INT. J. REMOTE SENSING, 2002, VOL. 23, NO. 2, 263–284

Spatial sampling design for monitoring the area of cultivated land

J. WANG†, J. LIU‡, D. ZHUAN‡, L. LI† and Y. GE†

†State Key Laboratory of Resources and Environmental Information System, Institute of Geography, Chinese Academy of Sciences, Beijing 100101, PR China; e-mail: wangjf@lreis.ac.cn
‡Institute of Remote Sensing Application, Chinese Academy of Sciences, Beijing 100101, PR China

(*Received 8 September 1999; in final form 24 july 2000*)

Abstract. Updated information on cultivated land is important for Chinese central and local governments. The data can be acquired using aerial photographs and Thematic Mapper (TM) images. But an exhaustive annual survey covering all of China's territory by these remote sensing images is too expensive, therefore a sampling technique has to be employed. Spatial sampling takes the spatial distribution characteristics of the object to be monitored into account. We propose both direct and indirect spatial sampling models for monitoring spatially discrete distributed objects. For the indirect method, each sampling domain is equal to a specified region but is not directly linked with the reporting unit, consequently, the report unit estimates may have few or perhaps even no samples within the report units. Therefore the indirect sampling model can provide sampling estimates for a large number of report units with a limited number of sample units and a limited sampling budget. The zoning of the monitored object is based on prior knowledge about the controlling factors and the spatial homogeneity of the variable. The method is used to develop a sampling solution for monitoring cultivated land dynamics. The models were tested in Shandon province and Zhaozhuang county.

1. Introduction

1.1. *Objectives*

The cultivated land per capita in China is about 1.6 mu (≈ 1 ha), less than one-third of the world average, and the limited gross cultivated land is further reduced by a large number of thin-non-cultivated features, such as small ridges, small wooded areas, graveyards, independent houses and yards, small roads, brooks, ditches, etc. which may occupy about 5–25% of the area within a cultivated area (Liu 1996, pp. 25–30). In addition, the cultivated land had changed rapidly due to urbanization and social–economic reform in the country in the last 20 years and the trend is expected to last for the coming decades (Liu 1996, pp. 262–275).

Therefore, true updated figures for the net area and the rate of change of cultivated land are concerns for Chinese governments at all three administrative levels of the country, province and county. An exhaustive investigation of the cultivated land throughout the country was begun in 1984 at a cost of 1 billion RMB (0.12 billion US$), which took 11 years to complete.

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160010025998

Our objective was to develop the necessary spatial sampling theories by which aerial photographs with high spatial resolution and Thematic Mapper (TM) images with lower spatial resolution can be jointly used to monitor the cultivated land and to report information to different administrative levels more rapidly with less expense and the required accuracy.

1.2. *Review*

Currently, two survey tools (administration upward reporting procedures and remote sensing) and two accounting methods (thorough population investigations and sampling surveys) are available for monitoring cultivated land area. Each has its own strengths and limitations (see table 1). The thorough population investigation is accurate but much more expensive and slower, furthermore, the complexity of geomorphic, climate, farming customs and the huge size of China increases the difficulty of the approach. The administration upward reporting procedure usually suffers from artificially modified figures because of lack of strict oversight and political reasons in a planned economic society. For example, in some cases, a local officer may be awarded if more cultivated land is developed according to his report, while in other cases, the local government may much more easily obtain supplements if less local resources are reported. Remote sensing is accurate and rapid but would be very expensive if images such as TM were used to cover all of China's territory yearly for a thorough investigation of the cultivated land area, although it is much cheaper than a thorough population investigation, which is based mainly on measuring land in the field. The sampling technique may lose some accuracy compared with the exhaustive investigation, but it has some significant advantages (Cochran 1977, pp. 1–2): (1) Reduced costs—if data is secured from only a small fraction of the aggregate, the total expenditure is smaller than if a complete census is attempted. (2) Increased speed—the data can be collected and summarized more quickly with a sample than with a complete count, which is a vital consideration when the information is urgently needed. (3) Extended accuracy—high quality personnel can be employed and given intensive training to assume more careful supervision of the field work and processing of the results becomes feasible when the work volume is reduced. Therefore, a sample may produce more accurate results than a complete enumeration.

Sampling techniques always include estimate variance, which depends on the spatial sampling scheme (Haining 1990, p. 176, Figure 5.1 and p. 180, Figure 5.2), the number of samples and the dispersion variance (Journel and Huijbregts 1978, pp. 48–51, pp. 61–68) of the monitored phenomena itself (Cochran 1977).

Table 1. Strengths and limitations of two survey tools and two accounting methods currently used in monitoring cultivated land.

		Population investigation	Sample survey
		Accurate + Expensive	Savings + Estimate variance
Administration	Inaccurate	Inaccurate + Expensive	Inaccurate + Estimate variance
Remote sensing	Rapid + Accurate	Accurate + Expensive	Accuracy + Estimate variance

Spatial sampling methods originated from classic sampling techniques (Cochran 1977) and benefit from the recent research on spatial statistics.

Classic sampling techniques (Morris and DeGroot 1975, Cochran 1977), include a relationship between the variance (σ^2) of the population estimate, the sampling density (n) and sampling scheme: $\sigma^2 = f(n, \text{scheme})$. But the correlations between samples, usually existing between spatially distributed objects, are neglected in the model, consequently, the estimated variance tends to be biased (Griffith 1988, Haining 1990, p. 183).

In spatial statistics, the spatial sample distribution and correlation are fundamental information for determining the estimate variance and sample support effects (Ghosh 1951, Journel and Huijbregts 1978, pp. 66–67, Isaaks and Srivastava 1989, pp. 476–480, Fotheringham *et al.* 1996, Bellehumeur and Legendre 1997).

The spacial sampling technique combines the classic technique and spatial statistics and has been investigated for decades (Rodgriguez-Iturbe and Mejia 1974, Langford and Kapinos 1979, O'Connell *et al.* 1979, Dunn and Harrison 1993). Most spatial sampling models consider objects continuously distributed in space such as rainfall with the estimating procedure directly extrapolated from the sample unit to report unit.

1.3. *Outline of the paper*

In the paper, remote sensing is used as the survey tool and the sampling technique is employed to summarize statistically the objects to be monitored, the cultivated land and its properties, which are distributed discontinuously but are mutually correlated in space. Two spatial sampling models have been developed for monitoring spatial discrete objects: the direct method by which the population in the report units is directly estimated from the sample units directly, and the indirect method by which the estimate for the report unit is formed from the sample unit data through using zones which are produced in advance by prior knowledge of the monitored object, especially that its value is relatively homogeneous in space within each zone (Institute of Agriculture Zonation Research of China 1995, Wang *et al.* 1997).

The indirect approach is a better choice than the direct approach and is indispensable when few or even no samples lay within the report units, which often occurs when the report units are very small or too numerous (such as counties) so that there are not enough samples for each report unit for statistical purposes. Measuring sufficient samples for each report unit would cause a huge increase in sample expenditures.

Both direct and indirect models are used in static (net farmland area estimated every 5 years) and dynamic (yearly farmland changes) estimates in large scale (province) and in small scale (county) units.

2. Measures of network performance

2.1. *Aims of a sampling network*

The sampling objectives can be categorized into three classes based on the different usages of the monitoring network: (1) estimating non-spatial characteristics of a spatial population; (2) forming a map or summary of features; and (3) obtaining observations for classifying data (Haining 1993, pp. 171–172). Alternatively, the spatial information infrastructure can be used to classify the sampling aims into three levels: Level 1: estimating background information for wide regional or national planning resource inventories; Level 2: providing general resource planning data;

and Level 3: collecting data for specific planning and management activities (Rodda *et al.* 1969, p. 12). In more quantitative language, the aims of a sampling network design can be to estimate:

- **Aim 1:** the population mean (Griffith *et al.* 1994) or total (Barnett 1974, pp. 35–36);
- **Aim 2:** the values at other places than the sampling sites;
- **Aim 3:** ratios (Barnett 1974, pp. 51–63, and Cochran 1977, pp. 30–34) or various other features of the region.

For Aim 1, the arithmetic average and the total of the sampling values divided by the sampling rate are unbiased estimates of the population mean and the population total (Cochran 1977, p. 22). For Aim 2, the values at other locations besides the sampling sites have to be estimated using spatial distribution based interpolation techniques in the light of information from the sampling (Cliff and Ord 1975, Haining 1990, pp. 296–307). Aim 3 mixes Aim 1 and Aim 2.

2.2. *Measures of sampling network performance*

The performance of a monitoring or sampling network can be measured using different indexes according to the different aims of the monitoring networks. There are at least three measures:

Measure 1: MSE (mean square error) of the mean

$$\sigma_n^2 = E\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i) - \frac{1}{A}\int_A f(x_i)\,\mathrm{d}x_i\right]^2 \tag{1}$$

where E denotes the mathematical expectation and $f(x_i)$ denotes a variable to be sampled at point x_i in the space domain A. The term $1/A \int_A f(x_i)\,\mathrm{d}x_i$ is the true mean over A. In practice, this mean is estimated by the arithmetic mean of n point samples and is given by $1/n \sum_{i=1}^{n} f(x_i)$. Expression (1) defines the mean square error of the estimate of the mean. The optimal sampling method minimizes (1) which is a function of the relationship between σ_n^2 and n.

Measurement 2: MSE of discrepancy

$$\mathrm{MSE}\,\hat{f}(x_i) = E[(\hat{f}(x_i) - f(x_i))^2 \mid A] \tag{2}$$

$f(x_i)$ denotes the true value at point i and $\hat{f}(x_i)$ is the estimated value of $f(x_i)$, which is interpolated at site i in the light of information from the sampling at a series of sites $\{x_i\}$.

Measure 3: Maximum site error

$$\left\{\max_i \| \hat{f}(x_i) - f(x_i)|, \forall i \in A\right\} \tag{3}$$

Measure 1 is appropriate when the sampling network is employed for monitoring the mean, or after a small modification, for monitoring the total spatial population; Measures 2 and 3 are suitable when the network is used for interpolation in space.

