International Journal of Geographical Information Science
Vol. 24, No. 4, April 2010, 523–543

Taylor & Francis
Taylor & Francis Group

Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning

Jinfeng Wang[a]*, Robert Haining[b] and Zhidong Cao[a]

[a]*Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences Beijing, China;* [b]*Department of Geography, University of Cambridge, Cambridge, UK*

(*Received 3 November 2008; final version received 5 March 2009*)

One of the major sources of uncertainty associated with geographical data in GIS arises when they are the outcome of a sampling process. It is well known that when sampling from a spatially autocorrelated homogeneous surface, stratification reduces the error variance of the estimator of the population mean. In this study, we evaluate the efficiency of different spatial sampling strategies when the surface is *not* homogeneous. When the surface is first-order heterogeneous (the mean of the surface varies across the map), we examine the effects of stratifying it into first-order homogeneous *zones* prior to the usual stratification for a systematic or stratified random sample. We investigate the effect of this form of spatial heterogeneity on the performance of different methods for estimating the population mean and its error variance. We do so by distinguishing between the real surface to be surveyed ($\Re$), the sampling frame ($\Im$) including the choice of zoning, and the statistical estimators (Ψ). The study shows that zoning improves estimator efficiency when sampling a heterogeneous surface. Systematic comparison provides rules of thumb for choice of sample design, sample statistics and uncertainty estimation, based on considering different spatial heterogeneities on real surfaces.

Keywords: spatial sampling; uncertainty; heterogeneity; grid strata; zoning strata; efficiency and strategy

1. Introduction

The next 10–15 years will see great advances in real-time environmental monitoring technologies. GIS together with spatial sampling theory and techniques are crucial to designing monitoring networks, drawing population inferences and assessing the accuracy of estimates such as the mean value of some attribute in an area. Compared with an exhaustive survey, the merits of sampling lie in requiring fewer observations resulting in lower overall cost while still being able to achieve levels of accuracy that are sufficient for purpose (Cochran 1977).

The mean is a critical parameter in many areas where GIS is used. The need to estimate the mean level of some attribute over a defined region when the mean is known not to be constant across the whole region arises in environmental science. For example: estimating crop yields across an administrative area where there are geographical differences in topography, soil type and perhaps even weather conditions; monitoring ecosystems where vegetation type is patchy (Huenneke *et al*. 2001). Air pollution displays considerable spatial

*Corresponding author. Email: wangjf@lreis.ac.cn

ISSN 1365-8816 print/ISSN 1362-3087 online

DOI: 10.1080/13658810902873512
http://www.informaworld.com

and temporal variability and is costly to monitor so it is important to design efficient sampling schemes to capture its spatial heterogeneity (Kumar 2009). In the case of large regions heterogeneity (non-stationarity) is often acknowledged, but size of area is not always the critical factor. Atmospheric elements such as coarse particulate matter, $PM_{10–2.5}$, have been found to be heterogeneously dispersed even across quite small areas (Ott *et al*. 2008).

Data quality is one of the major concerns in both GIScience and GIServices (Goodchild and Gopal 1989, Haining 2003, Leung *et al.* 2004, Shi 2005). The analyst has particular concerns when data are assembled from different sources and collected by different procedures (Lee *et al*. 2006, Brus and Heuvelink 2007, Villarini and Krajewski 2008). Evaluating the efficiency of different spatial data collection techniques and quantifying the uncertainty associated with attribute measurements remain critical elements in the GIS research agenda.

Two well known problems that arise when estimating the mean value of an attribute in a region is first the presence of spatial autocorrelation and second the presence of spatial heterogeneity in the attribute. Recognizing their presence has implications for the efficiency with which sampling is carried out – that is estimator error variance in relation to sample design and sample size (Ripley 1981, Haining 1988, Christakos 2005).

Heterogeneous geographical areas as these are becoming increasingly important in different areas of geographical data analysis and environmental science (Csillag *et al.* 1996, Goodchild and Haining 2004, Green and Plotkin 2007). The objective of the current research is to investigate the effect of spatial heterogeneity in the mean (or first-order heterogeneity), on the methodology of spatial sampling, and to suggest improved sampling strategies that have lower error variances than can be achieved with current methods. Stratified sampling is a conventional tool in surveying heterogeneous populations (Cochran 1977), but few studies, it seems, distinguish between the real strata of the population and the strata for sampling and statistics. It is often difficult in the spatial dimension to construct sampling strata that are concordant with real surface heterogeneity for physical, legal, logical, economic or cognitive reasons. This study identifies the difference and quantifies the uncertainty of a sample estimate arising from the difference between the real strata and the strata used for sampling and statistical estimation. We report our findings on the efficiency effects of there being a mismatch between real surface heterogeneity and sample stratification. Further, we explore the gains from employing two levels of stratification – a high level or macro scale of stratification reflecting surface heterogeneity thereby breaking an area into homogeneous subareas (zones) within which conventional grid square or micro scale stratification is then employed.

We first give a brief review to the existing studies in spatial sampling; then discuss spatial heterogeneity as a necessary precursor to investigating sampling efficiency and its determinants; third we discuss the implications of these findings for the choice of sampling strategy; fourth we consider theoretical and empirical examples.

2. Review

There are two approaches to spatial sampling: design-based and model-based (Brus and de Gruijter 1997). In design-based sampling, the population of values in a region is considered fixed and randomness enters through the process of selecting the locations to sample. The mean value for the region is a fixed but unknown quantity and the sample mean is an estimator of it. Repeated sampling according to a given scheme such as random sampling will generate a distribution of estimates of the (regional or population) mean. In model-based sampling, the set of values observed in a region represents one realization of some stochastic model. Unlike the design-based approach, the mean for the set of values is therefore a

random variable. Also the target of inference is not, as in the design-based approach, the regional mean but rather the mean of the stochastic model assumed to have generated the realized population. This conceptualisation of spatial data underlies geostatistics (Cressie 1993).

The model-based approach to sampling is most appropriate for tackling 'where' questions – for example, predicting values at particular locations, mapping and for estimating the parameters of the underlying stochastic model. The design-based approach is most often used for tackling 'how much' questions – estimating global properties such as the population mean or the proportion of an area under a particular land use. It is for this reason that the approach here is design-based because the target of inference is the mean, which is a global property of the specified attribute. For an extended review of these issues see also Haining (2003, pp. 96–99).

There are three main spatial sampling plans: random, stratified random and systematic sampling (for descriptions see, for example, Ripley 1981, pp. 19–22; Haining 2003, pp. 100–103). Figure 1 shows the pattern of sampling. A widely used and intuitively simple and robust estimator of the regional mean, $\bar{Y}$, is given by the sample mean:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{1}$$

where $\{y_1, \ldots, y_n\}$ represents the n sampled values. Under random sampling, where each individual in the population (location on the map) has an equal and independent chance of selection and the selection of any location has no effect on the selection of any other location on the map, Equation (1) is an unbiased estimator of $\bar{Y}$. Unbiasedness here means that the expected value of $\bar{y} = \bar{Y}$, which indicates that if we were to take several random samples from the population and calculate Equation (1) each time and then compute the average of

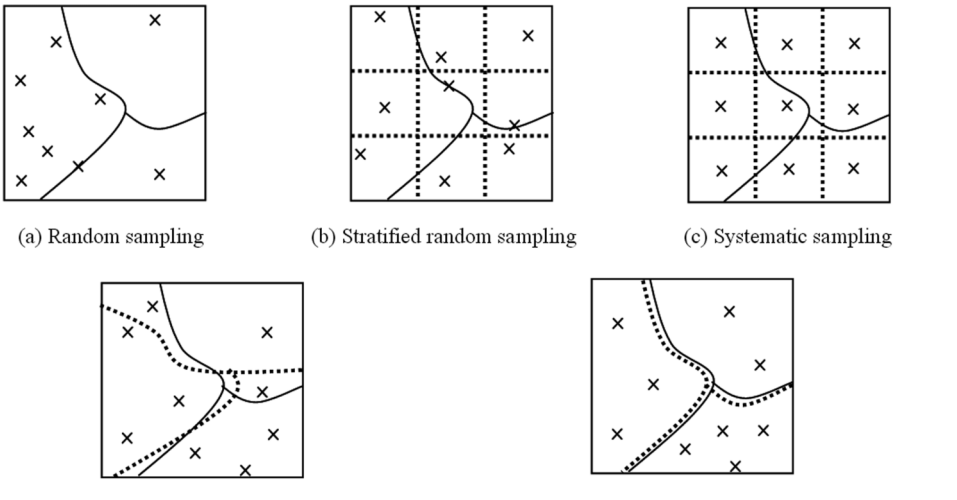

Figure 1. One heterogeneous population and different sampling and statistics. (The solid lines indicate the real zones of the surface (z) and dotted lines represent the zones used for sampling purposes and the calculation of statistics (h), the crosses indicate sampling sites.)

these means, it would equal $\bar{Y}$. If data values in the population are independent, the error variance of Equation (1) as an estimator of $\bar{Y}$ is given by σ^2/n where σ^2 is the population variance. The error variance calculation allows the user to determine the confidence interval associated with a single value of $\bar{Y}$ as an estimate of $\bar{Y}$. It also follows that $\hat{s}^2/n$ where

$$\hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{2}$$

is an unbiased estimator of this error variance (see, for example, Freund 1992).

