Mapping Multiple Variables for Predicting Soil Loss by Geostatistical Methods with TM Images and a Slope Map

Guangxing Wang, George Gertner, Shoufan Fang, and Alan B. Anderson

Abstract

Soil erosion is widely predicted as a function of six input factors, including rainfall erosivity, soil erodibility, slope length, slope steepness, cover management, and support practice. Because of the multiple factors, their interactions, and their spatial and temporal variability, accurately mapping the factors and further soil loss is very difficult. This paper compares two geostatistical methods and a traditional stratification to map the factors and to estimate soil loss. Soil loss is estimated by integrating a sample ground data set, TM images, and a slope map. The geostatistical methods include collocated cokriging and a joint sequential co-simulation model. With both geostatistical methods, local estimates and variances at any location where the factors and soil loss are unknown can be computed. The results showed that the two geostatistical methods performed significantly better than traditional stratification in terms of overall and spatially explicit estimates. Furthermore, the cokriging led to higher accuracy of mean estimates than did the co-simulation, while the latter provided decision makers with reliable uncertainties of the local estimates as useful information to assess risk when making decisions based on the prediction maps.

Introduction

Soil erosion in the United States is usually predicted by the Natural Resources Conservation Service (NRCS) using a Revised Universal Soil Loss Equation (RUSLE) (Renard *et al.*, 1997). In the equation, soil loss A ($t \cdot ha^{-1} \cdot year^{-1}$) is a function of six input factors, including rainfall-runoff erosivity R ($MJ \cdot mm \cdot ha^{-1} \cdot hour^{-1} \cdot year^{-1}$), soil erodibility K ($t \cdot ha \cdot hour \cdot ha^{-1} \cdot MJ^{-1} \cdot mm^{-1}$), slope length L, slope steepness S, cover management C, and support practice P: i.e.,

$$A = R \times K \times L \times S \times C \times P. \quad (1)$$

where the factors L, S, C, and P are dimensionless.

This is an empirical model and the factors are derived using regression models obtained with field measurements from experimental plots (Renard *et al.*, 1997). The annual R factor is a sum of erosivity index values for all rain-showers in an average year. The K factor is the soil loss rate per erosion index unit for a specified soil as measured on a standard plot defined as a 22.1-m length of uniform 9 percent slope in continuous clean-tilled fallow. It is determined by an empirical model that is a function of soil silt, sand, clay, organic matter, structure, and permeability. The product of slope length L and steepness S, called topographical factor LS, is the rate of soil loss from the field slope length and gradient to soil loss from a standard plot under otherwise identical conditions (Renard *et al.*, 1997). The LS factor depends on a set of regression models that are functions of slope and slope length measurements.

The C factor is the rate of soil loss from an area with specified cover. This factor reflects the effect of ground and vegetation cover on the reduction of soil loss by reducing rainfall and runoff. In the RUSLE, the C factor is a product of five sub-factors, including prior land use, canopy cover, surface cover, surface roughness, and soil moisture (Renard *et al.*, 1997). Each of the sub-factors depends on an empirical equation. In this study, the C factor was obtained according to the original Universal Soil Loss Equation (USLE) (Wischmeier and Smith, 1978), with measurements of ground cover, canopy cover, and minimum drip height. The P factor is the rate of soil loss with a support practice such as contouring, strip cropping, terracing, etc. Here it was assumed that $P = 1$ because of no support practices.

Generally, soil loss is most sensitive to the LS factor, and then C factor (Renard and Ferreira, 1993; Risse *et al.*, 1993; Benkobi *et al.*, 1994; Biesemans *et al.*, 2000). Erosion increases as slope length and steepness increase, and it increases more rapidly with slope steepness than with slope length. The better the ground and vegetation cover, the less the potential soil loss. Soil loss is also proportional to the R factor when other factors are held constant. For each specific soil, furthermore, a soil-loss tolerance value indicating a maximum permitted level of soil erosion for sustainable soil productivity has been derived for agricultural management by NRCS (Schertz, 1983). The tolerance values in the United States vary from 2.24 to 11.21 tonnes per hectare per year.

