ERDC/CERL TR-06-5

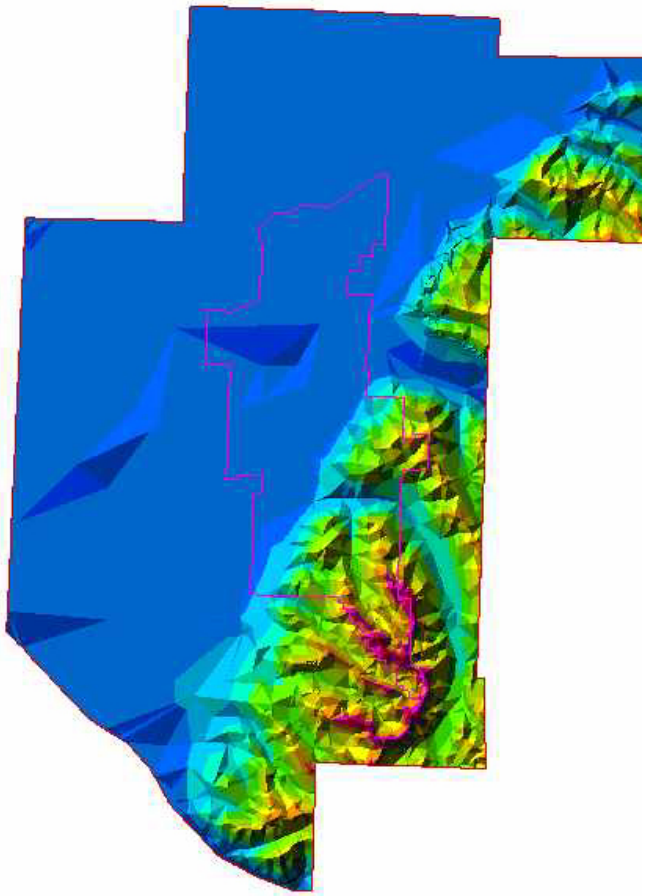
US Army Corps of Engineers®
Engineer Research and Development Center

Sampling and Mapping Soil Erosion Cover Factor for Fort Richardson, Alaska

Integrating Stratification and an Up-Scaling Method

Guangxing Wang, George Gertner,
Alan B. Anderson, and Heidi Howard

March 2006



Construction Engineering Research Laboratory

ERDC/CERL TR-06-5
March 2006

Sampling and Mapping Soil Erosion Cover Factor for Fort Richardson, Alaska

Integrating Stratification and An Up-Scaling Method

Guangxing Wang and George Gertner

Department of Natural Resources and Environmental Sciences
1102 S. Goodwin Avenue
University of Illinois at Urbana-Champaign
Urbana, IL 61801-4730

Alan B. Anderson and Heidi Howard

U.S. Army Engineer Research and Development Center
Construction Engineering Research Laboratory
PO Box 9005, Champaign, IL 61826-9005

Final Report

Prepared for U.S. Army Corps of Engineers
Washington, DC 20314-1000

Under Work Unit #7790DF

ABSTRACT: When a ground and vegetation cover factor related to soil erosion is mapped with the aid of remotely sensed data, a cost-efficient sample design to collect ground data and obtain an accurate map is required. However, the supports used to collect ground data are often smaller than the desirable pixels used for mapping, which leads to complexity in developing procedures for sample design and mapping. For these purposes, a sampling and mapping method was developed by integrating stratification and an up-scaling method in geostatistics — block cokriging with Landsat Thematic Mapper imagery. This method is based on spatial correlation and stratification of variables. It scales up not only the ground sample data but also the uncertainties associated with the data aggregation from smaller supports to larger pixels or blocks. This method uses the advantages of both stratification and block cokriging variance-based sample design, which lead to sample designs with variable grid spacing, and thus significantly increase the unit cost-efficiency of sample data in sampling and mapping. This outcome was verified by the results of this study.

Contents

List of Figures ..... **iv**

Preface ..... **v**

1 Introduction **1**

Background 1
Objective 3
Approach 4
Scope 4
Mode of Technology Transfer 4

2 Study Area and Data Sets **5**

3 Method **8**

Spatial Correlation 8
Linear Model of Coregionalization 9
Block Cokriging for Sampling and Mapping 10

4 Stratified Sample Designs **14**

5 Results **15**

6 Conclusions and Discussion **24**

References **26**

Report Documentation Page **28**

List of Figures

Figure **Page**

Figure 1. Digital elevation model and slope map of study area 6

Figure 2. Locations and cover factor values of sample plots, and ratio image (TM3+TM7)/TM4 that has highest correlation with cover factor 7

Figure 3. Block ordinary cokriging with TM imagery to scale up ground sample data from smaller supports of 1×1 m^2 to a coarser pixel of 30 × 30 m^2 in which the remotely sensed data are available 11

Figure 4. Experimental and modeled variograms of (a) cover factor, (b) TM ratio image (TM3+TM7)/TM4, and (c) cross variogram between the cover factor and ratio image..... 16

Figure 5. Average of cokriging variances versus grid spacing for sampling ground data with three different ratios of ground data to image data..... 16

Figure 6. Stratification of study area into steep and training subareas..... 18

Figure 7. Average of cokriging variances versus grid spacing for sampling ground data with a ratio 1:4 of ground-to-image data when the study area was divided into steep and training subareas..... 19

Figure 8. Block cokriging-based sample designs without stratification and with two strata, respectively, and the original sample design 20

Figure 9. Estimation maps of cover factor using three sample designs: original sample design, block cokriging variance-based sampling without stratification and with two strata (steep and training subareas), and their comparison with the reference map 21

Figure 10. Comparison of three sample designs for mapping the cover factor based on (a) relative root mean square error (RMSE) - (RMSE * 100 / mean) and (b) cost-efficiency per plot defined as (200 - relative RMSE0 / sample size)..... 23

Preface

This study was conducted for the U.S. Army Corp of Engineers under 622720A896, "Base Facility Environmental Quality," project "Land Based Carrying Capacity," 7790DF (P2-120198). The technical reviewer was Dr. William D. Severinghaus, CEERD-CVT.

The work was jointly performed by the Ecological Processes Branch (CN-N) of the Installations Division (CN), Construction Engineering Research Laboratory (CERL) and the University of Illinois (UI). The CERL Principal Investigator was Heidi Howard. The UI Principal Investigators were Dr. Guangxing Wang and Dr. George Gertner. Alan B. Anderson is Chief, CN-N, and Dr. John T. Bandy is Chief, CN. The associated Technical Director is Dr. William D. Severinghaus. The Acting Director of CERL is Dr. Ilker R. Adiguzel.

CERL is an element of the U.S. Army Engineer Research and Development Center (ERDC), U.S. Army Corps of Engineers. The Commander and Executive Director of ERDC is COL James R. Rowan, and the Director of ERDC is Dr. James R. Houston.

1 Introduction

Background

Soil erosion in the United States is usually predicted using a Revised Universal Soil Loss Equation (Renard et al. 1997). In the equation, soil loss is a function of six factors: rainfall-runoff erosivity, soil erodibility, topographic features (steepness and slope length), ground and vegetation cover, and support practice. For a specific area, the cover factor reflects the effect of ground and vegetation covers on the reduction of soil loss by reducing runoff and determines the dynamics of soil erosion (Benkobi et al. 1994). This factor is defined as a function of ground cover, canopy cover, and average minimum height for rain dripping down to the ground from the lowest canopy cover layer (Wischmeier and Smith 1978). At Fort Richardson, Alaska, various military training activities (e.g., off-road vehicular traffic) disturb ground and vegetation covers, and cause an increase of rainfall-related runoff and, thus, soil erosion. Therefore, sampling and mapping this cover factor in a cost-efficient way becomes critical to monitoring the dynamics of soil erosion.

Ground cover, vegetation cover, and soil erosion vary spatially, and their dynamics are a complex process. First, military training activities often take place unevenly in space and cause spatially variable disturbances of ground and vegetation covers. Furthermore, the high frequency of training activities in the same areas disturb ground and vegetation covers cumulatively and result in an increase in soil erosion over time. In other areas without such activities, soil erosion decreases due to recovery of the ground and vegetation covers. For these reasons, in order to monitor and map dynamics of soil erosion, ground data should be collected more in areas with high activities than in areas with low activities (Demers 2000). Therefore, the need is strong to develop a cost-efficient sample design method. This method should lead to a variable sampling distance between plots; that is, grid spacing that varies depending on the spatial variation of soil erosion.

On the other hand, the supports used to collect ground data are often smaller than the desirable units of mapping or pixels of images used because budget and effort are limited. For example, squares of 1 m x 1 m or transect lines of 30 to 100 m are usually used for collection of ground data for ground and vegetation covers, while the pixel size of Landsat Thematic Mapper (TM) imagery used for mapping is 30 m x 30 m. Another example is that global land use and land cover is often mapped

using Moderate Resolution Imaging Spectroradiometer or Advanced Very High Resolution Radiometer at the spatial resolution of 1 km x 1 km, while ground data are generally collected at sample plots that are smaller than 50 m x 50 m. The inconsistency of support sizes leads to difficulty and complexity in developing the sample design and mapping procedure. For these reasons, a sampling and mapping method was developed in this study.

"Sample design" deals with determining plot and sample size, plot shape, and spatial allocation of plots to collect ground data given a cost and desired precision of estimates. This study focused on determining a cost-efficient sample size given a plot size and map unit size for mapping the cover factor. The cost-efficient sample size is defined as the number of plots that minimizes cost given a desired precision of an estimate. Generally, a sample design based on random selection of data locations often leads to a larger sample and duplication of information because some plots are too close together. A sample design based on stratified random selection of data locations can increase cost-efficiency by subdividing the total military base into homogeneous strata such that variances are minimized within strata (Thompson 1992). In the traditional sample design and mapping methods, variables are assumed to have no spatial correlation and also the models to collect ground data obtained at smaller supports are often used directly to estimate the values of the variables at coarser spatial resolutions. The methods neglect the spatial correlation of variables and change across spatial resolutions.

