3B2v8.07f
XML:ver:5.0.1
10/7/06 17:36
DSS – V031 : 31011
Prod.Type:
pp.139–153(col.fig.:NIL)
ED:
PAGN: SCAN:

Developments in Soil Science, volume 31
P Lagacherie, A.B. McBratney and M. Voltz (Editors)

Chapter 11

OPTIMIZATION OF SAMPLE CONFIGURATIONS FOR DIGITAL MAPPING OF SOIL PROPERTIES WITH UNIVERSAL KRIGING

Gerard B.M. Heuvelink, Dick J. Brus and Jaap J. de Gruijter

Abstract

Digital soil mapping makes extensive use of auxiliary information, such as that contained in remote sensing images and digital elevation models. However, it cannot do without taking samples of the soil itself. Therefore, methods and guidelines need to be developed that assist users in designing spatial sample configurations for use in digital soil mapping. Existing geostatistical methods are insufficient because these typically have been developed for situations in which there is no auxiliary information. In this chapter, we explore how the existing methods may be extended to the case in which the auxiliary information is spatially exhaustive and where soil mapping is done using universal kriging. We develop and illustrate a methodology that optimizes the spatial configuration of observations by minimizing the spatially averaged universal kriging variance. The universal kriging variance incorporates trend estimation error as well as spatial interpolation error. Hence, the optimized sample configuration strikes a balance between an optimal distribution of observations in feature and geographic space. The results show that optimal distribution in feature space prevails over optimal distribution in geographic space when the stochastic component of the universal kriging model is weakly spatially autocorrelated. It also prevails when the total number of observations is small. In all other cases, the optimal configuration is close to that obtained with minimization of only the spatial interpolation error. Application to a variety of real-world cases with multiple predictors and different spatial dependence structures is needed to support and generalise these preliminary findings.

11.1 Introduction

Digital soil mapping aims at spatial prediction of soil properties by combining soil observations at points with auxiliary information, such as contained in remote sensing images, digital elevation models and climatological records. Direct observations of the soil are important for two main reasons. First, they are used to establish the character and strength of the relationship between the soil property of interest and the auxiliary information. Second, they are used to

improve the predictions based on the auxiliary information, by spatial interpolation of the differences between the observations and predictions. The twofold use of soil observations is nicely illustrated in universal kriging, which is commonly used for digital soil mapping when the auxiliary information is spatially exhaustive (Heuvelink and Webster, 2001; McBratney et al., 2003). Universal kriging treats the soil property of interest as the sum of a deterministic trend, which is taken as a regression on the auxiliary variables, and a spatially autocorrelated stochastic residual. The soil observations are used to estimate the regression coefficients as well as to interpolate the residual.

Fieldwork and laboratory analyses are labour-intensive and costly. It is therefore important that a sampling strategy is employed in which the available resources are used effectively. This means that the twofold use of soil observations in digital soil mapping should be incorporated in choosing an optimal sample configuration. However, the two uses of the soil observations generally impose conflicting requirements on the sample configuration. Estimation of the relationship between the soil property and the auxiliary information benefits from a large spread of the observations in feature space, while spatial interpolation of the differences between observations and predictions gains from a uniform spreading of the observations in geographic space (Lesch et al., 1995; Müller, 2001; Hengl et al., 2003). As yet, it is unclear how these two requirements should be weighed and how the weighing depends on the characteristics of a particular case. Existing methods for optimization of sample configurations have mainly focused on situations without auxiliary information, in which case the soil observations are only used for spatial interpolation. We review these methods briefly.

The first option for selecting sample locations in the context of spatial interpolation is sampling on a regular grid. Commonly used grid shapes are triangular and square lattices. These grid shapes can be evaluated by computing the maximum or average kriging variance, and choosing the shape for which the chosen criterion attains it minimum, given the number of sample points. Yfantis et al. (1987) show that for variograms with a small relative nugget, the maximum kriging variance is minimal for a triangular grid. Besides the configuration, we must also decide on the grid spacing (i.e., sampling density). Obviously, the smaller the grid spacing, the smaller is the average or maximum kriging variance. If a minimum precision is required and a reasonable variogram can be postulated, the variogram can be used to calculate the grid spacing required to achieve this minimum precision (McBratney and Webster, 1981; Olea, 1984; Christakos and Olea, 1992). So, for a range of grid spacings we may calculate the average or maximum kriging variance, plot these values against the grid spacings in a graph and determine the required grid spacing from this

graph. Lark (2000) showed how a fuzzy set of required grid spacings can be calculated in situations where we are uncertain about the variogram.