An average of soil loss for a specific area in the United States is usually provided to decision makers and land users by the NRCS. As sustainable and precision agricultural and environmental management develops, however, decision makers tend to need local estimates with their uncertainty measures defined as errors and variances of the estimates, and probabilities for the estimates being larger than a critical soil-loss tolerance value (Goovaerts, 1997; Wang *et al.*, 2000; Wang *et al.*, 2001a). This type of infor-

G. Wang, G. Gertner, and S. Fang are with the Department of NRES, University of Illinois, W503 Turner Hall, 1102 S. Goodwin Ave., Urbana, IL 61801 (wang12@uiuc.edu; gertner@uiuc.edu; f-shouf@uiuc.edu).

A.B. Anderson is with USACERL, P.O. Box 9005, Champaign, IL 61826–9005 (Alan.B.Anderson@erdc.usace.army.mil).

Photogrammetric Engineering & Remote Sensing
Vol. 69, No. 8, August 2003, pp. 889–898.

0099-1112/03/6908–889$3.00/0

mation can help decision makers make detailed plans and assess the risk of decisions being made at different space and time scales in terms of long-term annual averages. However, predicting soil erosion is very difficult because of the multiple factors, their interactions, the inputs from other systems, and spatial and temporal variability of the variables. Errors from the variables propagate and accumulate through the entire system when predicting soil loss. It is important for decision makers to consider these possible errors. Therefore, the uncertainty due to these errors should be provided in the final prediction of soil loss.

Using auxiliary data, including remotely sensed data and digital elevation models, can improve the efficiency and precision when predicting soil loss. The improvement depends on density of ground data, correlation between the ground data and auxiliary data, interactions between the variables, mixture pixels, and limitations of existing methods. The widely used methods for mapping are the so-called point-in-polygon or point-in-stratum based on supervised and unsupervised classification or stratification, and the integration of both methods (Campbell, 1996). The methods result in homogeneous polygons or strata of pixels. A polygon implies that the homogeneous cells adjoin with each other in space, while a stratum does not.

Regression modeling and k-nearest-neighbor methods are another two groups of the methods for mapping (Tomppo, 1996). Poso *et al.* (1999) compared the above methods for estimating forest variables such as diameter, height, stand volume, etc. The differences in mean square error between the methods were fairly small. The stratification methods lead to smoothing of estimates and disappearance of spatial heterogeneity, while regression modeling may result in illogical or extreme estimates because of extreme auxiliary data. A common assumption behind these methods is that the sample data are not spatially correlated, which makes it possible to provide unbiased estimates for populations. However, it is problematic that these methods will provide reliable local estimates and their estimation variances, and will reproduce cross-spatial variability and interactions of multiple variables.

Recently, geostatistical methods such as collocated cokriging and sequential co-simulation have been introduced for mapping. Collocated cokriging means that, in addition to neighboring ground sample data, auxiliary data that are available at each estimated location are used to derive estimates (Goovaerts, 1997). Sequential co-simulation means that neighboring ground sample data and previously simulated values, and auxiliary data that are available at each estimated location, are used to determine distributions of estimates before random drawing in simulation (Wang *et al.*, 2001b). For example, Hunner *et al.* (2000) modeled forest stand structure by cokriging, where the secondary variables included topography, TM images, and their vegetation indices. Barata *et al.* (1996) estimated forest areas using cokriging and remote sensing data. Wang *et al.* (2001b) integrated a sequential indicator co-simulation and Landsat TM data for vegetation classification. In addition to a vegetation map, they derived a misclassification probability map.

These geo-statistical methods assume that the variables of interest are spatially correlated with each other within a certain distance. Remotely sensed data can be considered to be models of ground characteristic variables. The spatial variability, that is, spatial dissimilarity of the ground characteristics that varies depending on the separation vector of data, is coded in the auxiliary data. Cross semivariograms used in the methods can capture the spatial dissimilarity and correlation between the ground characteristics and auxiliary data. These methods can thus result in unbiased estimates of the variables and reproduce the cross-spatial variability. They also can provide variances, covariances, and conditional probabilities of estimates at any locations as spatial uncertainties of the estimates. These spatial uncertainty estimates can represent spatial variation of mapping accuracy (Steele *et al.*, 1998).