In monitoring cultivated land area, the sampling network should be designed to meet Aim 1, and measure 1 should be used to indicate the network performance, based on the optimal relationship between the estimated error reduction, the sampling density and the sampling schemes.

3. Task specification

3.1. *Data and notations*

The relationships between the various layers are illustrated in figure 1. The bottom layer represents the cultivated land to be monitored. The second layer from the bottom represents sample units, i.e. aerial photographs or TM images. The knowledge or zoned layer above the sampling layer indicates the relative homogeneity of the variable to be monitored in space within each zone as compared with between zones. The variable, for instance, could be the thin-non-cultivated component coefficient, which is detectable in aerial photographs, or the non-cultivated land proportion, which is detectable by TM images. The layer above the zoned layer is the report units, such as provinces or counties which report the cultivated land area and its dynamics. The sampling layer, knowledge layer and reporting layer are overlaid together to form an integrative frame for spatial sampling design. The object population to be monitored is estimated from the sample units to the report units along the layered frame using the spatial sampling models developed in this paper.

In the following context, the TM cultivated land area means the cultivated land area determined from the TM images. The aerial photograph cultivated land area means the cultivated land area determined from the aerial photographs, which is equal to the net cultivated land area. The aerial photograph thin-non-cultivated component coefficient means the thin-non-cultivated component coefficient estimated from the aerial photographs.

For the static spatial sampling model in this paper, our objective is to estimate the net cultivated land area every 5 years using exhaustive data for the TM cultivated land area and the sampling data for the aerial photograph thin-non-cultivated component coefficients. The approach should be much cheaper than estimates exhaustive data for the aerial photograph cultivated land area. The notation used in the model is listed in table 2:

'#' or TNCC = thin-non-cultivated features within a cultivated land area, which are smaller than $30\,\mathrm{m} \times 30\,\mathrm{m}$ in size and are neglected by TM images, but should be identified by aerial photograph.

S_p = gross cultivated land area in p-th province or county obtained from the TM image, or the TM cultivated land area in the p-th report unit.

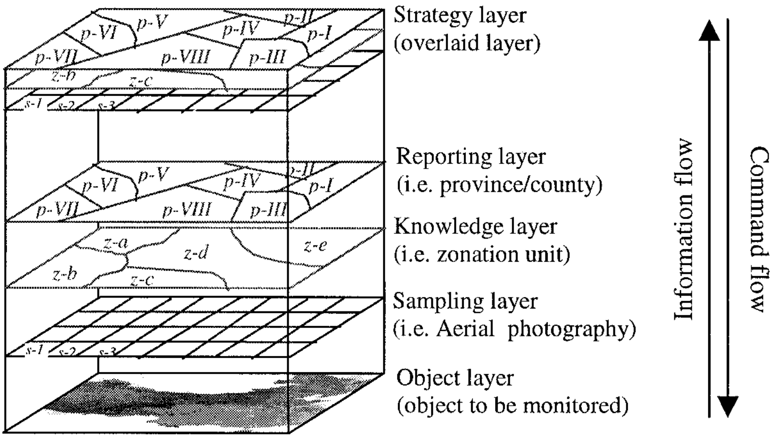

Figure 1. Spatial sampling strategy for monitoring

Table 2. Notation for static spatial sampling models.

	Variable									
	Gross cultivated land area, S	Net cultivated land area, S^0	Thin-cultivated component area, $S^{\#}$	TNCC coefficient, $\beta^{\#}=S^{\#}/S$	Standard variance, $\sigma_{\beta^{\#}}, \sigma_{S^{\#}}$	Relative variance, $\rho_{\beta^{\#}}, \rho_{S^{\#}}$	Sample number, n		Sampling weight, w	
Spatial unit	(1)	(2)	(3)	(4)	(5)	(6)	Real (7)	Max (8)	Real (9)	Max (10)
p-th province or county	$S_p=\Sigma_{a=1}^{N_{pa}} S_a$	$S_p^0=S_p-S_p^{\#}$	$S_p^{\#}$	$\beta_p^{\#}=S_p^{\#}/S_p$	$\sigma_{\beta_p^{\#}}, \sigma_{S_p^{\#}}$	$\rho_{\beta_p^{\#}}, \rho_{S_p^{\#}}$	n_{pa}	N_{pa} N_{pc}	$w_{pa}=$ $S_a/\Sigma_{a=1}^{n_{pa}} S_a$	$W_{pa}=$ $S_a/\Sigma_{a=1}^{N_{pa}} S_a$
z-th zone	$S_z=\Sigma_{a=1}^{N_{za}} S_a$	$S_z^0=S_z-S_z^{\#}$	$S_z^{\#}$	$\beta_z^{\#}=S_z^{\#}/S_z$	$\sigma_{\beta_z^{\#}}, \sigma_{S_z^{\#}}$	$\rho_{\beta_z^{\#}}, \rho_{S_z^{\#}}$	n_{za}	N_{za}	$w_{za}=$ $S_a/\Sigma_{a=1}^{n_{za}} S_a$	$W_{za}=$ $S_a/\Sigma_{a=1}^{n_{za}} S_a$
a-th sample	S_a	S_a^0	$S_a^{\#}$	$\beta_a^{\#}=S_a^{\#}/S_a$						

$S_p^{\#}$ = area of thin-non-cultivated features in p-th province or county
S_a = gross cultivated land area in a-th sample.
$S_a^{\#}$ = area of thin-non-cultivated features within TM cultivated land area in an aerial photograph sample a.
$S_p^0 = S_p - S_p^{\#}$, net cultivated land area in p-th province or county
a = a sampling unit, such as an aerial photograph in the static spatial sampling model or a TM image in the dynamic spatial sampling model
z = a zone unit
p = a report unit, such as a province or a county
$\beta_a^{\#}$ = thin-non-cultivated component coefficient within sample unit a, obtained by aerial photograph sampling or the aerial photograph thin-non-cultivated component coefficient
$\beta_{pa}^{\#}$ = thin-non-cultivated component coefficient within sample unit a in report unit p
$\beta_p^{\#}$ = thin-non-cultivated component coefficient in report unit p
n_{pa} = number of samples in report unit p, similarly, n_{za} = number of samples in zonation unit z
$\sigma_{\hat{\beta}_p^{\#}} = \sqrt{E_p(\hat{\beta}_p^{\#} - \beta_p^{\#})^2}$, standard estimate variance of $\hat{\beta}_p^{\#}$ of report unit p
$\sigma_{\beta_p^{\#}} = \sqrt{E_p(\beta_{pa}^{\#} - \beta_p^{\#})^2}$, standard dispersion variance of $\beta_a^{\#}$ within p
$\rho_{\hat{\beta}_p^{\#}} = \sigma_{\hat{\beta}_p^{\#}} / \beta_p^{\#}$, relative estimate variance of $\hat{\beta}_p^{\#}$

A variable without the cap, for instance β, refers to the true value and the one with a cap '$\hat{}$', for instance $\hat{\beta}$, refers to the estimate of the true value, E_2 refers to the mathematical expectation over z and '$\equiv$' denotes a definition. The other notation in table 2 is easily understood from these definitions.

For the dynamic spatial sampling model in this paper, our objective is to estimate the cultivated land area every year from the last exhaustive data for the TM cultivated land area and the current sampling TM non-cultivated-land proportion. The approach should be much cheaper than estimates using exhaustive aerial photograph or annual exhaustive TM. The notation used in the model are the same as the list in table 2 but with the following substitutions: 'a' represents a TM image rather than an aerial photograph in the static model, '#' is replaced by '##', referring to the non-cultivated land determined from the TM images. The areas of sample unit a and report unit p are denoted by A and P. $S_p^{\#\#}$ denotes the non-cultivated land area in p. $\beta_a^{\#\#}$ denotes non-cultivated land proportion in a TM image, which is equal to $S_p^{\#\#}/A$. $\beta_p^{\#\#}$ denotes the non-cultivated land proportion in p, which is equal to $S_p^{\#\#}/P$.

3.2. *Tasks*

Our first objective is to estimate the net cultivated land area S_p^0 every 5 years. $S_p^0 = S_p - S_p^{\#} = S_p - S_p \times \beta_p^{\#} = (1 - \beta_p^{\#}) \times S_p$. The monitored area is exhaustively measured ever 5 years by TM images to obtain the gross cultivated land area S_p. Aerial photographs are then used to estimate $\beta_p^{\#} = S_p^{\#}/S_p$. The value of $\beta_p^{\#}$ is mainly a function of geomorphology, climate, farmer's customs and the amount of development in the region, which are usually believed to change slowly (Liu and Yu 1990) and so are updated by aerial photographs sampling every 5 years and by exhaustive covering every 10 years.

Our second objective is to estimate the cultivated land area S_p, change every year. $S_p = P - S_p^{\#\#} = (1 - \beta_p^{\#\#}) \times P$. TM image samples are used to estimate $\beta_p^{\#\#} = S_p^{\#\#}/P$. The values are updated by TM image sampling every year and with exhaustive coverage every 5 years. The monitoring strategies are summarized in table 3.

S_p^0 can even be estimated every year using the values of yearly $\beta_p^{\#\#}$ and the every 5 years value of $\beta_p^{\#}: S_p^0 = (1-\beta_p^{\#}) \times S_p = (1-\beta_p^{\#}) \times (1-\beta_p^{\#\#}) \times P$.

Therefore, we seek to estimate $\beta_p^{\#}$, $S_p^{\#}$, $\beta_p^{\#\#}$, $S_p^{\#\#}$, and their standard and relative variances using aerial photograph and TM image sampling.

The theoretical basis for spatial sampling is to determine the optimal relationship between estimation error, sampling density and spatial arrangement of samples in the spatial context for objects that are distributed continuously or discretely in space through direct or indirect estimating approaches. The optimal relationship minimizes the estimate variance (Journel and Huijbregts 1978, pp. 48–61) considering spatial correlation.