In practice, regular grid sampling may be suboptimal for several reasons. Sampling may be hampered by enclosures that are inaccessible for sampling. The study area may have an irregular shape or the dimensions of the area may be small compared to the grid spacing. The latter may result in prediction error variances that are relatively large near the boundaries of the area. Also, we may want to use previously made observations at locations that cannot be matched with the grid. In these situations, minimization of the kriging variance criterion will yield a sample with irregular spacing. The criterion may be simplified in terms of the distance between sample points and prediction points, leading to the so-called spatial coverage samples, also referred to as space-filling samples (Royle and Nychka, 1998). Brus et al. (Chapter 14) propose the mean of the squared shortest distance as a minimization criterion, and minimize it with the well-known *k*-means algorithm from cluster analysis. Alternatively, we can define the minimization criterion in terms of the kriging variance. Sacks and Schiller (1988) and van Groenigen et al. (1999) minimized the average and maximum kriging variance using a simulated annealing algorithm.

Little work has been done on optimization of sample configurations when the observations are used both for estimating regression coefficients and for spatial interpolation. Lesch et al. (1995) selected sample sites for multiple linear regression. The fitted regression model was then used to predict the target variable at all locations with observations of the auxiliary variables. Also, the sample locations were required to be 'spatially representative of the entire survey area'. They proposed a site selection algorithm that starts from a sample that is closest to the optimum response surface (RS) design calculated from the experimental design theory. The second and third best RS designs are also calculated. Then points of the best RS design are swapped with corresponding points in the second and third best RS design. Swaps are accepted if they improve the spatial coverage. Hengl et al. (2003), struggling with the same problem, proposed an 'equal range' design. In this procedure, the study region is stratified on the basis of the frequency distribution of the auxiliary variables. Stratification limits are set at equal distances in feature space. From each stratum an equal number of sample points is selected randomly, thus ensuring that the entire sample has uniform spacing in feature space. Many samples are generated in this way, and the one with the best spatial coverage is retained. Müller (2001, Sections 5.4 and 5.5) avoids heuristic approaches by introducing information matrices and applying standard gradient algorithms to design measures defined on these matrices. However, approximations were needed to be able to apply this method in a situation where both a trend and a spatially autocorrelated residual are present.

In this chapter, we address the same problem but aim for solutions that are not heuristic and do not involve approximations. This requires that we make some assumptions. We assume that we have only one target soil property that is measured on a numerical scale, that the auxiliary information is spatially exhaustive, and that the relationship between the target variable and the auxiliary variables takes the form of a linear regression equation, with known structure but unknown coefficients. We also assume that the variogram of the residual is known. Especially, the first and last assumptions are quite restrictive, but in the 'Conclusions' section we outline how these might be relaxed for further development of the method. Under the above assumptions, the problem boils down to minimization of the maximum or spatially averaged universal kriging variance, which can be solved using numerical simulation. We illustrate the theoretical results with synthetic examples.

11.2 Universal kriging

We consider the following model (Christensen, 1990):

$$Z(\mathbf{s}) = \sum_{j=0}^{m} \beta_j \cdot g_j(\mathbf{s}) + \varepsilon(\mathbf{s}), \tag{11.1}$$

where $Z(\boldsymbol{s})$ is the target soil property; $\boldsymbol{s} = (x\ y)'$, a two-dimensional spatial coordinate; $g_j(\boldsymbol{s})$, explanatory variables (note that $g_0(\boldsymbol{s})\equiv 1$ for all $\boldsymbol{s}$); β_j, regression coefficients and $\varepsilon(\boldsymbol{s})$, a normally distributed residual with zero mean and constant variance $c(0)$. The residual ε is possibly spatially autocorrelated, as quantified through an autocovariance function or variogram.

