

A conditioned Latin hypercube method for sampling in the presence of ancillary information[☆]

Budiman Minasny*, Alex B. McBratney

Australian Centre for Precision Agriculture, Faculty of Agriculture, Food and Natural Resources, The University of Sydney, Australia

Received 9 September 2005; received in revised form 3 December 2005; accepted 16 December 2005

Abstract

This paper presents the conditioned Latin hypercube as a sampling strategy of an area with prior information represented as exhaustive ancillary data. Latin hypercube sampling (LHS) is a stratified random procedure that provides an efficient way of sampling variables from their multivariate distributions. It provides a full coverage of the range of each variable by maximally stratifying the marginal distribution. For conditioned Latin hypercube sampling (cLHS) the problem is: given N sites with ancillary variables (X), select x a sub-sample of size n ($n \ll N$) in order that x forms a Latin hypercube, or the multivariate distribution of X is maximally stratified. This paper presents the cLHS method with a search algorithm based on heuristic rules combined with an annealing schedule. The method is illustrated with a simple 3-D example and an application in digital soil mapping of part of the Hunter Valley of New South Wales, Australia. Comparison is made with other methods: random sampling, and equal spatial strata. The results show that the cLHS is the most effective way to replicate the distribution of the variables.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Sampling design; Spatial design; Optimisation; Simulated annealing; Soil survey

1. Introduction

The purpose of sampling is to obtain data that enable the estimation of some statistical parameter, or spatial predictions of some properties over an area. Surveys of soil (and other) materials at the earth's surface can be planned optimally by making the best use of the prior information at various

locations. Sampling is constrained by the financial and available resources, thus an efficient sampling strategy is sought. This has practical applications in soil survey for mapping, and for establishing sites for monitoring networks. Several sampling strategies have been developed in the earth sciences, namely:

- Estimating a variogram (Pettitt and McBratney, 1993).
- Estimating variables within an area using spatial interpolation. This includes methods that minimise the kriging variance (McBratney et al., 1981; van Groenigen et al., 1999).

[☆] Code available from server at <http://www.iamg.org/CGEditor/index.htm>

*Corresponding author. Tel.: +61 2 9036 9043;
fax: +61 2 9351 3706.

E-mail address: budiman@acss.usyd.edu.au (B. Minasny).

- Providing an optimal spatial coverage (Royle and Nychka, 1998).
- Establishing a prediction model from secondary or ancillary variables (Hengl et al., 2003; Lesch et al., 1995).

Brus and de Gruijter (1997) discussed the two fundamental approaches to soil sampling: namely design-based, which follows the classical survey, and model-based, following geostatistical analysis. Design-based sampling, which is mainly based on probability theory, allows unbiased estimates of the statistical parameters. Model-based approaches are more suited where spatial variation and its prediction are important. Most sampling are designed to provide a good spatial coverage of the area, or to cover the spatial variation. Royle and Nychka (1998) proposed space coverage filling designs whereby sampling sites are selected so as to minimise a criterion that is only a function of the distance between the sampling locations and non-sampling locations. Brus et al. (2004) used *k*-means clustering to partition the area of interest into geographically compact subregions, from which one sample point is selected purposively or randomly. This ensures a good geographical coverage of the area, especially for field with irregular shapes.

In many applications, sampling in an area can be assisted and guided by the presence of ancillary data or secondary information. For example an area may have information or maps from geophysical measurements, remote sensing images, vegetation maps, geology maps, digital elevation models (DEM) and their terrain attributes. Goovaerts (1997) called this exhaustive secondary information, where secondary information or ancillary data is available at all locations in the area of interest. The ancillary data are often useful for improving spatial estimation, particularly when the primary data are limited (Goovaerts, 1997). These ancillary variables occupy a hypercube in the feature, or variable, or attribute space. Thus rather than sampling the geographical space, we can sample the feature space. The challenge is to design a sampling scheme that selects sampling locations which can cover the whole feature space.

Lesch et al. (1995) developed an algorithm for the purpose of calibration of the electromagnetic induction (EMI) data to soil electrical conductivity, and clay content. The area had been previously surveyed with an EMI sensing equipment, and soil sampling was required to provide predictive functions relating EMI readings with soil properties.

The sampling strategy covered the range of EMI values and also ensured a geographical spread of the sampling sites. McBratney et al. (1999) introduced the variance quad-tree method, which sampled sparsely in uniform areas and more intensively in areas where variation is large.

In digital soil mapping (DSM), the prediction of soil properties and soil classes is based on forming relationships between observed soil attributes and ancillary soil and environmental variables (McBratney et al., 2003). These ancillary data usually can be obtained relatively cheaply over large areas, e.g. a DEM and its derivatives, and satellite images. The ancillary data are assumed to have some relationship with the soil variables. The samples are used to predict soil variables over the whole area either using regression, kriging with external drift, or cokriging (Goovaerts, 1997). There are few studies focussing on sampling techniques for this purpose. Gessler et al. (1995) provided a scheme that randomly samples along the whole range of compound topographic index (CTI) values. McKenzie and Ryan (1999) used terrain attributes, climatic and geological data to stratify an area into classes, and then random selection of samples within the class. The aim is to maximise the efficiency of the sampling scheme while ensuring that the variability within the sampling area is characterised adequately. Hengl et al. (2003) proposed sampling along the principal components of the environmental variables; the number of samples taken from each of the components is related to the proportion of the total variance described by each of the principal component. Heuvelink et al. (2004) designed a sampling scheme for spatial prediction using kriging. This is done by minimising the universal kriging prediction variance. However, the requirements are: known form of the trend, and the spatial structure of the residual of the model. This information is usually not available for areas that have not been sampled previously.