Geostatistical methods such as cokriging are based on the regionalized variable theory that values of a random function are similar when they are close to each other, and the similarity decreases as the distance between data locations increases (Curran and Atkinson 1998). This characteristic is also called spatial autocorrelation or spatial variability of a random function. Curran (1988) suggested the improvement in determining a sample size using minimized cokriging variance-based sample design. Atkinson et al. (1992, 1994, 2000) and McBratney and Webster (1983) demonstrated this method and its applications.

Sample designs for collecting ground data should be conducted to reduce the cost of mapping the ground and vegetation cover factor given a precision or to increase the accuracy of maps given a budget. Remote sensing imagery provides complete coverage of the ground surface. Because of significant correlation of this cover factor with spectral variables, remotely sensed data can be used to improve the estimation of unknown locations from the sample data (Wang et al. 2002). The widely used approaches include classification methods such as supervised and unsupervised classification or stratification, and mapping methods such as regression estimation, cokriging, and co-simulation in geostatistics (Campbell 1996; Goovaerts 1997). Wang et al. (2002) compared the methods for mapping this cover factor and

concluded that cokriging with TM imagery led to higher accuracy than unsupervised stratification and linear regression by strata.

Cokriging with remotely sensed data can be used to map variables of interest and cokriging variance-based sample design can be applied to design sampling strategies prior to mapping (Curran and Atkinson 1998). Cokriging makes use of spatial autocorrelation and cross correlation to explore and describe spatial variability of variables, and further to optimally estimate local values from data sampled elsewhere and provide the minimized error or cokriging variances (Goovaerts 1997). The cokriging variances depend on the spatial configuration of the data locations but not on the data values, and thus can be used to determine the sampling distance — grid spacing for a systematical sample design to collect ground data of a primary variable (Curran and Atkinson 1998).

Because of the limited budget, however, the size of the plots used to collect ground data is usually smaller than the pixels of images used for mapping. The inconsistency of support sizes means that calculated cokriging variances do not really reflect the uncertainties of estimates. An alternative is to use block cokriging to scale up the ground data at smaller supports for coarser spatial resolutions and calculate corresponding block cokriging variances (Deutsch and Journel 1998; Goovaerts 1997).

The block cokriging variance-based sample designs are not locally optimal, however, because the error variances do not depend on the data values and thus this method neglects the heterogeneity of a variable from area to area. The heterogeneity requires the variable grid spacing; that is, more ground measurements need to be collected in the parts of a study area with higher variation than in the parts with lower variation. The potential solution for this requirement may be to first stratify the study area into homogeneous strata and, within each stratum, the spatial variability of variables is then modeled and used for sample designs. This method will make use of the advantages of both stratification and minimized block cokriging variance for sample designs. Integrating stratification and block cokriging variance-based sampling design will thus increase the cost-efficiency for sampling and mapping.

Objective

The objective of this study was to develop a cost-efficient method to sample and map the cover factor when the size of supports used to collect ground data is smaller than the spatial resolution of map units. This method should lead to a sample

design with variable grid spacing that varies depending on spatial variation of the cover factor.

Approach

The method was developed by integrating stratification and block ordinary co-kriging with TM imagery. The TM images help increase the cost-efficiency of sampling and mapping the variable. The methods were then applied to Fort Richardson, where the cover factor was sampled and mapped for monitoring the dynamics of soil erosion due to military training activities.

Scope

This report analyzes to U.S. Army Integrated Training Area Management (ITAM) Range and Training Land Assessment (RTLA) program data from Fort Richardson. The methodology is applicable to all installations involved in the ITAM RTLA program. While different installations use various field methods to collect ground data, the data extrapolation methods described in this report are applicable to all of these installations.

Mode of Technology Transfer

The information in this report will be provided to U.S. Army ITAM personnel responsible for RTLA program implementation.

This report will be made accessible through the World Wide Web (WWW) at URL:

http://www.cecer.army.mil

2 Study Area and Data Sets

The study area consists of 25,184 ha (62,229 acres) located at Fort Richardson in south-central Alaska adjacent to the city of Anchorage and in a transitional zone between the maritime climatic zone to the south and the interior or continental climatic zone to the north (Colorado State University 2004). The area lies in an alluvial plain that is bordered on the east by the Chugach Mountains and on the north, south, and west by waters of Cook Inlet. Figure 1 shows a digital elevation model (DEM) and slope map of this area. Within this area, elevation decreases from southeast to northwest and varies from 16 to 5,329 m, and the range of slope at the spatial resolution of 30 m × 30 m changes from 0 to 82 degrees. Average annual precipitation is 388 mm (15.29 in.) with an average annual snowfall of 1,956 mm (77 in.). The area is dominantly covered by forests (55.3 percent), then scrub lands (23.7 percent), human disturbed lands (13.1 percent), barren lands (5.5 percent), bog and wetland (1.6 percent), meadow (0.7 percent), and water (0.5 percent) (Jorgensen et al. 2002). Lowland black spruce forests are common and bottomland riparian areas typically have spruce/poplar forests along with willow and alder shrubs. Moraines support white spruce forests. Cottonwood/tall bush communities are common on the floodplains. Ground and vegetation covers were disturbed due to various military training activities from 1997 to 2001, making soil erosion a big concern.

For the dynamics of soil loss, different numbers of sample plots were measured from 1997 to 2001, and the locations of the sample plots also varied from year to year, depending on the areas in which military training activities took place. In the southeast of the study area, there were no or few plots measured because of steepness, high cost of sampling, and no activities. The stratified random sample design was used. That is, the study area was first segmented into homogeneous polygons based on vegetation and soil types, and the polygons were randomly drawn. Within each of the selected polygons, several transects were again selected at random. Along each of the selected transects, a certain number of 1×1 m^2 square plots were designed and within each of the square plots, ground cover, canopy cover, and minimum rain drip vegetation heights were recorded. The ground and vegetation cover percentages of each plot were calculated. The soil erosion relevant ground and vegetation cover factor was then derived for each plot using the empirical models described by Wischmeier and Smith (1978).

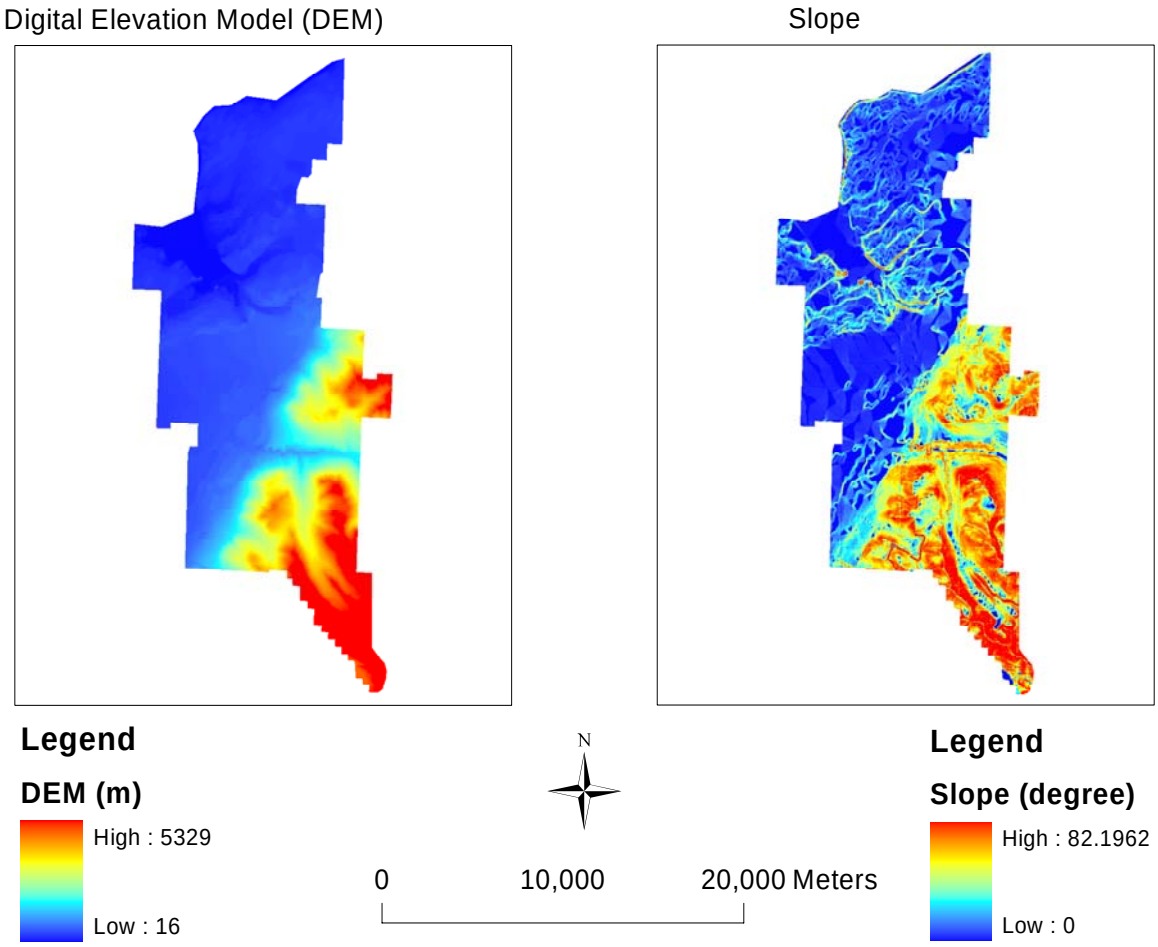


Figure 1. Digital elevation model (DEM) and slope map of study area.

In this study, a ground data set consisting of 774 plots measured in 1999 was used to develop the method described above. The locations and cover factor values of the sample plots are shown in Figure 2, and the values vary from 0.006 to 0.45. In addition, a scene of Landsat 7 ETM+ images at a spatial resolution of 30 m × 30 m was acquired for the year 1999. These images were geo-referenced to Universal Transverse Mercator (UTM) using orthophotographs and affin transformation. Six control points were selected and the root mean square error (RMSE) obtained was 8.98 m.

