Assessment of the Production and Economic Risks of Site-specific Liming using Geostatistical Uncertainty Modelling

P. Goovaerts[(1)], R.A. Viscarra-Rossel[(2)] and A.B. McBratney[(2)]

[(1)]Biomedware, Inc. and PGeostat, LLC
516 North State Street, Ann Arbor, MI 48104.
Telephone: +001 (734) 913-1098
FAX: +001 (734) 913-2201
Email: goovaerts@biomedware.com
[(2)] Australian Center for Precision Agriculture,
The University of Sydney, Australia

Abstract

This paper presents a geostatistical approach to model uncertainty about the spatial distribution of soil properties and to propagate it through a crop model, allowing for incorporation of uncertainty in the choice of management scenarios (no lime, single-rate liming and site-specific lime applications to acidic field soil). The methodology involved modeling the uncertainty about wheat yield, accounting for the local uncertainties about soil pH and lime requirement, and the uncertainties about crop model parameters used in the simulations. Indicator kriging (IK) was used together with stratified sampling of the probability distributions of variables and model parameters for the propagation of uncertainties through to the output yield and net profit maps. Comparison of the three scenarios showed that under the economic conditions of the analysis, the optimum was reached for a single-rate application of 3.5 Mg/Ha over the entire field instead of site-specific lime applications.

1. Introduction

In the last decade, farmers and agricultural engineers have paid more and more attention to the spatial variability of field soil properties and their impact on crop yields. The natural heterogeneity encountered in increasingly large fields have led them to question the common practice of fertilizing uniformly such fields. Significant resources misallocation, such as over or under-applications of fertilizers or pesticides, not only affects economical profits but cause adverse environmental conditions. Precision Agriculture (PA) aims to manage land by considering its heterogeneity in both space and time (McBratney et al., 1997).

PA requires vast amounts of information to delineate the variation of relevant soil and crop properties within a field. In the absence of field-deployed proximal soil sensing systems that can efficiently and economically collect fine-scale information, the reliability of management decisions depends on the sampling strategy, the analytical methodology, the geostatistical techniques used for map production, and the uncertainty of the system (McBratney, 1992). Geostatistics is commonly used for the characterisation of geo-referenced spatial data, and interpolation by kriging. However it is important to keep in mind that kriging estimates are not true values. Hence the prediction error, or more generally the uncertainty about the soil attribute at an unsampled location, must be assessed and accounted for in any decision-making process. Geostatistics provide tools for deriving the set of possible outcomes at any single location as well as the associated probability of occurrence (Goovaerts, 1999, 2001). Even if actual soil property values were known everywhere, the prediction of yield would still be uncertain because of our imperfect knowledge of the underlying mechanisms. Therefore, the uncertainty about crop yield arises from both the uncertainty about the soil properties and the uncertainty about the parameters of the model used to derive crop yield. While this concept of propagation of uncertainty is well known for GIS operations (Heuvelink et al., 1989; French et al., 2003), it has only recently received attention in soil science (Pebesma and Heuvelink, 1999; Van Meirvenne and Goovaerts, 2000; Hansen et. al., 1999), and very little in PA.

This paper presents a procedure whereby indicator kriging (IK) (Journel, 1983) and stratified sampling of the resultant probability distributions are used together with transfer functions for the propagation of uncertainties about crop response to uniform and site-specific management. The technique is applied to the assessment of the production and economic risks of two liming management scenarios (single-rate and site-specific liming) and no lime application for a 17 ha agricultural field in southeastern Australia.

2. Theory: Modelling and Propagation of Uncertainty

Consider the problem of assessing the uncertainty attached to the response value *y* (*e.g.* pH, crop yield) of a model, such as a regression used to determine $pHCaCl_2$ values following application of a given amount of lime, or a crop model. Two sources of uncertainty may be distinguished: uncertainty arising from the incomplete knowledge of the value of the input variable *z* at **u**, and uncertainty about the parameters of the transfer function itself, say:

$$\mathrm{Y}(\mathbf{u}) = \mathrm{f}(\mathrm{Z}(\mathbf{u})) = \mathrm{a} + \mathrm{b} \cdot \mathrm{Z}(\mathbf{u}) + \varepsilon. \qquad (1)$$

where *a* and *b* are model parameters and *e* follows a normal distribution with a zero mean and a variance σ^2.