In what follows, it will be convenient to use matrix notation, so that Eq. (11.1) may be rewritten as

$$Z(\mathbf{s}) = g'(\mathbf{s}) \cdot \beta + \varepsilon(\mathbf{s}), \tag{11.2}$$

where g and β are column vectors of the m+1 explanatory variables and m+1 regression coefficients, respectively. The universal kriging prediction at an unobserved location s_0 from n observations $Z(s_i)$ is given by

$$\tilde{Z}(\mathbf{s}_0) = (\mathbf{c} + \mathbf{G}(\mathbf{G}'\mathbf{C}^{-1}\mathbf{G})^{-1}(\mathbf{g}_0 - \mathbf{G}'\mathbf{C}^{-1}\mathbf{c}))'\mathbf{C}^{-1}\mathbf{z}, \tag{11.3}$$

where $\mathbf{G}$ is the $n \times (m+1)$ matrix of predictors at the observation locations; $\mathbf{g}_0$, the vector of predictors at the prediction location; $\mathbf{C}$, the $n \times n$ variance–covariance matrix of the n residuals; $\mathbf{c}$, the vector of covariances between the residuals at the observation and prediction locations; and $\mathbf{z}$, the vector of observations $z(s_i)$. $\mathbf{C}$ and $\mathbf{c}$ are derived from the variogram of ε.

The universal kriging prediction-error variance at s_0 is given by

$$\sigma^2(\mathbf{s}_0) = \mathbf{c}(0) - \mathbf{c}'\mathbf{C}^{-1}\mathbf{c} + (\mathbf{g}_0 - \mathbf{G}'\mathbf{C}^{-1}\mathbf{c})'(\mathbf{G}'\mathbf{C}^{-1}\mathbf{G})^{-1}(\mathbf{g}_0 - \mathbf{G}'\mathbf{C}^{-1}\mathbf{c}) \tag{11.4}$$

Although Eqs. (11.3) and (11.4) present well-known results (Christensen (1990, Section VI.2); Müller (2001, Section 2.2)), it is instructive to take a closer look. The universal kriging variance incorporates both the prediction error variance of the residual (first two terms on the right-hand side of Eq. (11.4), i.e., $c(0)–\mathbf{c}'\mathbf{C}^{-1}\mathbf{c}$) as well as the estimation error variance of the trend (third term on the right-hand side of Eq. (11.4)). Comparison of these two components of the prediction variance yields insight into what the main source of error is in a specific situation. If the first component is much larger than the second, then trend estimation has a negligible effect on the prediction error. If, on the other hand, the second component is much larger than the first, then uncertainty about the trend coefficients will be the main contributor to the universal kriging variance. Although it cannot be easily deduced from Eq. (11.4), the contribution of the second component tends to be small when the number of observations is large compared to the number of predictors. Note also from Eq. (11.4) that the universal kriging prediction variance does not depend on the data values themselves, but merely on the spatial pattern of the predictors and the covariance structure of the residual. This enables computation of the prediction variance prior to collecting the data, provided that the predictors and the variogram of the residual are known. This attractive property will be used in the next section to compute optimal sample configurations.

Universal kriging with externally defined predictors is sometimes also referred to as *kriging with external drift* (Goovaerts, 1997; Bishop and McBratney, 2001). In fact, some authors reserve the term universal kriging exclusively for those cases where the trend functions $g_j(s)$ are polynomials in the coordinates (Deutsch and Journel, 1998; Wackernagel, 1998). *Regression kriging* (Odeh et al., 1995; Hengl et al., 2004) uses, as a starting point, the same model as given in Eq. (11.1), and provided the same variogram of residuals is used to estimate the trend and krige the residuals, it yields the same results as presented in Eqs. (11.3) and (11.4).

11.3 Calculation of the optimal sample configuration

Given the model defined through Eq. (11.1), we now want to optimize the sample configuration. For this we first need to define a criterion, which we will take as the spatially averaged universal kriging variance. Thus, given the sample size, our objective is to locate the sample points such that the average universal kriging variance is minimized. In theory, it is very easy to compute the optimal

configuration. For each possible configuration, simply compute the criterion and select the one that has the smallest value from all the configurations. In practice, the problem is that there is a huge (or infinite) number of possible configurations, meaning that evaluation of all possible configurations becomes impossible. Instead, some efficient search algorithm will have to be employed. Here, we will use spatial simulated annealing, following van Groenigen et al. (1999, 2000).

