2036

Statistical techniques for calibrating regional (multi-field) EM survey data.

S. M. Lesch
USDA-ARS, George E. Brown Jr. Salinity Laboratory, Riverside, CA
slesch@ussl.ars.usda.gov

1. INTRODUCTION

Electrical conductivity (EM) surveys have become a common technique for the assessment of various near surface soil properties. Both direct contact and non-invasive conductivity surveying techniques are now regularly employed in many precision agriculture applications (Corwin & Lesch, 2003), and these same techniques have been used for many years in soil salinity assessments (Rhoades et al., 1999). Although large amounts of EM data can now be readily collected (due to the mobilization of various conductivity sensors), this data still must be calibrated to the target soil property of interest (such as soil salinity or % clay content, etc.). In order to calibrate conductivity survey data, a limited set of "reference" or "ground-truth" soil samples are typically also collected during the survey process. The target soil properties of interest are measured on these calibration samples and then a suitable (geo)statistical prediction model is employed, such as a cokriging model or spatial regression model (Lesch et al., 1995).

The purpose of this research is to review and compare two commonly used spatial prediction techniques, cokriging and hierarchical spatial regression, specifically with respect to predicting both soil salinity and soil texture from calibrating regional EM survey data. The regional survey data that will be discussed in the oral presentation is the 1991 Broadview water district EM/salinity study (Corwin et al., 1995). This EM-38 grid survey consisted of 2378 locations across 2350 ha of irrigated agricultural land. Forty three adjacent fields were surveyed during this study and a calibration set of 315 co-located 1.2 m deep soil sample cores were also acquired (at about 15% of the EM survey sites) and analyzed for soil salinity (EC_e, dS/m), saturation percent (SP, %) and gravimetric soil water content (2g, %). One of the primary survey goals was to use the acquired EM data to predict the spatial salinity distribution throughout the district. Note that the modeling of this data set from the geostatistical perspective has been previously discussed in detail by Vaughan et al., (1995) and Bourgault et al., (1997).

2. STATISTICAL METHODOLOGY

A brief overview of both ordinary cokriging and hierarchical spatial regression modeling is presented in this section (for the case of a single covariate).

Cokriging (CoK): using a single covariate

The ordinary cokriging (CoK) estimator can be written as a linear combination of the neighboring primary (target) and secondary (covariate) data; i.e.,

In: Proceedings of the International Salinity Forum, Managing Saline Soils and Water: Science, Technology, and Soil Issues. April 25-27, 2005. Riverside, CA pp:289-292.

$$Y(s_0) = \sum_{i=1}^{m} \delta_i Y(s_i) + \sum_{j=1}^{n} \omega_j X(s_j), \ \ where \sum \delta_i = 1 \ and \sum \omega_j = 0 \cdot \qquad (1)$$

In matrix formulation, this estimator can be expressed as

$$Y(s_0) = \mathbf{D}^{\mathbf{T}}\mathbf{y} + \mathbf{W}^{\mathbf{T}}\mathbf{x}, \ \ where\ \mathbf{1}^{\mathbf{T}}\mathbf{D} = 1 \ and\ \mathbf{1}^{\mathbf{T}}\mathbf{W} = 0. \qquad (2)$$

The optimal vector weights **D** and **W** are found by solving the following matrix system

$$\begin{bmatrix} \mathbf{C_{YY}} & \mathbf{C_{YX}} & \mathbf{1} & \mathbf{0} \\ \mathbf{C_{XY}} & \mathbf{C_{XX}} & \mathbf{0} & \mathbf{1} \\ \mathbf{1^T} & \mathbf{0^T} & 0 & 0 \\ \mathbf{0^T} & \mathbf{1^T} & 0 & 0 \end{bmatrix} * \begin{bmatrix} \mathbf{D} \\ \mathbf{W} \\ -\mu_y \\ -u_x \end{bmatrix} = \begin{bmatrix} \mathbf{c_{Y0}} \\ \mathbf{c_{X0}} \\ 1 \\ 0 \end{bmatrix} \qquad (3)$$

where **C_{YY}** and **C_{XX}** represent the assumed covariance structures of the primary and secondary data, $\mathbf{C_{YX}}$ represents the assumed cross-covariance structure, and $\mathbf{c_{Y0}}$ and $\mathbf{c_{X0}}$ represent the primary and secondary covariance vectores between the observed data and new prediction site (Wackernagel, 1998)

