A PURPOSIVE SAMPLING SCHEME FOR PRECISION AGRICULTURE[♦]

A.B. McBRATNEY, B.M. WHELAN

Australian Centre for Precision Agriculture, McMillan Building A05, University of Sydney, NSW 2006, Australia.

D.J.J. WALVOORT

Survey Methods Department, Winand Staring Centre, SC-DLO & Laboratory of Soil Science & Geology, Agricultural University, Wageningen, The Netherlands

B. MINASNY

Department of Agricultural Chemistry and Soil Science, Ross St Building A03, University of Sydney, NSW 2006, Australia.

ABSTRACT

Precision Agriculture brings the information revolution to agriculture. However, it is a fundamental requirement for the provision of quality information that a sufficient quantity relevant data has been previously obtained and analysed. For Precision Agriculture, the data sets will be required to embrace a range of crop, soil, landscape and other environmental attributes over significant areas. While there is no doubting the quantity/density of data obtainable from modern crop yield sensors, the quantity and density of data gathered on other important variables is less than optimal. The ultimate solution is the development of real-time sensors for these attributes. In the mean time, the conflict between manual sampling expense and sampling scheme resolution has lead to a grid sampling density for most attributes that fails to capture the true variability within a field. This dilemma must be overcome while the community awaits more technological solutions.

Here a new formal method of spatial sampling, the variance quad-tree (VQT) method, is introduced. This scheme samples sparsely in uniform areas and more intensively where variation is large. This is motivated by a non-stationary covariance structure of the variables of interest. Previous work has shown this to be the case for spatial variation of crop yield. The method is equivalent to sampling on a regular grid in a transformed space, the method implicitly defining the transformation. Applications for the VQT method should include sampling for Digital Elevation Models (DEM), soil sampling and weed density estimation. The method is compared with grid sampling and 'smart' sampling methods.

[♦] McBratney, A.B., Whelan, B.M., Walvoort, D.J.J., Minasny, B., 1999. A purposive sampling scheme for precision agriculture. In: Precision Agriculture '99 (J.V. Stafford Ed.), pp 101-110, Sheffield Academic Press, Sheffield.

INTRODUCTION

Field observation has been traditionally based on discrete sampling procedures using either a grid-based or statistically-based random sampling strategy. Sampling by grid is at present a laborious procedure if large areas are to be tested. For the production of accurate maps, the appropriate sampling scheme and minimum lag must be determined, and as highlighted by McBratney & Whelan (these proceedings), the inherent variability expected in most attributes would suggest the principal sampling lag should be as small as possible. This inevitably leads to a conflict between accuracy and sampling cost.

Much of the soil and crop attribute sampling for Precision Agriculture has been conducted manually on grids of 100m or larger. Birrell et al. (1996) graphically depict an observed increase in the confidence range for the spatial representation of soil pH, K and P associated with an increase in sampling grid from 25m to 100m. The common choice of grid size appears to indicate that reducing sampling cost has triumphed over accurate spatial resolution. Such economic rationality will always restrict the detail in information obtainable from discrete sampling procedures. While the discrete sampling procedure will continue to be employed until remote and proximal sensing is commonplace, it is imperative that more appropriate and efficient methods of sampling scheme design are developed.

Here we introduce a new technique for sampling scheme design based on the variance quad-tree (VQT) method. This scheme samples sparsely in uniform areas and more intensively where variation is large. The aim is to maximise efficiency of the sampling scheme (to minimise costs) while ensuring that the variability within the sampling area is characterised effectively.

METHODS

The VQT-algorithm

1. **Encapsulate the area of interest in a rectangle;**
2. **Split the rectangle into four equally sized areas, that have the same shape as the first rectangle;**
3. **Compute for each area (*A*) the following quantity**

$$\bar{\gamma}^*(A_h, A_h) \approx \sqrt{\frac{1}{2}\sum_{i=1}^{n_h}\sum_{j=1}^{n_h}\left[z(x_i) - z(x_j)\right]^2} \qquad (1)$$

where
$\gamma(\cdot)$ is the semivariance;
$x, x' \in A$;
n is the number of discrete points x_i and x_j in A;
$z(\cdot)$ is the variable of interest.