The spatial sampling models developed in this paper are listed in appendix 1, 2 and 3. The applications of the models are given in the following.

4. Examples

Both Shandon province, with an area of 150 000 km^2, in Eastern China, and Zhaozhuang county, with an area of 4550 km^2, in Shandon Province, were used as pilot regions to test the models developed in this paper. The analysis illustrates the static direct sampling model (appendix 1), the static indirect sampling model (appendix 2), the dynamic direct sampling model and the dynamic indirect sampling model (appendix 3).

The static and dynamic samplings are illustrated in figures 2 and 3; (*a*)–(*d*) are for province and (*e*) and (*f*) are for the county. 'Radio' refers to the thin-non-cultivated component coefficient in the static models and to the non-cultivated land proportion in the dynamic models. The 'thin feature' in figure 2 refers to the thin-non-cultivated components within the cultivated land. The 'non-farm' in figure 3 refers to the non-cultivated land in each region. 'Non-zero' in the brackets denotes the direct model while 'zone' denotes the indirect model. The analysis of the county data shows that the direct sampling model tends to be invalidated and the indirect model must be used because the report units are too small and too many so that there are not enough aerial photographs or TM images as samples in each county to provide confident statistics.

4.1. *Static spatial sampling (figure 2)*

In static spatial sampling, aerial photographs are used to identify the thin-non-cultivated features within the cultivated area which are identified from the exhaustive TM images. 'Static' means that the thin-non-cultivated component coefficients change

Table 3. Strategies for monitoring cultivated land area in China (revised from Liu (1996)).

Strategies for monitoring cultivated land	TM images ('*a*' refers to TM image)	Aerial photographs ('*a*' refers to aerial photograph)
Complete coverage (every 10 years)		Exhaustive coverage for S^0, $S^{\#}$ and $\beta^{\#}$
Static survey (every 5 years)	Exhaustive coverage for S_a	Sampling to $\beta_a^{\#}$
Dynamic survey (every year)	Sampling for $\beta_a^{\#\#}$ (every year) of non-cultivated land	Static $\beta_a^{\#}$ (every 5 years) for thin-non-cultivated feature

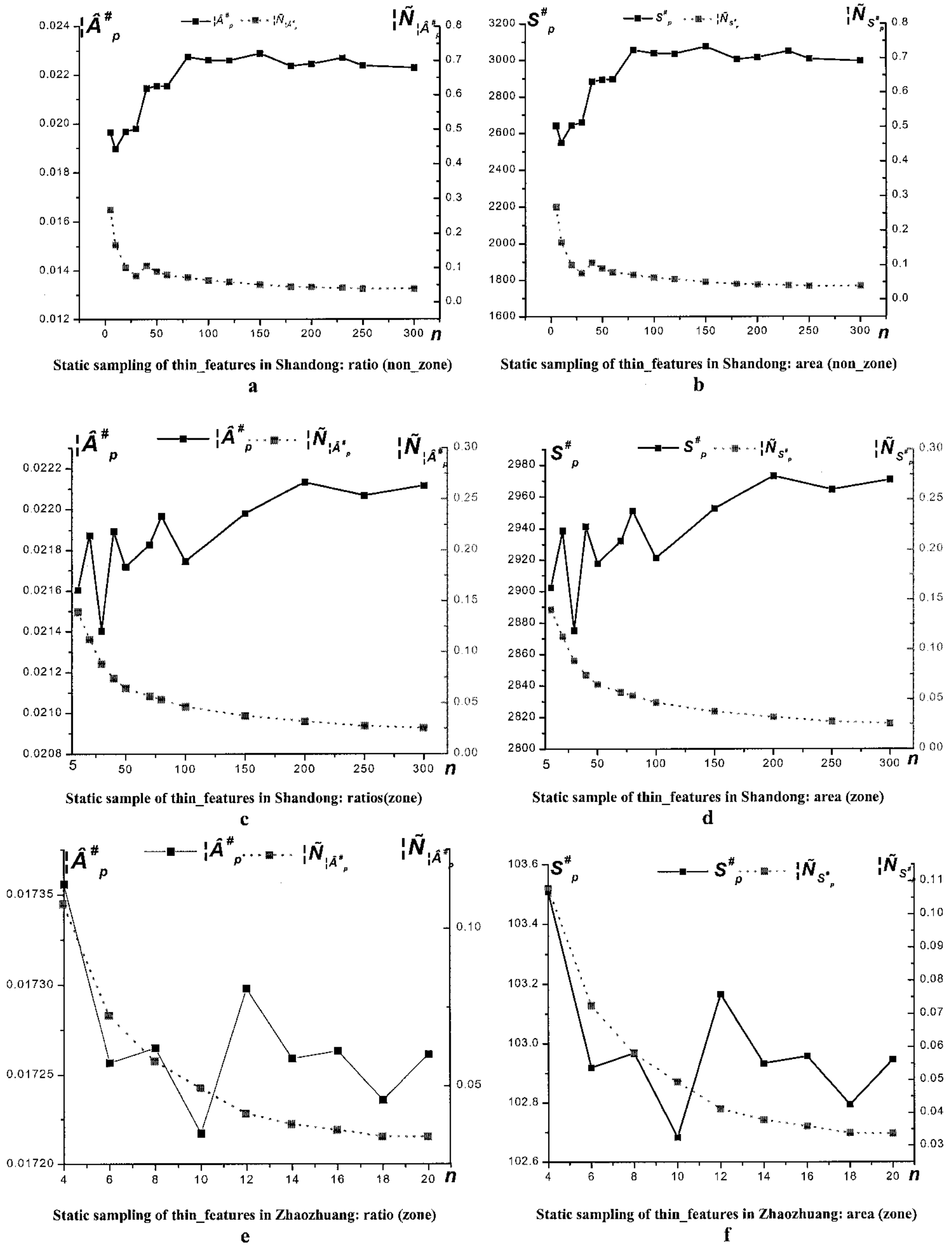

Figure 2. Using aerial photograph for sampling static thin-non-cultivated components within cultivated land in Shandon province and Zhaozhuang county (unit of $S_p^{\#}$ is 10 000 mu, 100 mu = 6.6666 ha).

little over several years and are updated by aerial photograph sampling every 5 years. The notation in table 2 is used for the following data:

- Aerial photographs, as samples for $\beta_a^{\#}$, were distributed randomly in space in Shandon in 1996.

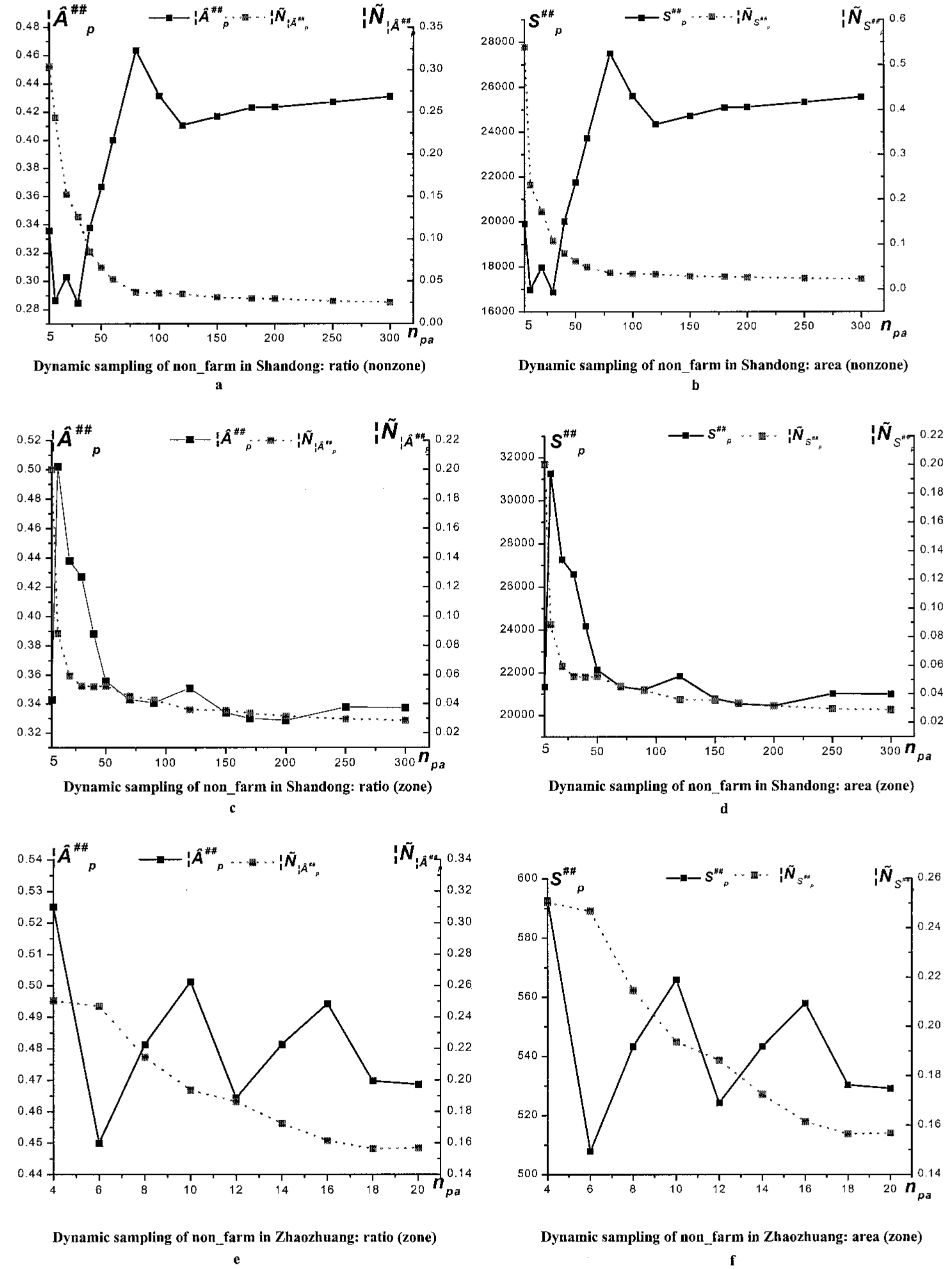

Figure 3. Using TM imges for sampling the dynamic non-cultivated land area in Shandon province and Zhaozhuang county (unit of $S_p^{\#\#}$ is 10 000 mu, 100 mu = 6.6666 ha).