In addition to original bands, a normal difference vegetation index and other four-ratio images were calculated. The ratio images were TM5/TM4, TM7/TM4, (TM3+TM5)/TM4, and (TM3+TM7)/TM4. The coefficients of correlation between the cover factor and the images were then calculated. Figure 2 shows the ratio image (TM3 + TM7)/TM4. This image had the highest correlation with the cover factor, 0.47166, and it was combined with the set of ground data for sample designs and mapping of the cover factor using block ordinary cokriging.

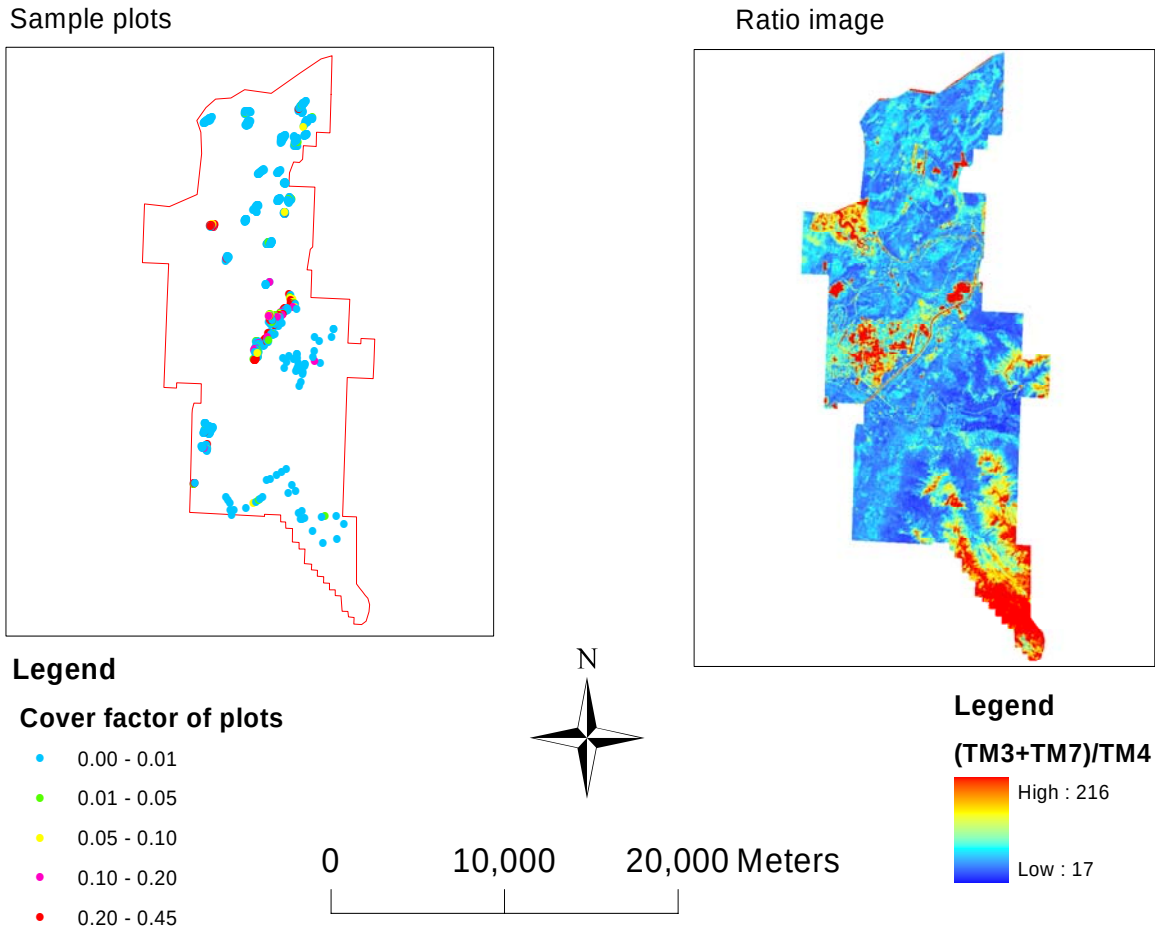


Figure 2. Locations and cover factor values of sample plots, and ratio image (TM3+TM7)/TM4 that has highest correlation with cover factor.

3 Method

Spatial Correlation

Cokriging with TM images for mapping and use of cokriging variance for sample design are based on the theory of regionalized variables (Goovaerts, 1997; Curran and Atkinson 1998). In this study, the cover factor and a spectral variable ratio image are denoted with Z_1 and Z_2, respectively. Suppose that **u** is a location in 2-dimensional space, $Z_1(\mathbf{u})$ and $Z_2(\mathbf{u})$ are random functions. The random functions satisfy the intrinsic hypothesis (Matheron 1971) if, in addition to $E[Z_i(u) - Z_i(u+h)] = 0$ for all **u** and **h** separation vector of data locations, the following condition is met:

$$\gamma_{Z_i Z_i}(h) = 0.5 Var[Z_i(u) - Z_i(u+h)] \quad \text{"}Var\text{"} \quad means \quad \text{var}\,iance \qquad \text{(Eq 1)}$$

$\gamma_{Z_i Z_i}(h)$ exists and is a function of **h** only and not a function of **u**. The expected squared difference $\gamma_{Z_i Z_i}(h)$ that quantifies spatial autocorrelation of the random function is called the variogram. In practice, the variogram is finite for all **u** and **h**, and can be estimated using the experimental variogram:

$$\hat{\gamma}_{Z_i Z_i}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} (z_i(u_\alpha) - z_i(u_\alpha + h))^2 \qquad \text{(Eq 2)}$$

where *N(h)* is the number of all pair-wise Euclidean distances, and $z_i(u_\alpha)$ and $z_i(u_\alpha + h)$ are observations of the variable Z_i at spatial locations u_α and $u_\alpha + h$, respectively. If the random function is not only intrinsically stationary but also second-order stationary (i.e., has both constant mean and variance), then its spatial covariance $C_{Z_i Z_i}(h)$ function exists and has this relationship with the variogram:

$$C_{Z_i Z_i}(h) = Cov\{Z_i(u), Z_i(u+h)\} \quad \text{"Cov" means covariance} \qquad \text{(Eq 3)}$$

$$C_{Z_iZ_i}(h) = C_{Z_iZ_i}(0) - \gamma_{Z_iZ_i}(h) \quad \text{(Eq 4)}$$

Moreover, the cover factor is spatially cross-correlated with the spectral variable, and the spatial cross-correlation can be quantified by a cross-variogram defined as:

$$\gamma_{Z_1Z_2}(h) = 0.5Cov\{Z_1(h) - Z_1(u+h), Z_2(u) - Z_2(u+h)\} \quad \text{(Eq 5)}$$

It is symmetric if $\gamma_{Z_1Z_2}(h) = \gamma_{Z_1Z_2}(-h)$ and $\gamma_{Z_1Z_2}(h) = \gamma_{Z_2Z_1}(h)$. The cross-variogram can be estimated using the experimental cross-variogram:

$$\hat{\gamma}_{Z_1Z_2}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} (z_1(u_\alpha) - z_1(u_\alpha + h))(z_2(u_\alpha) - z_2(u_\alpha + h)) \quad \text{(Eq 6)}$$

Under the second-order stationary assumption, there will also be a cross-covariance function $C_{Z_1Z_2}(h)$, and its relationship with the cross variogram $\gamma_{Z_1Z_2}(h)$ when $C_{Z_1Z_2}(h) \neq C_{Z_2Z_1}(h)$ is:

$$C_{Z_1Z_2}(h) = Cov\{Z_1(u), Z_2(u+h)\} \quad \text{(Eq 7)}$$

$$\gamma_{Z_1Z_2}(h) = 0.5[C_{Z_1Z_2}(0) + C_{Z_2Z_1}(0)] - 0.5[C_{Z_1Z_2}(h) + C_{Z_2Z_1}(h)] \quad \text{(Eq 8)}$$

Linear Model of Coregionalization

For cokriging, a matrix variogram in which the diagonal entries are variograms and the off-diagonal entries are cross-variograms must be strictly conditionally negative definite. Meeting this requirement is not easy for a specific function. In practice, a limited list of known valid models (i.e., nugget effect, spherical, exponential, and Gaussian models), together with positive linear combinations of the models (i.e., nested models), is used (Goovaerts 1997). This practice does not work in the multivariate case (i.e., cokriging). A compromise is to use the linear coregionalization model (LCM). If $Y(u) = [Y_1(u), ..., Y_L(u)]$ is a vector-valued random function from L orthogonal random functions, then an LCM would be of the form (Goovaerts 1997):

$$\gamma(\mathbf{h}) = \sum_{l=1}^{L} b_l \gamma_l(\mathbf{h}) \quad \text{(Eq 9)}$$

where each b_l is an L×L positive definite matrix, and each function $\gamma_l(\mathbf{h})$ is a valid (scalar) variogram. If at least one of the b_l's is strictly positive definite, and all the $\gamma_l(\mathbf{h})$'s are strictly conditionally negative definite, then $\gamma(\mathbf{h})$ is a strictly conditionally negative definite matrix function. Thus, the LCM is a generalization of the nested model in the scalar case.

Mathematically, the entries $b_l(i, j)$ of each matrix b_l correspond to the sill or slope values of a positive definite model $\gamma_l(\mathbf{h})$. For each l, $b_l(i, j) = b_l(j, i)$. Thus, it is sufficient that $b_l(i, i) \geq 0$ and $b_l(j, j) \geq 0$, and the determinant of the matrix is positive or zero. In practice, a basic component $\left[b_l(i, j)\right]$ involved in the cross variogram must appear in the variograms and the coefficient matrices associated with the coregionalization need to be checked for positive definiteness. For details of the LCM, refer to Goovaerts (1997).

Block Cokriging for Sampling and Mapping

Suppose that a set of ground data obtained for the cover factor (primary variable Z_1 in the study area), and the data of the ratio image (secondary variable Z_2) is available at each location. The ratio image has a spatial resolution of 30 × 30 m², while the ground data set of cover factors was collected at the plots of 1 × 1 m². The cover factor needs to be estimated at all pixels of 30 × 30 m². An alternative for this purpose is to use a block cokriging estimator that can scale up ground data from a finer spatial resolution to a coarse one by combining the ground and image data for estimations of each location (Deutsch and Journel 1998; Goovaerts 1997).