This does not require the fitting of a parametric variogram model.

4. **The area with the largest $\bar{\gamma}^*(A_h, A_h)$ is split into four equally sized areas.**

Equation (1) is derived from Neyman's theorem (after Cochran, 1977, p. 98-99) as follows:

In stratified random sampling the sampling variance is minimised for a fixed total size of sample n if:

$$n_h = n \frac{A_h S_h}{\sum A_h S_h} \tag{2}$$

where:

n_h is the number of samples in stratum h;
n is the total number of samples;
S_h is the standard deviation within stratum h.

If we want to select the same number of samples in each stratum, i.e. $n_i = n_j \ \forall i, j$ then A_iS_i should be equal to A_jS_j for all i and j. Therefore, the algorithm based on Neyman's theorem splits strata for which A_hS_h is largest. Splitting continues until all strata have approximately the same product of A_h and S_h. Substituting the square root of the average semivariance for S_h gives:

$$\begin{aligned} & A_h S_h \equiv \\ & A_h \sqrt{\bar{\gamma}(A_h, A_h)} = \\ & A_h \sqrt{\frac{1}{A_h^2} \int\limits_{A_h} \int\limits_{A_h} \gamma(x - x')\mathrm{d}x\mathrm{d}x'} \approx \\ & n_h \sqrt{\frac{1}{n_h^2} \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} \gamma(x_i - x_j)} \approx \\ & \sqrt{\frac{1}{2} \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} \left[z(x_i) - z(x_j) \right]^2} \end{aligned} \tag{3}$$

Therefore, sampling simply amounts to taking the same number of random samples within each stratum. A_h is approximated by n_h, because the VQT algorithm is applied to data on a fixed grid.

5. **Repeat steps 3 to 4 for all areas until some predefined number of iterations has been completed, or** $\max(\bar{\gamma}(A, A)) < \varepsilon$**, where** ε **is the maximum** $\bar{\gamma}(A, A)$ **allowed.** If i is the iteration index, then the total number of areas to evaluate at iteration step i is given by $n_i = 3i + 1$, but only 4 of these are new areas which require step 3.
6. **Sampling locations are placed at random within the** $n_i = 3i + 1$ **cells (strata) of the approximately equal-variance partition.** According to Neyman's theorem, samples should be taken at random for design based inference. Alternatively, model-based sampling designers might place the sampling locations at the centre of the cells.

This algorithm results in an approximately equal variance partition of the field. This alogorithm is similar in effect to the methods discussed by Monestiez et al. (1997) but the computational approach is quite different.

RESULTS & DISCUSSION

The capabilities of the VQT algorithm will be shown using crop yield data obtained from a real-time crop yield monitor. Block kriging using local variograms and 20m blocks (Whelan & McBratney (in press)) has been used to predict wheat yield from one field (48 ha) in successive seasons (1996, 1997) which are shown in Figure 1. There is greater variability expressed in the 1997 season where the mean wheat yield is smaller.

Using the number of iterations of the VQT algorithm to control the partitioning we can examine the variance expected within each strata as the iterations increase. Figure 2 shows how the variance within the strata decreases rapidly as the number of iterations increases. The within-strata variance begins to plateau after 67 partitions in 1996 and 76 partitions in 1997. Such graphs could be used to calculate the number of strata required to ensure a desired maximum variance is encapsulated within each strata.

(a) (b)