For sampling in the presence of ancillary data, it would be beneficial to cover the range of values of each of the ancillary variable. The challenge is to select sampling sites which can cover the hypercube of the feature space. One possibility is to use Latin hypercube sampling (LHS) (McKay et al., 1979), a constrained Monte-Carlo sampling scheme. This paper outlines the theoretical background of LHS, describes the method and provides a recipe for conditioned Latin hypercube sampling (cLHS). This paper then describes the algorithm and implementation in Matlab (MathWorks, 2005), and illustrates it

with examples. We provide an example from soil survey, though recognising that this algorithm applies equally in other fields.

2. Latin hypercube sampling

LHS is a stratified random procedure that provides an efficient way of sampling variables from their multivariate distributions. It was initially developed for the purpose of Monte-Carlo simulation, efficiently selecting input variables for computer models (McKay et al., 1979; Iman and Conover, 1980). It has been used in soil science and environmental studies for assessing the uncertainty in a prediction model (Minasny and McBratney, 2002), and in geostatistics for simulation of Gaussian random fields (Pebesma and Heuvelink, 1999; Zhang and Pinder, 2004).

LHS follows the idea of a Latin square where there is only one sample in each row and each column. Latin hypercube generalises this concept to an arbitrary number of dimensions. In LHS of a multivariate distribution, a sample size n from multiple variables is drawn such that for each variable the sample is marginally maximally stratified. A sample is maximally stratified when the number of strata equals the sample size n and when the probability of falling in each of the strata is n^{-1} (McKay et al., 1979).

LHS works as in the following manner: given k variables X_1, \dots, X_k the range of each variable X is divided into n equally probable intervals (strata), then for each variable a random sample is taken at each interval (stratum). The n values obtained for each of the variables are then paired with each other either in a random way or based on some rules.