Figure 3 shows the block ordinary cokriging that scale up the ground sample data (marked by "x") from smaller supports of 1 × 1 m² to coarser pixels (solid lines) of 30 × 30 m². Within each pixel the remotely sensed data are available and a pixel can be regarded as a block consisting of 900 1 × 1 m² supports (dash lines). Each of the smaller supports is first estimated, then the estimates are averaged and used as an estimate of the block. In practice, the estimate of the block can be directly calculated by assigning the weight of the block to each location of data at smaller ground supports and coarser image pixels within a given neighborhood such as a radius of 5,000 m.

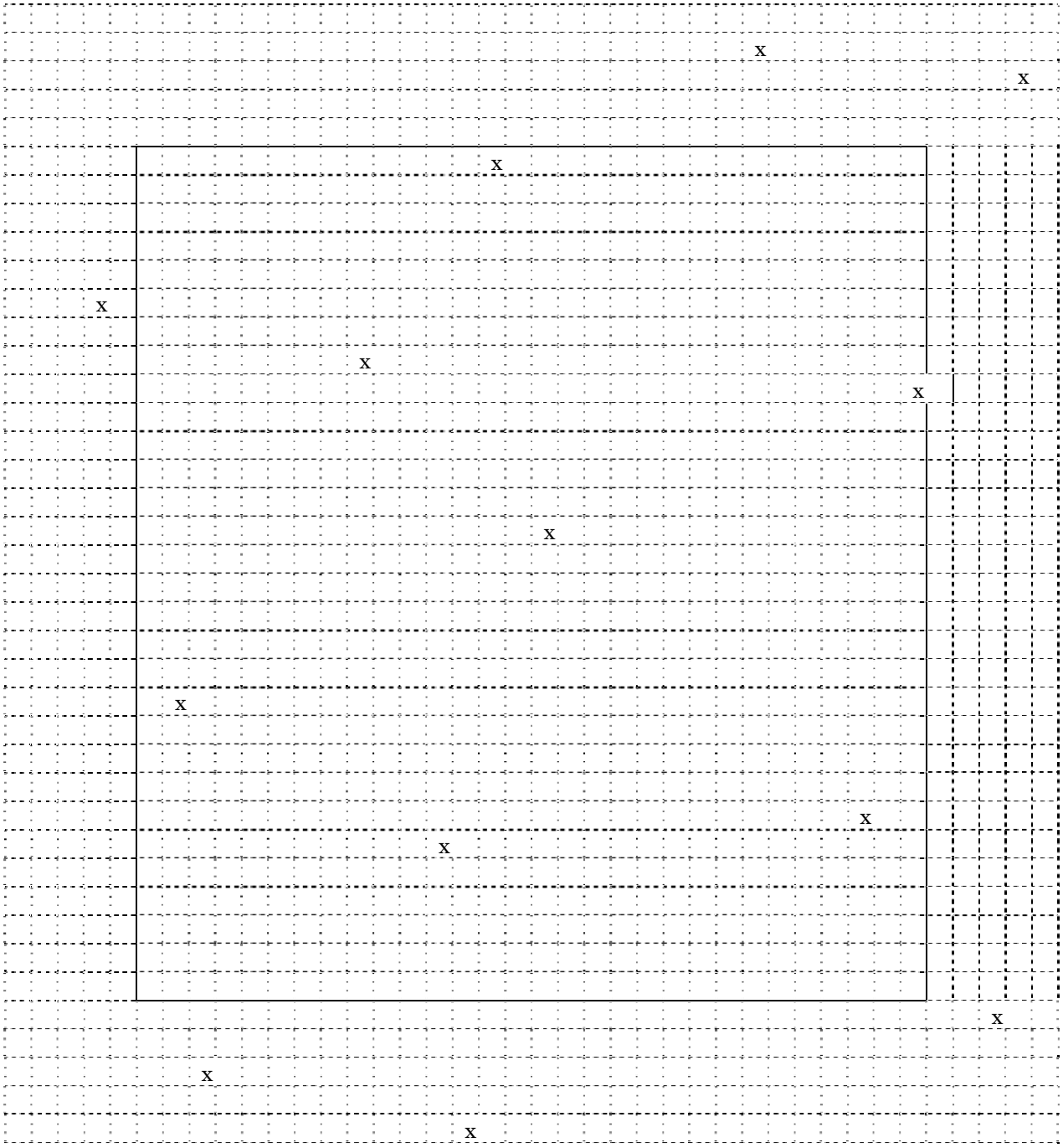


Figure 3. Block ordinary cokriging with TM imagery to scale up ground sample data (marked by "x") from smaller supports of 1×1 m^2 to a coarser pixel (solid line square) of 30 × 30 m^2 in which the remotely sensed data are available. The coarser pixel can be regarded as a block consisting of 900 smaller 1×1 m^2 supports (dash lines) whose estimates are obtained and then averaged as an estimate of the block.

Let *v(u)* denote a block centering at location *u*; $n_1(v(u))$, the number of ground sample data of the cover factor; and $n_2(v(u))$, the number of image data of the spectral variable, the ordinary cokriging estimator for the block *v(u)* is:

$$\hat{z}_1(v(u)) = \sum_{i=1}^{2} \sum_{\alpha=1}^{n_i(v(u))} \lambda_{i\alpha}(v(u)) z_i(u_\alpha) \quad \text{(Eq 10)}$$

where $\lambda_{i\alpha}(v(u))$ are the weights of the block to data. This estimator is unbiased if $\sum_{\alpha=1}^{n_1(v(u))} \lambda_{1\alpha}(v(u)) = 1$ and $\sum_{\alpha=1}^{n_2(v(u))} \lambda_{2\alpha}(v(u)) = 0$. The constraint $\sum_{\alpha=1}^{n_1(v(u))} \lambda_{1\alpha}(v(u)) = 1$ implies that not all of the weights for the cover factor are zero and at least one observation must be used for estimation of each block. The system of the equations to determine the weights in the block cokriging estimator is obtained by imposing two constraints: (1) that the estimator is unbiased, and (2) that the variance of the error of estimation is minimal. Using the solution of this system, the block cokriging variance is calculated (Deutsch and Journel 1998; Goovaerts 1997):

$$\sigma^2_{z_1}(v(u)) = C_{z_1 z_1}(v(u), v(u)) - \sum_{i=1}^{2} \sum_{\alpha=1}^{n_i(v(u))} \lambda_{i\alpha}(v(u)) C_{z_1 z_i}(u_\alpha, v(u)) - \mu_1 \quad \text{(Eq 11)}$$

Where $C_{z_1 z_1}(v(u), v(u))$ is the block covariance of the cover factor and is approximated by the arithmetic average of the covariances between any two discretizing points within the block *v(u)*. $C_{z_1 z_i}(u_\alpha, v(u))$ is the covariance (i = 1) or cross covariance (i = 2) between data support u_α and the block *v(u)* and is approximated by the arithmetic average of the covariances between the data support and the points discretizing the block *v(u)*. μ_1 is a Lagrange multiplier.

As kriging variance is for a single variable, cokriging variance is used for sample designs of multiple variables (McBratney and Webster 1983). According to Eq. 11, the block cokriging variance varies depending on only the variogram and cross variogram, spatial configuration of data locations in relation to the block to be estimated, and on the distances among the data locations. It does not depend on the data values. If (1) the variogram and cross variogram are known; (2) a systematical and square grid pattern is used; and (3) the numbers of ground and image data used for estimation of each block are fixed, changing grid spacings, then sampling distances of ground and image data, respectively, will lead to different block cokriging variances (Atkinson et al. 1992; McBratney and Webster 1983). Then, the maximum block cokriging variance can be used as the measure of quality of the sample design. From the diagram of maximum block cokriging variance against grid spacing, a sampling distance for the desirable precision requirement can be determined.

The block cokriging variance-based sample design will lead to a systematic sample that ensures that neighboring sample points are as far apart as possible for a fixed area. Thus, this method minimizes the duplication of information by maximizing the distance between sample points given a desired precision, and it is more cost-effective than the sample design based on random selection of locations (Atkinson et al. 1992, 1994, and 2000; Curran and Williamson 1986).

Cokriging variance-based sample design for multiple variables is more complicated than the kriging variance-based method for a single variable (McBratney and Webster 1983). The former requires models of spatial cross-correlation between variables, in addition to spatial autocorrelation functions, and deals with different sampling intensities of the primary and secondary (spectral) variable and the ratio between them. In the univariate case, the maximum kriging variance occurs at the locations that are most away from the data locations, and is located at the center of spatial configuration of observations given a spatial pattern such as square grid. In the multivariate case, however, the position of the maximum cokriging variance changes depending on the form of variograms and on the strength of cross correlation, the sampling interval, and the ratio of the numbers of data locations for the two variables inside the search neighborhood that matters. Wang et al. (2005) suggested that the average of cokriging variances could be applied as the measure of uncertainty for sample design in order to avoid searching for the maximum cokriging variance. In addition, the authors also found that using neighboring pixels for image data to map the cover factor led to more reliable estimates than using the pixels that were drawn by a systematic selection.

4 Stratified Sample Designs

In this study, the average of cokriging variances within the study area was calculated and plotted against grid spacing given a square grid. With a desirable precision, a sampling distance was then determined from the developed diagrams. Moreover, the neighboring pixels around a location to be estimated were directly used instead of the pixels systematically selected for the image data used for calculating cokriging variance. In addition, three sampling intensities were compared with ratios of ground-to-image data from 1:1 to 1:4 and 1:9 to explore the effect of the image data to reduce cost or increase map accuracy for sampling and mapping the cover factor. For estimation of each location, 12 ground data were used. The number of image data for three ratios varied from 12 to 48 and 108.

Using remotely sensed data, a study area usually can be stratified into homogeneous subareas or strata. Within each of the subareas or strata, the spatial variability of cover factor is similar, and a sample design can be made. This will lead to variable grid spacing that is optimal within the corresponding subarea or stratum. In this study, the southeast part of the area has a slope greater than 45 degrees, and no or few military training activities took place. The off-road vehicular traffic mainly occurred in the southwest, northwest, and northeast flat areas. Instead of using remotely sensed data, therefore, this study area was directly divided into training and steep subareas, and a sample design was then carried out within each of them.

