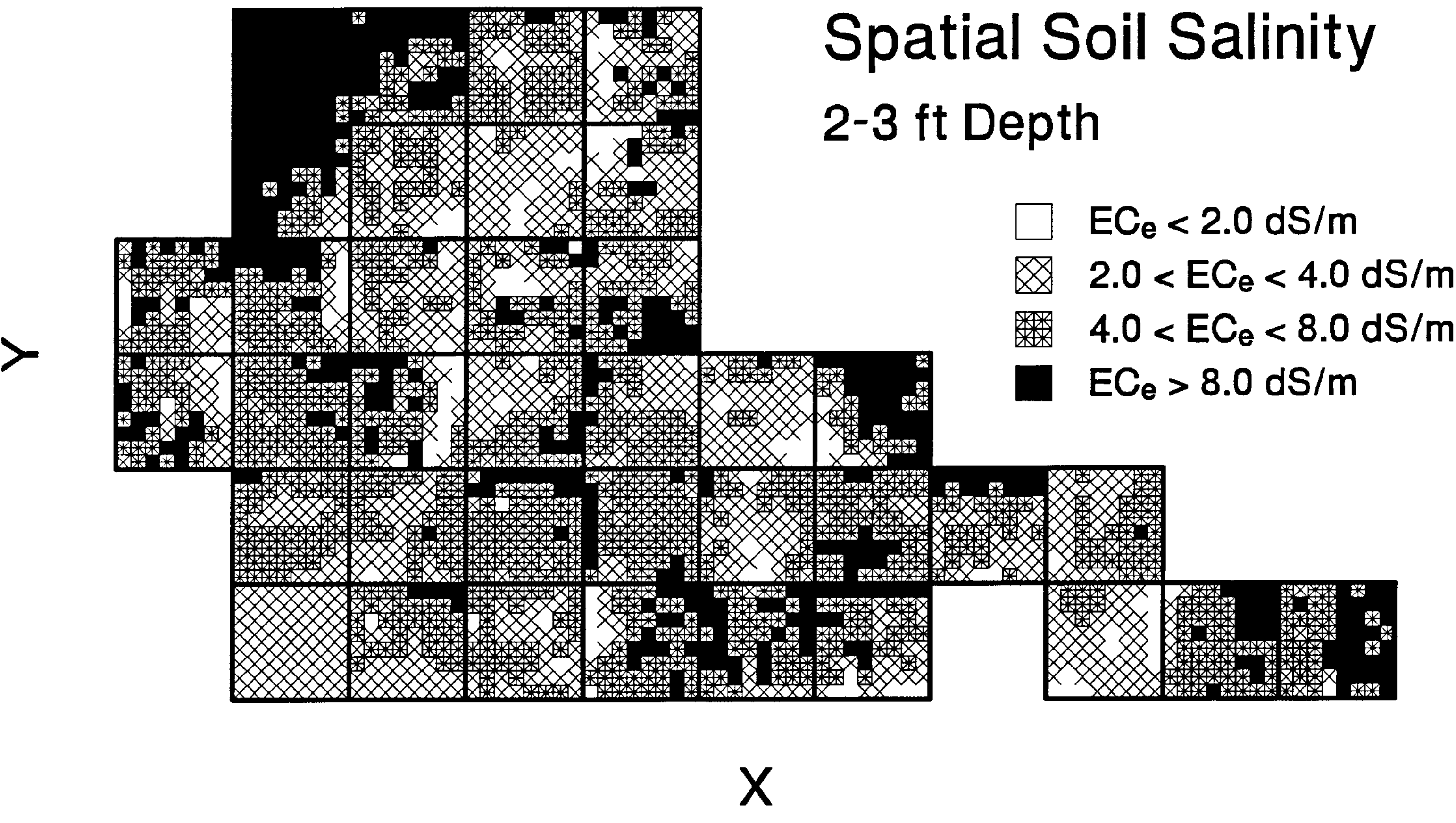
Broadview Water District Area

U.S.S.L. Salinity Survey: May 1991



Broadview Water District Area

U.S.S.L. Salinity Survey: May 1991