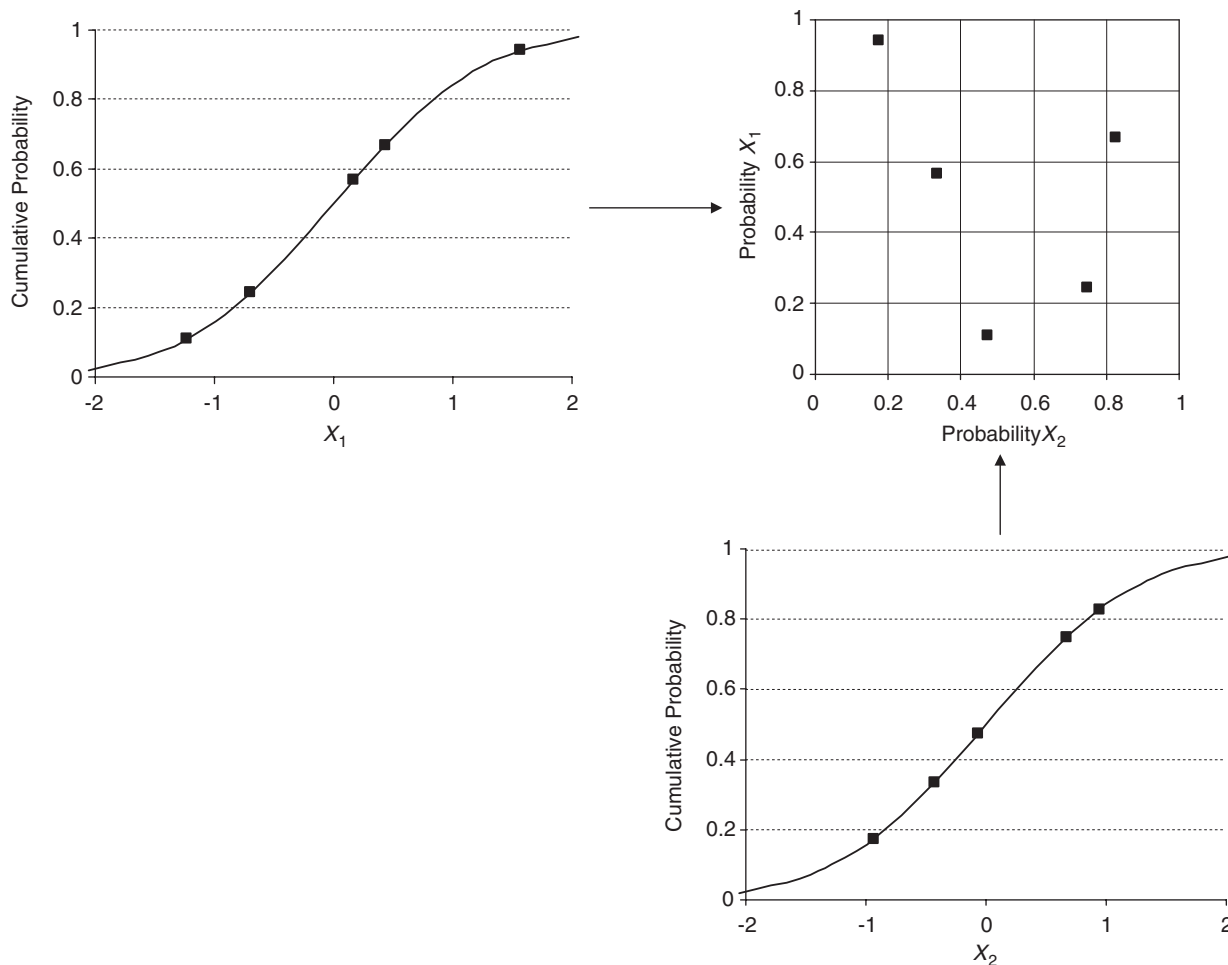


Fig. 1. Example of LHS for 2 variables with normal distribution. Cumulative probability for each variable is divided into 5 equal strata, and a random sample is taken at each strata. Five samples from each variable are then paired in random manner forming a Latin square.

Finally we have n samples, where the samples cover the n intervals for all variables. Thus, this sampling scheme does not require more samples for more dimensions (variables). This method ensures that each of the variable in X is represented in a fully stratified manner. An example for two-dimensional data is given in Fig. 1.

The LHS algorithm is as follows:

- divide the distribution of each variable X into n equiprobable intervals;
- for the i th interval, the sampled cumulative probability can be written as:

$$\text{Prob}_i = (1/n)r_u + (i-1)/n,$$

where r_u is a uniform random number ranging from 0 to 1;

- transform the probability into the sampled value x using the inverse of the distribution function F^{-1} ;

$$x = F^{-1}(\text{Prob});$$

- the n values obtained for each variable x are paired randomly or in some prescribed order with the n values of the other variables.

3. Conditioned Latin hypercube sampling

For sampling of existing ancillary data, we cannot directly apply LHS to the multivariate distribution. The samples selected by the conventional LHS represent a combination of the multivariate variables that may not exist in the real world. For example, consider the elevation and slope data from a site in the Hunter Valley (further example will be given in Section 4.2). Fig. 2 shows the scatter plot of the cumulative probability of the elevation and slope data as grey dots. We used the LHS algorithm as provided in Section 2 to draw 10 Latin hypercube samples. The resulting 10 samples are shown as circles in Fig. 2, we can see that only 4 out of the 10 samples coincide with data points, and the other 6 samples do not exist in the real world. This problem will be more pronounced when dealing with larger number of variables and sample sites. A solution is to keep repeating the LHS from the distribution and search through the data and identify the data that are taxonomically most similar to the combination of values chosen. Another better solution is to pick sites which can form a Latin hypercube in the feature space.

This becomes an optimisation problem: given N sites with ancillary data (X), select n sample sites ($n \ll N$) so that the sampled sites x form a Latin hypercube, or the multivariate distribution of X is maximally stratified. This requires a search procedure, in this paper we propose the cLHS with search algorithm based on heuristic rules combined with an annealing schedule (Metropolis et al., 1953). This algorithm works on both continuous and categorical data.

For continuous variables, first we define the sample size n . Each component of X (size $N \times k$) is divided into n equally probable strata based on their distribution, and x (size $n \times k$) is a sub-sample of X . We define matrix η which counts the number of x that fall into each of the defined stratum:

$$\eta = \begin{bmatrix} \eta_{11} & \eta_{12} & \cdot & \eta_{1k} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \eta_{n1} & \cdot & \cdot & \eta_{nk} \end{bmatrix},$$

where the rows represent the strata s_1, \dots, s_n and the columns represents variable X_1, \dots, X_k . A true LHS will have values of 1 for all cells of matrix η . The challenge is to find x that has the value of η close or equal to 1.

We present the cLHS algorithm as follows:

- (1) Divide the quantile distribution of X into n strata, and calculate the quantile distribution for each of the variables, q_j^i, \dots, q_j^{n+1} , for $j = 1, \dots, k$, and $i = 1, \dots, n+1$. Calculate \mathbf{C} , the correlation matrix for X .
- (2) Pick n random samples from N , x ($i = 1, \dots, n$) are the sampled sites and r ($i = 1, \dots, N - n$) is the reservoir or unsampled sites. Calculate \mathbf{T} , the correlation matrix of x .
- (3) Calculate the objective function:
For continuous variables:

$$O_1 = \sum_i^n \sum_{j=1}^k |\eta(q_j^i \leq x_j < q_j^{i+1}) - 1|,$$

where $\eta(q_j^i \leq x_j < q_j^{i+1})$ is the number of x_j that fall in between quantiles q_j^i and q_j^{i+1} .

For categorical data, the objective function is to match the probability distribution for each of the class