By combining the ground data and ratio image mentioned above, an estimation map of the cover factor was generated with a sequential Gaussian collocated co-simulation (Anderson et al. 2005; Wang et al. 2002). In the co-simulation, 500 realizations of the cover factor map were generated, and a sample average map was calculated as the estimation of the cover factor surface. The map was then used as a reference surface, based on which a stratified sample design and a sample design without stratification were constructed using the block cokriging variance-based method described above. Using these two samples together with the original sample, the cover factor was mapped and the resulting maps were compared with this reference map.

5 Results

The experimental variograms $\hat{\gamma}_{CF}(|h|)$ and $\hat{\gamma}_{TM}(|h|)$ of the cover factor and ratio image (TM3 + TM7) / TM4, and experimental cross-variogram $\hat{\gamma}_{CF,TM}(|h|)$ between them were calculated for the entire area using VARIOWIN software (Pannatier 1996). These variograms were modeled using the spherical models:

$$\hat{\gamma}_{CF}(|\mathbf{h}|) = 0.0055 + 0.0069[1.5\frac{|\mathbf{h}|}{12000} - 0.5(\frac{|\mathbf{h}|}{12000})^3] \quad 0 < |h| \le 12000 \qquad \text{(Eq 12)}$$

$$\hat{\gamma}_{TM}(|\mathbf{h}|) = 140.0 + 180.0[1.5\frac{|\mathbf{h}|}{12000} - 0.5(\frac{|\mathbf{h}|}{12000})^3] \quad 0 < |h| \le 12000 \qquad \text{(Eq 13)}$$

$$\hat{\gamma}_{(CF,TM)}(|\mathbf{h}|) = 0.272 + 0.792[1.5\frac{|\mathbf{h}|}{12000} - 0.5(\frac{|\mathbf{h}|}{12000})^3] \quad 0 < |h| \le 12000 \qquad \text{(Eq 14)}$$

where $|h|$ is the magnitude of separation vector of data locations. Because the coefficient matrices in the linear coregionalization model must be positive definite, it may be necessary to adjust the entries in one or more of the matrices. In that case, the resulting variograms may not fit the experimental variograms as well as the original fits. Figure 4 presents these experimental and modeled variograms and cross-variograms. The semivariances increased over the separate distance given a direction, and reached their maximum values up to 12,000 m, a common range of spatial correlation.

With the models, block cokriging variance of each pixel was calculated when different grid spacings from 120 m to 2,580 m (corresponding sample size varied from 17,489 to 38 plots) and three ratios (mentioned above) of ground-to-image data were used. Figure 5 presents the change in average value of the block cokriging variances versus grid spacing for sampling the cover factor. For the same ratio of ground-to-image data, the average of block cokriging variances increased as the sampling distance increased.

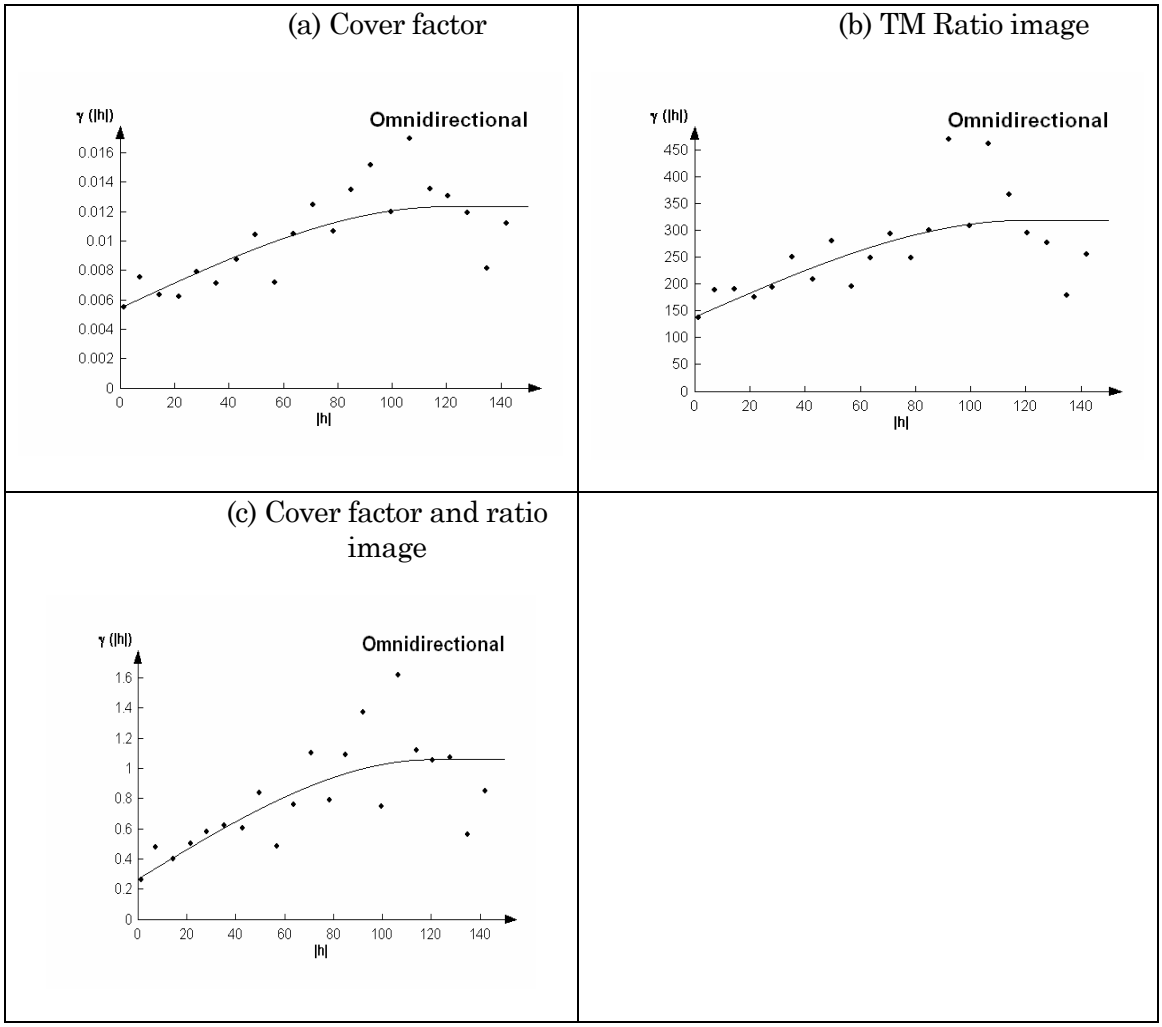


Figure 4. Experimental and modeled variograms of (a) cover factor, (b) TM ratio image (TM3+TM7)/TM4, and (c) cross variogram between the cover factor and ratio image. $\gamma(|\mathbf{h}|)$ is the semivariance and |h| is the separation distance of data locations with the unit of 100 m.

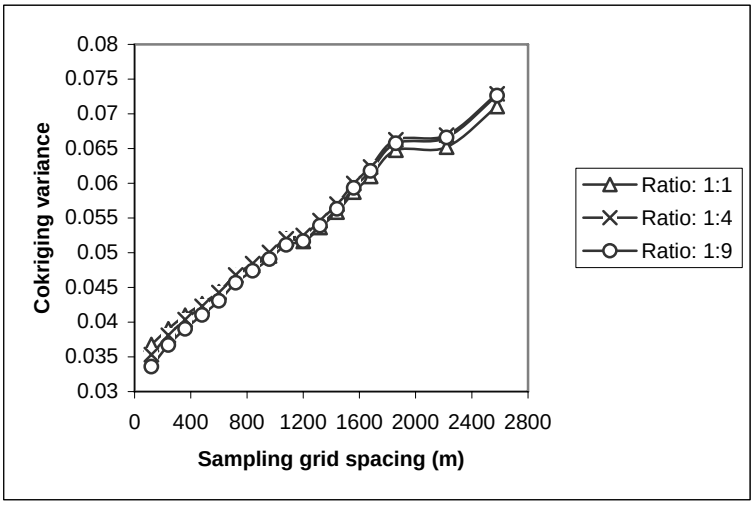


Figure 5. Average of cokriging variances versus grid spacing for sampling ground data with three different ratios of ground data to image data.

For the same sampling distance, the average of block cokriging variances in Figure 5 was smaller for the ground-to-image data ratio of 1:9, that is, where 12 ground data and 108 image data were used, than for the ground-to-image data ratio 1:4, where 12 ground data and 48 image data were used. However, the difference became smaller as the grid spacing increased. For the ground-to-image data ratio 1:1 where 12 ground data and 12 image data were used, the average of block cokriging variances was larger than those from both ratios 1:4 and 1:9 when the grid spacing was less than 720 m, and from the grid spacings of 720 m to 1,200 m, smaller than that from the ratio 1:4 but still larger than that from the ratio 1:9, and, after that, smaller than those from both ratios 1:4 and 1:9. The reasons for the phenomena above may include (1) as a greater grid spacing was used to sample the cover factor, a larger block cokriging variance was obtained from the spatial configuration of data locations, and the effect of image data of neighboring locations in the reduction of block cokriging variance relatively decreased; and (2) as the separation distance of data locations increased, the spatial cross correlation between the cover factor and spectral variable decreased, that is, spatial variability increased, which implied that using more image data than needed might lead to more noise.