One should note that a fundamental goal in any CoK analysis is to facilitate predictions "off the grid"; i.e., to make predictions at locations where no covariate data exists. When the covariate data exists everywhere (or has been acquired on a dense enough grid so that interpolated predictions are unnecessary) then other kriging techniques such as kriging with external drift (KED) can typically be used in place of a full CoK analysis (Deutsch & Journel, 1992; Wackernagel, 1998).

Hierarchical Spatial Regression (HSR): using a single covariate

Hierarchical spatial regression (HSR) models have become quite popular within the last 10 years. These models can be used for either classical (frequentist) and bayesian inference and nearly all linear geostatistical modeling techniques can be equivalently recast into the hierarchical regression setting (Royle & Berliner, 1999). In its most basic form, the HSR modeling approach represents a two-stage, conditional regression modeling technique. In the first stage, the primary (target) variable is modeling as a conditional linear function of the secondary covariate variable plus a spatially dependent random error component. In matrix notation, the equation can be expressed as:

$$\mathbf{y}(s)|\mathbf{x} = \mathbf{Z}\beta + \eta(s), \ \ \eta(s) \sim MSG(\mathbf{0}, \Sigma(\theta)) \qquad (4)$$

where **Z** represents a suitable design matrix that specifies the postulated y / x relationship, *∃* represents the regression model parameters, $\boldsymbol{0}$(s) represents the spatially correlated error component that is assumed to follow a some form of multi-variate spatial gaussian error structure, in turn defined (indexed) by the *2* parameter vector (Royle & Berliner, 1999; Schabenberger & Gotway, 2005) . Strictly speaking, this equation only predicts target data levels at locations where the covariate data exists; i.e., it represents a conditional model of y|x. Hence, in the second stage, the covariate (x) data is also assumed to exhibit spatial structure; for example

$$\mathbf{x}(s) \sim MSG(\mathbf{u}, \Sigma(\pi)) \tag{5}$$

This second stage (or hierarchical) assumption allows one to make predictions across the entire domain of interest; i.e., one uses Eqn. (5) to generate the best unbiased linear predictor of $x(s_0)$ at the non-surveyed location s_0 and then this predictor is substituted into Eqn. (4) to generate the corresponding $y(s_0)$ prediction. If predictions are only needed where known covariate data is available, then the second stage of the analysis need not be performed. In other words, Eqn. (5) only enters into the HSR model when predictions are made "off the grid", otherwise the HSR model functions just like an ordinary spatial regression model (i.e., similar to a geostatistical KED equation).

Royle & Berliner (1999) give an excellent overview of the HSR modeling approach, including how this approach can be adapted to mimic either KED and CoK models or additionally provide for a more comprehensive class of regression model structures in the first stage of the analysis.

3. DISCUSSION & CONCLUSION

In general, the following four features make the HSR modeling approach a more attractive alternative to an ordinary CoK analysis when modeling regional EM survey data:

Flexibility in modeling the covariate relationship.

It is well known that the ordinary CoK model assumes a simple linear relationship between the target and covariate data; i.e., y = b0 + b1[x]. However, in many EM surveys the true y/x relationship may often be (a) curve-linear, (b) contaminated by global trend and/or, (c) contaminated by between-field (blocking) effects. This latter issue of systematic, between-field variation in the EM data represents a particularly important issue in regional surveys. Global trends, blocking effects, and/or curve-linear response patterns can all be easily incorporated into HSR models (using standard software packages), while modifying CoK models to handle these effects tends to be much more mathematically and computationally demanding.

More efficient parameter estimation.