$$O_2 = \sum_{j=1}^c \left| \frac{\eta(x_j)}{n} - \kappa_j \right|,$$

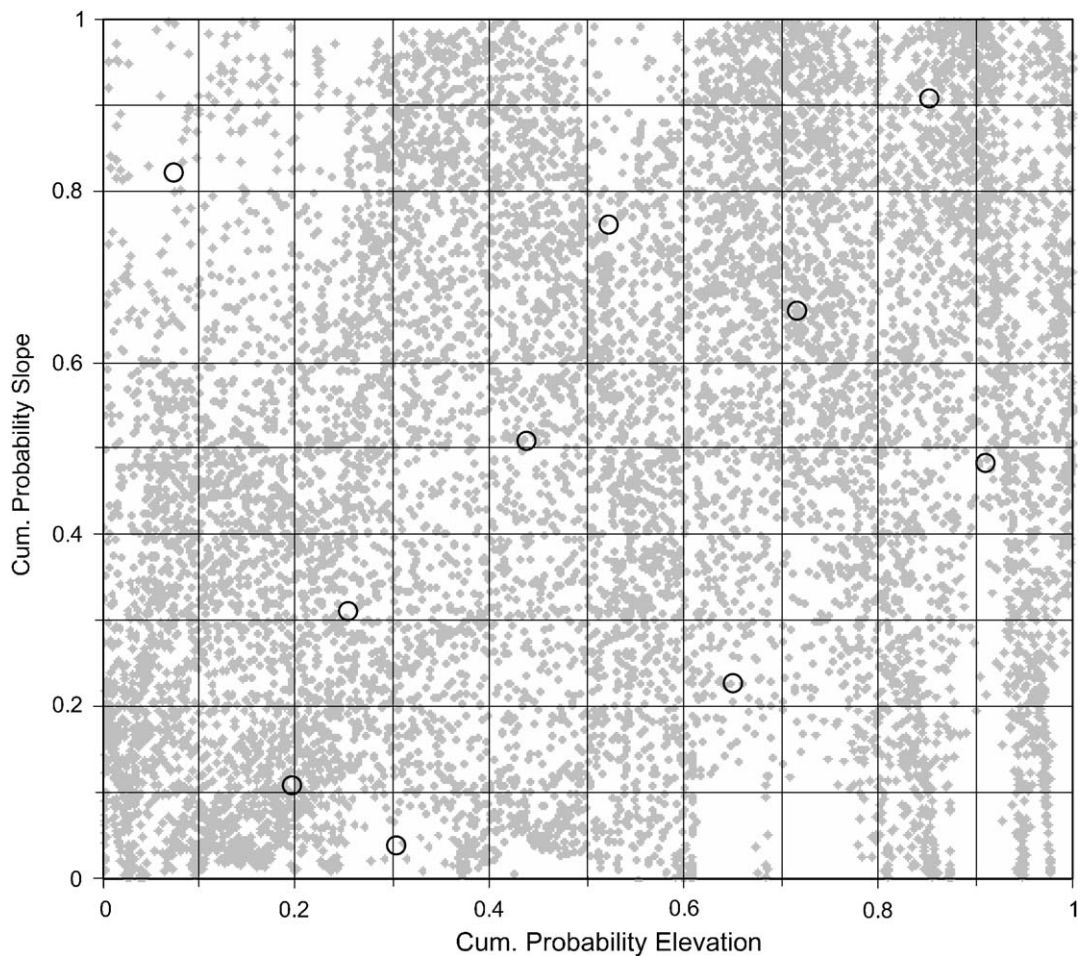


Fig. 2. Scatter plot of cumulative probability of elevation and slope for Hunter Valley area. Grey dots represent data in area, and black circles represent samples drawn using conventional LHS.

where $\eta(x_j)$ is the number of x that belongs to class j in sampled data, and κ_j is the proportion of class j in X .

To ensure that the correlation of the sampled variables will replicate the original data, we add another objective function

$$O_3 = \sum_{i=1}^k \sum_{j=1}^k |c_{ij} - t_{ij}|,$$

where c is the element of \mathbf{C} , correlation matrix of X , and t is the equivalent element of \mathbf{T} , correlation matrix of x .

The overall objective function is

$$O = w_1 O_1 + w_2 O_2 + w_3 O_3,$$

where w is the weight given to each component of the objective function, for general applica-

tions w is set to 1 for all components of the objective function.

- (4) Perform an annealing schedule (Press et al., 1992):

$Metro = \exp[-\Delta O/T]$, where ΔO is the change in objective function, and T is a cooling temperature (between 0 and 1) which is decreased by a factor d at each iteration.

- (5) Generate $rand$ a uniform random number between 0 and 1, if $rand < Metro$, accept the new values, otherwise discard changes.
- (6) Try to perform changes: Generate a uniform random number $rand$.

If $rand < p$,

Pick a sample randomly from x and swap it with a random site from reservoir r .
Else,

Remove sample(s) from x which has the largest $\eta(q_j^i \leq x_j < q_j^{i+1})$ and replace it with random site(s) from the reservoir r .
End if.

The value of p is between 0 and 1 showing the probability of the search being a random search or systematically replacing the samples that worst fit the strata.

(7) Go to step (3)

Repeat steps (3)–(7) until the objective function value falls beyond a given stopping criterion or for a defined number of iterations.

The final sample will represent a true or approximate Latin hypercube of the feature space whereby the distribution and multivariate correlation will be preserved. The algorithm described above is coded in Matlab (Mathworks, 2005); the code is available from author's website www.usyd.edu.au/su/agric/acpa. This technique has several advantages, namely:

1. Continuous as well as categorical variables can be incorporated.
2. The sampling is based on the empirical distribution of the original data, thus it is nonparametric.
3. Spatial coordinates can be incorporated to ensure a good spread of the sampling points if this is required.
4. Additional constraints can be imposed on the objective function, e.g. distance from roads and field boundaries.