- The zone map for $\beta_a^{\#}$ was produced by prior knowledge on $\beta_a^{\#}$, which is a function of geomorphic, weather, economic level and farming customs, is relatively homogeneous within each zone.
- Report unit map was based on Shandon Province and Zhaozhuang county boundaries in GIS vector format.

For the static direct spatial sampling model, the thin-non-cultivated component coefficients ($\hat{\beta}_p^{\#}$) and their corresponding relative errors ($\hat{\rho}_{\hat{\beta}_p^{\#}}$) were calculated using equations (A 1.3) and (A 1.7), respectively. The thin non-cultivated component area ($\hat{S}_p^{\#}$) and their corresponding relative errors ($\hat{\rho}_{\hat{S}_p^{\#}}$) were calculated using equations (A 1.8) and (A 1.10), respectively.

For the static indirect spatial sampling model, $\hat{\beta}_p^{\#}$, $\hat{\rho}_{\hat{\beta}_p^{\#}}$, $\hat{S}_p^{\#}$ and $\hat{\rho}_{\hat{S}_p^{\#}}$ were calculated using equations (A 2.7), (A 2.10), (A 2.11) and (A 2.13).

4.2. *Dynamic spatial sampling (figure 3)*

In dynamic spatial sampling, the TM images were used to identify the non-cultivated land within the report units. 'Dynamic' means that the non-cultivated land changes yearly.

The notation in table 2 for the static models and the variable substitution rules from static to dynamic models at the end of §3.1 were used for the following data:

- TM images, as samples for $\beta_a^{\#\#}$, were distributed randomly in space in Shandon in 1996.
- The zone map of $\beta_a^{\#\#}$ was produced by prior knowledge. N_{pz} and N_{za} were determined naturally.
- The map of $\Delta\beta_a^{\#\#}$ is the difference between the values of $\beta_a^{\#\#}$ in 1992 and 1996, which had two thorough investigations using TM images.
- The report unit map used Shandon Province and Zhaozhuang county boundaries in GIS vector format.

First, the variable substitution rules were used for the dynamic spatial sampling models. Then, for the direct models, the non-cultivated land proportions ($\hat{\beta}_p^{\#\#}$) and their corresponding relative errors ($\hat{\rho}_{\hat{\beta}_p^{\#\#}}$) were calculated using equations (A 1.3) and (A 1.7), respectively. The non-cultivated land areas ($\hat{S}_p^{\#\#}$) and their corresponding relative errors ($\hat{\rho}_{\hat{S}_p^{\#\#}}$) were calculated using equations (A 1.8) and (A 1.10). For the indirect model, $\hat{\beta}_p^{\#\#}$, $\hat{\rho}_{\hat{\beta}_p^{\#\#}}$, $\hat{S}_p^{\#\#}$ and $\hat{\rho}_{\hat{S}_p^{\#\#}}$ were calculated using equations (A 2.7), (A 2.10), (A 2.11) and (A 2.13).

5. Discussion and conclusions

5.1. *Objectives*

The net cultivated land area and the cultivated land area yearly change in the country, in provinces and counties are important information for both central and local governments. Although remote sensing is a cheaper, quicker and more accurate than human surveys and administrative reporting procedures in obtaining information, it is still expensive if all of China's territory is covered exhaustively by aerial photographs or by TM images, which are the minimum possible methods that can be used to obtain the net cultivated land area and its dynamics. Accordingly, sampling techniques are needed to obtain the information while balancing the monitoring operation expense and the estimate accuracy.

5.2. *Methodology*

The spacial sampling models developed in this paper are suitable for objects that are discretely distributed objects in space, such as cultivated land.

To obtain the net cultivated land area, serial photographs were used as samples to determine the thin-non-cultivated component coefficients $\beta_p^{\#}$. The values are believed to remain relatively stable over several years in relatively larger areas such as a county and a zone defined for the variables. Consequently, the net cultivated land area can be estimated every 5 years by multiplying the gross cultivated land area, which is obtained by thorough TM measurements once every 5 years, by $1-\beta_p^{\#}$, where $\beta_p^{\#}$ is also updated by aerial photographic sampling area once every 5 years. The same approach can be used with an easy substitution of the variable meanings to annually determine the cultivated land area by yearly TM image sampling.

In this paper, the sampling domain and report units are considered to be two separate layers which do not necessarily coincide at their boundaries. Therefore, the estimate in the report units do not necessarily use samples completely falling within the report unit as with implicit existing sampling techniques (Rodriguez-Iturbe and Mejia 1974, Jones *et al.* 1979, Cochran 1977, Burgess *et al.* 1981, Haining 1988, Dunn and Harrison 1993, Griffith *et al.* 1994, Csillag *et al.* 1996, Henry and Hope 1998). With more than 2700 counties in China, a complete survey would be too expensive even if with only a few samples in each county for statistical purpose. Consequently, an indirect sampling model with data derived from sampling layer to the reporting layer through a zone layer is necessary. The zones are produced by prior knowledge of the object to be monitored.

The thin-non-cultivated component coefficient is controlled by geomorpholigic, climate and farmers' customs, the zone for this coefficient is constructed using knowledge of these factors with less than 400 zones for all of China. Similarly, less than 300 zones were selected for China for the cultivated land dynamic degree, with the zones mainly controlled by the urbanization and government agriculture policies. These two zone distributions were used to develop the indirect sampling models to estimate the results from the sample units to the report units through the zones.

5.3. *Philosophy*

Multiresolution techniques are drawing more and more attention in monitoring and estimating procedures (Lee 1994, Csillag *et al.* 1996, Yang and Merchant 1997). A joint use of fine and coarse remote sensing images can save money while meeting the required accuracy. For instance, $\hat{S}_p^{\#}=\hat{\beta}_p^{\#}\times\hat{S}_p$, where $\hat{\beta}_p^{\#}$ is an intensity variable, obtained by fine resolution (or accurate) aerial photographic samples and $\hat{S}_p$ is an extension variable, obtained by coarse resolution (or cheaper) TM images. The strategies to estimate $\hat{S}_p^0$ are summarized in table 4, where $\hat{\beta}_p^{\#}(\mathrm{AP^{sp}})$ means the $\hat{\beta}_p^{\#}$ obtained by aerial photograph sample and $S_p(\mathrm{TM^{ec}})$ means the S_p obtained by exhaustively distributed TM images.

5.4. *Total errors*

The estimated total cultivated land area comes from three types of TM pixels (first column in figure 4): (1) pure pixels, formed by net cultivated land or by net non-cultivated land; (2) mixed pixels 1, cultivated land mixed with non-cultivated land; and (3) mixed pixels 2, formed by cultivated land with thin-non-cultivated features inside.

The farmland area based on the three types of TM pixels is denoted by $_1(S_p(\mathrm{TM^{ec}})$, $_2S_p(\mathrm{TM^{ec}})$ and $_3S_p(\mathrm{TM^{ec}})$, which all include real values and errors: $_1S_p^0\pm\Delta_1 S_p$, $_2S_p^0\pm\Delta_2 S_p$ and $_3S_p^0+S_p^{\#}$. The first two $\Delta_1 S_p$ and $\Delta_2 S_p$ most likely occur at the boundary between the cultivated land and the non-cultivated land, therefore they are much smaller than the area S_p. These two types of errors are random errors caused by the uncertainty of the artificial recognition method and under/over predic-

Table 4. Joint use of fine resolution samples and coarse resolution covering data in monitoring.

Joint use of aerial photographs and TM images to estimate S_p^0		Aerial photograph (a refers to an aerial photograph)		
		Exhaustic covering to estimate $S_a^{\#}$, S_a^0, $\beta_a^{\#}$	Used as sample to estimate $\beta_a^{\#}$ $(=S_a^{\#}/S_a)$	
		Updated every 10 years	Updated every 5 years	None
		TM image ('a' refers to a TM image)		
None		Exhaustive determination of $\beta_p^{\#}$ ($\mathrm{AP^{ec}}$) used as steering value for next 10 years	Every 5 years updated sampling $\beta_p^{\#}$ ($\mathrm{AP^{sp}}$) is used for next 5 years	
Used as sample to estimate $\beta_a^{\#\#}=S_a^{\#\#}/A$	Updated every year		Every year, $S_p^{\#\#}=P\times\hat{\beta}_p^{\#\#}(\mathrm{TM^{sp}})$ $S_p=P-S_p^{\#\#}(\mathrm{TM^{sp}})$ $S_p^0=(1-\hat{\beta}_p^{\#}(\mathrm{AP^{sp}}))\times S_p(\mathrm{TM^{sp}})$	Yearly updated sampling $\beta_p^{\#\#}(\mathrm{TM^{sp}})$ is used for the current year
Exhaustive coverage to estimate S_a, $S_a^{\#\#}$, S_a^0, $\beta_a^{\#\#}$	Updated every 5 years		Every 5 years, $S_p^{\#}=S_p(\mathrm{TM^{ec}})\times\hat{\beta}_p^{\#}(\mathrm{AP^{sp}})$ $S_p^0=(1-\hat{\beta}_p^{\#}(\mathrm{AP^{sp}}))\times S_p(\mathrm{TM^{ec}})$	Exhaustive determination of $\beta_p^{\#\#}(\mathrm{TM^{ec}})$ is used as steering value for next 5 years