However, in the case of spatial data, although members of the sample are independent by construction, data values that are near to one another in space are unlikely to be independent because of a fundamental property of attributes in space, which is that they show spatial structure or continuity (spatial autocorrelation). In this case the error variance of Equation (1) is closely approximated by (Ripley 1981, Dunn and Harrison 1993, Griffith *et al.* 1994):

$$\frac{\sigma^2 - \overline{\operatorname{cov}(y_i, y_j)}}{n} \tag{3}$$

where n is the size of the sample and the second term in the square brackets is the average autocovariance between all pairs of individuals (i, j) in the population (sampled and unsampled). Since the expected value of $\hat{s}^2$ is given by (Haining 1988):

$$\sigma^2 - \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j<1}^{n-1}\operatorname{cov}(y_i, y_j) \tag{4}$$

where the second term is the average autocovariance between all pairs of individuals (i, j) in the sample, it follows that $\hat{s}^2/n$ again provides an unbiased estimator of the error variance of (1).

In the case of stratified random sampling the error variance of Equation (1) is given by (Ripley 1981, Dunn and Harrison 1993):

$$\frac{\sigma^2 - \overline{\operatorname{cov}(y_i, y_j)}}{n} \tag{5}$$

where the second term inside the square brackets is the average autocovariance between all possible pairs (k, l) within a stratum. In the case of systematic sampling the error variance of Equation (1) is given by (Ripley 1981, Dunn and Harrison 1993):

$$\overline{\operatorname{cov}(y_u, y_v)} - \overline{\operatorname{cov}(y_i, y_j)} \tag{6}$$

where $\overline{\operatorname{cov}(y_u, y_v)}$ is the average autocovariance between members of the systematic sample.

The sample mean is an unbiased estimator of the population mean irrespective of whether random, stratified random or systematic sampling is employed but different sampling strategies have different error variances and hence can be compared in terms of their relative efficiency under different assumptions about the spatial autocorrelation in the population.

Early work evaluating spatial sampling strategies in two dimensions include Quenouille (1949), Das (1950), Zubrzycki (1958) and Matern (1960). These theoretical studies based on surfaces that were the outcomes of spatially homogeneous processes (Ripley 1981) showed that systematic sampling outperforms other sampling schemes except where there is periodicity in the attribute and these findings have been generally endorsed by empirical studies (Matern 1960, Milne 1959, Payandeh 1970, Dunn and Harrison 1993).

The work by Dunn and Harrison (1993) showed that whilst systematic sampling was the most efficient of the three methods of sampling (and random sampling consistently the least efficient), the gains in efficiency, relative to stratified random sampling, were highly variable. They also compared two different methods of estimating the error variance of the mean from a single systematic sample (Ripley 1981, p. 27). Their work was based on sampling real land use maps with complex and varied spatial autocorrelation structures and their findings suggested that the presence of non-stationarities and anisotropies in real maps could have a severe effect on the efficiency of systematic sampling.

Map stratification, whether for stratified random or systematic sampling, usually involves square strata and the sampler needs to decide on strata size as well as the orientation and starting point for the overlaid grid. Typically the starting point is chosen at random and the effect of the grid starting position on sample variability is assumed to be small. This is probably true for the choice of grid orientation unless there are strong directionalities in the surface as in the case for example of a repetitive ridge and valley topography. The stronger the spatial autocorrelation on the map, with high levels over long distances, the more redundancy will be present in a sample if the inter-point sampling distance is small (Griffith 2005). Ripley (1981) remarks that the gain from stratification will be most when spatial autocorrelation is large for all distances up to the scale of the strata used but becomes negligible beyond that scale. 'This suggests that for monotonically decreasing correlation functions we should take small strata; hence the number of sample points per strata will be small' (Ripley 1981, p.24–5). If an area to be sampled is second-order heterogeneous, that is there are different spatial autocorrelation structures in different sections of the map, then presumably there will be efficiency gains to be achieved by adapting strata size to that heterogeneity (Berry and Baker 1968, Dunn and Harrison 1993).

3. Spatial heterogeneity and prior information

An attribute measured on a geographical surface comprises two elements of second-order variation: a global variance (population variance) and the spatial structure of that variation (population spatial autocovariance or autocorrelation). Both elements of geographical variation need to be recognized in designing and evaluating sample designs including the sampling plan and the choice of estimator (Cochran 1977, Haining 1988, Griffith 2005). For example areas showing greater variance will need to be more intensively sampled than areas showing lesser variance to achieve the same level of error variance. However, neither of these two elements of second-order variation may be independent of location on the map in which case the attribute displays second-order spatial heterogeneity (or second-order non-stationarity). A measure of population variance for an attribute may differ significantly between different parts of a map (e.g. a mountain area compared to a lowland area) and the spatial structure of that variation may also be location dependent. Second-order spatial heterogeneity is a frequently occurring property of attributes across a geographical area, particularly areas that are physically large. In addition there may be first-order spatial heterogeneity (or first-order non-stationarity), that is the mean is also location dependent.

One way to represent heterogeneity is to partition a map into zones (Wang *et al.* 1997). Because of the continuity of spatial variation (but with the possible exception of relatively sharp boundaries as between for example land and sea), there is usually no 'true' or 'correct' zonation, but at any given scale of geographical detail some zonations (regional classifications) will be better than others. The best zonations create areas that are first- and second-order homogenous although in practice, there may be limits to how good any zonation can be because of lack of knowledge about the study area or the complexity of the surface.

Constructing zones may be based on prior knowledge, pre-sampling, an effective proxy variable (Rodeghiero and Cescatti 2008), or on the distribution of other variables that are known to affect the value of the attribute of interest (e.g. altitude in the case of estimating crop yields or vegetation cover). Purposive sampling is generally more efficient than probability sampling (Brus and de Gruijter 1997, de Gruijter *et al*. 2006). Usually sampling is more intensive in areas important for human society: more signal relay stations should be located in areas with more frequent traffic accidents (Rogerson *et al*. 2004), the seismic monitoring network is denser in Beijing than in other areas of China, major crop production areas are covered by intensive surveillance of meteorological conditions, and police are allocated more densely in urban areas. Below we list some more examples where prior information can be called upon to capture the heterogeneity of a surface:

- *Theoretical assumptions* about surfaces based on our understanding of earth processes (Rodriguez-Iturbe and Mejia 1974, Haining 1988, Christakos 2005, Sen 2008);
- *Adjunct knowledge* such as a Digital Elevation Model (DEM) for land use classification, mapping noise using distance decay from source points (Stoter *et al*. 2008), various ancillary variables (soil series, relative elevation, slope, electrical conductivity and soil surface reflectance) to estimate soil carbon stock (Simbahan and Dobermann 2006), environmental variables are interpolated using covariates for which more detailed information are available (Brus and Heuvelink 2007);
- *Regression models or data adaptive algorithms*: increasing the accuracy of design-based sampling strategies (Brus and Te Riele 2001, Almeida *et al.* 2008); predicting soil distribution using explanatory variables including classical terrain factors, land cover and lithology maps and various channels from LANDSAT ETM imagery (Grinand *et al*. 2008); and a flexible multi-source spatial-data fusion system (Li *et al*. 2008).
- *Mechanistic modelling*. Both mathematical models, such as random field models (Haining 1988, Christakos 2005) and location allocation models (Kumar 2009) seek to characterize the features of the object under study and may provide complementary prior knowledge.

We argue here that if used carefully, zoning may improve the efficiency of spatial sampling when spatial heterogeneity is present in the surface. We can derive a population estimate, based on a sample, which has a smaller error variance.