4. Examples and application

4.1. A simple 3-D example

To illustrate the concept, we generate a 3-D grid: $[x, y, z]$ each component with values from 1 to 50 with an increment of unity. So we have an array of data $50 \times 50 \times 50 = 125\,000$ points. We wish to select $n = 5$ points out of $N = 125\,000$ points which will form a Latin hypercube of $[x, y, z]$. Although this problem can be solved easily using the conventional LHS, this example is used to illustrate how the cLHS works. After 165 iterations we obtain 5 points which represent the Latin hypercube of the 3-D data (Fig. 3).

4.2. Soil sampling application

A part of the Pokolbin area in the Hunter Valley is used for this study. The Hunter Valley is situated approximately 140 km North of Sydney, New South Wales, Australia. The area of this study is about 11 km^2 ; a GIS of the area is derived from a variety of sources. A DEM for the area was produced from digitised topographic maps, originally produced from ground surveying and aerial photograph interpretation. The DEM forms the basis of the calculation of the terrain attributes. Digital data captured on 19 August 1999 by the Landsat 7 satellite was acquired for the study area. Land use is obtained by supervised classification of an aerial photo of the area. The main land uses are: water,

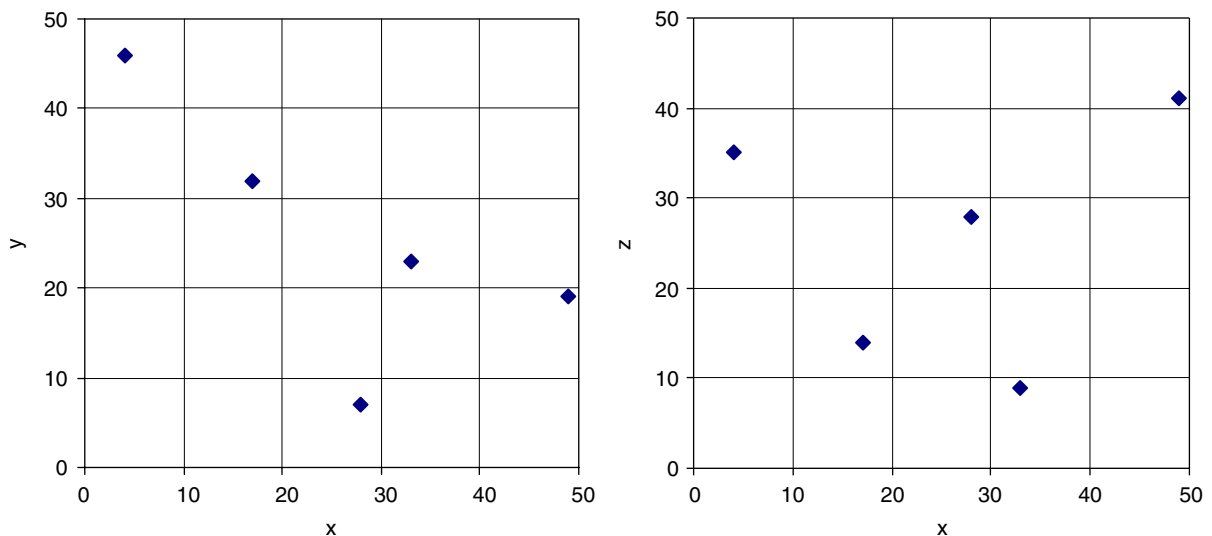


Fig. 3. A simple 3-D example of 5 samples selected from 125 000 points using cLHS.

buildings and roads, vineyard, pasture, native vegetation, and disturbed (cultivated) areas. All data layers were recorded on a common grid of 25 m × 25 m, a total of 17 292 pixels. Pixels where the land cover are water, roads and buildings were excluded, giving us 15 021 pixels.

We wish to sample 100 sites (out of 15 021 points) in this area for digital soil mapping, and to select sites which will cover the following soil-environmental variables:

- slope angle (in degrees) obtained from the DEM,
- CTI, which is derived from DEM and computed using

$$\text{CTI} = \ln\left(\frac{A_s}{\tan \beta}\right),$$

where A_s is the flow area and β is slope angle. Large CTI values indicate an increased likelihood of saturated conditions, the larger values are usually found in the lower parts of watersheds and convergent hollow areas associated with soils with small hydraulic conductivity or areas of low slope.

- normalised difference vegetation index calculated from Landsat band 3 (visible red) and

band 4 (near infrared): $\text{NDVI} = (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$.

- land use, as the soil properties in this area are heavily influenced by their uses.

We compare three sampling methods:

- cLHS with 4 variables: CTI, slope angle, NDVI, and land use.
- Simple random sampling, where n points are randomly selected from X .
- Stratified spatial sampling (Brus and de Gruijter, 1997; Brus et al., 2004), where the area is divided into n polygons using the k -means clustering, and the sampling points are the centre of the polygon. This method ensures a good spatial coverage of the whole area.