When the study area was divided into two strata (steep and training subareas in Figure 6), the experimental variograms and cross-variograms were calculated and modeled for each stratum. These variograms include:

Within the steep subarea:

$$\hat{\gamma}_{CF}(|h|) = 0.000115 + 0.00041[1.5\frac{|h|}{6020} - 0.5(\frac{|h|}{6020})^3] \quad 0 < |h| \le 6020 \qquad \text{(Eq 16)}$$

$$\hat{\gamma}_{TM}(|h|) = 115.176 + 136.792[1.5\frac{|h|}{6020} - 0.5(\frac{|h|}{6020})^3] \quad 0 < |h| \le 6020 \qquad \text{(Eq 17)}$$

$$\hat{\gamma}_{(CF,TM)}(|h|) = 0.105 + 0.213[1.5\frac{|h|}{6020} - 0.5(\frac{|h|}{6020})^3] \quad 0 < |h| \le 6020 \qquad \text{(Eq 18)}$$

Within the training subarea:

$$\hat{\gamma}_{CF}(|h|) = 0.0057 + 0.0081[1.5\frac{|h|}{12040} - 0.5(\frac{|h|}{12040})^3] \quad 0 < |h| \le 12040 \qquad \text{(Eq 19)}$$

$$\hat{\gamma}_{TM}(|h|) = 128.1 + 146.993[1.5\frac{|h|}{12040} - 0.5(\frac{|h|}{12040})^3] \quad 0 < |h| \leq 12040 \quad \text{(Eq 20)}$$

$$\hat{\gamma}_{(CF,TM)}(|h|) = 0.294 + 0.700[1.5\frac{|h|}{12040} - 0.5(\frac{|h|}{12040})^3] \quad 0 < |h| \leq 12040 \quad \text{(Eq 21)}$$

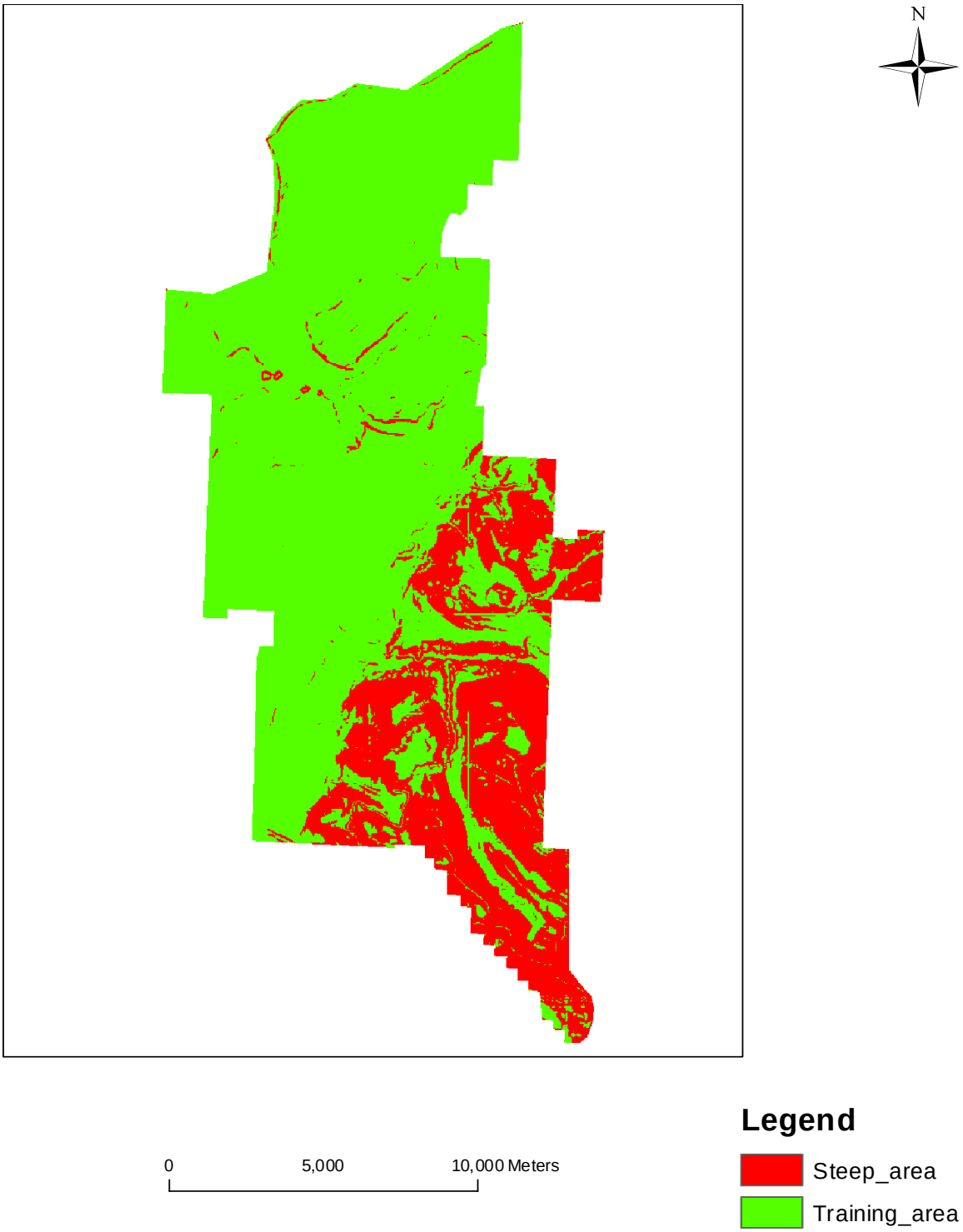


Figure 6. Stratification of study area into steep and training subareas.

The parameters of the variogram models differed between the steep area to the training area. For example, the range of spatial correlation is 6,020 m and 12,040 m for the steep and training subareas, respectively. Figure 7 shows the averages of block cokriging variances versus grid spacing for sampling the cover factor with these two strata. The ratio 1:4 of ground-to-image data was used to save computational time. For comparison, the global curve without stratification was also presented in Figure 7. As the grid spacing increased, the average of block cokriging variances increased slowly for the training subarea and the entire area, but rapidly for the steep subarea.

Given a precision-average of block cokriging variance equal to 0.04, Figure 7 was used to calculate the grid spacing for the entire area, training subarea, and steep subarea, respectively. The grid spacing obtained was 295 m when the sample design was carried out using block cokriging variance-based sampling without stratification, 322 m for the steep subarea, and 765 m for the training subarea using block cokriging variance-based sampling with two strata. The corresponding sample size was 2,890 without stratification and 938 with stratification. The number of the original sample plots was 747. Figure 8 presents these three samples. Using the reference map of cover factor by co-simulation, the effects of these three samples to map the cover factor were compared.

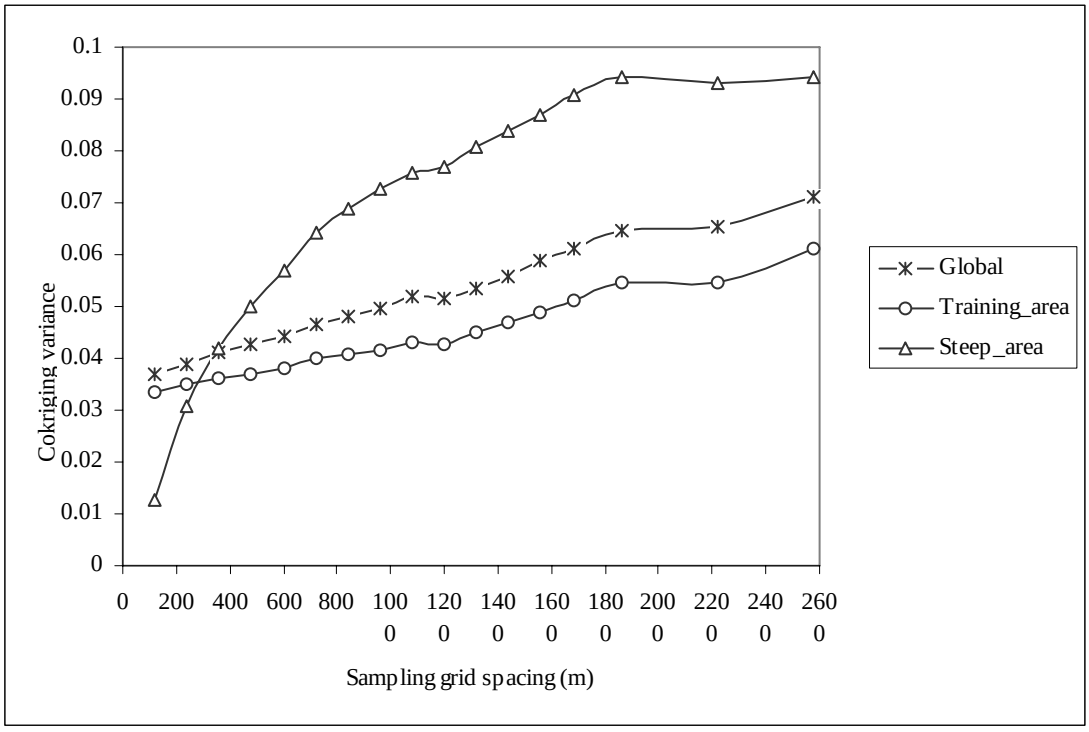


Figure 7. Average of cokriging variances versus grid spacing for sampling ground data with a ratio 1:4 of ground-to-image data when the study area was divided into steep and training subareas.