4. Spatial sampling efficiency

Spatial sampling can be described with reference to a triple (Ψ, $\Im$, $\Re$): the real surface to be surveyed ($\Re$), the spatial sampling frame ($\Im$) that is laid down over the area and typically takes the form of either a random, stratified random or systematic distribution of n point locations where data are recorded and the statistic (Ψ) used to estimate the quantity of

interest using the sample data. Spatial sampling efficiency depends on the choice of $\Im$ and Ψ given the properties of $\Re$.

4.1. Specification

We are interested in the efficiency gains from creating statistical strata $\Im$ ($h1$, $h2$, ...,) that correspond with the strata on a real heterogeneous surface ($\Re$), which we denote ($z1$, $z2$, ...,). To avoid confusion with the term 'strata' commonly used to refer to the subdivisions of an area associated with stratified random and systematic sampling, we shall henceforth refer to these as 'zones' – statistical zones (h) and real zones (z). We shall call the set of such zones that partition the surface a 'zonation' (Figures 1(d) and (e).

To help fix ideas, Figure 2 shows a rectangular region divided into two real zones where the northern half of the map is a surface of +s ($z1$) and the southern half is a surface of 0s ($z2$) – a sharp boundary between two areas with no intra-zonal variation. The +s and 0s on the maps are sample points taken by a systematic sample taken from within two statistical zones ($h1$ and $h2$) so a + has a sample value of 1 and a 0 has a sample value of 0. So each of the zones $h1$ and $h2$ are partitioned into strata. The 14 different cases in Figure 2 show different statistical strata ($h1$, $h2$), shown by light background shading, laid onto the map where it is only in the case of 1(i) there is perfect correspondence between the z and h zonations. All the other cases represent different forms of misalignment and different types of boundaries between the two z zones (sharp linear, crenulated and 'fuzzy').

We start by presenting some general results for estimator variance. As noted by Ripley (1981, p. 27) there are two approaches to estimate the sampling or error variance of Equation (1) in the case of a systematic sample (and in the case of a stratified random sample where there is only one sample point per stratum). One approach uses Equation (2) and treats $\hat{s}^2/n$ as the estimate of the error variance of the sample mean, the other approach ('post hoc stratification') groups the strata into blocks (typically of size two, but three if boundary conditions require it) uses Equation (2) on each of the blocks, takes the average over all the blocks and then divides by n. Ripley (1981) reports a study by Milne (1959) that suggests that either method gives a 'good idea of the true sampling variance'. The empirical work reported in Dunn and Harrison (1993) shows both methods overestimating the true sampling variance but with the second method being the better of the two.

Let $V(.)$ denote variance. Let $\bar{y}_{\text{zone\_}\{h\}}$ denote the sample mean calculated from a statistical zonation $\{h\}$ and $\bar{y}_h$ is the mean calculated for any one of the zones of the zonation. Then:

$$V\left(\bar{y}_{\text{zone\_}\{h\}}\right) = V\left[\sum_{h=1}^{L_h} W_h \bar{y}_h\right] = \sum_{h=1}^{L_h} W_h^2 V(\bar{y}_h) = \frac{1}{n}\sum_{h=1}^{L_h} W_h s_h^2 \tag{7}$$

where L_h is the number of statistical zones in the zonation h, W_h is the proportion of the total sample (n) in zone h (n_h/n) and S_h^2 is the sample variance for the data from zone h.

Now it can be shown (see Appendix substituting zonation h for z):

$$V(\bar{y}) = \frac{s^2}{n} = \frac{1}{n}\left[\sum_{h=1}^{L_h} W_h s_h^2 + \sum_{h=1}^{L_h} W_h (\bar{y}_h - \bar{Y})^2\right] \tag{8}$$

where s^2 denotes the sample variance based on all the data. The quantity, $V(\bar{y}) - V(\bar{y}_{\text{zone\_}\{h\}}) = \frac{1}{n}\sum_{h=1}^{L_h} W_h (\bar{y}_h - \bar{Y})^2$, is the efficiency difference between a zoned and

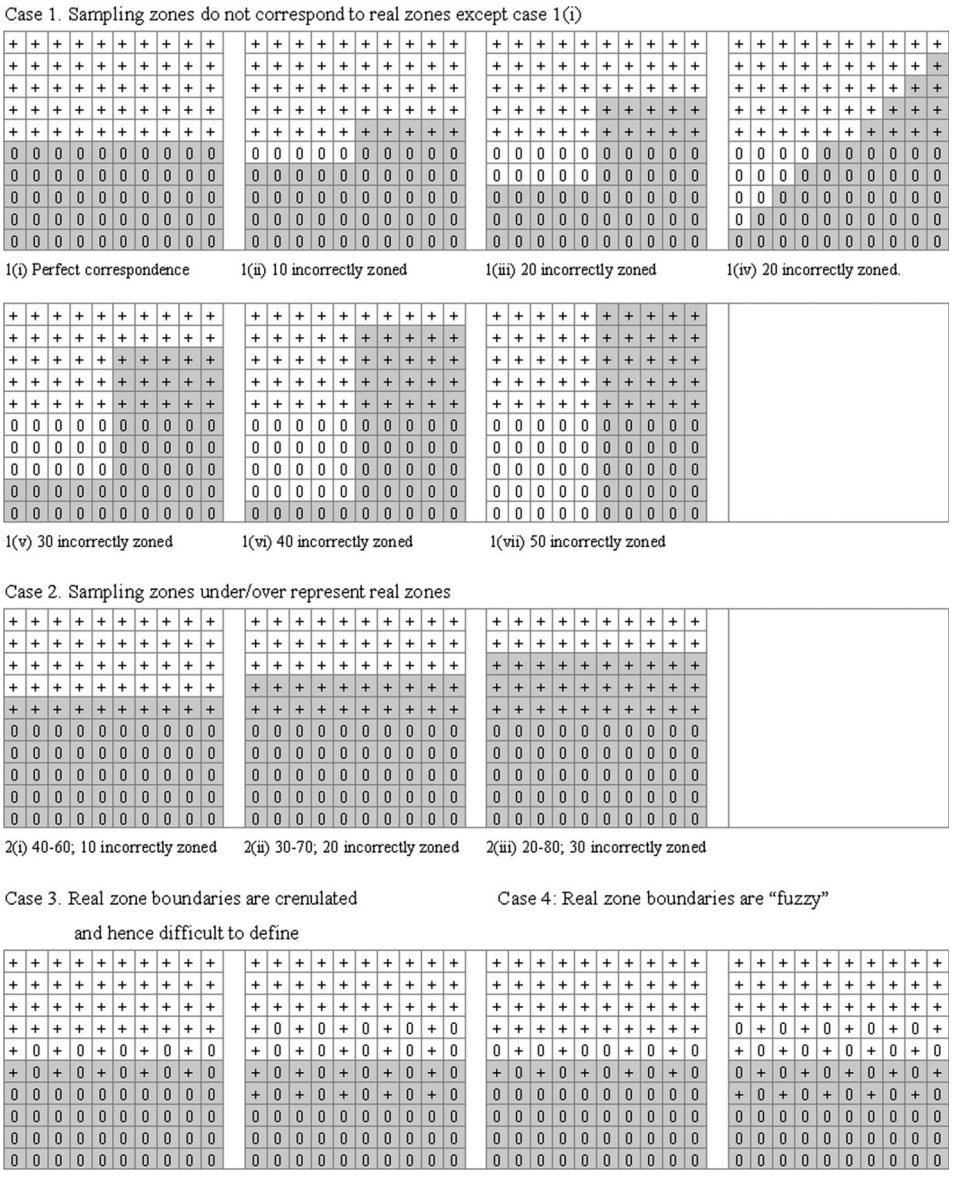

Figure 2. Maps showing relationship between real (z) and statistical (h) zones.

an unzoned estimator for the sample mean. It is proportional to the zone-weighted sum of squared differences between the zone means and the population mean. The first key observation from these results is that the bigger the difference of the zone means $\{\bar{y}_h\}$ from the population mean $\bar{Y}$, the larger the gain from zoning; the gain vanishes if the zone means are all equal and hence equal to the population mean.