The evolution of the objective function with number of iterations is given in Fig. 4. Starting from random sampling this algorithm search through the data to find points that can represent a Latin hypercube. The objective function value (O) decreases with increasing iterations with perturbations (iterations < 4000) introduced by the annealing process. For this example we use 50 000 iterations for searching

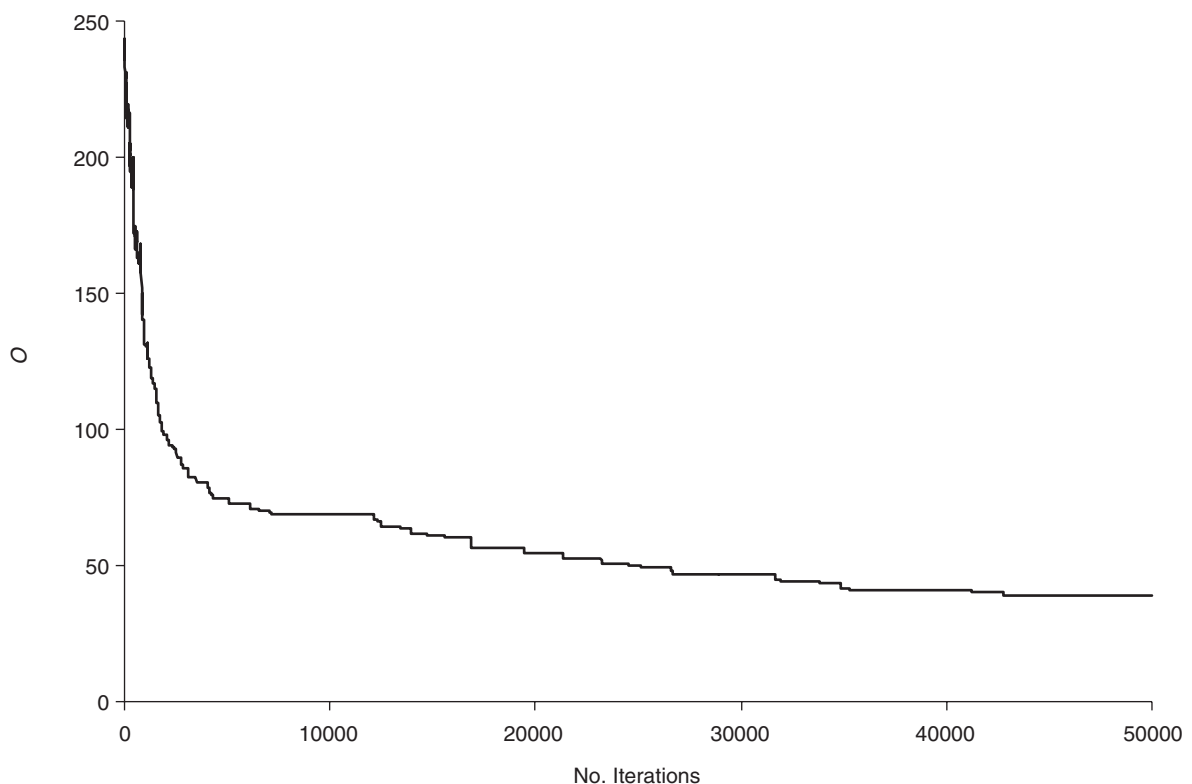


Fig. 4. Evolution of objective function (O) with number of iterations.

an optimal solution with computing time 20 min using a PC with Pentium IV 2.0 GHz processor and 1.0 GB of RAM.

The sampling sites chosen by LHS and stratified spatial sampling are given in Fig. 5. Note that areas that are water and buildings and roads have been excluded in the sampling strategy. We can see that the stratified spatial sampling select sites that provide a good coverage of the area, nevertheless the sites selected using cLHS are also well spread in the area and not clustered.

The distribution of the variables for different sampling techniques is presented in Fig. 6. The statistical distribution of the data and the sampled sites are given in Tables 1, 2, and 3. LHS of the target variables best mimics the original distribution (Fig. 6). The distribution at each interval of the histogram is adequately sampled by cLHS, while other methods slightly over- and under-sampled some areas in the distribution. Simple random sampling and stratified spatial sampling are slightly biased in the selection of the values (especially NDVI). The stratified spatial sampling has good geographical coverage, but does not adequately represent the distribution of the variables. The land use is not sampled proportionally. Since the objective of this sampling is for calibration, sampling a good cover of the environmental variables is the most appropriate.

The ability to select samples that represent a hypercube of the original data in the cLHS algorithm enables us to build a model to predict soil classes or soil attributes. This model then can be used to predict over the whole area. We note that

the solution given by the cLHS algorithm may not be an exact Latin hypercube, as the real data might not contain a combination that will give a true Latin hypercube. Nevertheless, this approximate Latin hypercube generally gives a good representation of the ancillary variables. However, if a Latin hypercube exists, the algorithm should be able to select them as seen in the simple 3-D example.

The example we give here pertains to soil sampling for model calibration, it can be used for different applications where ancillary data are available. Another good application is sampling for monitoring networks where there is prior knowledge of the area. Sampling that covers the geographical space may not necessary give us a good representation of the properties of the area. Thus, we propose the conditioned LHS as an alternative to sampling strategy that ensures a good coverage of the ancillary data.