Simulated annealing is an iterative procedure that works as follows. Starting from an arbitrary initial configuration of sample points for which the criterion has been computed, a small perturbation in the configuration is brought about by moving one sample point in a random direction. The magnitude of the movement is random as well but it cannot be larger than a chosen value, which is gradually decreased as the iteration continues. The criterion is recomputed for the new configuration. If it is smaller than the previous criterion then the movement is accepted and a new perturbation is made, using the new configuration as a starting point. If, on the other hand, the movement has led to an increase in the criterion's value then there is still a chance that the new (worse) configuration is accepted. This is done to be able to escape from local minima. The probability of accepting a worsening configuration is also gradually decreased as the iteration continues. The iterative procedure is repeated until a maximum number of iterations is reached or until acceptance of new configurations has become very rare. In practice, thousands of iterations or even more are needed to obtain satisfactory results.

There is no guarantee that the optimal configuration will be obtained but confidence can be gained if redoing the iteration with different starting configurations yields approximately the same solutions. Also, the algorithm may be checked by running simplified examples for which it is known that the optimal configuration must have certain symmetry properties.

11.4 Synthetic examples

We use synthetic examples to analyse how the balance between optimization in feature and geographic space works out in practice. We examine how the balance is affected by three types of parameter settings. These are the number of observations, the structure of the trend and the degree of spatial autocorrelation of the stochastic residual. The number of observations is taken as 4, 9 and 16. In practice, usually there will be more observations, but a limited number of observations are used here because it facilitates the interpretation of the results. Three different trends are evaluated (see Eq. (11.1)):

(1) $m = 0$
(2) $m = 1$, $g_1(\mathbf{s}) = x$
(3) $m = 2$, $g_1(\mathbf{s}) = x$ and $g_2(\mathbf{s}) = x^2$

In other words, we consider a constant trend, a trend that is linear in the x-coordinate and a trend that is quadratic in the x-coordinate. Of course, in digital soil mapping one would use 'real' auxiliary variables, but here we take polynomials of a spatial coordinate as an example, again in order to facilitate the interpretation of the results.

Three variograms for the residual ε are distinguished:

(1) exponential variogram with zero nugget variance and spatial range equal to the diagonal of the spatial domain (case of strong spatial autocorrelation),
(2) exponential variogram with 50%-nugget variance and spatial range equal to one quarter of the diagonal of the spatial domain (case of weak spatial autocorrelation),
(3) pure nugget variogram (case of no spatial autocorrelation).

In total, this gives 27 cases. The spatial domain is taken as a square area, with the x-axis in the horizontal and the y-axis in the vertical. The sample configurations resulting from running the simulated annealing optimization are given in Figures 11.1–11.3.

We also computed the gain in precision that is obtained by including trend estimation in the optimization of the sample configuration. We did this by comparing the average universal kriging variance of the optimized configuration with that obtained with the configuration optimized for when there is no trend. The results are presented in Tables 11.1–11.3.

11.5 Discussion of results

Optimization of the sample configuration in case of a constant trend and a spatially autocorrelated residual (top rows of Figures 11.1 and 11.2) yields a regular square grid. It is square and not triangular, which can be explained from the fact that the numbers of observations chosen are squares of whole numbers. Note that the iterations have not fully converged because there are some irregularities in the configurations, notably for the 16 observation cases. Increasing the number of iterations would resolve this problem, but the changes made to the configurations would only be marginal. Comparison of the strong and weak spatial autocorrelation case (top rows of Figures 11.1 and 11.2 again) shows that for the former, observations are placed somewhat closer to the boundaries of the spatial domain. The 'optimized' configuration for the case of a constant trend and no spatial autocorrelation (top row of Figure 11.3) yields a random pattern. This is because, in this case, it really does not matter where the sample points are located. All configurations yield the same value for the average kriging variance and are therefore equally good. The configurations shown in the top row of Figure 11.3 are therefore completely arbitrary.