Presented here are a collocated cokriging, a joint sequential co-simulation, and their results to map the factors and predict soil loss. Sample ground data, Landsat TM images, and a slope map derived from a DEM are integrated together in the generation of the maps. With both geostatistical methods, local estimates and variances at any location where the factors and soil loss are unknown can be computed. A specific objective is to compare these two geostatistical methods to a traditional stratification for predicting soil loss in terms of statistical parameter estimates and spatial distribution of the estimates. The methods were assessed in a case study area where a prediction map of annual soil loss with uncertainties is needed for decision making for the purpose of land management and planning. Mapping the input factors and predicting soil loss was done at a spatial resolution of 90 m by 90 m. Based on the assessment, suggestions for applications of the two geostatistical methods for remote sensing aided mapping are given.

Study Area and Data Sets

This study was carried out in an area of 87,890 ha, located at Fort Hood, Texas (Figure 1). The climate is characterized by long hot summers and short mild winters, and the average annual precipitation is 810 mm. Soils are generally shallow to moderately deep and clayey, underlain by limestone bedrock. The dominant vegetation type at the east and northeast is oak-juniper woodlands. West and south parts are savannah type, dominated by grasses with scattered stands of live oak. In the center there is a mixture of the savannah type and oak-juniper woodlands.

The ground data sample required to predict soil loss was based on a land-condition trend analysis plot inventory and field methods design (Tazik *et al.*, 1992). Based on vegetation and soil types, 211 plots were selected in a stratified random fashion and measured in the spring and summer of 1989 (Price et al., 1995). The locations of the

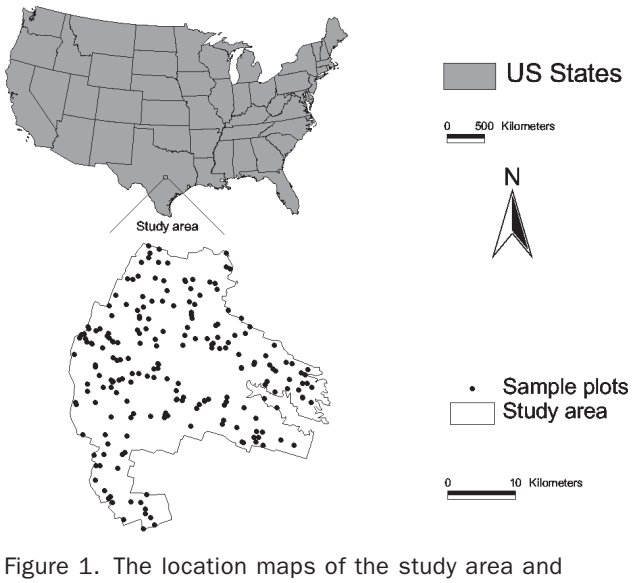

Figure 1. The location maps of the study area and ground sample plots.

sample plots are shown in Figure 1. The number of these plots at each vegetation and soil type was proportional to its percent of land area. The plot width was 6 m and the length was 100 m. The measured variables included slope length and steepness, soil silt, sand, clay, organic matter, structure and permeability, ground cover, canopy cover, and average vegetation height.

For each plot, the values of the *K*, *LS*, and *C* factors were calculated using the empirical models of RUSLE (Renard *et al.*, 1997). There are no rainfall stations within the study area. The values of the *R* factor can be obtained from an isoerodent map (Renard *et al.*, 1997). However, Wang *et al.* (2002) found that the *R* factor at the study area has greatly increased compared to that from the old map, and they derived a new *R* factor map covering this study area using a sequential Gaussian simulation. This new map was used to extract the *R* factor values of this study area (Figure 2). The estimates of the *R* factor within the study area varied from 5848 to 6263 with an average and standard deviation of 6061 and 35, respectively. The coefficient of variation for the estimates, less than 1 percent, was very small. Higher values were located at the north and central areas.

A scene of Landsat TM images dated 16 October 1989 with a spatial resolution of 30 m by 30 m was acquired. These images consisted of TM1: 0.45 to 0.53 μm, TM2: 0.52 to 0.60 μm, TM3: 0.63 to 0.69 μm, TM4: 0.76 to 0.90 μm, TM5: 1.55 to 1.75 μm, and TM7: 2.08 to 2.35 μm. They were geo-referenced to the Universal Transverse Mercator (UTM) projection and re-sampled to a resolution of 90 m by 90 m by averaging spectral reflectance values of nine pixels within a 3 by 3 window and coarsening the pixel size.