5. Conclusions

Sampling in the presence of ancillary data is designed to select a limited number of sites in an area. An efficient sampling method is therefore needed to cover the whole range of the ancillary variables. LHS conditioned by the ancillary data offers an attractive solution to this problem. The cLHS becomes an optimisation process, i.e. select x (n points) from X (N data) (where $n \ll N$) so that x form a Latin hypercube of X . We approach this method by using swapping rules combined with an annealing schedule. This optimisation process can be used to sample not only continuous but also

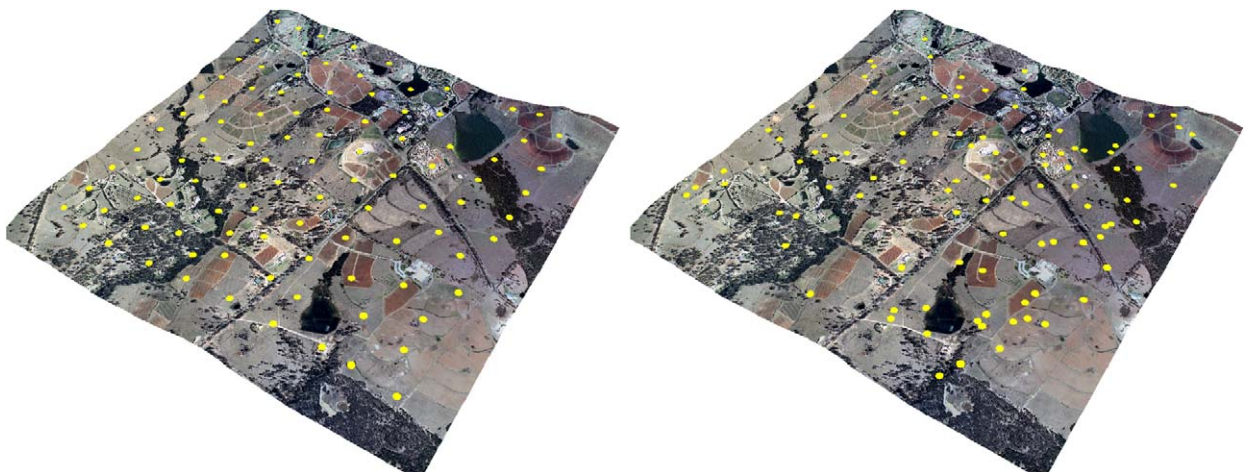


Fig. 5. Location of sample points in landscape: equal spatial strata sampling (left), cLHS (right).

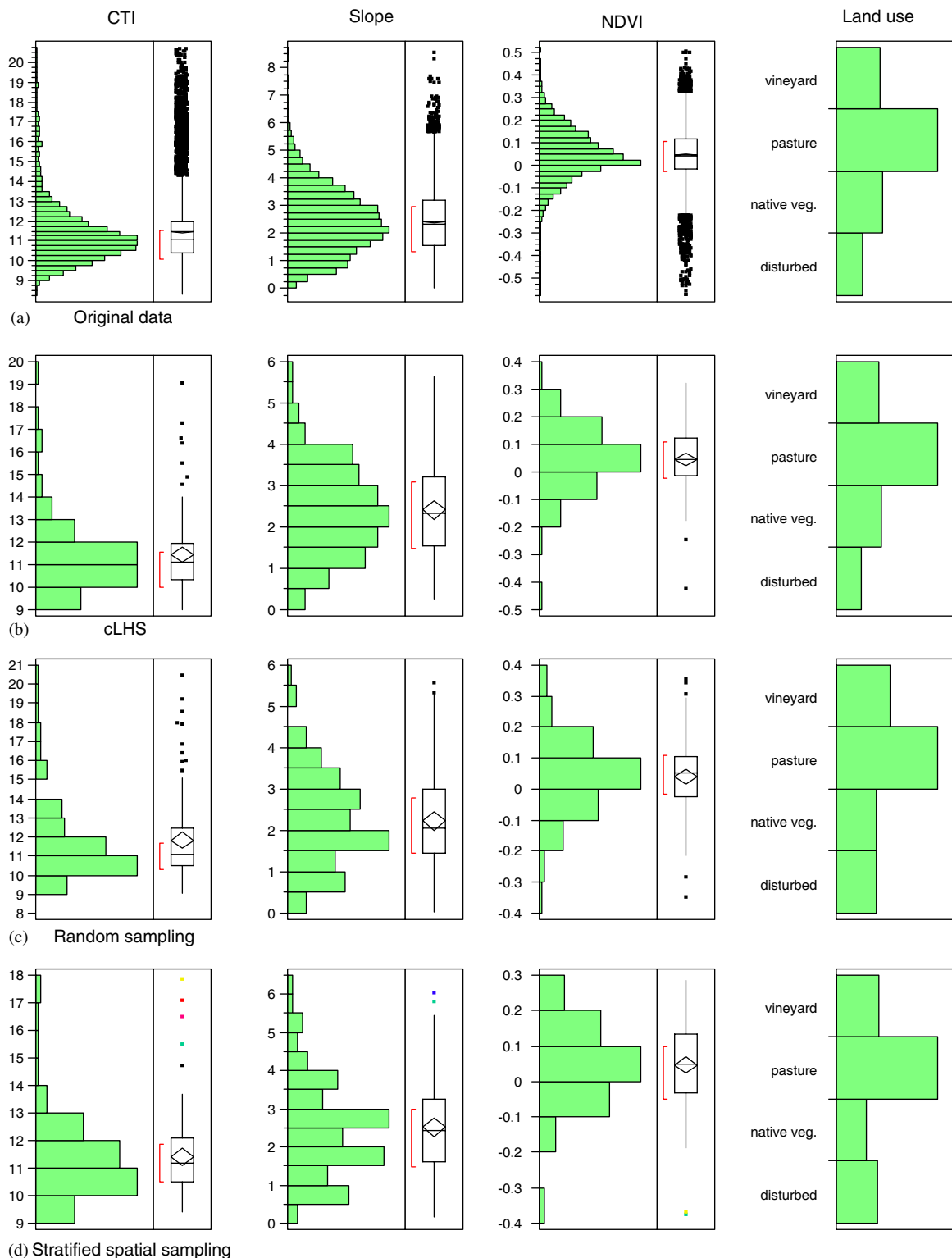


Fig. 6. Histogram and box plot of variables (a) original data, (b) cLHS, (c) simple random sampling, and (d) equal spatial strata. For box plot, ends of box are 25th and 75th quantiles, line across middle of box identifies median sample value and means diamond indicates sample mean and 95% confidence interval. Bracket along edge of box identifies shortest half, which indicates most dense 50% of observations.