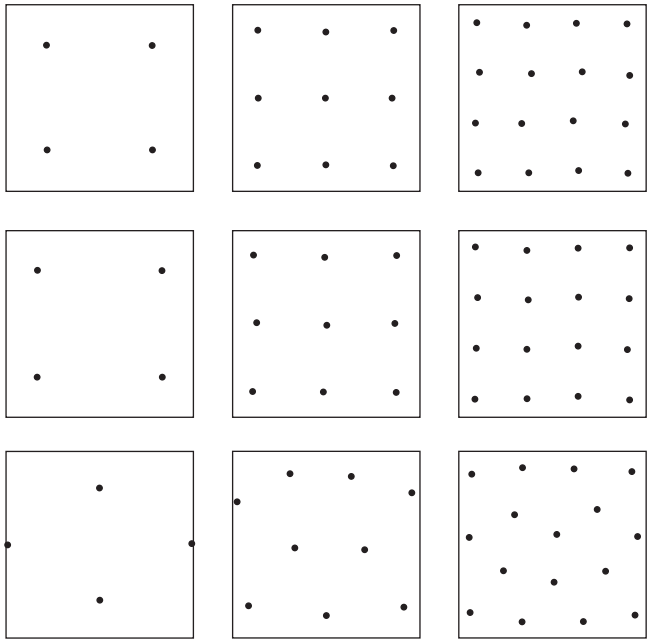

Figure 11.1. *Optimized sample configurations in the case of strong spatial autocorrelation. Cases considered are constant trend (top row), linear trend in x-coordinate (middle row) and quadratic trend in x-coordinate (bottom row). Sample sizes are 4 (left column), 9 (middle column) and 16 (right column).*

Comparison of the resulting configurations for the constant trend and linear trend case shows that including trend estimation in the optimization causes the sample points to move towards the extreme values of the x-coordinate (i.e., the left and right boundaries of the spatial domain). The effect is not as strong for the strong autocorrelation case as for the weak autocorrelation case. This is because the spatial interpolation error increases more strongly in the strong autocorrelation case when sample points are moved to the boundaries of the spatial domain. A balance is sought between reduction of trend estimation error and spatial interpolation error, and this turns out to yield a configuration closer to the constant trend configuration when the regression residual is strongly autocorrelated. If the regression residual is spatially uncorrelated (Figure 11.3), then all efforts can be directed to the minimization of the trend estimation error variance because the spatial interpolation error is the same, regardless of which configuration is used. For the trend that is linear in the x-coordinate (middle row of Figure 11.3) this yields solutions in which half of the sample points are placed at locations with the minimum value for x and half at locations with the maximum value for x. This is the minmax D-optimal design (Hengl et al., 2003). The

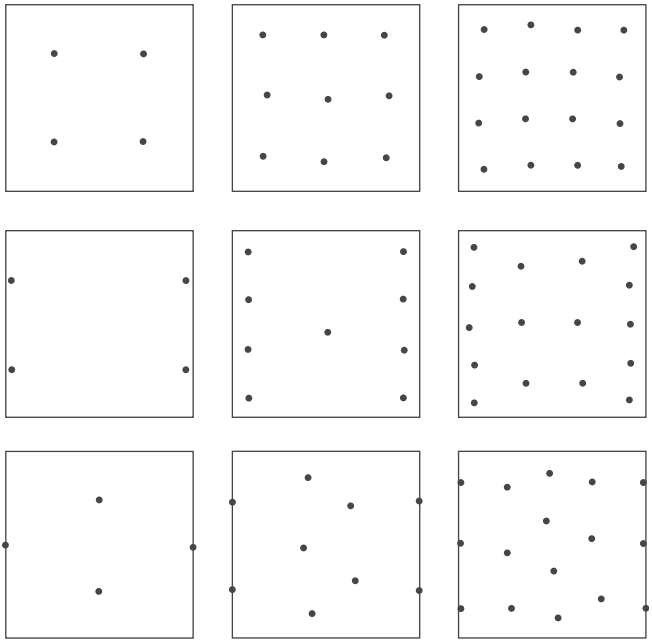

Figure 11.2. *Optimized sample configurations in the case of weak spatial autocorrelation. Cases considered are constant trend (top row), linear trend in x-coordinate (middle row) and quadratic trend in x-coordinate (bottom row). Sample sizes are 4 (left column), 9 (middle column) and 16 (right column).*

y-coordinates of these locations have no effect on the criterion and are again arbitrarily chosen. It is interesting to observe that in the case of weak spatial autocorrelation, including a linear trend distorts the structure of the optimal configuration in the case of 9 and 16 observations. The rectangular grid configuration is no longer optimal and is replaced by a configuration in which more sample points are placed at extreme values of x. For the strong autocorrelation case the grid structure remains intact but is only stretched in the x-direction.