We now want to compare efficiencies arising from the difference between a true zonation (z) and a statistical zonation (h). In the case of the zonation z, the heterogeneous surface ($\Re$) can be partitioned into L_z zones within each of which the mean is constant but differs from

the mean in any adjacent zone (so that it is not possible to merge adjacent zones and have the mean remain constant within the new zone). It follows that:

$$V\left(\bar{y}_{\text{zone\_}\{h\}}\right)\frac{1}{n}\sum_{h=1}^{L_h} W_h s_h^2 \quad \text{and} \quad V\left(\bar{y}_{\text{zone\_}\{z\}}\right) = \frac{1}{n}\sum_{z=1}^{L_z} W_z s_z^2$$

and the difference is

$$V\left(\bar{y}_{\text{zone\_}\{h\}}\right) - V\left(\bar{y}_{\text{zone\_}\{z\}}\right) = \frac{1}{n^2}\left(\sum_{h=1}^{L_h} n_h s_h^2 - \sum_{z=1}^{L_z} n_z s_z^2\right) = \frac{1}{n^2}\sum_{k=1}^{L_k} n_k\left(s_{h=k}^2 - s_{z=k}^2\right) \tag{9}$$

under the additional assumption that $L = L_z = L_k$ (the same number of real and statistical zones), $n_h = n_z = n_k$ (the same size of sample).

The results show that the efficiency difference between biased zoned statistics (zonation $h \neq z$) and unbiased zoned statistics ($h = z$) is proportional to the difference of the sum of the sample size-weighted sampling variances. From Equations (7–9), we have

$$\frac{s^2}{n} = V(\bar{y}) \geq V\left(\bar{y}_{\text{zone\_}\{h\}}\right) \geq V\left(\bar{y}_{\text{zone\_}\{z\}}\right) = \frac{1}{n}\sum_{z=1}^{L_z} W_z s_z^2 \tag{10}$$

Zonation improves estimator efficiency because it removes the variance associated with spatial variation in the mean. The second key observation is that the gains from zoning will decrease as the statistical zonation $\{h\}$ becomes less and less aligned with the real zonation $\{z\}$. In order to explore this quantitatively, we return to the cases shown in Figure 2.

4.2. *Numerical example*

In order to illustrate the effect of zoning $\{h\}$ on estimator efficiency, we calculated the sampling variance for the mean $(\bar{y})$ where the sample data have been obtained from a centric systematic sample on a heterogeneous surface that comprises two homogeneous regions (each with a constant mean and zero intra-zone variance) as shown in Figure 2. We use the formula $\hat{s}^2/n$ for calculating the sampling variance. The benchmark is $\hat{s}^2/n$ calculated for all 100 sample points without any zoning $\{h\}$. Figure 3 shows the graph of the ratio of the sampling variance obtained after zoning to the benchmark value and expressed as a percentage. The sampling variance has been obtained by calculating $\hat{s}^2$ for each of the two zones separately, then averaging the two values and then dividing by $n = 100$. In this simple situation, the relative efficiency calculations are determined by the number of 'misallocated' samples. However, even in those cases where there has been a substantial level of misallocation the gains from stratification are evident.

4.3. *Spatial sampling strategy*

Using the triple (Ψ, $\Im$, $\Re$), Figure 4 depicts four critical stages in a sampling survey and the sources of error variance. Clearly, the earlier spatial heterogeneity can be recognized in the design of a sample, the more likely it is that efficiency gains can be realized. For example, if as illustrated in Section 2, there is prior knowledge about the heterogeneity in $\Re$ sufficient to construct a statistical zonation approximating to the true zonation of the real surface, then

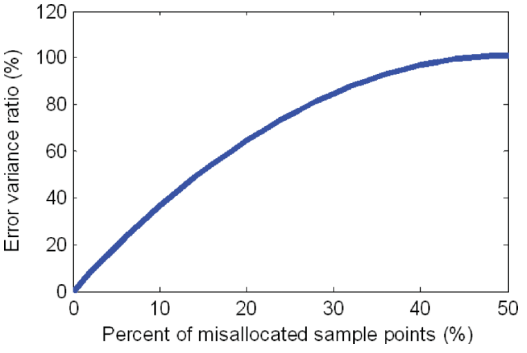

Figure 3. Graph of efficiency gains as a result of zoning but with different percentages of misallocation. Calculations using the cases shown in Figure 2, based on a systematic sample and using the random sampling formula for calculating estimator sampling variance. The case of perfect correspondence between the real (z) and statistical (h) zones is taken as the benchmark.

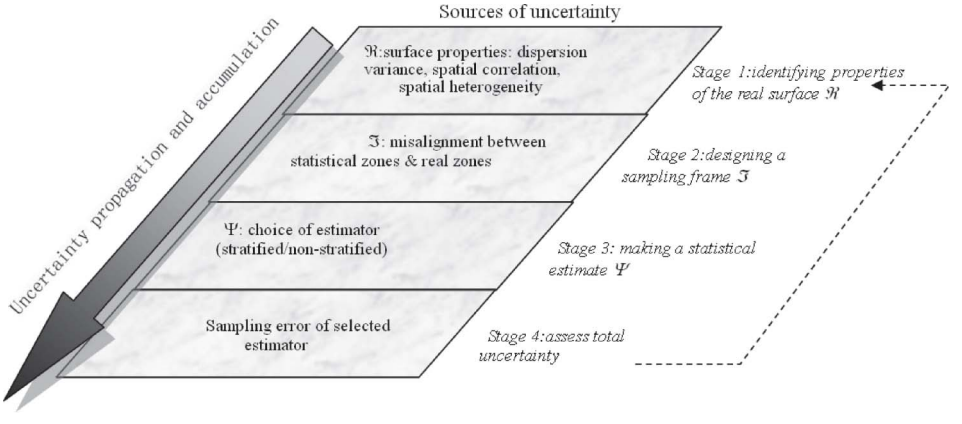

Figure 4. Four stages associated with spatial sampling and the accumulation of uncertainty. If given a report with sample estimates (stage 4), they need to be assessed against the properties of the real surface (stage 1), the distribution of the sample zones (stage 2) and the choice of estimator (stage 3). Error and uncertainty are accumulated through the process.

sampling should be conducted using that zonation in order to try to maximize sampling efficiency. Even if there is only partial knowledge then some efficiency gains might be achieved as suggested by the numerical examples in Section 4.2.

In the next section we turn to two empirical case studies in order to explore empirically the effects of constructing zones on sampling variance.

5. Empirical examples

5.1. Sample surveys of non-cultivated land in Shandong province, N. E. China, 1985 and 1995

We examine, in the case of a spatially heterogeneous surface, how different sampling strategies perform and the gains that accrue in terms of error variance of the sample mean from partitioning the area into homogeneous (or quasi-homogeneous) zones. The study area

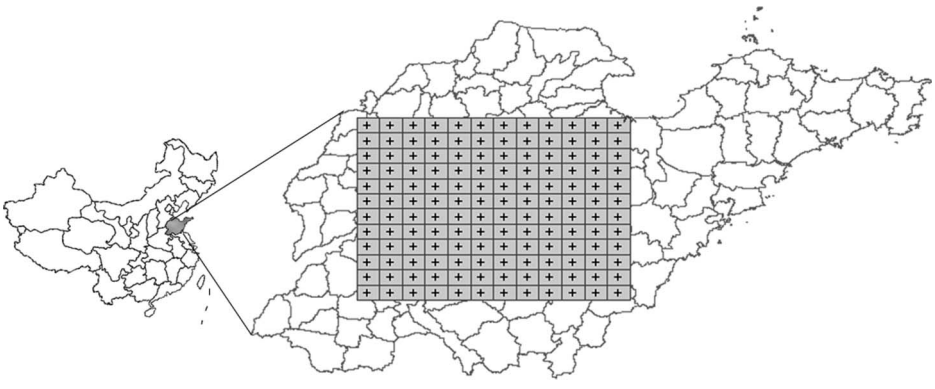

Figure 5. Study area (grid system) in Shandong province, East China.

is the rectangle within Shandong province (see Figure 5), which was covered by aerial photos in 1985 and 1995. The population is 144 units (12 × 12 grid) each about 520 km^2. The aerial photos are used to pick up the proportion of the non-cultivated land in each unit, and the true values of each unit in 1985 and 1995 are shown in Figures 6(a) and (b)respectively.

Using simple random sampling as the baseline, systematic sampling and stratified random sampling were applied together with two types of additional knowledge about underlying spatial heterogeneity: zones defined by elevation 1985 and 1995 and zones defined for the 1995 case study based on the evidence of heterogeneity from the 1985 survey.

Let $\Re 85$ and $\Re 95$ denote the surface of the true proportion of non-cultivated area in Shandong province in 1985 and 1995, respectively. The evidence for heterogeneity in the mean comes from Figures 6(a) and (b), which suggest an area with much higher levels of non-cultivated land in the central area of the map in both 1985 and 1995.