Table 1
Statistical distribution of ancillary variables from original and sampled data

	No. of points	25% Quartile	Median	75% Quartile	Mean	Std. dev.
<i>CTI</i>						
Original data	15021	10.406	11.089	11.957	11.454	1.724
cLHS	100	10.328	11.114	11.963	11.459	1.719
Random	100	10.481	11.095	12.459	11.829	2.241
Stratified spatial	100	10.492	11.201	12.073	11.415	1.521
<i>Slope</i>						
Real	15021	1.573	2.346	3.203	2.428	1.172
cLHS	100	1.553	2.346	3.205	2.410	1.141
Random	100	1.460	2.059	3.001	2.238	1.154
Stratified spatial	100	1.607	2.444	3.252	2.514	1.273
<i>NDVI</i>						
Real	15021	−0.015	0.043	0.122	0.048	0.118
cLHS	100	−0.015	0.044	0.123	0.046	0.116
Random	100	−0.127	0.053	0.104	0.040	0.123
Stratified spatial	100	−0.031	0.047	0.133	0.046	0.116

Table 2
Probability distribution of land use using different sampling strategies

	Disturbed	Native vegetation	Pasture	Vineyard
Original data	0.120	0.212	0.467	0.200
LHS	0.120	0.210	0.470	0.200
Random	0.170	0.170	0.430	0.230
Stratified spatial	0.190	0.140	0.470	0.200

Table 3
Correlation coefficient between ancillary variables using different sampling techniques

	CTI and Slope	CTI and NDVI	Slope and NDVI
Original data	−0.480	0.034	0.001
LHS	−0.474	0.066	−0.001
Random	−0.592	0.099	−0.057
Stratified spatial	−0.452	0.072	−0.006

categorical variables, and other criteria can also be added to the objective function. This kind of sampling should produce a reasonably efficient way of sampling soil and its environment so that

the range of conditions are encountered, ensuring a good chance of fitting relationships if they exist.

Acknowledgement

This work is funded by an Australian Research Council Discovery project on Digital Soil Mapping.

Appendix A. Supplementary data

The online version of this article contains additional supplementary data. Please search for doi:10.1016/j.cageo.2005.12.009.

References

- Brus, D.J., de Gruijter, J.J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma* 80, 1–59.
- Brus, D.J., de Gruijter, J.J., Van Groenigen, J.W., 2004. Designing purposive and random spatial coverage samples by the k-means clustering algorithm. Global Workshop on Digital Soil Mapping, Montpellier, France, September 2004.
- Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil-landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems* 9, 421–432.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, Oxford, NY 496pp.
- Hengl, T., Rossiter, D.G., Stein, A., 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Australian Journal of Soil Research* 41, 1403–1422.
- Heuvelink, G., Brus, D., de Gruijter, J., 2004. Optimisation of sample configurations for digital soil mapping with universal kriging. Global Workshop on Digital Soil Mapping, Montpellier, France, September 2004.
- Iman, R.L., Conover, W.J., 1980. Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Communications in Statistics Theory and Methods* A9, 1749–1874.
- 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 Resources Research* 31, 387–398.
- Mathworks, 2005. *Matlab Release 14*. The Mathworks Inc., Natick, MA.
- McBratney, A.B., Webster, R., Burgess, T.M., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. *Computers & Geosciences* 7 (4), 331–334.
- McBratney, A.B., Whelan, B.M., Walvoort, D.J.J., Minasny, B., 1999. A purposive sampling scheme for precision agriculture. In: Stafford, J.J. (Ed.), *Precision Agriculture '99*. Sheffield Academic Press, Sheffield, pp. 101–110.
- McBratney, A.B., Mendonça-Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52.

- McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21, 239–245.
- McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma* 89, 67–94.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 21, 1087–1092.
- Minasny, B., McBratney, A.B., 2002. Uncertainty analysis for pedotransfer functions. *European Journal of Soil Science* 53, 417–430.
- Pebesma, E.J., Heuvelink, G.B.M., 1999. Latin hypercube sampling of Gaussian random fields. *Technometrics* 41, 303–312.
- Pettitt, A.N., McBratney, A.B., 1993. Sampling designs for estimating spatial variance components. *Applied Statistics* 42, 185–209.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1992. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, second ed. Cambridge University Press, Cambridge, UK 963pp.
- Royle, J.A., Nychka, D., 1998. An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Computers and Geosciences* 24 (5), 479–488.
- van Groenigen, J.W., Siderius, W., Stein, A., 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma* 87, 239–259.
- Zhang, Y., Pinder, G.F., 2004. Latin-hypercube sample-selection strategies for correlated random hydraulic-conductivity fields. *Water Resources Research* 39 (Art. No. 1226).