Additionally, a digital elevation model (DEM) with a grid spacing of 10 m and vertical resolution of 1 m for this area was acquired from the U.S. Geological Survey. This DEM was derived from contour maps and classified as Level-2 with a vertical accuracy of 1.5 contour lines. From the DEM, a slope map was derived using the partial derivative of a regularized spline function with tension developed by Mitášová *et al.* (1996) and re-sampled to a resolution of 90 m by 90 m by the averaging process mentioned above to match the TM image data.

As an example of auxiliary data used, the DEM, slope map, and Landsat TM7 image are shown in Figure 3. Elevations range from 183 m to 373 m above sea level and decrease from the northwest to the southeast. Most slopes are in the 1- to 6-degree range with few slopes in excess of 30 degrees. TM7 reflectance values are higher at the southwest areas than at the east and northeast areas because the vegetation cover is lower at the former areas than at the latter.

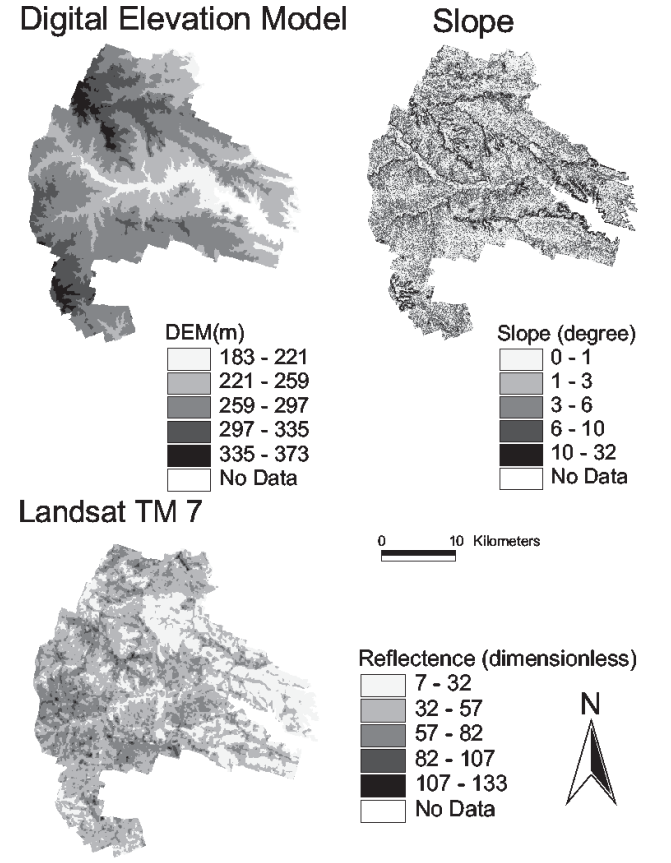

Figure 3. A digital elevation model (DEM), slope map from the DEM, and Landsat TM7 image for the study area.

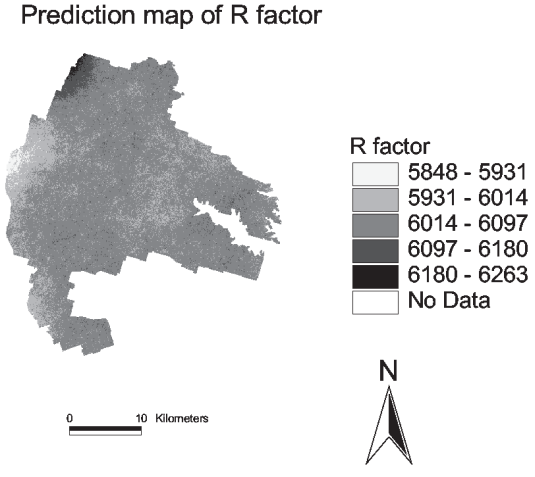

Figure 2. Prediction map of rainfall-runoff erosivity factor *R* (MJ · mm/ha · h · y) for Fort Hood area.