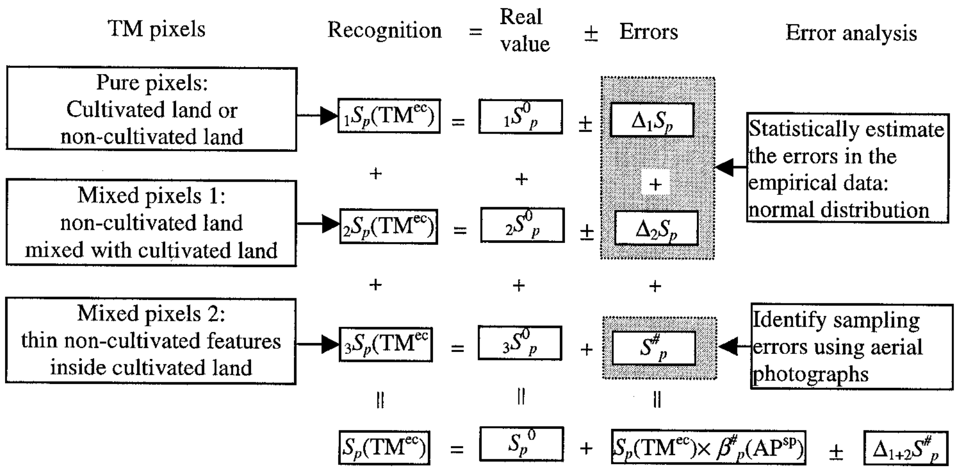

Figure 4. Error propagation and solutions in estimating the cultivated land using the sampling technique.

tion at the boundary. There is much research and experience on this issue (Haack *et al.* 1987, Russel 1991, Fisher 1994, Janssen and van Der 1994, Jean-Fromcois and Sabel 1996, Curtis and Gopal 2000). The errors usually consist of a series of small real numbers with random positive and negative signs that are normally distributed around a zero mean. In our work, the recognition procedure caused about 5–15% error compared with real values obtained by a careful field examination (Liu 1996, pp. 39–40). The last error $S_p^{\#}$ reduces S_p inside the cultivated land, so it is a systematic error and which can be identified by the sampling technique in this paper.

The total percentage error for $\hat{\beta}_p$ and $\hat{S}_p$ with this technique can be estimated by integrating the artificial recognition error of the samples into the sampling estimate variance $\hat{\sigma}_{\hat{\beta}_p}$ and $\hat{\sigma}_{\hat{S}_p}$. $\hat{\sigma}_{\hat{\beta}_p}$ and $\hat{\sigma}_{\hat{S}_p}$ are functions of the sampling scheme (e.g. random sampling in space in this paper) and the number of samples, shown by the theoretical analyses in appendixes 1, 2 and 3 and the illustrations in figures 2 and 3. If the artificial sample recognition error for β_a is assumed to be normally distributed, the extra error added to $\hat{\sigma}_{\hat{\beta}_p}$ and $\hat{\sigma}_{\hat{S}_p}$ caused by the artificial sample recognition error for β_a would be reduced after the errors are filtered through the equations of the forms in (A 1–5) and (A 1–9).

S_p^0 is estimated using $\hat{S}_p^0 = (1-\hat{\beta}_p^{\#})S_p$

$$\begin{aligned}
\hat{S}_p^0 &= (1-\hat{\beta}_p^{\#})S_p = {}_1S_p + {}_2S_p + {}_3S_p - \frac{S_p^{\#}}{S_p} S_p \\
&= {}_1S_p^0 + {}_2S_p^0 \pm \Delta_1 S_p \pm \Delta_2 S_p + {}_3S_p - ({}_3S_p - {}_3S_p^0) \\
&= {}_1S_p^0 + {}_2S_p^0 + {}_3S_p^0 \pm \Delta_1 S_p \pm \Delta_2 S_p \\
&= S_p^0 \pm \Delta_1 S_p \pm \Delta_2 S_p
\end{aligned}$$

Therefore, $E(\hat{S}_p^0) = S_p^0$.

The total error of $\hat{S}_p^0$ is $(\pm\Delta_1 S_p \pm \Delta_2 S_p)$, a series of small real values with random positive and negative signs that are normally distributed around a zero mean theoretically and which have a 5–15% spread in practice, as discussed above.

5.5. *Further work*

The increasing interest in uncertainty (Goodchild and Gopal 1989, Altman 1994, Journel 1996, Atkinson 1999, Nishii 1999) had led to uncertainty models which could be embedded into the spatial sampling models.

There are three quantity levels in survey reporting according to whether the attribute and location properties are included in the final report. The lowest level estimate is reporting attributes without location, or reporting the total amount of cultivated land for the whole country with no specific spatial units. The attributes could be estimated using an Input and Output table model from ecnomics or similar models. The middle level estimate is reporting attributes linked with spatial report units, as in this paper. The highest level estimate is reporting attributes with spatial distribution maps. The sampling model to produce this report for different types of objects needs further research (Csillag *et al.* 1996).

A first solution for sampling and monitoring the cultivated land dynamics is proposed in this paper and a much better model is under development.

Besides the direct and indirect spatially random sampling models for discrete distributed objects, as presented in this paper, stratified and systematic sampling models and models based on both statistical and spatial process mechanisms of the monitored objects are under consideration.

Acknowledgments

This research was supported by grant KZ95T-03-04 and KZ951-B1-302 from the Chinese Academy of Sciences and grant 69896250 from the Natural Science Foundation of China. The authors thank the anonymous reviewers for their comments.

Appendix 1. Static direct sampling model: directly estimates from sample unit *a* to report unit *p*

Given: N aerial photographs as samples of $\beta_a^{\#}$;
Report unit map: province and county boundary;
Assume: various n_{pa}
Objective: estimate #, $\{\hat{\beta}_p^{\#}, \hat{\sigma}_{\hat{\beta}_p^{\#}}, \hat{\rho}_{\hat{\beta}_p^{\#}}, \hat{S}_p^{\#}, \hat{\sigma}_{\hat{S}_p^{\#}}, \hat{\rho}_{\hat{S}_p^{\#}}\}$
Input: sample a $(N_{pa}, \beta_a^{\#}, S_p, w_{pa}, W_{pa})$, with different n_{pa}
Output: report p $(\hat{\beta}_p^{\#}, \hat{\sigma}_{\hat{\beta}_p^{\#}}, \hat{\rho}_{\hat{\beta}_p^{\#}}, \hat{S}_p^{\#}, \hat{\sigma}_{\hat{S}_p^{\#}}, \hat{\rho}_{\hat{S}_p^{\#}})$

(final result: static net cultivated land area, $\hat{S}_p^0 = S_p - \hat{S}_p^{\#}$)

- To estimate $\hat{\beta}_p^{\#}$,

By definition: $\hat{S}_p^{\#} \equiv \hat{\beta}_p^{\#} \times \hat{S}_p$, where $\hat{S}_p$ is obtained by a thorough TM investigation, therefore $\hat{S}_p = S_p$. According to Cochran (1977, p. 22)

$$E\left(\sum_{a=1}^{n_{pa}} S_a\right) = \frac{n_{pa}}{N_{pa}} \sum_{a=1}^{N_{pa}} S_a \tag{A 1.1}$$

so, use $N_{pa}/n_{pa} \sum_{a=1}^{n_{pa}} S_a$ to estimate $\sum_{a=1}^{N_{pa}} S_a = S_p$, and

$$W_{pa} \equiv \frac{S_a}{\sum_{a=1}^{N_{pa}} S_a} = \frac{n_{pa}}{N_{pa}} \frac{S_a}{\sum_{a=1}^{n_{pa}} S_a} \equiv \frac{n_{pa}}{N_{pa}} w_{pa} \tag{A 1.2}$$

$$\because\ \hat{\beta}_p^{\#} \equiv \frac{\hat{S}_p^{\#}}{\hat{S}_p} \equiv \frac{\text{thin-non-cultivated land area in } p\text{-th report unit}}{\text{gross cultivated land area in } p\text{-th report unit}}$$

$$= \frac{\dfrac{N_{pa}}{n_{pa}} \sum_{a=1}^{n_{pa}} S_a^{\#}}{\dfrac{N_{pa}}{n_{pa}} \sum_{a=1}^{n_{pa}} S_a} = \frac{\sum_{a=1}^{n_{pa}} S_a^{\#}}{\sum_{a=1}^{n_{pa}} S_a}$$

$$\therefore\ \sum_{a=1}^{n_{pa}} S_a^{\#} = \hat{\beta}_p^{\#} \times \sum_{a=1}^{n_{pa}} S_a$$

in addition

$$\because\ \sum_{a=1}^{n_{pa}} S_a^{\#} = \sum_{a=1}^{n_{pa}} (\beta_a^{\#} S_a) = \sum_{a=1}^{n_{pa}} \left(\beta_a^{\#} \frac{S_a}{\sum_{a=1}^{n_{pa}} S_a} \right) \times \sum_{a=1}^{n_{pa}} S_a$$

$$\therefore\ \hat{\beta}_p^{\#}(n_{pa}) = \sum_{a=1}^{n_{pa}} \left(\beta_a^{\#} \frac{S_a}{\sum_{a=1}^{n_{pa}} S_a} \right) = \sum_{a=1}^{n_{pa}} (\beta_a^{\#} w_{pa}) \tag{A 1.3}$$

similarly and by (A 1.2)

$$\beta_p^{\#}(N_{pa}) = \int_{N_{pa}} \beta_a^{\#}(a) \frac{S_a(a)}{\int_{N_{pa}} S_a(a)\,\mathrm{d}a} \mathrm{d}a = \int_{N_{pa}} \beta_a^{\#}(a) W_{pa}(a)\,\mathrm{d}a$$

$$= \frac{1}{N_{pa}} \int_{N_{pa}} [n_{pa} \beta_a^{\#}(a) w_{pa}(a)]\ \mathrm{d}a$$