The uncertainty about the input variable is modeled by the conditional cumulative distribution function (ccdf) $F(\mathbf{u};z|(n))$, which gives the probability that the random variable Z does not exceed any given value z at **u**. In this paper, ccdfs were derived using indicator kriging (Journel, 1983). Modeling the parameter uncertainty may be approached by building for each parameter *a* and *b* a probability distribution that is assumed to be Gaussian. If the model is a regression function, the mean and variance of each probability distribution, as well as the covariance between parameters, are determined automatically. Both sources of uncertainty can then be incorporated numerically by sampling randomly the different distributions of input values, model parameters and regression errors and feeding each combination of sampled values ($z^{(l)}(\mathbf{u})$, $a^{(l)}$, $b^{(l)}$, $e^{(l)}$) into the function *f*(.) to retrieve the corresponding *y*-value $y^{(l)}(\mathbf{u})$. Practical implementation of the so-called Monte-Carlo simulation faces two problems. First, the random sampling of probability distributions is not very efficient and the total number of combinations required to sample the full space of uncertainty becomes prohibitive quickly as more parameters are considered. Second, the probability distributions of correlated parameters can not be sampled independently.

Stratified sampling is an efficient way to generate random samples that include all the range of possible values. The procedure involves stratifying the probability distribution into M disjunct equiprobable classes, then drawing randomly a value within each class. In this way, the input distributions (in particular the lower and upper tails) are represented in their entirety. If only two parameters are considered and M = 100, then 10,000 combinations of possible values can be created. If the parameters are correlated, such correlation should be reproduced by the 10,000 pairs of simulated values. This is performed by the following procedure (Davis, 1987):

1. Sample the standard normal probability distribution (zero mean and unit variance) using stratified sampling (M = 100 classes). Repeat the procedure, as many times as there are parameters. In the above example, one would need to repeat the procedure twice, yielding two sets of 100 random numbers that can be combined into 10,000 pairs, $[v^{(l)}, w^{(l)}]$, l = 1,…,10,000.
2. Decompose the variance-covariance matrix of parameters, **C**, into the product of a lower and an upper triangular matrix: **C** = **L U** (lower-upper decomposition).
3. Premultiply each vector of 2 random numbers $[v^{(l)}, w^{(l)}]$ by the lower triangular matrix **L**, and add the vector of means of the two parameters $[m_a, m_b]$:

$$\begin{bmatrix} a^{(l)} \\ b^{(l)} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11} & 0 \\ \mathbf{L}_{12} & \mathbf{L}_{22} \end{bmatrix} \cdot \begin{bmatrix} v^{(l)} \\ w^{(l)} \end{bmatrix} + \begin{bmatrix} m_a \\ m_b \end{bmatrix} \qquad (2)$$

The assumption here again is that the parameters follow a Gaussian distribution. The two sets of 10,000 values $[a^{(l)}, b^{(l)}]$, l = 1,…,10,000 will be such that the mean of each parameter as well as their variance-covariance structures are reproduced.

3. Case Study

3.1 Modelling Uncertainty about pH_{buffer}

Figure 1 (left graph). For each of the 244 samples, soil pH was measured in 0.01M $CaCl_2$ (pH_{CaCl_2}) and in Mehlich buffer (pH_{buffer}), and paired measurements were averaged, resulting in 122 observations. The *ccdf* of pH_{buffer} was derived at the nodes of a 40 × 50 grid (with 10 m spacing) using IK. The mean (E-type estimate) and variance of these ccdfs are mapped in Figure 1, which indicates that the the lowest values occurred on the westernmost part of the field while the uncertainty about pH_{buffer} estimates, as measured by the *ccdf* variance, is smallest in the central and eastern parts of the field.