The Moran test for spatial autocorrelation (using GeoDA, with a first-order, 0/1, neighbour weight matrix) applied to the 1985 and 1995 surfaces gives values of 0.5476 and 0.4325 respectively, demonstrating the presence of significant positive spatial autocorrelation on both maps ($p < 0.05$) although spatial autocorrelation is stronger in 1985 than in 1995. This finding is also reflected in the parameters of the fitted spherical semi-variogram models where for 1985: range = 136 km, nugget $C_0 = 0.00156$, sill $C = 0.05442$ and $C_0/C = 0.0287$; and for 1995: range = 136 km, nugget $C_0 = 0.00582$, sill $C = 0.03622$ and $C_0/C = 0.1608$.

Figures 7(a) and (b) show the frames ($\Im$) used for systematic and stratified random sampling, where r equals the dimensions of the square strata and $f = 1/r^2$ equals the sampling proportion because one sample unit is drawn from each strata. Figure 7(c) shows three zones defined by equalizing three intervals of altitude because altitude is believed to be one of the important determinants of cultivation; and Figure 7(d) shows four zones used on the 1995 data defined by the results of the 1985 sample survey. Minimizing the variance within each of the zones and maximizing the variance between the zones and keeping spatial connectivity within each zone produce the zonation. In detail, we first order the values of the proportion of non-cultivated land in 1985, delimit the series into three equal intervals, then smooth the values over space: the value in each grid cell is replaced by the average of a window centred at the grid cell and calculated from its surrounding 3 by 3 set of grid values in order to guarantee spatial connectedness for each of the zones.

In the cases where randomization is involved in the sampling (simple random and stratified random), 1000 Monte Carlo (using a Matlab program compiled by the authors) repeated samples of size 36, 16 and 9, corresponding to sampling proportions of 1/4, 1/9 and

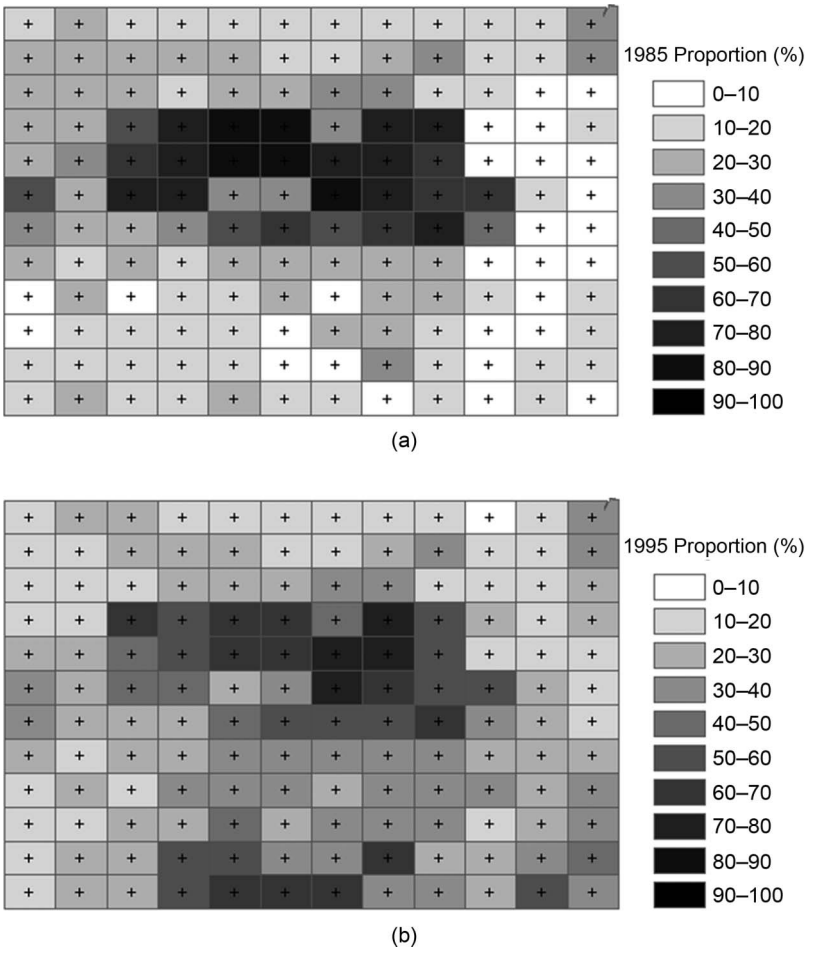

Figure 6. (a) The true proportion of non-cultivated land in Shandong province in 1985. (b) The true proportion of non-cultivated land in Shandong province in 1995.

1/16, respectively, are obtained for each of the sampling frames. To maintain comparability, Monte Carlo sampling was also used in the case of systematic sampling even though this is not needed since all the sampling designs can be listed.

The following sampling frames disregard any spatial heterogeneity in $\Re$:

- $\Im 1$. Simple random sampling: drawing a sample of specified size at random from the set of 144 cells. Sampling was done without replacement. This sampling design is used as a benchmark.
- $\Im 2$. Stratified random sampling: the surface is stratified as shown in Figures 7(a) and (b) and one sample unit is randomly drawn from each strata.
- $\Im 3$. Systematic sampling: as for $\Im 2$ but one sample unit is taken from the same position within each strata.

The following sampling frames allow for spatial heterogeneity in $\Re$:

- $\Im 4$. The surface is zoned by elevation, then the zones are stratified and stratified sampling carried out.

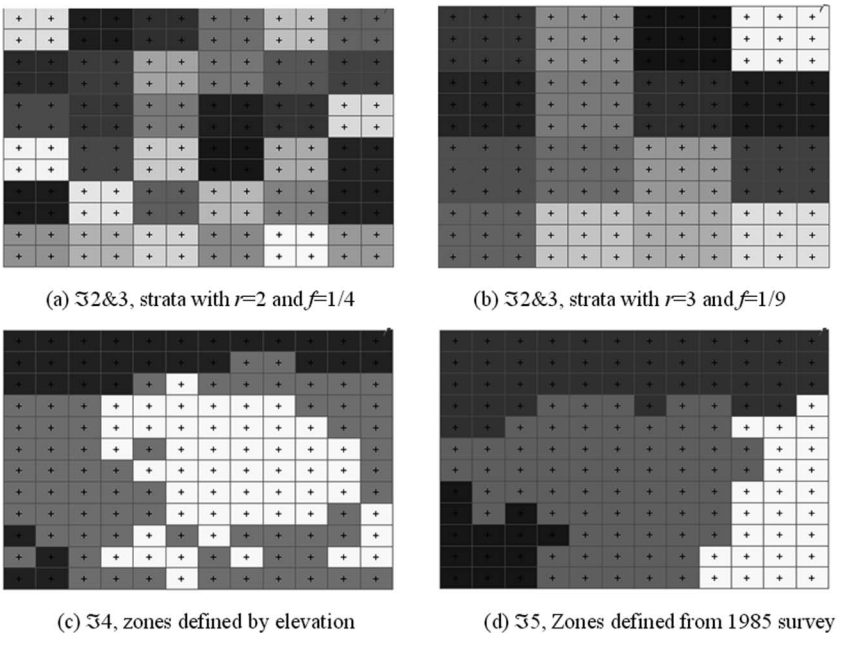

Figure 7. Grids and strata used for different systematic and stratified random sampling and the zones used to allow for heterogeneity.

ℑ5. The 1995 map is zoned using the evidence on heterogeneity from the 1985 sample and classification algorithms (Li *et al*. 2008), then the zones are stratified and stratified sampling carried out.

For each of these sampling designs (ℑ1–ℑ5), two statistics (Ψ) are calculated for each of the three sampling fractions. The two statistics are the sample mean and the variance of the sample mean. Since random sampling (ℑ1) is known to have the lowest efficiency it is used as the baseline to evaluate the efficiency of the other five sampling plans (ℑ2–ℑ5), which is quantified by the design effect (deff) statistic:

$$\text{deff}(\Im i) = \frac{V(\Im 1)}{V(\Im i)}, \quad i = 2, 3, \ldots, 5 \tag{11}$$

the bigger the deff, the higher the efficiency gain of the sampling design.

Tables 1 and 2 are based on the results of 1000 repeated Monte Carlo samplings. For random sampling ℑ1 f = 1/4, there are 36 sample units in one simple random sampling, which are averaged to get a mean value. This sampling is repeated 1000 times to yield 1000 mean values and these are averaged to get the mean value and the variance of the mean.

We draw the following conclusions from these results.