- To estimate $\sigma^2_{\beta_p^{\#}}$

By Rodriguez-Iturbe and Mejia (1974) and Ripley (1981, pp. 19–23)

$$\hat{\sigma}^2_{\hat{\beta}_p^{\#}}(n_{pa}) \equiv E_p[\hat{\beta}_p^{\#} - \beta_p^{\#}]^2$$

$$\equiv E\left[\frac{1}{n_{pa}} \sum_{a=1}^{n_{pa}} (n_{pa} \beta_a^{\#} w_{pa}) - \frac{1}{N_{pa}} \int_{N_{pa}} (n_{pa} \beta_a^{\#} w_{pa})\,\mathrm{d}a \right]^2$$

$$= \frac{1}{n_{pa}} \{1 - E_p[r(a-a')]\} \hat{\sigma}^2_{\hat{\beta}^{\#}_{pa}(N_{pa})} = F(n_{pa}) \hat{\sigma}^2_{\hat{\beta}^{\#}_{pa}(N_{pa})}$$

where $F(n_{pa}) = (1/n_{pa})\{1 - E_p[r(a-a')]\}$

$$E_p[r(a-a')] = \frac{1}{N_{pa}} \times \frac{\sum_{a=1}^{N_{pa}} \sum_{a'=1}^{N_{pa}} \left[N_{pa} \beta_a^{\#} w_{pa} - \sum_{a=1}^{N_{pa}} (\beta_p^{\#} W_{pa}) \right] \left[N_{pa'} \beta_{a'}^{\#} w_{pa'} - \sum_{a=1}^{N_{pa}} (\beta_p^{\#} W_{pa}) \right]}{\sum_{a=1}^{N_{pa}} \left[N_{pa} \beta_a^{\#} w_{pa} - \sum_{a=1}^{N_{pa}} (\beta_p^{\#} W_{pa}) \right]^2} \tag{A 1.4}$$

$$\hat{\sigma}^2_{\hat{\beta}^{\#}_{pa}(N_{pa})} \equiv \sum_{a=1}^{N_{pa}} \left\{ \left[\beta^{\#}_a N_{pa} W_{pa} - \sum_{a=1}^{N_{pa}} (\beta^{\#}_a W_{pa}) \right]^2 W_{pa} \right\} \tag{A 1.5}$$

or

$$\hat{\sigma}_{\hat{\beta}^{\#}_p}(n_{pa}) = \sqrt{F(n_{pa})}\,\hat{\sigma}_{\hat{\beta}^{\#}_{pa}(N_{pa})} \tag{A 1.6}$$

- To estimate $\hat{\rho}_{\hat{\beta}^{\#}_p}(n_{pa})$

$$\hat{\rho}_{\hat{\beta}^{\#}_p}(n_{pa}) \equiv \hat{\sigma}_{\hat{\beta}^{\#}_p}(n_{pa}) / \beta^{\#}_p(N_{pa}) \tag{A 1.7}$$

- To estimate $S^{\#}_p$

$$\hat{S}^{\#}_p(n_{pa}) \equiv \hat{\beta}^{\#}_p(n_{pa}) \times S_p \quad \text{and} \quad S^{\#}_p(N_{pa}) \equiv \beta^{\#}_p(N_{pa}) \times S_p \tag{A 1.8}$$

- To estimate $\sigma^2_{S^{\#}_p}$

$$\begin{aligned} \hat{\sigma}_{\hat{S}^{\#}_p}(n_{pa}) &\equiv \sqrt{E_p[\hat{S}^{\#}_p(n_{pa}) - S^{\#}_p(N_{pa})]^2} = S_p \times \sqrt{E_p[\hat{\beta}^{\#}_p(n_{pa}) - \beta^{\#}_p(N_{pa})]^2} \\ &= \hat{\sigma}_{\hat{\beta}^{\#}_p}(n_{pa}) \times S_p \end{aligned} \tag{A 1.9}$$

- To estimate $\hat{\rho}_{\hat{S}^{\#}_p}(n_{pa})$

$$\hat{\rho}_{\hat{S}^{\#}_p}(n_{pa}) \equiv \hat{\sigma}_{\hat{S}^{\#}_p}(n_{pa}) / S^{\#}_p(N_{pa}) = [\hat{\sigma}_{\hat{\beta}^{\#}_p}(n_{pa}) \times S_p] / [\beta^{\#}_p(N_{pa}) \times S_p] = \hat{\rho}_{\hat{\beta}^{\#}_p}(n_{pa}) \tag{A 1.10}$$

Appendix 2. Static indirect sampling model: estimate from sample unit *a* to report unit *p* through zone *z*

Given N aerial photographs as samples of $\beta^{\#}_a$;
Zone map of $\beta^{\#}_a$: $\beta^{\#}_a$ is homogeneous in space within each zone, with N_{pz}, N_{za};
Report unit map: province and county boundary;
Assume: various n_{pa}
Objective: estimate #, $\{\hat{\beta}^{\#}_p, \hat{\sigma}_{\hat{\beta}^{\#}_p}, \hat{\rho}_{\hat{\beta}^{\#}_p}, \hat{S}^{\#}_p, \hat{\sigma}_{\hat{S}^{\#}_p}, \hat{\rho}_{\hat{S}^{\#}_p}\}$

A2.1. *Estimate from sample units to zone units:* a *to* z
Input: $(N_{pa}, N_{za}, \beta^{\#}_a, S_z, w_{za}, W_{za})$ and let $n_{za} = (S_z/S_p)n_{pa}$
Output: $z(\hat{\beta}^{\#}_z, \hat{\sigma}_{\hat{\beta}^{\#}_z}, \hat{\rho}_{\hat{\beta}^{\#}_z}, \hat{S}^{\#}_z, \hat{\sigma}_{\hat{S}^{\#}_z}, \hat{\rho}_{\hat{S}^{\#}_z})$, final result: static cultivated land area $\hat{S}^0_z = S_z - \hat{S}^{\#}_z$
Replace p by z in the Appendix 1 static direct sampling model:

- $$\hat{\beta}^{\#}_z(n_{za}) = \sum_{a=1}^{n_{za}} (\beta^{\#}_a w_{za}) \quad \text{and} \quad \beta^{\#}_z(N_{za}) = \frac{1}{N_{za}} \int_{N_{za}} [n_{za}\beta^{\#}_a(a) w_{za}(a)]\, \mathrm{d}a \tag{A 2.1}$$
- $$\hat{\sigma}_{\hat{\beta}^{\#}_z}(n_{za}) = \sqrt{F(n_{za})}\,\hat{\sigma}_{\hat{\beta}^{\#}_z(N_{za})} \tag{A 2.2}$$
- $$\hat{\rho}_{\hat{\beta}^{\#}_z}(n_{za}) \equiv \hat{\sigma}_{\hat{\beta}^{\#}_z}(n_{za}) / \beta^{\#}_z(N_{za}) \tag{A 2.3}$$

where $F(n_{za}) = (1/n_{za})\{1 - E_z[r(a-a')\}$

$$E_z[r(a-a')] = \frac{1}{N_{za}} \times \frac{\displaystyle\sum_{a=1}^{N_{za}} \sum_{a'=1}^{N_{za}} \left[N_{za}\beta^{\#}_a w_{za} - \sum_{a=1}^{N_{za}} (\beta^{\#}_z W_{za}) \right] \left[N_{za'}\beta^{\#}_{a'} w_{za'} - \sum_{a=1}^{N_{pa}} (\beta^{\#}_z W_{za}) \right]}{\displaystyle\sum_{a=1}^{N_{za}} \left[N_{za}\beta^{\#}_z w_{za} - \sum_{a=1}^{N_{za}} (\beta^{\#}_z W_{za}) \right]^2}$$

$$\hat{\sigma}_{\beta^{\#}_{za}(N_{za})} = \sqrt{\sum_{a=1}^{N_{za}} \left\{ \left[\beta^{\#}_a N_{za} W_{za} - \sum_{a=1}^{N_{za}} (\beta^{\#}_a W_{za}) \right]^2 W_{za} \right\}}$$

- $\hat{S}^{\#}_z(n_{za}) \equiv \hat{\beta}^{\#}_z(n_{za}) \times S_z$ and $\hat{S}^{\#}_z(N_{za}) \equiv \beta^{\#}_z(N_{za}) \times S_z$ (A 2.4)
- $\hat{\sigma}_{\hat{S}^{\#}_z}(n_{za}) = \hat{\sigma}_{\hat{\beta}^{\#}_z}(n_{za}) \times S_p$ (A2.5)
- $\hat{\rho}_{\hat{S}^{\#}_z}(n_{za}) = \hat{\rho}_{\hat{\beta}^{\#}_z}(n_{za})$ (A 2.6)

A2.2. *Estimate from zone units to report units:* z *to* p

Input: $z(\hat{\beta}^{\#}_z, \hat{\sigma}^2_{\hat{\beta}^{\#}_z}, S^{\#}_z, \hat{\sigma}^2_{\hat{S}^{\#}_z}, N_{pz}, S_p, w_{pz})$ and let $n_{za} = (S_z/S_p)n_{pa}$

Output: $p(\hat{\beta}^{\#}_p, \hat{\sigma}_{\hat{\beta}^{\#}_p}, \hat{\rho}_{\hat{\beta}^{\#}_p}; \hat{S}^{\#}_p, \hat{\sigma}_{\hat{S}^{\#}_p}, \hat{\rho}_{\hat{S}^{\#}_p})$

Let

$$n_{pa} = n_{z=1,a} + n_{z=2,a} + \dots + n_{z,a} + \dots + n_{z=n_{pz},a} = \sum_{z=1}^{N_{pz}} n_{z,a}$$

where n_{pa} = number of samples a within p, n_{za} = number of samples a within z, $n_{pz} = N_{pz}$ = number of z in p. Similarly, let

$$N_{pa} = N_{z=1,a} + N_{z=2,a} + \dots + N_{z,a} + \dots + N_{z=N_{pz},a} = \sum_{z=1}^{N_{pz}} N_{za}$$

where N_{pa} = population number of sample a in p, N_{za} = population number of sample a in z, $N_{pz} = n_{pz}$ = number of z in p.