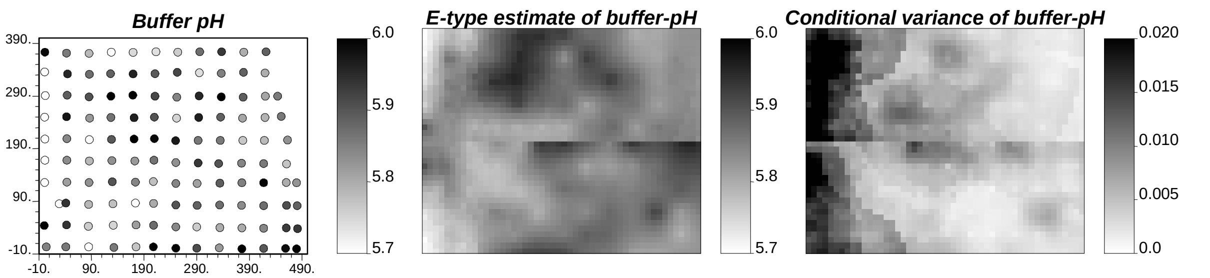

Figure 1. Location map of pH data, and maps of mean (E-type estimate) and variance (conditional variance) of IK-based ccdfs.

3.2 Propagation of uncertainty: $pHCaCl_2$

At each node, the pH_{buffer} *ccdf* was discretised into 100 equiprobable classes and one value was then randomly drawn from within every class. The set of 100 values were fed into the following equation (Viscarra Rossel and McBratney, 2001) that regresses soil pH_{CaCl_2} as a function of pH_{buffer} and the amount of lime applied:

$$pH_{CaCl_2} = 35.532 - 12.815q - 0.379p + 1.282q^2 + 0.124qp - 0.004p^2 + e \tag{3}$$

where q is the equilibrated Mehlich pH_{buffer}, p is the amount of lime applied, and $e \sim N(0,\sigma^2)$. Twenty-one rates of lime, ranging from no lime to 5 Mg/Ha (0.25 Mg/Ha increments), were considered. Equation 3 assumes that the target pH is attained after lime is applied and incorporated, usually 3 to 6 months prior to planting. The six distributions of regression parameters and the error term (e) were jointly sampled using a stratification into 5 equiprobable classes, leading to 78,125 (5^7) possible combinations of simulated parameters. Each of these combinations was used to propagate the uncertainty about the pH_{buffer} values, as represented by the distribution of 100 simulated values, through Equation 3. Hence the result at each node was a distribution of 7812,500 simulated pH_{CaCl_2} values for each amount of lime.

The three following liming management scenarios were considered:

1. No lime application (NL), which amounts at ignoring the acidity problem in the field, i.e. p=0 in expression 3.
2. Single-rate application (SR-L). For this traditional method of land management, one considered an application of 3.5 Mg/Ha corresponding to the average lime requirement (LR) of the field.
3. Variable-rate liming (VR-L). For the site-specific management scenario, at each grid node we selected out of the 21 different rates of lime, the one that yields the largest expected net profit, see next section for details about computation of this profit.

Figure 2 shows the maps of E-type estimates of pH_{CaCl_2} obtained for each of the three management scenarios. Both single-rate and variable-rate liming produced similar field-average pH_{CaCl_2} values (5.62 versus 5.59), which is expected since the field-average LR is 3.6 Mg/Ha, a value close to the single-rate application of 3.5 Mg/Ha. However, variable-rate liming reduced the spatial heterogeneity of pH_{CaCl_2}.