The results of the quadratic trend case confirm those obtained for the linear trend case. Again a balance is sought between optimization in feature and geographic space, whereby optimization in geographic space is more important when spatial autocorrelation is strong. The only difference with the linear trend case is that in this case, optimization in feature space aims at placing half of the sample points at locations with a central x-value, one quarter at locations with a minimum x-value, and one quarter at locations with a maximum x-value. This is most evident from the configurations presented in the bottom row of Figure 11.3, where optimization in feature space is the only consideration. This also explains why the optimal configuration in case of a quadratic trend and four observations

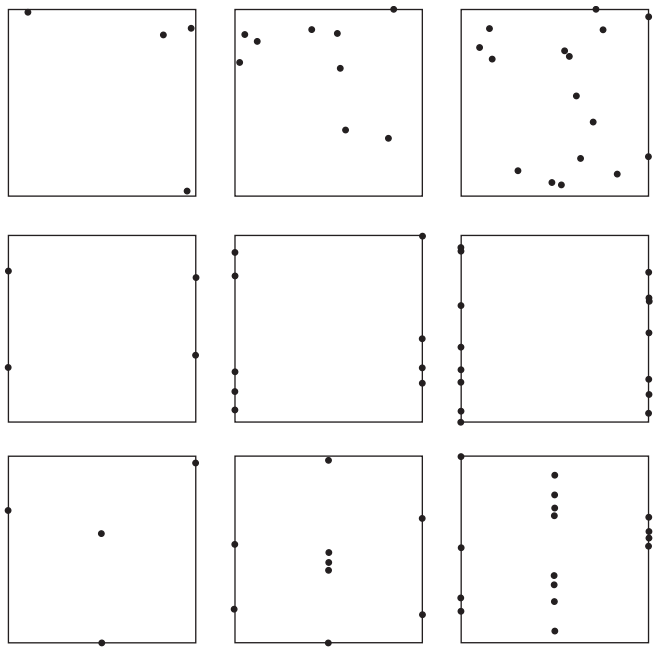

Figure 11.3. *Optimized sample configurations in the case of no spatial autocorrelation. Cases considered are constant trend (top row), linear trend in x-coordinate (middle row) and quadratic trend in x-coordinate (bottom row). Sample sizes are 4 (left column), 9 (middle column) and 16 (right column).*

Table 11.1. *Gain in precision by including trend estimation error in sample configuration optimization. Case of strong spatial autocorrelation. The gain is expressed as the ratio of the average kriging variance obtained with the 'ordinary kriging sample configuration' (i.e. top row of Figure 11.1) and that obtained with the 'universal kriging sample configuration' (i.e. appropriate row of Figure 11.1).*

Trend model	$n = 4$	$n = 9$	$n = 16$
Constant (no trend)	1.0000	1.0000	1.0000
Linear in x-coordinate	1.0282	1.0029	1.0008
Quadratic in x-coordinate	∞	1.0339	1.0020

(strong and weak spatial autocorrelation case) yields a diamond-shaped configuration instead of a rectangle. The advantage of the diamond-shaped configuration is that it satisfies the optimal conditions for trend estimation, while still maintaining a fairly uniform spreading in geographic space. Note also that

Table 11.2. *Gain in precision by including trend estimation error in sample configuration optimization. Case of weak spatial autocorrelation. The gain is expressed as the ratio of the average kriging variance obtained with the 'ordinary kriging sample configuration' (i.e. top row of Figure 11.2) and that obtained with the 'universal kriging sample configuration' (i.e. appropriate row of Figure 11.2).*

Trend model	$n = 4$	$n = 9$	$n = 16$
Constant (no trend)	1.0000	1.0000	1.0000
Linear in x-coordinate	1.1722	1.0391	1.0121
Quadratic in x-coordinate	∞	1.1711	1.0469

Table 11.3. *Gain in precision by including trend estimation error in sample configuration optimization; case of no spatial autocorrelation. The gain is expressed as the ratio of the average kriging variance obtained with the 'ordinary kriging sample configuration' (i.e. top row of Figure 11.3).*

Trend model	$n = 4$	$n = 9$	$n = 16$
Constant (no trend)	1.0000	1.0000	1.0000
Linear in x-coordinate	1.1324	1.0935	1.0516
Quadratic in x-coordinate	2.0454	1.1319	1.0935

these results are in agreement with well-known results from the theory of experimental design (Atkinson and Donev, 1992; Hengl et al., 2003).