- To estimate $\hat{\beta}^{\#}_p$

$$\begin{aligned}
\hat{\beta}^{\#}_p(n_{pa}) &= \sum_{a=1}^{n_{pa}} [w_{pa}\hat{\beta}^{\#}_a] = \sum_{a=1}^{n_{z=1,a}} \left(\frac{S_a}{S_p}\frac{S_{z=1}}{S_{z=1}} \beta^{\#}_{z=1,a} \right) + \sum_{a=1}^{n_{z=2,a}} \left(\frac{S_a}{S_p}\frac{S_{z=2}}{S_{z=2}} \beta^{\#}_{z=2,a} \right) \\
&\quad + \dots + \sum_{a=1}^{n_{za}} \left(\frac{S_a}{S_p}\frac{S_z}{S_z} \beta^{\#}_{za} \right) + \dots + \sum_{a=1}^{n_{pz,a}} \left(\frac{S_a}{S_p}\frac{S_{pz}}{S_{pz}} \beta^{\#}_{pz,a} \right) \\
&= w_{p,z=1} \sum_{a=1}^{n_{z=1,a}} (w_{z=1,a}\beta^{\#}_{z=1,a}) + w_{p,z=2} \sum_{a=1}^{n_{z=2,a}} (w_{z=2,a}\beta^{\#}_{z=2,a}) \\
&\quad + \dots + w_{p,z} \sum_{a=1}^{n_{z,a}} (w_{z,a}\beta^{\#}_{z,a}) + \dots + w_{p,pz} \sum_{a=1}^{n_{pz,a}} (w_{pz,a}\beta^{\#}_{pz,a}) \\
&= w_{p,z=1}\hat{\beta}^{\#}_{z=1} + w_{p,z=2}\hat{\beta}^{\#}_{z=2} + \dots + w_{p,z}\hat{\beta}^{\#}_z + \dots + w_{p,pz}\hat{\beta}^{\#}_{pz} \\
&= \sum_{z=1}^{n_{pz}} (w_{pz}\hat{\beta}^{\#}_z), \quad \text{where } w_{pz} = \frac{S_z}{S_p} = \frac{S_z}{\sum_{z=1}^{N_{pz}} S_z}
\end{aligned} \qquad \text{(A 2.7)}$$

- To estimate $\hat{\rho}^2_{\hat{\beta}^{\#}_p}$

$$\begin{aligned}
\because\ \hat{\rho}^2_{\hat{\beta}^{\#}_{pa}(N_{pa})} &= \sum_{a=1}^{N_{pa}} \left\{ \left[\beta^{\#}_a N_{pa} W_{pa} - \sum_{a=1}^{N_{pa}} (\beta^{\#}_a W_{pa}) \right]^2 W_{pa} \right\} \\
&= \sum_{z=1}^{N_{pz}} \sum_{a=1}^{N_{za}} \left\{ \left[\beta^{\#}_a N_{pa} \frac{N_{za}}{N_{za}} \frac{S_a}{S_p} \frac{S_z}{S_z} - \sum_{z=1}^{N_{pz}} \sum_{a=1}^{N_{za}} \left(\beta^{\#}_a \frac{S_a}{S_p} \frac{S_z}{S_z} \right) \right]^2 \frac{S_a}{S_p} \frac{S_z}{S_z} \right\} \\
&= \sum_{z=1}^{N_{pz}} \sum_{a=1}^{N_{za}} \left\{ \left[\beta^{\#}_a N_{za} \frac{N_{pa}}{N_{za}} \frac{S_z}{S_p} \frac{S_a}{S_z} - \sum_{z=1}^{N_{pz}} \sum_{a=1}^{N_{za}} \left(\beta^{\#}_a \frac{S_a}{S_z} \frac{S_z}{S_p} \right) \right]^2 \frac{S_a}{S_z} \frac{S_z}{S_p} \right\}
\end{aligned}$$

$$= \sum_{z=1}^{N_{pz}} \sum_{a=1}^{N_{za}} \left\{ \left[\beta_a^{\#} N_{za} W_{za} - \sum_{z=1}^{N_{pz}} w_{pz} \sum_{a=1}^{N_{za}} (\beta_a^{\#} W_{za}) \right]^2 W_{za} W_{pz} \right\}$$

$$= \sum_{z=1}^{N_{pz}} W_{pz} \sum_{a=1}^{N_{za}} \left\{ \left[\beta_a^{\#} N_{za} W_{za} - \sum_{z=1}^{N_{pz}} (\beta_a^{\#} W_{za}) \right]^2 W_{za} \right\}$$

$$= \sum_{z=1}^{N_{pz}} W_{pz} \sigma F^2_{\beta_{za}^{\#}(N_{za})} \tag{A 2.8}$$

Using the same approach with the assumption that the data is spatially stationary (Ripley 1981, p. 9), we can prove

$$E_p[r(a-a') = E_z[r(a-a')]$$

Then

$$\hat{\sigma}^2_{\beta_p^{\#}}(n_{pa}) \equiv E_p[\hat{\beta}_p^{\#} - \beta_p^{\#}]^2 = \frac{1}{n_{pa}} \{1 - E_p[r(a-a')]\} \hat{\sigma}^1_{\beta_{pa}^{\#}(N_{pa})}$$

$$= \sum_{z=1}^{N_{pz}} \frac{1}{n_{pa}} \{1 - E_z[r(a-a')]\} W_{pz} \hat{\sigma}^2_{\beta_{za}^{\#}(N_{za})}$$

$$= \sum_{z=1}^{N_{pz}} \frac{n_{za}}{n_{pa}} \frac{1}{n_{za}} \{1 - E_z[r(a-a')]\} W_{pz} \hat{\sigma}^2_{\beta_{za}^{\#}(N_{za})}$$

$$= \sum_{z=1}^{N_{pz}} W^2_{pz} \frac{1}{n_{za}} \{1 - E_z[r(a-a')]\} \hat{\sigma}^2_{\beta_{za}^{\#}(N_{za})}$$

$$= \sum_{z=1}^{N_{pz}} W^2_{pz} \hat{\sigma}^2_{\beta_{za}^{\#}(n_{za})}$$

or

$$\hat{\sigma}_{\hat{\beta}_p^{\#}}(n_{pa}) = \sqrt{\sum_{z=1}^{N_{pz}} [W_{pz} \hat{\sigma}_{\hat{\beta}_z^{\#}(n_{za})}]^2} \tag{A 2.9}$$

- To estimate $\hat{\rho}_{\hat{\beta}_p^{\#}}(n_{pa})$, referring to (A 1.7), (A 2.9) and (A 2.7)

$$\hat{\rho}_{\hat{\beta}_p^{\#}}(n_{pa}) \equiv \hat{\sigma}_{\hat{\beta}_p^{\#}}(n_{pa}) / \beta_p^{\#}(N_{pa})$$

$$= \sqrt{\sum_{z=1}^{N_{pz}} [W_{pz} \hat{\sigma}_{\hat{\beta}_z^{\#}(n_{za})}]^2} \bigg/ \sum_{z=1}^{N_{pz}} (W_{pz} \beta_z^{\#}) \tag{A 2.10}$$

- To estimate $\hat{S}_p^{\#}$, referring to (A 1.8) and (A 2.7)

$$\hat{S}_p^{\#}(n_{pa}) = S_p \times \hat{\beta}_p^{\#}(n_{pa}) = S_p \times \sum_{z=1}^{n_{pz}} (w_{pz} \hat{\beta}_z^{\#}) \tag{A 2.11}$$

- To estimate $\hat{\sigma}^2_{\hat{S}_p^{\#}}$, referring to (A 1.9) and (A 2.9)

$$\hat{\sigma}_{S_p^{\#}}(n_{pa}) = S_p \times \hat{\sigma}_{\hat{\beta}_p^{\#}}(n_{pa}) = S_p \times \sqrt{\sum_{z=1}^{N_{pz}} W^2_{pz} \hat{\sigma}^2_{\hat{\beta}_z^{\#}(n_{za})}} \tag{A 2.12}$$

- To estimate: $\hat{\rho}^{1}_{S^{\#}_{p}}(n_{pa})$, referring to (A 1.10) and (A 2.10)

$$\hat{\rho}_{\hat{S}^{\#}_{p}}(n_{pa}) \equiv \hat{\sigma}_{\hat{S}^{\#}_{p}}(n_{pa}) / S^{\#}_{p}(N_{pa}) = \hat{\rho}_{\hat{\beta}^{\#}_{p}}(n_{pa})$$
$$= \sqrt{\sum_{z=1}^{N_{pz}} [W_{pz}\hat{\sigma}_{\hat{\beta}^{\#}_{z}(n_{za})}]^{2}} \Big/ \sum_{z=1}^{N_{pz}} (W_{pz}\beta^{\#}_{z}) \qquad \text{(A 2.13)}$$

Appendix 3. Dynamic sampling model

The dynamic model is used for sampling to monitor the yearly change of the cultivated land area. As in appendix 1 and 2, both direct and indirect sampling models could be developed for the dynamic monitoring.

An initial solution to the problem is proposed which reflects the limitations of the available information. A much more sophisticated model is under development.

The following two assumptions are made in this paper: (1) The cultivated land surrounding large cities dramatically changes with time so it is inappropriate to use sampling techniques, instead, such areas should be exhaustively monitored by remote sensing images, so they are not considered in this paper. (2) Cultivated land change in past years reflected the cultivated land dynamics in the region, while the variation of $\beta^{\#\#}_{a}$ in the past represented its dynamics for the next several years and can be updated by a sampling survey. The later assumption is enhanced by the Chinese government land policy that the amount of cultivated land should be kept in dynamic equilibrium, that is, newly occupied cultivated land should be replaced later somewhere else.

A3.1. *First solution*

Given N TM images as samples of $\beta^{\#\#}_{a}(t_2)$, where t_2 indicates data for year t_2;
Zone map of $\beta^{\#\#}_{a}(t_2)$: $\beta^{\#\#}_{a}(t)$ is homogeneous in space within each zone, with $N^{\#\#}_{pz}$, $N^{\#\#}_{za}$, and the value $\Delta\beta^{\#\#}_{a}(t_2 - t_1) = |\beta^{\#\#}_{a}(t_1)|$ attached to each zone;
Report unit map: province and county boundary.