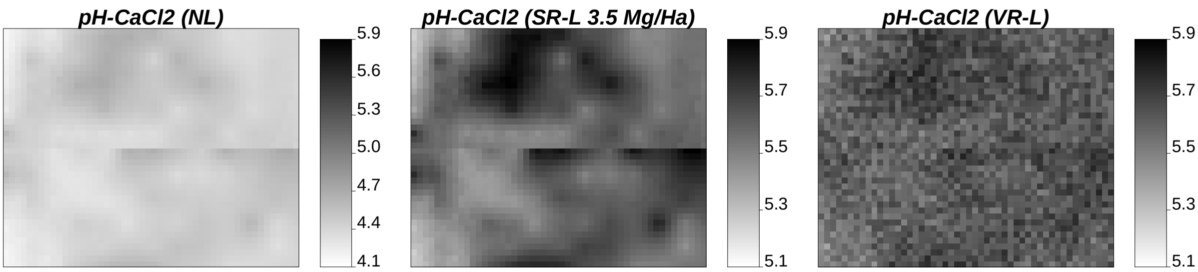

Figure 2. Maps of the mean (E-type estimate) of probability distributions of $pHCaCl_2$ values obtained for each of the three management scenarios.

3.3 Propagation of uncertainty: yield values and profitability

For each amount of lime, the simulated pH_{CaCl_2} values were then fed into the following production function (R^2 = 0.85) to derive wheat yield:

$$Y(Mg/Ha) = -10.411 + 3.648w - 0.246w^2 + e \qquad (4)$$

where w is pH_{CaCl_2}, and $e \sim N(0,\sigma^2)$. Equation 4 was derived from a wheat trial conducted by Viscarra Rossel (2000). As for Equation 3, the distributions of regression parameters and the error term (e) in Equation 4 were jointly sampled using 8 equiprobable classes, leading to 4,096 (8^4) possible combinations of simulated parameters and regression errors. The uncertainties were propagated by using the 100 pH_{CaCl_2} values as input into Equation 4 for all 4,096 combinations of the simulated parameters, producing at each node and for each amount of lime applied (0 - 5 Mg/Ha) a distribution of 409,600 simulated wheat yield values.

The profitability of lime application at each node was determined by considering the cost of lime and its application, and the price for wheat in Australia. Information and lime application costs were not considered as this study was conducted only to provide a basic insight into the current economic effects of site-specific liming. The cost of lime used in the study was AUD$55 per Mg/Ha applied (cost of lime and transport AUD$45 and application AUD$10). The price for wheat was valued at AUD$200 /Mg. The simple equation for net profit (NP) was:

$$NP = (Y \times 200) - (p \times 55) \qquad (5)$$

For each amount of lime p, the probability distribution of the net profit was numerically approximated by the distribution of 409,600 NP values obtained from the set of simulated wheat yield values. The probability of making a positive NP was estimated, for each management scenario, by the proportion of 409,600 wheat yield values that provides a revenue exceeding the cost of liming.

Figure 3 shows, for the three management scenarios discussed in Figure 2, the maps of the expected net profit (i.e. average simulated NP value) as well as the probability of making a profit. The benefits of liming acid soil are clearly illustrated: a single-rate application of 3.5 Mg/Ha caused the average NP of the field to increase by almost AUD$80 /Ha. Likewise the range of probabilities of making a profit increased to almost 100% for all but the western portions of the field where pH_{CaCl_2} values were lowest at around 5 pH units. Optimal variable-rate liming produced maps of expected NP and probability of positive NP which are very similar to single-rate maps. Therefore it appears that in this instance site-specific lime applications are not economically warranted since almost optimal profits may be obtained by applying the average 3.5 Mg/Ha rate of lime homogeneously across the field. These results seem consistent with the fact that optimal application rates vary around the average optimal LR value of 3.6 Mg/Ha.