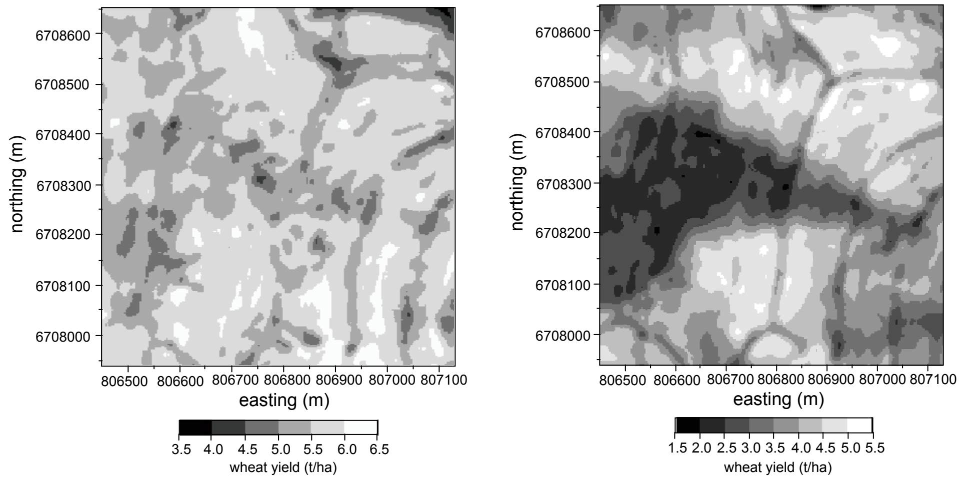

FIGURE 1. Wheat yield maps for (a) the 1996 season and (b) the 1997 season in a single field in north-western NSW, Australia.

(a) (b)

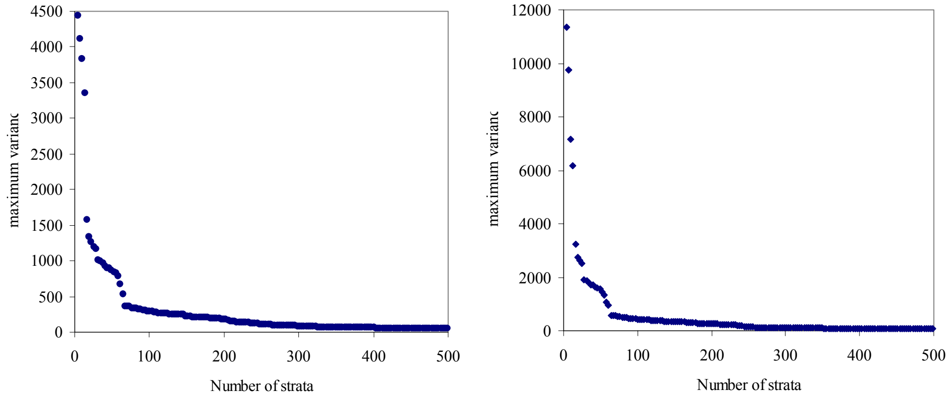

FIGURE 2. Maximum variance per strata as a function of the number of partitions requested for (a) the 1996 season and (b) the 1997 season.

Calculating 500 partitions, the VQT algorithm should identify areas within the field where the yield changes greatest over a small area. Figure 3 shows the 500 strata as calculated for the two yield maps. This analysis highlights the suitability of the VQT for designing sampling schemes for continuously variable attributes based on intensive covariate data. In this instance, crop yield variability could be used to direct sampling for soil attributes that are considered to influence crop yield. By ensuring that each sample is representative of a uniform support variance, the information gathered should be of greater value than merely grid sampling the same area.

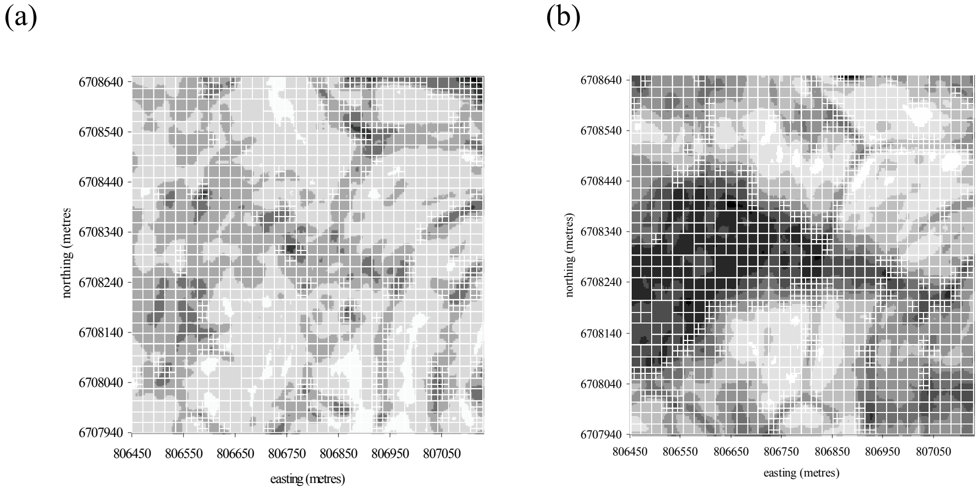

FIGURE 3. 500 partitions using the VQT algorithm on wheat yield for (a) the 1996 season and (b) the 1997 season.