(1) As expected stratified random and systematic sampling have consistently lower sampling variances than random sampling. Also, the 1985 map has the higher level of spatial autocorrelation compared to 1995 and the design effects arising from stratification (ℑ2 and ℑ3) are higher for 1985 relative to 1995 in three out of the four cases. This conclusion is consistent with the theoretical comparison between the error variance, Equation (5), for stratified random sampling, the error variance,

Table 1. Means and variances obtained from sampling the proportion of non-cultivated land in Shandong province in 1985 (ℜ85) and 1995 (ℜ95).

	1985								1995							
	$r = 2,\ f = 1/4$				$r = 3,\ f = 1/9$				$r = 2,\ f = 1/4$				$r = 3,\ f = 1/9$			
Sampling methods	Min	Max	Mean	Var.	Min	Max	Mean	Var.	Min	Max	Mean	Var.	Min	Max	Mean	Var.
Random sampling ℑ1	0.177	0.384	0.270	0.033	0.129	0.458	0.268	0.052	0.268	0.413	0.329	0.023	0.219	0.460	0.329	0.040
Stratified random sampling ℑ2	0.196	0.343	0.270	0.023	0.232	0.336	0.270	0.033	0.308	0.357	0.328	0.016	0.261	0.417	0.328	0.026
Systematic sampling ℑ3	0.258	0.289	0.270	0.011	0.163	0.372	0.270	0.032	0.280	0.383	0.329	0.018	0.279	0.362	0.328	0.024
Zoned by elevation sampling ℑ4	0.175	0.385	0.267	0.033	0.141	0.429	0.272	0.048	0.254	0.402	0.329	0.020	0.247	0.435	0.327	0.031
Zoned using 1985 data ℑ5									0.304	0.352	0.329	0.007	0.294	0.363	0.329	0.010

Note: The true mean value is 0.2704 in 1985 and 0.3291 in 1995.

Table 2. Design effects due to different methods of sampling the proportion of non-cultivated land in Shandong in 1985 and 1995.

	1985		1995	
Design effect	$r = 2, f = 1/4$	$r = 3, f = 1/9$	$r = 2, f = 1/4$	$r = 3, f = 1/9$
deff($\Im 2$) = $V(\Im 1)/V(\Im 2)$ stratified random	1.4439	1.5818	1.4409	1.5385
deff($\Im 3$) = $V(\Im 1)/V(\Im 3)$ systematic	2.8389	1.5866	1.2818	1.6129
deff($\Im 4$) = $V(\Im 1)/V(\Im 4)$ zoned by elevation	1.0152	1.0675	1.1101	1.2862
deff($\Im 5$) = $V(\Im 1)/V(\Im 5)$ zoned using 1985 survey results			3.1351	3.8462

Equation (6), for systematic sampling, with the error variance σ^2/n of simple random sampling.

(2) Systematic sampling is more efficient than stratified random sampling in three out of the four cases. This also corresponds to earlier results that have suggested that systematic sampling should outperform stratified random sampling on spatially autocorrelated maps with no periodicities (Dunn and Harrison 1993). Maintaining a fixed distance between sample points on a spatially autocorrelated map tends to reduce the information redundancy in the sample since there are no sample points that are close together which can occur with stratified random sampling.

(3) Zoning the map by elevation ($\Im 4$) has not improved the efficiency of the estimator relative to stratified random ($\Im 2$) or systematic ($\Im 3$) sampling. In all cases the estimator based on $\Im 4$ produces notably higher variances although it is better than random sampling. Presumably, and contrary to earlier expectations, the distribution of non-cultivated land is not associated with elevation. This zoning produces a less efficient estimator since the statistical zonation corresponds poorly to the true heterogeneity of the real surface, see Equation (10).

(4) In the case of the 1995 map, constructing zones using the experience gained from the 1985 sampling survey has produced significant improvements in efficiency relative to stratified random and systematic sampling in both cases. This could be theoretically confirmed by comparing Equation (7) for the case of stratified random (Equation (5)) and systematic (Equation (6)) sampling. Intuitively, a good zonation that perfectly matched the spatial heterogeneity of the real surface would have a very small dispersion variance, and so a very small error variance for the sample mean.

There may be a sampling proportion effect here with the benefits of zoning on the basis of previous experience becoming more apparent when the sampling fraction is small. This empirical example has compared findings on maps with different levels of spatial autocorrelation but used a small data set. To explore the effects of zoning further we now consider a second example using a much larger data set.

5.2. Sample survey of irrigated cultivated land in Shandong province, N. E. China, 2000

The study area is a rectangle of size 4320 km^2 in Shandong province (see Figure 8). The thematic mapper (TM) images in year 2000 are used to record the proportion of the cultivated area that is irrigated. The spatial resolution is a 30 m by 30 m^2 grid accumulated

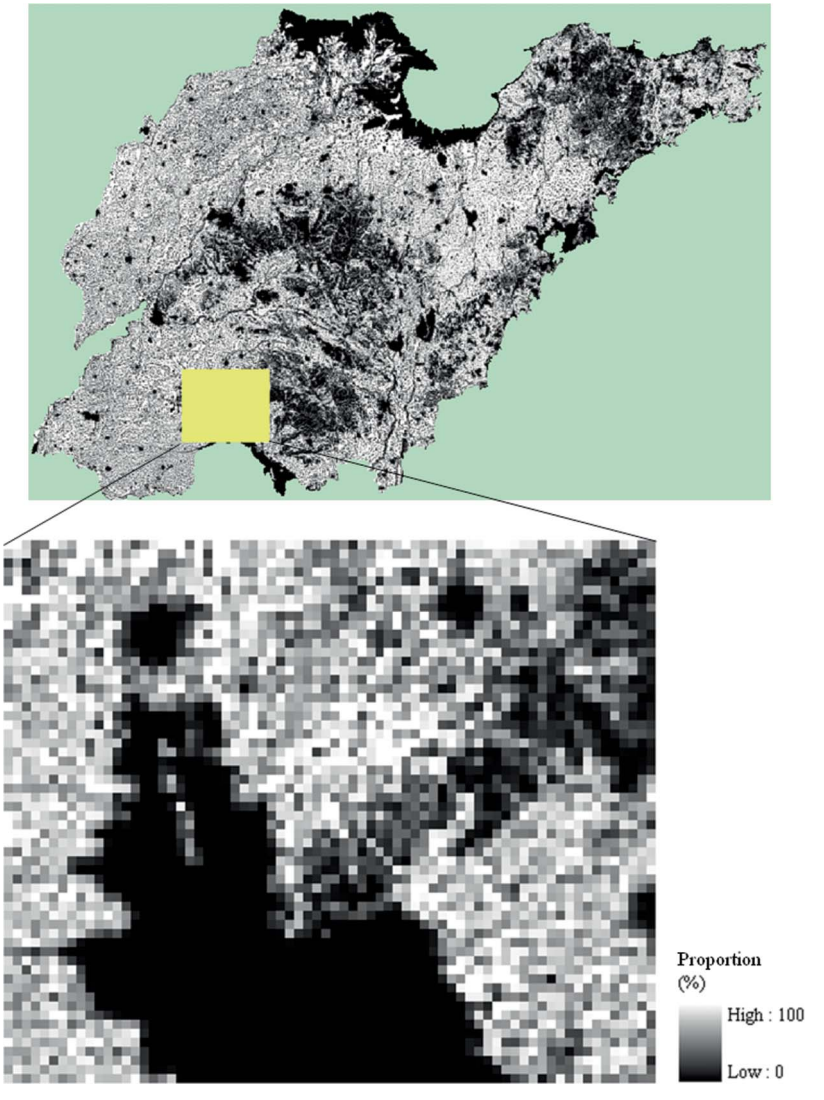

Figure 8. Shandong province (above) and true proportion of irrigated cultivated land (below): 2000.

into a 1 by 1 km^2 grid system. The population is composed of 4320 units (1 × 1 km^2) and the true value of each grid unit is shown in Figure 8 as a choropleth map.

We let ℜ00 denote the map of the true proportion of irrigated cultivated land in Shandong province in 2000. Spatial heterogeneity in the mean is evident in Figure 8, which suggests an area with much lower levels of irrigated cultivated land in the south central area of the map. Moran's statistic for spatial autocorrelation is 0.7434, evidence of significant positive spatial autocorrelation ($p < 0.05$), stronger than either of the two maps used in Example 1. The parameters of the fitted spherical semi-variogram model are range = 28 km, nugget $C_0 = 3.960$, sill $C = 15.403$ and $C_0/C = 0.257$ for 2000.