The figures in Tables 11.1–11.3 present the gains in precision. A value equal to 1 means that there is no gain in precision at all. This is the case for all constant trend models because in these cases there is no trend and the two optimum sample configurations are the same. These 'gains' are included in the tables for completeness only. The tables show that the gain is relatively large when the spatial autocorrelation is weak and when the sample size is small. Note that the gain is infinitely large for the case of a quadratic trend and sample size 4. This can be explained as follows. The square configurations as given in the top left of Figures 11.1 and 11.2 yield only two effective observations for trend estimation because two observations that have the same x-coordinate provide the same information as one observation. Thus, two effective observations are available, and these are used to estimate a trend with three unknown regression coefficients. This yields an ill-posed problem, leading to infinite estimation variance. Hence, the gain in precision by employing the configurations as given in the bottom left of Figures 11.1 and 11.2 is infinitely large.

11.6 Conclusions

The synthetic examples presented in this chapter show that taking trend estimation into account can have a marked effect on the optimized sample configuration. However, the effect decreases as the number of sample points increases. Here, we used only 16 sample points as a maximum, which is in fact a small number for real-world situations. In future work, it should be analysed how a further increase in the number of sample points will affect the optimized sample configuration and the associated precision gain. Gathering from the results presented here, one might suspect that in case of many sample points, little gain is to be expected from including trend estimation in optimizing the sample configuration. After all, gains in precision that are 5% or less are not very interesting, and extrapolation of the gains presented in Tables 11.1–11.3 suggests that for larger datasets (tens or hundreds of sample points) the gain will be much smaller than that. However, we cannot be certain about this until we have done the analysis because there are also reasons that suggest that the gain in precision may still be substantial in real-world situations. First, it is important to realise that in many practical cases, much of the variation in the target soil property is explained by the predictors. As a result, the residual will often be weakly spatially autocorrelated. We have seen in the synthetic examples that the effect on the optimal sample configuration and the precision gain is larger in case of weak spatial autocorrelation. Second, we should also bear in mind that in the practice of universal and regression kriging for digital soil mapping, it frequently occurs that a large number of predictors are used. Here we used a maximum of two predictors, but real-world cases typically use between five and ten predictors or even more (e.g. Bishop and McBratney (2001); Hengl et al. (2004)). Including a large number of predictors implies that more regression coefficients need to be estimated, making trend estimation more important. In addition, using more predictors also increases the risk of running into multicollinearity problems. How multicollinearity affects the optimization of the sample configuration is unclear and needs to be investigated as well.

It is remarkable that the optimal sample configuration is not dependent upon the strengths of the relationships between the predictors and the target soil variable. Intuitively, one would expect that predictors that explain a large portion of the variation in the target variable should have a stronger influence than predictors that have a small predictive power. This turns out to be not the case. Even when a predictor does not explain a single bit of variation in the target soil variable, it will influence the sample configuration as much as a predictor with a strong explanatory power. This can be explained as follows. Once we have decided to include a predictor in the universal kriging model, we need to estimate the associated regression coefficient. If we do a poor job then we might

end up with an estimate that is far off from its true value. In the example of a predictor that has no explanatory power, we might end up with an estimate that differs substantially from zero while the true (but unknown) regression coefficient is in fact zero. This is potentially as harmful as wrongly estimating a regression coefficient that is in reality nonzero. Thus, in both cases, we should choose the sample configuration such that the estimation error variance is as small as possible (of course also taking the effect on the spatial interpolation error into account).

In order to avoid heuristic or approximate solutions, we made several assumptions that simplified the problem of optimizing the sample configuration for digital soil mapping. First, we assumed that the auxiliary information is spatially exhaustive and used in a universal kriging procedure. If the auxiliary information is not spatially exhaustive then other methods will have to be employed to take advantage of the auxiliary information. One possibility would be cokriging (see also McBratney and Webster (1983)). Optimization of the sample configurations of the primary and secondary variables used in cokriging is not fundamentally different from the optimization problem tackled in this chapter, although some adaptations would be required. Second, we also assumed that we had only one target soil variable. In practice, many soil properties must usually be mapped in the same project. Different soil variables will yield different optimal sample configurations. A reasonable solution to come up with a single configuration seems to be to find a compromise between these configurations, or perhaps better to use a criterion that is a combination of the criteria used for the separate soil variables. In our analysis, we also assumed that the variogram of the stochastic residual of the universal kriging model is known. This is not a very realistic assumption. One solution to circumvent this assumption is to first conduct a preliminary fieldwork that is specifically aimed at estimation of the variogram. The result is then used to develop an optimal sample configuration for minimization of the universal kriging variance. Note that the algorithm presented in this chapter can easily optimize sample configurations subject to fixed locations of previously collected observations. Another solution would be to include the uncertainty in the variogram parameters in the optimization. This would make an elegant solution, but it would require the use of so-called model-based geostatistics (Diggle et al., 1998), which is, as yet, not easy. Sample configuration optimization in combination with model-based geostatistics may also be used in the case where the target soil variable is measured on a categorical scale, where the regression is nonlinear or when the residual follows a non-Gaussian distribution.