In Figure 3b, where greater variance is exhibited in the crop yield, the VQT partitioning seeks out and delineates the boundaries of possible soil management units. This would be expected as the borders between discontinuities in yield-influencing soil attributes should be reflected in greater local variance in crop yield. The increased variance attributed to contour bank lines and harvest artefacts are also recognised. While harvest artefacts such as headland effects are undesirable for accurate variance partitioning, for this example the fact that they are identified is very useful in demonstrating the operation of the VQT algorithm.

Figure 4 helps confirm that the variance partitioning based on crop yield maps could be of use in identifying areas in the field which may have high variability in soil attributes. Figure 4a is an aerial photograph of bare soil in the field. A lighter coloured, lighter textured soil unit can be traced entering from the centre left of the field boundary.

The influence of this soil unit is evident in the yield map in Figure 1b and when the bare soil photograph is overlain with the VQT partitioning (Figure 4b), the correspondence between higher yield variability and the soil unit boundary, contour banks and headlands is evident.

With the present cost of soil sampling and analysis (especially in Australia), designing a sampling scheme such as this with 10 samples per hectare would be unsuitable for widespread application. Soil sampling schemes, in the USA particularly, are currently designed on a minimum 1 ha grid basis. Evidence from a number of studies (Franzen & Peck, 1995; Haneklaus et al., 1997; McBratney & Pringle, 1997; McBratney & Whelan (this proceedings)) suggests that this resolution is too coarse and that sample site separation should be between 30m (0.1ha) and 80m (0.6 ha) for reliable site characterisation of soil attributes.

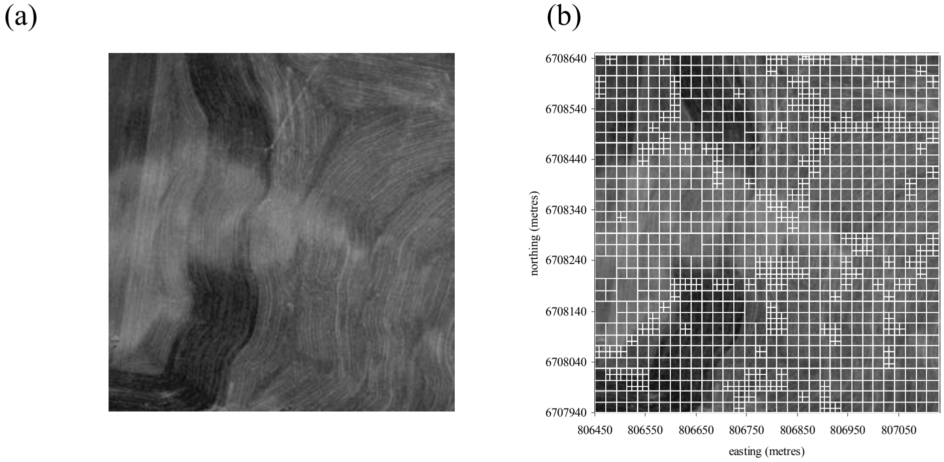

FIGURE 4. Bare soil aerial photograph of the field (a) and overlain VQT partitions for the 1997 season (b).

Figure 5 shows the VQT set to 82 strata, which would equate to the larger grid sampling scale of 0.6 hectares. The VQT approach has identified a number of areas within the field that warrant investigation at a finer scale than the rest of the field. This would be missed using a regular grid approach. Obviously, with a smaller sample number, the VQT algorithm would provide a less spatially uniform sampling design. In such a case the process of quadrant division would incorporate more variance within each partition and division would begin in the more variable areas of the field.