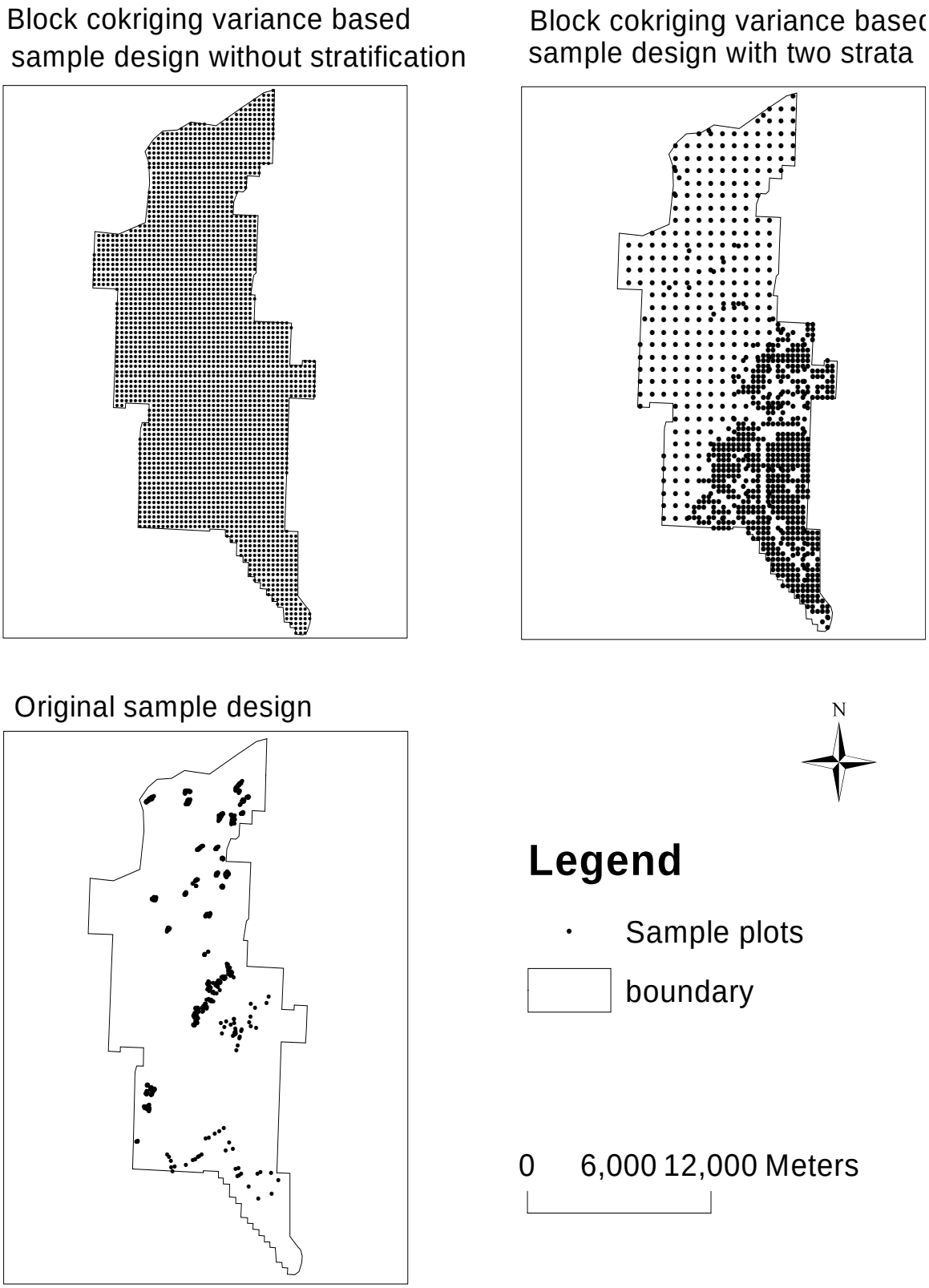


Figure 8. Block cokriging-based sample designs without stratification and with two strata, respectively, and the original sample design.

Figure 9 shows the estimation maps of cover factors from the three sample designs just described using the block ordinary cokriging estimator with the ratio image (TM3 + TM7) / TM4 and their comparison with the reference map. The estimates of the map by the original samples were obviously smoothed. Two maps by the block cokriging variance-based sample designs with and without stratification were similar to each other and also appeared close to the reference map in terms of spatial distribution and pattern of the values.

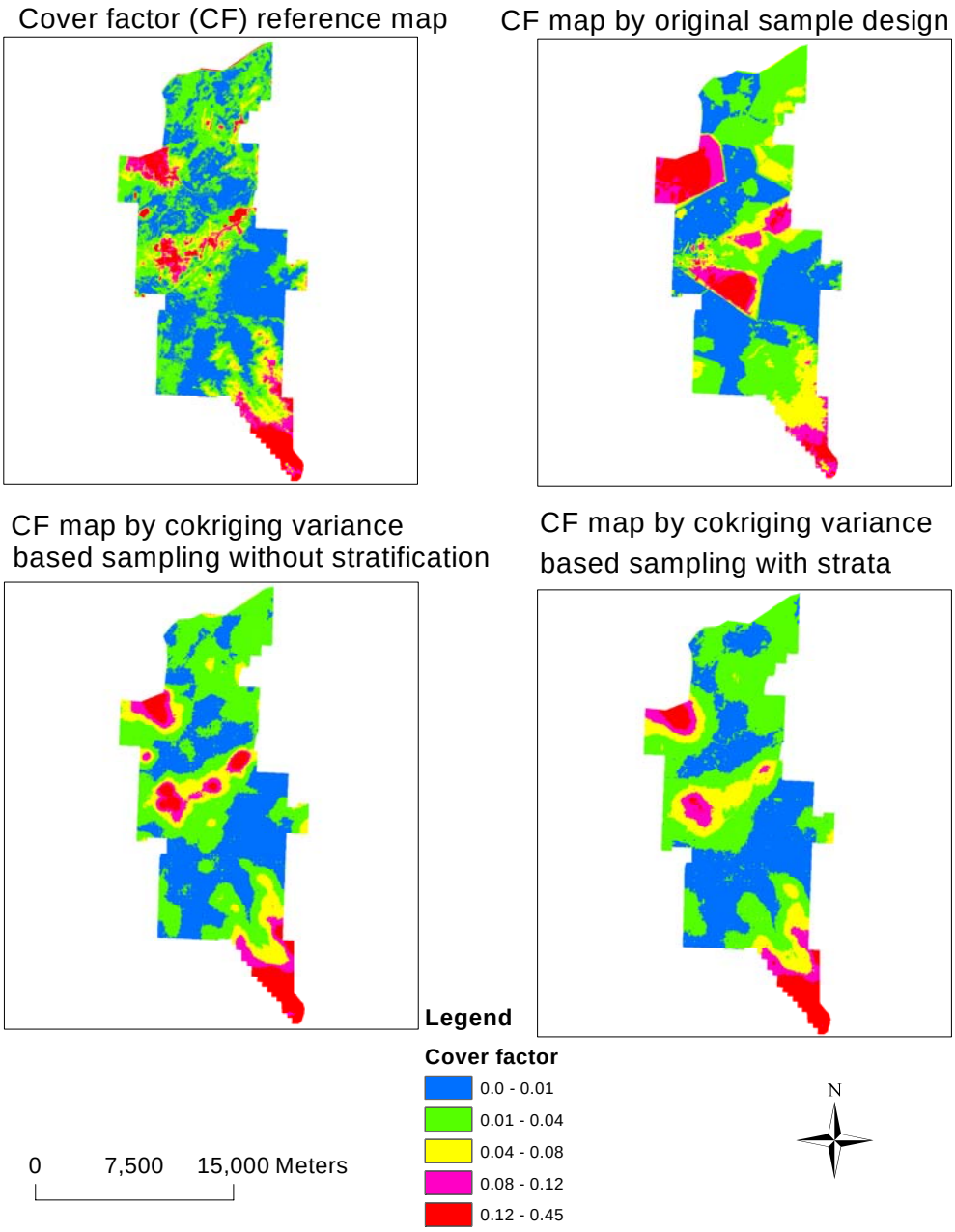


Figure 9. Estimation maps of cover factor (CF) using three sample designs: original sample design, block cokriging variance-based sampling without stratification and with two strata (steep and training subareas), and their comparison with the reference map.

In Figure 10, the three sample designs were further compared by calculating relative root mean square errors (RMSE * 100 / sample mean) of these estimation maps from the reference map using an independent sample of 384 randomly drawn plots. The original sample had the largest relative RMSE, 144.3 percent, but had the smallest sample size, 747 plots. The error of the sample design by the block cokriging variance-based sampling with stratification was slightly larger, 73.2 percent, than that without stratification, 62.3 percent, while the sample size of the latter was three times larger than that of the former. Thus, the cost-efficiency per sample plot significantly differed between these sample designs.

Because the largest relative RMSE was 144.3, 200 - relative RMSE was defined as the total efficiency of sampling and mapping and (200 - relative RMSE)/sample size as the unit cost-efficiency. The block cokriging variance-based sample design with stratification had the highest unit cost-efficiency, followed by the original sample design and the block cokriging variance-based sample design without stratification. The reason that the original sample design had a higher cost-efficiency than the block cokriging variance-based sample design without stratification was that the former had a different precision requirement and a significantly smaller sample size than the latter. If the unit cost-efficiency of the original sample was regarded as 1, the unit cost-efficiencies by the block cokriging variance-based sample designs with and without stratification were 1.81 and 0.64, respectively. That is, the stratification increased the unit cost-efficiency by three times when compared with no stratification.