3.4 Sensitivity Analysis

The decision to implement PA management techniques by farmers will depend mainly on two basic factors: their production objectives and their perceived consequences of alternative operations. Undoubtedly both of these will vary amongst farmers, however one can speculate that their decision to adopt will be greatly influenced by how much the implementation of technology will increase yields (hence NP) and lower their production risk (or uncertainty)**.**

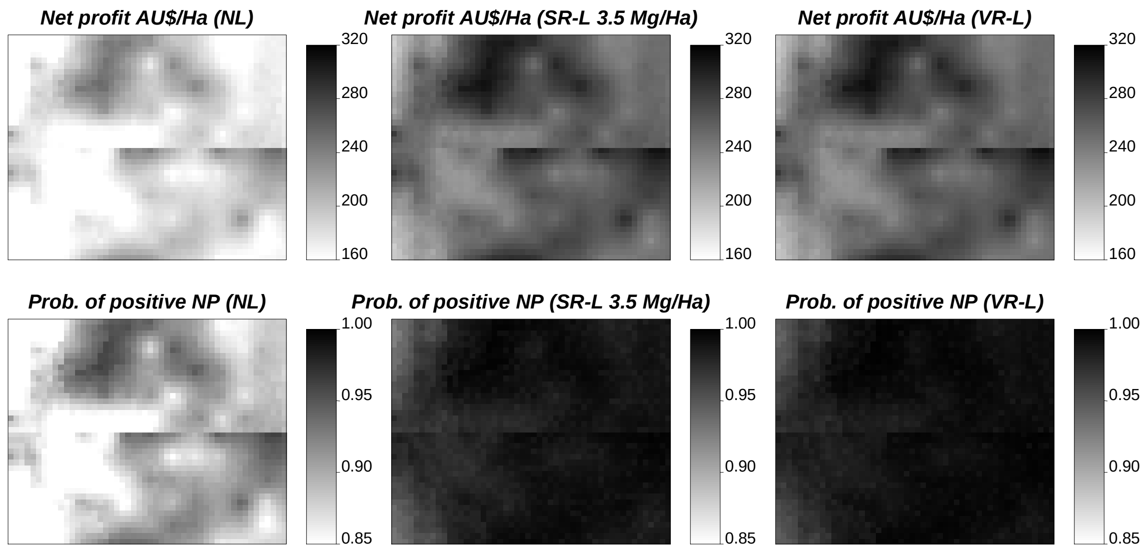

Figure 3. Maps of the expected net profit (top row) and probability of making a profit obtained for each of the three management scenarios.

The above results suggest that site-specific liming of the experimental field may not be economically justified under the cost / price conditions of the analysis, i.e., AUD$200 for the price of wheat and AUD$55 for liming. Since NP (and hence the decision of whether to adopt a site-specific liming practice) depends on the ratio of cost of lime to price of wheat, a number of different scenarios were compared to make inferences about when site-specific liming may be profitable to adopt, see Figure 4. For example, if the cost of liming drops to AUD$47 /Mg/Ha and the price of wheat remains at AUD$200 /Mg, then site-specific liming will produce a NP of AUD$290 /Ha, which is AUD$32 /Ha higher than the single-rate NP. Net profits below the critical curve (Figure 4, broken lines) do not warrant site-specific liming since the profit obtained from a single-rate application across the field is AUD$258 /Ha. Conversely, cost / price combinations above the critical curve (Figure 4, solid lines) produce profits that are higher than that resulting from a single-rate application of lime, hence under such economic conditions site-specific liming may be justified.

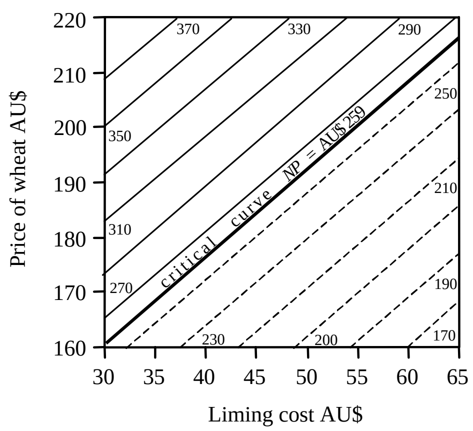

Figure 4. Sensitivity analysis for the adoption of site-specific liming in the experimental field at Kelso, NSW. Above the critical *NP* curve site-specific liming is economically justified, below critical *NP* curve a single-rate application of 3.5 Mg/Ha is more profitable resulting in a *NP* of AU$ 258 / Ha.