In this chapter, we used the average universal kriging variance as a criterion, but other criteria may also be used. A frequently used alternative is the maximum kriging variance. Quite different criteria, such as trend estimation and

variogram parameter estimation criteria, or weighed combinations of multiple criteria (Müller, 2001), may also be considered. This can all be done with little modification to the methodology. In the synthetic examples, we used trends that are linear or quadratic functions in the x-coordinate, but it should be stressed that the methodology presented works for any number and any type of trend functions. We now need to apply the methodology to real-world situations with trend functions that are derived from relevant auxiliary information. The tool is there and is ready to be used. It will help us to gain more insight into how the optimal sample configuration depends on the characteristics of a particular situation and to design optimal sample configurations for practical applications, thus encouraging the efficient use of fieldwork resources.

References

Atkinson, G.L., Donev, A.N., 1992. Optimum Experimental Design. Clarendon Press, Oxford.

Bishop, T.F.A., McBratney, A.B., 2001. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma 103, 149–160.

Christakos, G., Olea, R.A., 1992. Sampling design for spatially distributed hydrogeologic and environmental processes. Adv. Water Resour. 15, 219–237.

Christensen, R., 1990. Linear Models for Multivariate, Time, and Spatial Data. Springer, New York.

Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software and User's Guide, 2nd Ed.. Oxford University Press, New York.

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. Appl. Stat. 47, 299–350.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75–93.

Hengl, T., Rossiter, D.G., Stein, A., 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. Australian J. Soil Res. 41, 1403–1422.

Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: past, present, and future. Geoderma 100, 269–301.

Lark, R.M., 2000. Designing sampling grids from imprecise information on soil variability, an approach based on the fuzzy kriging variance. Geoderma 98, 35–59.

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

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

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

McBratney, A.B., Webster, R., 1983. Optimal interpolation and isarithmic mapping of soil properties. 5. co-regionalization and multiple sampling strategy. J. Soil Sci. 34, 137–162.

Müller, W.G., 2001. Collecting Spatial Data: Optimum Design of Experiments for Random Fields, 2nd Ed.. Physica-Verlag, Heidelberg.

Odeh, I., McBratney, A.B., Chittleborough, D., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67, 215–226.

Olea, R.A., 1984. Sampling design optimization for spatial functions. Math. Geol. 16, 369–392.

Royle, J.A., Nychka, D., 1998. An algorithm for the construction of spatial coverage designs with implementation in SPLUS. Comp. Geosci. 24, 479–488.

Sacks, J., Schiller, S., 1988. Spatial designs. In: S. Gupta and J. Berger (Eds.), Statistical Decision Theory and Related Topics IV, Vol. 2. Springer Verlag, New York.

van Groenigen, J.W., Pieters, G., Stein, A., 2000. Optimizing spatial sampling for multivariate contamination in urban areas. Environmetrics 11, 227–244.

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

Wackernagel, H., 1998. Multivariate geostatistics: an introduction with applications, 2nd Ed.. Springer-Verlag. QA :1

Yfantis, E.A., Flatman, G.T., Behar, J.V., 1987. Efficiency of kriging estimation for square, triangular and hexagonal grids. Math. Geol. 19, 183–205.

AUTHOR QUERY FORM

ELSEVIER

Developments in Soil Science

JOURNAL TITLE:	**DSS-V031**
ARTICLE NO:	**31011**

Queries and / or remarks

Query No	Details required	Author's response
AQ1	Please provide the publisher location for ref. Wackernagel.	
UC REF	If references appear under section “Uncited References”, then cite at relevant places in the text. In case of nonavailability of citation, the corresponding references will be deleted from the reference list. Please ignore this query if there is no 'Uncited Reference' section.	