As before, random (ℑ1), stratified random (ℑ2) and systematic (ℑ3) sampling are carried out. Figure 9 displays the strata for systematic and stratified random sampling,

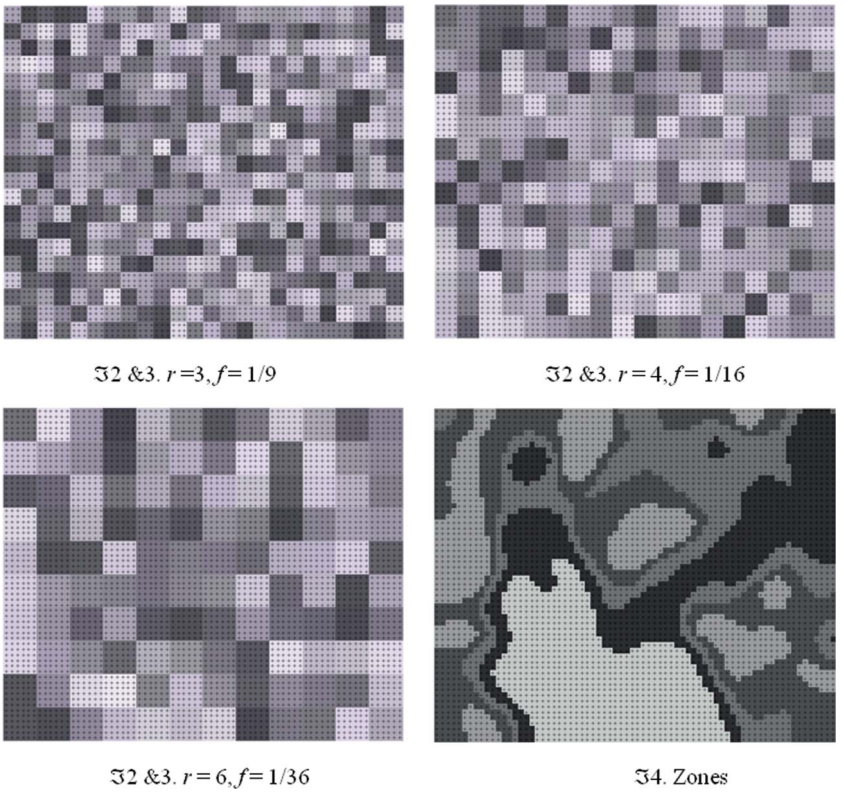

Figure 9. Systematic and stratified sampling frames and the zones used to allow for heterogeneity.

where r and f are defined as before and the zoning for (ℑ4) is constructed by dividing elevation into six levels, each with an equal number of cells, then applying spatial smoothing to keep each of the zones compact.

The sampling mean and variance are calculated as before under 1000 Monte Carlo replications. As expected random sampling (ℑ1) has the lowest efficiency. It is used as the baseline to evaluate the efficiency of the other three sampling plans (ℑ2–ℑ4) quantified by the design effect (deff), which is calculated as before. Tables 3 and 4 report results and we draw the following conclusions:

(1) Given the same size of sample, random sampling is the least efficient. Systematic sampling is more efficient than stratified random sampling in all three cases (see Section 5.1(2)).
(2) Zoning is more efficient than systematic sampling in two of the three cases (see Section 5.1(4)). It is less efficient than systematic sampling when f = 1/9 but its efficiency becomes more marked as the sampling fraction decreases. As with the first example, the benefits of zoning become apparent when the sampling fraction falls, that is the sampling points become sparse.

In conclusion it is worth stressing that these results show that different sampling methodologies are not distinguished from one another by the point estimates of the mean that they yield. All the methods provide similar, statistically unbiased estimates. They are

Table 3. Means and variances obtained from sampling the proportion of irrigated cultivated land in Shandong province in 2000 (ℜ00).

Sampling methods	$r = 3, f = 1/9$				$r = 4, f = 1/16$				$r = 6, f = 1/36$			
	Min	Max	Mean	Var.	Min	Max	Mean	Var.	Min	Max	Mean	Var.
Random sampling ℑ1	0.470	0.567	0.519	0.016	0.454	0.600	0.519	0.022	0.409	0.623	0.521	0.033
Stratified random sampling ℑ2	0.506	0.537	0.519	0.009	0.487	0.547	0.519	0.017	0.466	0.593	0.520	0.026
Systematic sampling ℑ3	0.488	0.544	0.519	0.008	0.478	0.559	0.519	0.012	0.453	0.597	0.519	0.021
Zoned sampling ℑ4	0.495	0.547	0.520	0.009	0.484	0.557	0.519	0.011	0.455	0.572	0.520	0.017

Note: the true proportion of irrigated cultivated land is 0.5195.

Table 4. Design effects due to different methods of sampling the proportion of irrigated cultivated land in Shandong province in 2000 (ℜ00).

Design effect	$r = 3, f = 1/9$	$r = 4, f = 1/16$	$r = 6, f = 1/36$
Var(ℑ1)/Var(ℑ2)	1.7692	1.3005	1.2472
Var(ℑ1)/Var(ℑ3)	1.9167	1.8595	1.6009
Var(ℑ1)/Var(ℑ4)	1.8089	1.8908	1.8920

differentiated by the error variance or uncertainty that is attached to those point estimates. One implication of this is that zonation allows an analyst to achieve a desired level of estimator precision with a smaller sample size – important if sampling is expensive.

6. Conclusions and discussion

Spatial sampling is one of the core techniques in both GIScience and GIServices, concerned with the efficient collection of population data and making inferences with smaller error variance. Understanding how to achieve high levels of sampling efficiency is likely to be of greater importance to GIS analysts in the years ahead for the reasons discussed.

The effects of spatial structure (spatial autocorrelation) on the error variance of sampling schemes have been the subject of systematic study dating back to the early work of Milne (1959) and Matern (1960). The benefits of stratification for the purpose of improving the error variance of estimators of the mean when spatial autocorrelation is present have been well established. These results apply to homogeneous surfaces and when the semi-variogram or the autocovariance function is available. Spatial heterogeneity in the mean, the variance and in the structure of spatial autocorrelation are also properties of spatial populations but appear not to have received the same systematic attention in the sampling literature.

Our results indicate that spatial heterogeneity in the mean of the real surface to be surveyed impacts on the sampling error associated with the sample mean as an estimator of the population mean. As the numerical examples have illustrated, if the nature of the underlying heterogeneity is understood it can be used to improve the sampling efficiency of the estimator of the mean because sampling efficiency is sensitive to spatial heterogeneity in the mean of the surface to be sampled.

Systematic sampling has been recommended for spatial surveys by previous studies. Its superiority to random sampling is well established and although its superiority to stratified random sampling is not as clear-cut, much of the evidence (both theoretical and empirical) indicates that it is to be preferred. In this study, we have shown both the benefits and dangers of introducing a higher level of stratification which we call zoning and which is designed to match the spatial heterogeneity in the mean of the real surface. The benefits are most evident when the choice of zones coincides closely with that heterogeneity. From the empirical studies reported here it would seem that knowledge acquired from earlier surveys, and possibly pilot surveys, are the best guide to that heterogeneity and it may be dangerous to depend too much on what may appear to be plausible indicators of heterogeneity – such as altitude in the first example.

Extra effort may be required to implement zoning so any extra costs will need to be weighed against the benefits. The sampler needs to be sure that the underlying heterogeneity is properly understood and captured in the choice of zones. In that case the extra cost (in terms of constructing the zones and implementing the sampling strategy in accordance with that partitioning) relative to the improvement in sampling efficiency seems to be a price worth paying. We noted that the gains from zoning were most evident when the sampling fraction was small, implying a sparse distribution of sample points. This suggests that zoning may be particularly useful if the costs of taking a sample are high so that the number of sample points needs to be kept as small as possible whilst still ensuring that the aims of the survey are met. It would be interesting to compare results obtained by this method with, for example, the method of Van Groenigen *et al*. (1999), which combines systematic sampling with a few additional clustered observations taken around a few randomly selected points where the additional sampling is based on pre-zoning.

The results presented here are based on raster data sets. GIS researchers work with other types of data as well. Although data in vector format for example might present some additional complications in terms of sample selection we conjecture that the main results identified here will still apply. There may also be value in exploring the extension of these methods to other types of problems such as map reconstruction through spatial interpolation where the distribution of the attribute is heterogeneous. This reflects one of the core interests of GIS researchers, namely capturing spatial differentiation. An attribute may be expensive to collect and so efficient methods are needed to produce maps from a comparatively sparse network of sample sites.