Both the mean and covariance parameters in an HSR model are commonly estimated using either maximum likelihood (ML) or restricted maximum likelihood (REML) techniques (for example, using the SAS MIXED procedure). Although in theory the same type of estimation techniques can be applied to the CoK model, few commercial or share-ware CoK software packages support ML or REML techniques. Note that ML (or REML) estimation techniques are more efficient than weighted least squares procedures when the joint (spatial) multivariate Normality assumption holds.

Parameter tests can be easily carried out within the HSR modeling framework.

In many surveys explicit parameter estimates (of the y/x relationship) may be desired, and/or one may wish to explicitly test for statistically significant between-field blocking effects or global trend, etc. HSR models facilitate such tests, via either asymptotic likelihood ratio test statistics or conditionally specified F-tests. In contrast, CoK equations (regardless of how they might be specified) do not facilitate any type of formal parameter testing methodologies.

The hierarchical approach to model specification is more succinct.

In regional EM surveys, the spatial density of the EM covariate(s) may be sufficient to preclude the need of covariate interpolation. Hence, from the geostatistical perspective one would naturally be inclined to employ some type of simpler KED model in place of the CoK model. However, such a decision must be made before the data modeling begins and if the KED approach is adopted then no interpolation can be performed (i.e., no predictions off the EM grid can be computed). In contrast, such problems do not arise within the HSR framework. If one wishes to restrict attention to just locations having known covariate information, then only the first stage of the HSR modeling approach needs to be carried out. If the need for interpolation arises later on, then the second stage of the modeling approach can be performed. No modification to the (stage 1) conditional regression model is necessary in this latter scenario, thus no analysis effort is wasted.

An analysis of the Broadview EM survey and soil sample calibration data will be given in the oral presentation. Some of the key results from this analysis show that (a) the magnitude of the 1991 EM survey data changes substantially from field to field, (b) due to this fact, there is significant between-field variation that must be accounted for when modeling these data, and (c) spatial analysis of covariance (ANOCOVA) models can be used to adjust out these block effects and accurately predict both the soil SP and log salinity levels from the (log transformed) EM survey data. Spatial ANOCOVA models represent one type of model structure that is available within the HSR modeling framework; our results suggest that these models may be especially well suited for calibrating regional EM survey data.

4. REFERENCES

Bourgault, G., A.G. Journel, J.D. Rhoades, D.L. Corwin and S.M. Lesch. 1997. Geostatistical analysis of a soil salinity data set, (Ed. D.L. Sparks) Advances in Agronomy, Vol. 58, Academic Press, N.Y., N.Y.

Corwin, D.L., J.D. Rhoades, P.J. Vaughan and S.M. Lesch. 1995. An integrated methodology for assessing soil salinity and salt-loading to the groundwater at a regional scale. ASA-CSSA-SSSA Bouyoucos Conf, Riverside, CA. 356-370.

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

Deutsch, C.V. and A.G. Journel. 1992. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, N.Y, N.Y.

Lesch, S.M., D.J. Strauss and J.D. Rhoades. 1995. 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.

Rhoades, J.D., F. Chanduvi and S.M. Lesch. 1999. Soil salinity assessment: Methods and interpretation of electrical conductivity measurements. FAO Irrigation & Drainage Paper 57. FAO, Rome, Italy.

Royle, J.A. and L.M. Berliner. 1999. A hierarchical approach to multivariate spatial modeling and predisction. J. Ag. Bio. Env. Statistics. 4:1-28.

Schabenberger, O. and C.A. Gotway. 2005. Statistical Methods for Spatial Data Analysis. Chapman & Hall / CRC Press, N.Y., N.Y.

Vaughan, P.J., S.M. Lesch, D.L. Corwin and D.G. Cone. 1995. Water content effect on soil salinity prediction: a geostatistical study using cokriging. Soil Sci. Soc. Am. J. 59:1146-1156.

Wackernagel, H. 1998. Multivariate Geostatistics, 2nd Ed. Springer-Verlag, Berlin, Germany.