4. Conclusions

Non-parametric geostatistics coupled with Monte-Carlo simulation allows one to model the uncertainty attached to spatial extrapolation of soil properties and propagate it through a series of complex transfer functions. Although it can become computationally intensive, the method is very flexible in that limited assumptions are made regarding the shape of local probability distribution of input values as well as the type (i.e. linear vs non-linear) of transfer function. Uncertainty assessment can then be combined with economical functions to derive the expected net profit, and sensitivity analysis is used to explore under which conditions specific management scenarios would be more profitable than others.

Under the constraints of the analysis, single-rate liming was determined to be the best management option. The implications of these results on the null hypothesis of PA, (i.e. that given the variation in yields relative to the scale of a single field, then the optimal risk aversion strategy is uniform management), are that where variability is low and knowledge-based uncertainty high, the optimal risk aversion strategy is uniform management. A simple economic sensitivity analysis was presented to provide only an indication of when site-specific liming may be economically feasible in the Kelso field. The methodology presented may be used to help decide which type of management to implement for maximum profit. As such, this research is not specific to this particular scenario, but may be generally applied to assess the risk associated with the adoption of site-specific soil and crop management techniques.

5. References

DAVIS, M., 1987, Production of conditional simulations via the LU decomposition of the covariance matrix. *Mathematical Geology*, **19**, 91-98.

FRENCH, N.H.F., GOOVAERTS, P., KASISCHKE, E.S. 2003, Uncertainty in estimating carbon emission from Boreal forest fires. *Journal of Geophysical Research-Atmospheres*, in press.

GOOVAERTS, P., 1999, Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, **89**, 1-45.

GOOVAERTS, P., 2001, Geostatistical modelling of uncertainty in soil science. *Geoderma*, **103**, 3-26.

HANSEN, S., THORSEN, M., PEBESMA, E.J., KLEESCHULTE, S., SVENDSEN, H., 1999, Uncertainty in simulated nitrate leaching due to uncertainty in input data. A case study. *Soil Use and Management*, **15**, 167-175.

HEUVELINK, G.B.M., BURROUGH, P.A., STEIN, A., 1989, Propagation of errors in spatial modelling in GIS. *International Journal of Geographical Information Systems*, **3**, 303-322.

JOURNEL, A.G., 1983, Non-parametric estimation of spatial distributions. *Mathematical Geology*, **15**, 445-468.

McBRATNEY, A.B., 1992, On variation, uncertainty and informatics in environmental soil management. *Australian Journal of Soil Research*, **30**, 913-935.

McBRATNEY, A.B., WHELAN, B.M., SHATAR, T.M., 1997, Variability and uncertainty in spatial, temporal and spatiotemporal crop-yield and related data. In Lake, J.V., Bock, G.R. and Goode, J.A. (Eds.). *Precision Agriculture: Spatial and temporal variability of environmental quality*. Ciba Foundation Symposium 210. John Wiley & Sons Ltd., England.

PEBESMA, E.J., HEUVELINK, G.B.M., 1999, Latin hypercube sampling of Gaussian random fields. *Technometrics*, **41**, 303-312.

VAN MEIRVENNE, M., GOOVAERTS, P., 2000, Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. *Geoderma*, **102**, 75-100.

VISCARRA ROSSEL, R.A., 2000, *Managing soil acidity using precision agriculture: towards an on-the-go proximal soil pH and lime requirement sensing system*. Unpublished PhD thesis, Department of Agricultural Chemistry and Soil Science, The University of Sydney.

VISCARRA ROSSEL, R.A., McBRATNEY, A.B., 2001, A response-surface calibration model for rapid and versatile site-specific lime-requirement predictions in south-eastern *Australia. Australian Journal of Soil Research*, **39**, 185-201.