The VQT approach provides a means for locating the support area for subsequent sampling. Exact location of the sampling site within these areas can be undertaken in a number of ways. The two most common approaches are shown in Figure 6. Figure 6a displays the sampling points located at the central points of each partition, 6b is a random allocation.

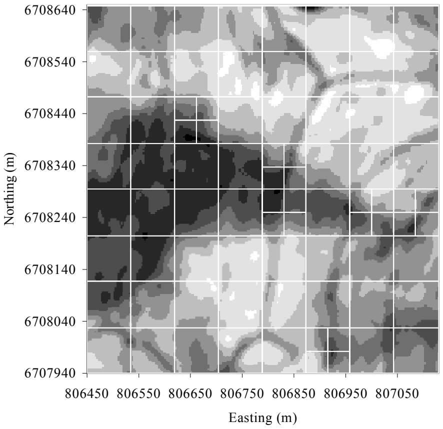

FIGURE 5. VQT algorithm applied to the 1997 season yield data using 82 partitions.

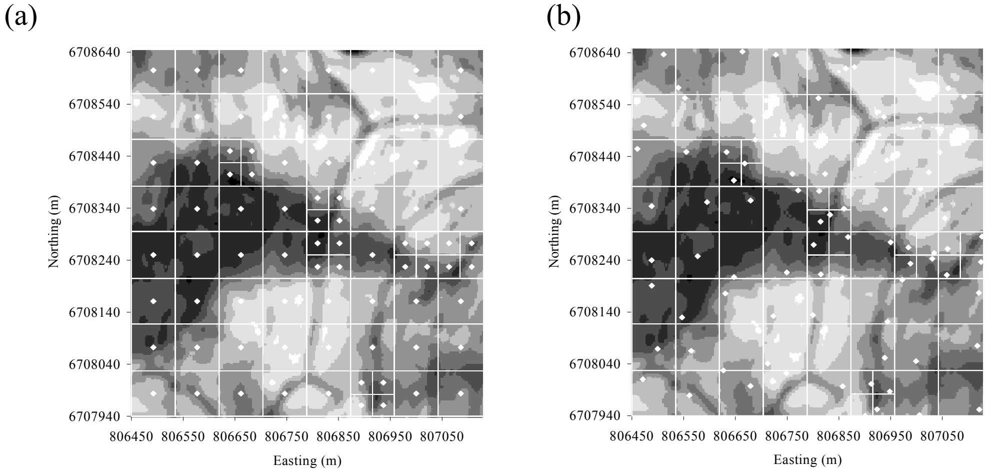

FIGURE 6. Soil sampling schemes based on the VQT partitioning and (a) partition centre point (b) random point allocation.

It is also possible to combine the information from the two seasons yield maps through a clustering process. An in-house fuzzy clustering program (FuzMe, 1998) was applied to the data and identified two significant clusters. By allocating final membership based on a fuzzy membership greater than 0.5, the underlying cluster map in Figure 7 has been constructed. The VQT algorithm was applied to the fuzzy cluster membership values and sample sites have been allocated based on partition central points (Figure 7a) and random point allocation (Figure 7b). The partitioning is different to that shown in Figure 6, but the clustering is obviously dominated by the yield response pattern in 1997.