Broadview Water District Area

U.S.S.L. Salinity Survey: May 1991

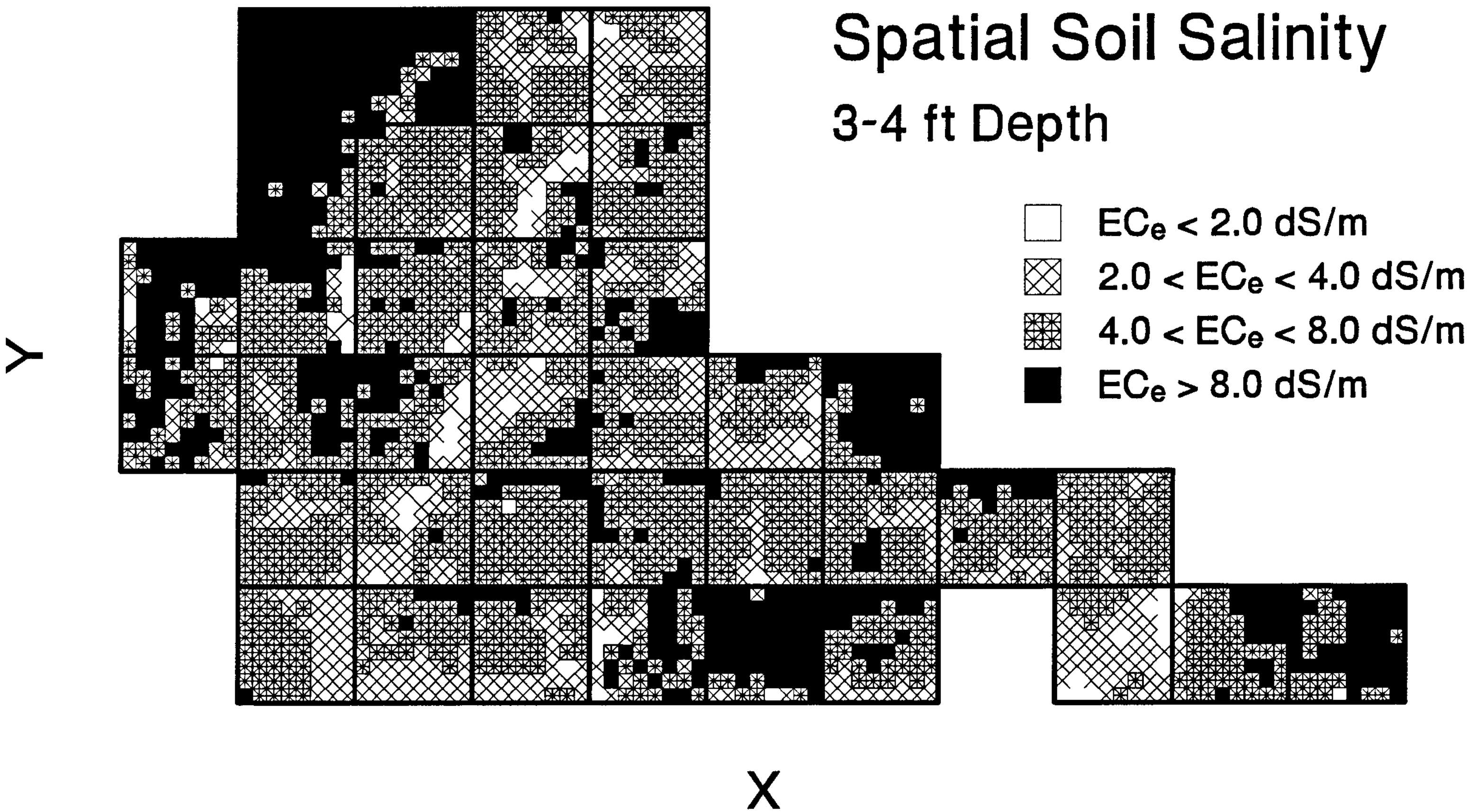
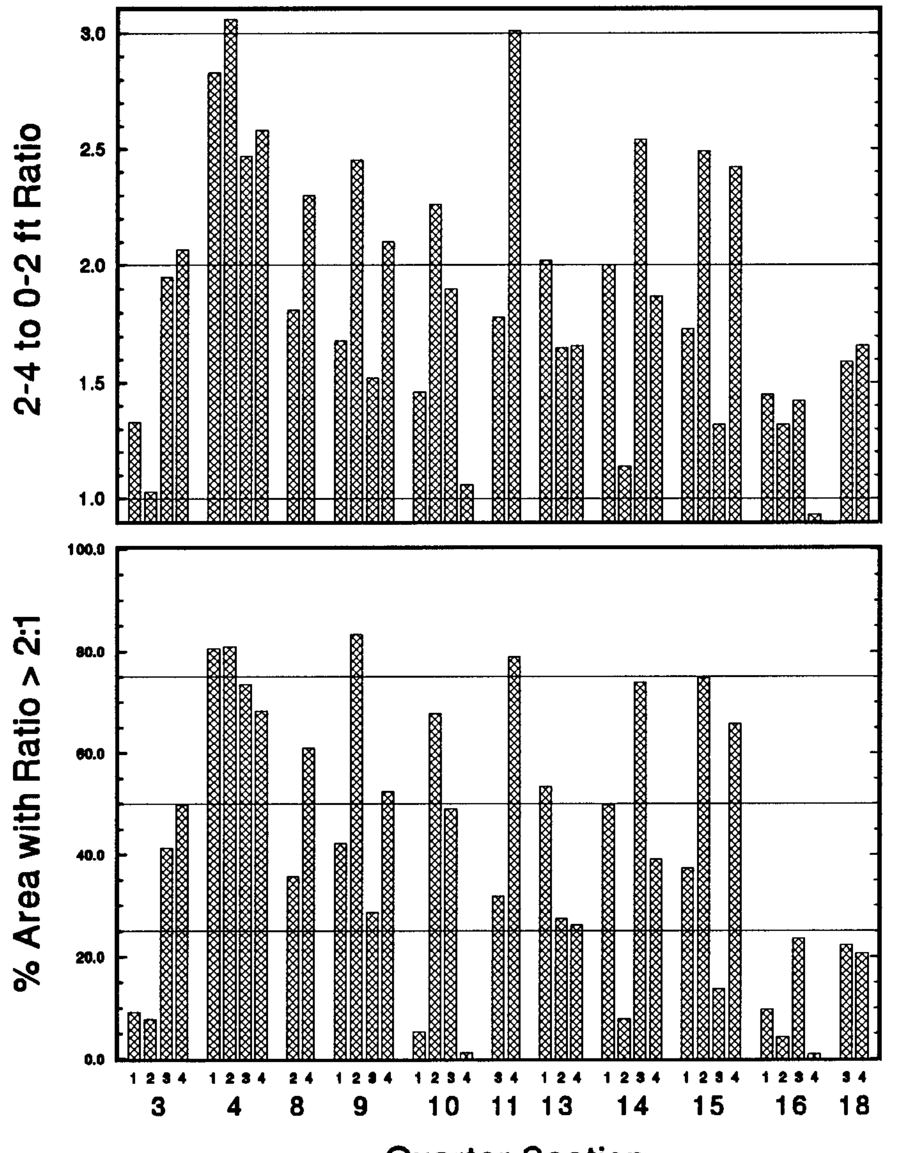


FIGURE 11

Broadview Water District Area

U.S.S.L. Salinity Survey: May 1991

Quarter Section Salinity Profile Characteristics



Broadview Water District Area

U.S.S.L. Salinity Survey: May 1991