Acknowledgments

This material is based on work supported by the NSFC (40471111, 70571076), CAS (KZCX2-YW-308) and the MOST (2006AA12Z215, 2007DFC20180 and 2007AA12Z241). The authors thank Iain Macleod for his help in preparing the manuscript and also wish to record their thanks to three anonymous referees who provided helpful comments on an earlier draft of this paper.

References

Almeida, C.M., *et al*., 2008. Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science*, 22 (9), 943–963.

Berry, B.L. and Baker, A.M., 1968. Geographic sampling. *In*: B.L. Berry and D.F. Marbles, eds. *Spatial analysis*. Englewood Cliffs, NJ: Prentice-Hall, 91–100.

Brus, D.J. and 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. and Te Riele, W.J.M., 2001. Design-based regression estimators for spatial means of soil properties: the use of two-phase sampling when the means of the auxiliary variables are unknown. *Geoderma*, 104, 257–279.
Brus, D.J. and Heuvelink, G.B.M., 2007. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma*, 138, 86–95.
Christakos, G., 2005. *Random field models in earth sciences*. New York: Dover Publications.
Cochran, W.G., 1946. Relative accuracy of systematic and stratified random samples for a certain class of populations. *Annals of Mathematical Statistics*, 17, 164–177.
Cochran, W.G.., 1977. *Sampling techniques*. 3rd ed. Chichester: John Wiley.
Cressie, N., 1993. *Statistics for spatial data*. Chichester: John Wiley.
Csillag F., Kertesz M., and Kummert, A., 1996. Sampling and mapping of heterogeneous surfaces: multi-resolution tiling adjusted to spatial variability. *International Journal of Geographical Information Systems*, 10, 851–875.
Das, A.C., 1950. Two-dimensional systematic sampling and the associated stratified and random sampling. *Sankhya*, 10, 95–108.
de Gruijter, J.J., *et al*., 2006. *Sampling for Natural Resource Monitoring*. Berlin: Springer.
Dunn, R. and Harrison, A.R., 1993. Two-dimensional systematic sampling of land use. *Applied Statistician*, 42, 585–601.
Goodchild, M. and Gopal, S. 1989. *Accuracy of spatial databases*. London: Taylor & Francis.
Goodchild, M.F. and Haining, R.P., 2004. GIS and spatial data analysis: converging perspectives. *Papers Regional Science*, 83, 363–385.
Green, J.L. and Plotkin, J.B., 2007. A statistical theory for sampling species abundances. *Ecology Letters*, 10, 1037–1045.
Griffith, D.A., 2005. Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the Association of American Geographers*, 95, 740–760.
Griffith, D.A., Haining, R.P., and Arbia, G.., 1994. Heterogeneity of attribute sampling error in spatial data sets. *Geographical Analysis*, 26, 300–320.
Grinand, C., *et al*., 2008. Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. *Geoderma*, 143, 180–190.
van Groenigen, J.W., Siderius, W., and Stein, A., 1999. Constrained optimization of soil sampling for minimization of the kriging variance. *Geoderma*, 87, 239–59.
Freund, J.E., 1992. *Mathematical statistics*. 5th ed. Englewood Cliffs, NJ: Prentice-Hall.
Haining, R.P., 1988. Estimating spatial means with an application to remote sensing data. *Communication Statistics – Theory and methodology*, 17, 537–597.
Haining, R.P., 2003. *Spatial data analysis: theory and practice*. Cambridge: Cambridge University Press.
Huenneke, L.F., Clason, D., and Muldavin, E., 2001. Spatial heterogeneity in Chihuahuan Desert vegetation: implications for sampling methods in semi-arid ecosystems. *Journal of Arid Environments*, 47, 257–270.
Kumar, N., 2009. An optimal spatial sampling design for intra-urban population exposure assessment. *Atmospheric Environment*, 43, 1153–1155.
Lee, C., Moudon, A.V., and Courbois, J.-Y.P., 2006. Built environment and behavior: spatial sampling using parcel data. *Annals of Epidemiology*, 16, 387–394.
Leung Y., Ma, J.H., and Goodchild, M., 2004. A general framework for error analysis in measurement-based GIS Part 1: the basic measurement-error model and related concepts. *Journal of Geographical Systems*, 6, 325–354.
Li, L.F., *et al*., 2008. An information-fusion method to regionalize spatial heterogeneity for improving the accuracy of spatial sampling estimation. *Stochastic Environmental Research and Risk Assessment*, 22, 689–704.
Matern, B., 1960. *Spatial variation*. 2nd ed. *Lecture notes in statistics* 36. Berlin: Springer.
Milne, A., 1959. The centric systematic area-sample treated as a random sample. *Biometrics*, 15, 270–297.
Ott, D.K., Kumar, N., and Peters, T.M., 2008. Passive sampling to capture spatial variability in $PM_{10-2.5}$. *Atmospheric Environment*, 42, 746–756.
Payandeh, B. 1970. Relative efficiency of two-dimensional systematic sampling. *Forestry Science*, 16, 271–276.
Quenouille, M.H., 1949. Problems in plane sampling. *Annals of Mathematical Statistics*, 20, 355–375.
Ripley, B., 1981. *Spatial statistics*. New York: Wiley.
Rodeghiero, M. and Cescatti, A. 2008. Spatial variability and optimal sampling strategy of soil respiration. *Forest Ecology and Management*, 255, 106–112.

Rodriguez-Iturbe, I. and Mejia, J.M., 1974. The design of rainfall networks in time and space. *Water Resources Research*, 10, 713–728.
Rogerson, P.A., *et al*., 2004. Optimal sampling design for variables with varying spatial importance. *Geographical Analysis*, 36, 177–194.
Sen, S., 2008. Framework for probabilistic geospatial ontologies. *International Journal of Geographical Information Science*, 22, 825–846
Simbahan, G.C. and Dobermann, A., 2006. Sampling optimization based on secondary information and its utilization in soil carbon mapping. *Geoderma*, 133, 345–362
Shi, W.Z., 2005. *Principles of modeling uncertainties in spatial data and spatial analyses*. Beijing: Science Press.
Stoter, J., de Kluijver, H., and Kurakula, V., 2008. 3D noise mapping in urban areas. *International Journal of Geographical Information Science*, 22, 907–924.
Villarini, G. and Krajewski, W.F., 2008. Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges. *Advances in Water Resources*, 31, 1015–1023.
Wang, J.F., Wise, S., and Haining, R., 1997. An integrated regionalization of earthquake, flood and drought hazards in China. *Transactions in GIS*, 2, 25–44.
Wang, J.F., *et al*., 2002. Spatial sampling design for monitoring the area of cultivated land. *International Journal of Remote Sensing*, 13, 263–284.
Zubrzycki, S., 1958. Remarks on random, stratified and systematic sampling on a plane. *Colloquium Mathematicum*, 6, 251–264.

Appendix

$$s^2 = \sum_{z=1}^{L_z} W_z s_z^2 + \sum_{z=1}^{L_z} W_z (\bar{y}_z - \bar{Y})^2$$

Proof:

$$\begin{aligned}
s^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{Y})^2 = \frac{1}{n}\sum_{z=1}^{L_z}\sum_{i=1}^{n_z}(y_i - \bar{y}_z + \bar{y}_z - \bar{Y})^2 \\
&= \frac{1}{n}\sum_{z=1}^{L_z}\sum_{i=1}^{n_z}[(y_i - \bar{y}_z)^2 + (\bar{y}_z - \bar{Y})^2 + 2(y_i - \bar{y}_z)(\bar{y}_z - \bar{Y}) \\
&= \frac{1}{n}\sum_{z=1}^{L_z}\sum_{i=1}^{n_z}(y_i - \bar{y}_2)^2 + \frac{1}{n}\sum_{z=1}^{L_z}\sum_{i=1}^{n_z}(\bar{y}_z - \bar{Y})^2 + \frac{2}{n}\sum_{z=1}^{L_z}\sum_{i=1}^{n_z}(y_i - \bar{y}_z)(\bar{y}_z - \bar{Y}) \\
&= \frac{1}{n}\sum_{z=1}^{L_z} n_z s_z^2 + \frac{1}{n}\sum_{z=1}^{L_z} n_z (\bar{y}_z - \bar{Y})^2 \\
&= \sum_{z=1}^{L_z} W_2 s_z^2 + \sum_{z=1}^{L_z} W_z (\bar{y}_z - \bar{Y})^2.
\end{aligned}$$