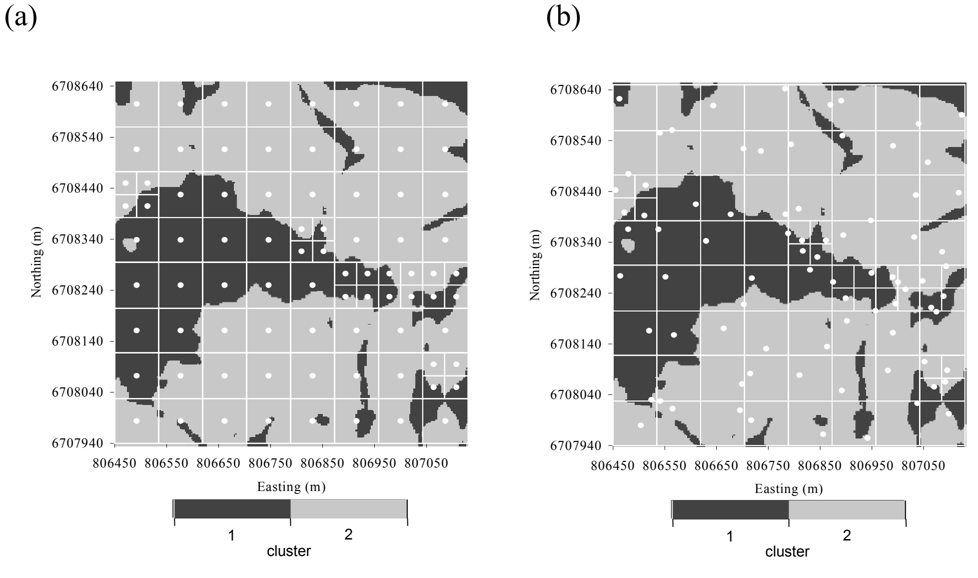

FIGURE 7. Soil sampling schemes based on fuzzy class membership and subsequent VQT partitioning. Sampling sites designated by (a) partition centre point (b) random point allocation.

CONCLUSIONS

The VQT algorithm provides a means for partitioning the variance in continuously variable attributes so that the resulting strata contain data of equal variability. When data is available at the required resolution, this technique is expected to provide a more efficient method of designing sampling schemes for covariates than a regular grid. The resulting sampling scheme is also likely to allow a more informed description of the variability in the covariate by purposely increasing sampling density in areas believed to be more variable.

Here crop yield has been used to direct within-field soil sampling. The technique would also be highly suitable for soil/crop sample scheme design based on elevation data, aerial or satellite imagery, or for the reverse operation when detailed soil data is available. While continuous sampling and analysis systems for the important variables contributing to crop yield variability remain in development, such a method for sampling scheme design should prove valuable.

ACKNOWLEDGMENT

We thank Dr Jaap de Gruijter (Staring Centre, DLO, Wageningen, The Netherlands) for suggesting the Neyman criterion as a basis for partitioning.

REFERENCES

Birrell, S.J., Sudduth, K.A. & Kitchen, N.R. (1996). Nutrient mapping implications of short-range variability. In P.C. Robert, R.H. Rust & W.E. Larson (ed.) *Precision Agriculture: Proceedings of the 3rd International Conference on Precision Agriculture,* ASA, Madison, pp. 207-216.

Cochran, W.G. (1977). *Sampling Techniques.* Third Edition, John Wiley and Sons, New York. 428p.

Franzen, D.W. & Peck, T.R. (1995). Sampling for site-specific application. In P.C. Robert, R.H. Rust & W.E. Larson (ed.) *Site-Specific Management for Agricultural Systems: Proceedings of the 2nd International Conference on Site-Specific Management for Agricultural Systems,* ASA, Madison, pp. 535-552.

Haneklaus, S., Ruehling, I, Schroeder, D. & Schnug, E. (1997). Studies on the variability of soil and crop fertility parameters and yields in different landscapes of northern Germany. In J.V. Stafford (ed.) *Precision Agriculture '97: Proceedings of the 1st European Conference on Precision Agriculture,* Bios, London, pp. 785-792.

McBratney, A.B. & Pringle, M.J. (1997). Spatial variability in soil - implications for precision agriculture. In J.V. Stafford (ed.) *Precision Agriculture '97: Proceedings of the 1st European Conference on Precision Agriculture,* Bios, London, pp. 3-31.

McBratney, A.B. & Whelan, B.M. (these proceedings). The ‘Null Hypothesis’ of precision agriculture. *Proceedings of the 2nd European Conference on Precision Agriculture,* Bios, London.

Monestiez, P., Meiring, W., Sampson, P.D. & Guttorp, P. (1997). Modelling non-stationary spatial covariance structure from space-time monitoring data. In J.V. Lake, Bock, G. & Goode, J. (eds) *Precision agriculture: spatial and temporal variability of environmental quality.* Ciba Foundation Symposium 210. John Wiley & Sons, Chichester. pp. 38-51.

Whelan, B.M. & McBratney, A.B. (in press). Prediction uncertainty and implications for digital map resolution. *Proceedings of the 4th International Conference on Precision Agriculture,* ASA, Madison.