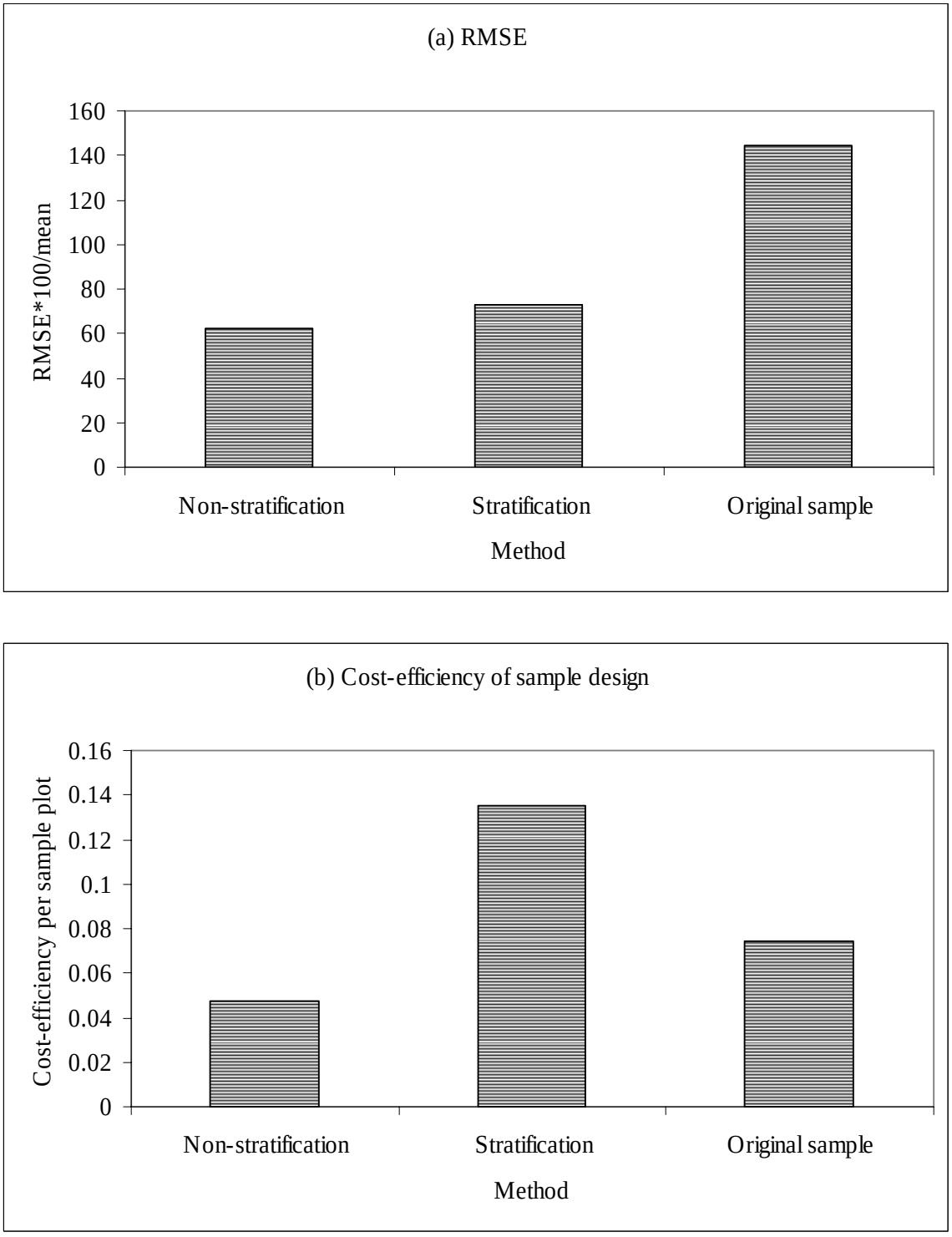


Figure 10. Comparison of three sample designs for mapping the cover factor based on (a) relative root mean square error (RMSE) - (RMSE * 100 / mean) and (b) cost-efficiency per plot defined as (200 - relative RMSE0 / sample size).

6 Conclusions and Discussion

When natural resources, and environmental and ecological systems are modeled and mapped with the aid of remotely sensed data, cost-efficient sample designs for collecting ground data and obtaining accurate maps of variables are widely required. Because of limited budgets, however, very often the supports used to collect ground data are smaller than the map units and pixels of images used. This inconsistency leads to difficulty and complexity in developing the procedure of sample design and mapping. For these purposes, a sampling and mapping method was developed by integrating stratification and an up-scaling method – block cokriging variance-based sample design with TM imagery. This method can lead to sample designs with variable grid spacing and can be used to significantly increase the unit cost-efficiency of sample data in mapping. This method is especially effective in cases where a study area obviously contains heterogeneous strata. The method was validated in this study to sample and map a ground and vegetation cover factor for a monitoring system of soil erosion by comparing different sample designs with and without stratification.

The block cokriging variance-based sample design with TM imagery is based on the variograms and cross-variograms that provide the link between aspatial statistics of a variable and geographic space, and the link between a mapped variable and a spectral variable geographically. The block cokriging with TM imagery scales up not only the ground sample data but also the uncertainties of spatial configuration of locations of ground and image data used and variogram models from smaller supports to larger pixels or blocks. For the same grid spacing of ground data, the average of block cokriging variances generally decreases as the number of image data used increases. The decrease becomes slighter and even disappears with a further increased number of image data because the uncertainties and noise from the image data may cancel out the effect of spatial cross-correlation between the primary and secondary variables. In addition, using too many image data will lead to a large increase in computational time and may also cause the cokriging equation system to become unstable (Goovaerts 1997).

The disadvantage of using block cokriging variance to determine the grid spacing of a sample is that the block cokriging variances do not reflect the uncertainties from variation of the ground and image data used. Thus, the grid spacing obtained is not locally optimal. However, the estimation of a population mean will be most precise

if the population can be partitioned into strata in such a way that within each stratum, the units are as similar as possible. That is, stratified sampling addresses homogeneity of units within stratum. Thus, integrating stratification and cokriging variance-based sampling theoretically can lead to the most cost-efficient sample design with variable grid spacing. This integration improves the accuracy and cost efficiency of sampling and mapping and is applicable for sample designs of the variables for which spatial variation varies from area to area. This outcome was verified by the results of this case study in which the unit cost-efficiency of the sample using the block cokriging variance-based sample design with stratification was, respectively, 1.8 times higher than that of the original sample using traditionally stratified sampling and 3 times higher than that of the sample using the block cokriging variance-based sampling without stratification.

References

Anderson, A.B., G. Wang, S. Fang, G.Z. Gertner, G. Burak, and D. Jones. 2005. Assessing and predicting changes in vegetation cover associated with military land use activities using field monitoring data at Fort Hood, Texas. Journal of Terramechanics 42(3&4):207-229.

Atkinson, P.M., R. Webster, and P.J. Curran. 1992. Cokriging with ground-based radiometry. Remote Sensing of Environment 41:45-60.

Atkinson, P.M., R. Webster, and P.J. Curran. 1994. Cokriging with airborne MSS imagery. Remote Sensing of Environment 50:335-345.

Atkinson, P.M., G.M. Foody, P.J. Curran, and D.S. Body. 2000. Assessing the ground data requirements for regional scale remote sensing of tropical forest biophysical properties. Intern. J. of Remote Sensing 21:2571-2587.

Benkobi, L., M.J. Trlica, and J.L. Smith. 1994. Evaluation of a refined surface cover subfactor for use in RUSLE. Journal of Range Management 47:74-78.

Campbell, J.B. 1996. Introduction to Remote Sensing. The Guilford Press, New York.

Colorado State University. 2004. Transformation of U.S. Army Alaska, Final Environmental Impact Statement. Center for the Environmental Management of Military Lands, Colorado State University, Fort Collins, CO.

Curran, P.J. 1988. The variogram in remote sensing: An introduction. Remote Sensing of Environment, 24:493-507.

Curran, P.J., and P.M. Atkinson. 1998. Geostatistics and remote sensing. Progress in Physical Geography, 22:61-78.

Curran, P.J., and H.D. Williamson. 1986. Sample size for ground and remotely sensed data. Remote Sensing of Environment, 20:31-41.

Demers, M.N. 2000. Fundamentals of Geographic Information Systems. John Wiley and Sons, Inc., New York.

Deutsch, C.V., and A.G. Journel. 1998. Geostatistical Software Library and User's Guide. Xford University Press, Inc., New York.

Goovaerts, P. 1997. Geostatistics for natural resources evaluation. Oxford University Press, New York.

Jorgensen, M. Torre, J.E. Roth, S.F. Schlentner, E.R. Pullman, M. Macander, and C.H. Racine. 2003. An ecological land survey for Fort Richardson, Alaska. ERDC/CRREL TR-03-19, ADA417494.

Matheron, G. 1971. The Theory of Regionalized Variables and Its Applications. Ecole des Mines de Paris, Fontainebleau.

McBratney, A.B., and R. Webster. 1983. Optimal interpolation and isarithmic mapping of soil properties V. Co-regionalization and multiple sampling strategy. Journal of Soil Science 34:137-162.

Pannatier, Y. 1996. VARIOWIN Software for spatial data analysis in 2D. Springer-Verlag New York, Inc., New York.

Renard, K.G., C.R. Foster, G.A. Weesies, D.K. McCool, and D.C. Yoder. 1997. Predicting soil erosion by water: A guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE). U.S. Department of Agriculture, Agriculture Handbook Number 703, U.S. Government Printing Office, SSOP Washington, DC.

Thompson, S. 1992. Sampling. John Wiley and Sons, Inc., New York.

Wang, G., S. Wente, G. Gertner, and A.B. Anderson. 2002. Improvement in mapping vegetation cover factor for universal soil loss equation by geo-statistical methods with Landsat TM images. International Journal of Remote Sensing, 23:3649-3667.

Wang, G., G.Z. Gertner, and A.B. Anderson. 2005. Remote sensing towards optimization of sampling and mapping a soil erosion ground and vegetation cover factor by cokriging. ISPRS J. of Photogrammetry and Remote Sensing (in review).

Wischmeier, W.H., and D.D. Smith. 1978. Predicting rainfall-erosion losses from cropland east of the Rock Mountains: Guide for selection of practices for soil and water conservation. USDA, Agriculture Handbook. No. 282, U.S. Government Printing Office, SSOP Washington, DC.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
03-2006	Final	

4. TITLE AND SUBTITLE
Sampling and Mapping Soil Erosion Cover Factor for Fort Richardson, Alaska: Integrating Stratification and An Up-Scaling Method

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Guangxing Wang, George Gertner, Alan B. Anderson, and Heidi Howard

5d. PROJECT NUMBER
622720A896

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Engineer Research and Development Center (ERDC)
Construction Engineering Research Laboratory (CERL)
PO Box 9005
Champaign, IL 61826-9005

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDC/CERL TR-06-5

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Corps of Engineers
441 "G" Street NW
Washington, DC 20314-1000

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Copies are available from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

14. ABSTRACT

When a ground and vegetation cover factor related to soil erosion is mapped with the aid of remotely sensed data, a cost-efficient sample design to collect ground data and obtain an accurate map is required. However, the supports used to collect ground data are often smaller than the desirable pixels used for mapping, which leads to complexity in developing procedures for sample design and mapping. For these purposes, a sampling and mapping method was developed by integrating stratification and an up-scaling method in geostatistics – block cokriging with Landsat Thematic Mapper ™ imagery. This method is based on spatial correlation and stratification of variables. It scales up not only the ground sample data but also the uncertainties associated with the data aggregation from smaller supports to larger pixels or blocks. This method uses the advantages of both stratification and block cokriging variance based sample design, which leads to sample designs with variable grid spacing, and thus significantly increases the unit cost-efficiency of sample data in sampling and mapping. This outcome was verified by the results of this study.

15. SUBJECT TERMS
Fort Richardson, AK
soil erosion
military training
land rehabilitation
satellite imagery

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Heidi Howard
Unclassified	Unclassified	Unclassified	SAR	36	19b. TELEPHONE NUMBER (include area code) (217) 373-5865