Assume: various n_{pa}

Objectives: estimate ##, $\{\hat{\beta}^{\#\#}_{p}(t_3)\hat{\sigma}_{\hat{\beta}^{\#\#}_{p}}(t_3), \hat{\rho}_{\hat{\beta}^{\#\#}_{p}}(t_3), SF^{\#\#}_{p}(t_3), \hat{\sigma}_{\hat{S}^{\#\#}_{p}}(t_3), \hat{\rho}_{\hat{S}^{\#\#}_{p}}(t_3)\}$

With two assumptions:

- Do not consider the cultivated land surrounding larger cities which should be monitored by exhaustive TM images;
- $|\Delta\beta^{\#\#}_{a}(t_2 - t_1)|$ reflects the spatial–temporal dynamic degree of the cultivated land in a region.

With these above two assumptions, the first solutions of the dynamic models are all the same as the static models in appendixes 1 and 2 except for the following substitutions of variable meanings.

1. 'a' represents the TM images in the dynamic model rather than the aerial photographs in the static model.
2. Replace # by ##, the non-cultivated land subtracted from a TM image.
3. For calculating the standard error σ and the relative error ρ of proportion $\beta^{\#\#}_{p}$ and the $S^{\#\#}_{p}$, replacing $\beta^{\#\#}_{a}(t)$ by $\beta^{\#\#}_{a}(t) + |\Delta\beta^{\#\#}_{a}(t_2 - t_1)|$, where $|\Delta\beta^{\#\#}_{a}(t_2 - t_1)|$, called the cultivated land dynamic degree, is the absolute value of the difference between the proportion of cultivated land area in year t_2 and that in year t_1. TM images covering the whole territory for two separate years are used to form the values in different regions.

4. For calculating $\beta_p^{\#\#}$ and $S_p^{\#\#}$, $\beta_a^{\#\#}(t)$ is used directly (rather than $\beta_a^{\#\#}(t)+|\Delta\beta_a^{\#\#}(t_2-t_1)|$).

With the above substitutions, the direct model in appendix 1 and the indirect model in appendix 2 for static sampling design can be used for dynamic sampling design.

A3.2. *Advanced solution (in development)*

Besides directly determining the dynamic degree of the cultivated land by the difference between two years' TM images, auto-regression or regression on the controlling factors by multitemporal imagery of the cultivated land can better reflect the changes (Henry and Hope 1998, Tokola *et al.* 1999). In the dynamic model, σ and ρ based on $[\beta_a^{\#\#}(t)+|\Delta\beta_a^{\#\#}(t_2-t_1)|]$ tend to be, of course, larger than that based on $\beta_a^{\#\#}(t)$. A higher dynamic degree $|\Delta\beta_a^{\#\#}(t_2-t_1)|$ leads to higher estimate variances $\hat{\sigma}^2_{S_p^{\#\#}}$ and $\hat{\sigma}^2_{S_p^{\#\#}}$. Therefore, given a desired estimate variance, extra Δn_{pa} are needed to eliminate the effect of the dynamic degree $|\Delta\beta_a^{\#\#}(t_2-t_1)|$ on the estimate variance. Although it is reasonable to take into account the dynamic degree of the cultivated land when calculating the estimate variances, much better solutions to β_p and S_p with smaller σ and ρ are expected when more samples are put into the more rapidly changing areas than into the less rapidly changing areas, given the total number of samples. The sampling efficiency is expected to improve if the sampling model is based on several factors driving the cultivated land change (Skinner *et al.* 1994). Theoretical formulas for the optimal dynamic spatial sampling, which directly includes the dynamic degree of the cultivated land into the whole estimation procedure rather than only into the estimate variance, are being developed.

References

Atkinson, P., 1999, Geographical information science: geostatistics and uncertainty. *Progress in Physical Geography*, **23**, 134–142.

Altman, D., 1994, Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographic Information Systems*, **8**, 271–289.

Barnett, V., 1974, *Elements of Sampling Theory* (London: Edward Arnold).

Bellehumeur, C., and Legendre, P., 1997, Aggregation of sample units: an analytical solution to predict variance. *Geographical Analysis*, **29**, 258–266.

Burgess, T. M., Webster, R. and McBratney, B., 1981, Optimal interpolation and isarithmic mapping of soil properties, IV sampling strategy. *Journal of Soil Science*, **32**, 643–659.

Csillag, F., Kertesz, M., and Kummert, A., 1996, Sampling and mapping of heterogeneous surfaces: multi-resolution tiling adjusted of spatial variability. *International Journal of Geographic Information Systems*, **10**, 851–875.

Cliff, A. D., and Ord, J. K., 1975, The comparison of means when samples consist of spatially autocorrelated observations. *Environment and Planning A*, **7**, 725–734.

Cochran, W. G., 1977, *Sampling Techniques*, 3rd edn (New York: John Wiley & Sons).

Curtis, E., and Gopal, S., 2000, Fuzzy set theory and thematic maps: accuracy assessment and area estimation. *International Journal of Geographic Information Sciences*, **14**, 153–172.

Dunn, R., and Harrison, ?., 1993, Two dimensional systematic sampling of land use. *Applied Statistics*, **42**, 585–601.

Fisher, P., 1994, Visualization of the reliability in classified remotely sensed images. *Photogrammetric Engineering and Remote Sensing*, **60**, 905–915.

Fotheringham, A., Charlton, M., and Brunsdon, C., 1996, The geography of parameter space: an investigation of spatial non-stationary. *International Journal of Geographical Information Systems*, **10**, 605–627.

Ghosh, B., 1951, Random distances within a rectangle and between two rectangles. *Bulletin of the Calcutta Mathematical Society*, **43**, 17–24.

Goodchild, M., and Gopal, S., 1989, *The Accuracy of Spatial Databases* (London: Taylor and Francis).

Griffith, D., 1988, *Advanced Spatial Statistics* (Dordrecht: Kluwer), pp. 94–99.

Griffith, D., Haining, R., and Arbia, G., 1994, Heterogeneity of attribute sampling error in spatial data sets. *Geographical Analysis*, **26**, 300–320.

Haack, B., Nevin, B., and Steven, A., 1987, An assessment of landsat MSS and TM data, for urban and near-urban land-cover digital classification. *Remote Sensing of Environment*, **21**, 201–213.

Haining, R., 1988, Estimating spatial means with an application to remotely sensed data. *Communications in Statistics, Theory and Methods*, **17**, 573–579.

Haining, R., 1990, *Spatial Data Analysis in the Social and Environmental Sciences* (Cambridge: Cambridge University Press).

Henry, M., and Hope, A., 1998, Monitoring post-burn recovery of chaparral vegetation in southern California using multi-temporal satellite data. *International Journal of Remote Sensing*, **19**, 3097–3107.

Institute of Agriculture Zonation Research, 1995, *The Cultivated Land in China* (Beijing: China Agriculture Science & Technology Press).

Isaaks, E., and Srivastava, R., 1989, *Applied Geostatistics* (Oxford: Oxford University Press), pp. 476–499.

Janssen, L., and van Der, W., 1994, Accuracy assessment of satellite derived land-cover data: a review. *Photogrammetric Engineering and Remote Sensing*, **60**, 419–426.

Jean-Fromcois, M., and Sabel, R., 1996, Comparison of land use classification obtained by visual interpretation and digital processing, *ITC Journal*, **3**, 278–283.

Jones, D. A., Gurney, R. J., and O'Connell, P. E., 1979, Network design using optimal estimation procedures. *Water Resources Research*, **15**, 1801–1812.

Journel, A., and Huijbregts, C. H., 1978, *Mining Geostatistics* (London: Academic Press).

Journel, A., 1996, Modeling uncertainty and spatial dependence: stochastic imaging. *International Journal of Geographic Information Systems*, **10**, 517–522.

Langford, R., and Kapinos, S., 1979, The national water data network: a case history. *Water Resources Research*, **15**, 1687–1691.

Lee, D., 1994, Simulating and mapping spatial complexity using multi-scale techniques. *International Journal of Geographical Information Systems*, **8**, 411–427.

Liu, J., and Yu, Z., 1990, The study and practice on the application of color infrared aerial remote sensing techniques to non-cultivation coefficient calculation in Tibet. *Remote Sensing of Environment* [in Chinese], **5**, 27–38.

Liu J. (ed.), 1996, *Macro-scale Survey and Dynamic Study of Natural Resources and Environment of China by Remote Sensing* (Beijing: China Science & Technology Press).

Morris, H. and DeGroot, 1975, *Probability and Statistics*, 2nd edn (Reading, MA: Addison-Wesley), pp. 226–227.

Nishii, R., 1999, Accuracy and inaccuracy assessments in land-cover classification. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 491–498.

O'Connell, P. E., Gurney, R. J., Jone, A. A., Miller, J. B., Nicholass, C. A., and Senior, M. R., 1979, A case study of rationalisation of a rain gate network in southwest England. *Water Resources Research*, **15**, 1813–1822.

Ripley, B. D., 1981, *Spatial Statistics* (New York: John Wiley & Sons).

Rodda, J. C., and Mandeville, A. N., 1969, Hydrologic network design—needs, problems and approaches. Report 12, World Meteorological Organization, Geneva.

Rodriguez-Iturbe, I., and Mejia, J. M., 1974, The design of rainfall networks in time and space. *Water Resources Research*, **10**, 713–728.

Russell, G., 1991, A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, **37**, 35–46.

Skinner, C., Holmes, D., and Holt, D., 1994, Multiple frame sampling for multivariate stratification. *International Statistical Review*, **62**, 333–347.

Tokola, T., Lofman, S., and Erkkila, A., 1999, Relative calibration of multitemporal landsat data for forest cover change detection. *Remote Sensing of Environment*, **68**, 1–11.

Wang, J., Haining, R., and Wise, S., 1997, An integrated regionalization of earthquakes, floods and drought hazards in China. *Transactions in GIS*, **2**, 25–44.

Yang, W., and Merchant, J., 1997, Impact of upscaling techniques on land cover representation in Nebraska, U.S.A. *Geocarto International*, **12**, 27–39.