Methods

A collocated cokriging and a joint sequential co-simulation were separately developed to create prediction maps of soil erodibility factor *K*, topographical factor *LS*, and vegetation cover factor *C*. In the cokriging and co-simulation, the *K*, *LS*, and *C* factors were defined as primary variables, while the secondary or auxiliary variables selected were Landsat TM images for the *C* and *K* factors, and the slope map for the *LS* factor. The primary variable data are available only at the sampled ground locations and the auxiliary data are available over the study area at a grid spacing of 90 m by 90 m. The prediction maps of the *K*, *LS*, and *C* factors were also produced by a traditional stratification with the same TM images and slope map. Soil loss was then calculated using the prediction maps of the *K*, *LS*, and *C* factors derived by each of the three methods. In the calculation of soil loss for the different methods, the same *R* factor map (Figure 2) was used. These methods were finally compared based on confidence intervals of mean estimates of the *K*, *LS*, and *C* factors, correlation and root-mean-square error between the estimated and observed values, and spatial distribution and uncertainties of the local estimates.

Image Data Analysis

The image data were transformed to generate spectral variables that were most correlated with the factors. The transformations considered include

(1) Normal difference vegetation index (NDVI) = (TM4 + TM3)/(TM4 − TM3),
(2) TM Ratio 1 = TM3/TM4,
(3) TM Ratio 2 = TM7/TM4,
(4) TM Ratio 3 = (TM3 * TM7)/TM4,
(5) TM Ratio 4 = (TM3 * TM5)/TM4,
(6) TM Ratio 5 = (TM3 + TM7)/TM4, and
(7) TM Ratio 6 = (TM2 + TM3 + TM7)/TM4.

Traditional Stratification

The traditional method used was the integration of supervised and unsupervised stratification (Campbell, 1996). It was assumed that soil erosion is homogeneous at some locations that may not be neighbors. The cells that have similar estimates of soil loss regardless of their locations belong to the same stratum. The auxiliary data of the sample plots were extracted from TM images and a slope map, and were used to create a classification model with initial seeds and variances of homogeneous strata. Using this model, the cells of the area grid were then assigned into the strata according to their image and slope data. Within each stratum, an average and variance for each factor of soil loss was calculated and assigned to each cell that belongs to the stratum.

Spatial Variability and Semivariograms

Suppose that a study area is divided into N cells of a grid and P primary variables are estimated. In this area, a sample is drawn and the sample data set $\{z_p(u_\alpha), u_\alpha = 1, 2, ..., n, p = 1, 2, ..., P\}$ is obtained for P variables, and n is the number of ground sample data. The datum of a variable p at location u_α is $z_p(u_\alpha)$. The expectation and variance for the variable p are m_p and σ_p^2, respectively. The auxiliary data $y_q(u)$ $(q = 1, 2, ..., Q)$ for Q secondary variables are available at each location to be estimated.

In geostatistics, two sample data $z_p(u_\alpha)$ and $z_p(u_\alpha + h)$ separated by a distance h given a direction are assumed to be only slightly dissimilar at the beginning, and become more and more different as h increases and finally become independent. This is called spatial variability of a variable. Collocated cokriging and joint sequential co-simulation are based on auto- and cross-spatial variability of the variables, measured by $\gamma_{pp'}(h)$ cross-spatial semivariogram of two different primary variables or a primary variable and a secondary variable, or auto-semivariogram when $p = p'$. Using the sample data above, the experimental semivariograms can be obtained (Goovaerts, 1997): i.e.,

$$\gamma_{pp'}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z_p(u_\alpha) - z_p(u_\alpha + h)] \cdot [z_{p'}(u_\alpha) - z_{p'}(u_\alpha + h)] \quad (2)$$

where N(h) is the number of data pairs. The experimental auto- and cross-semivariograms can often be fitted using a spherical, exponential, and Gaussian model. For example, a spherical model is

$$\gamma_{pp'}(h) = c_0 + c_1 \times Spherical\left(\frac{h}{a}\right) = \begin{cases} c_0 + c_1\left[1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right] & \text{if } h \le a \\ c_0 + c_1 & \text{if } h > a \end{cases} \quad (3)$$

where c_0 and c_1, respectively, are called nugget and structure parameters, and $c = c_0 + c_1$ is called a sill parameter. The structure parameter accounts for the structured variance of the spatial variability and the nugget parameter implies the nugget variance when $h = 0$. The parameter a is the range parameter, providing the range of spatial variability. Within the range, observations can be considered spatially dependent, and beyond the range, observations can be considered essentially independent. As the distance h increases, the auto- and cross-semivariograms vary and approach their limit values.

Collocated Cokriging

Collocated cokriging estimators are alternatives to improve the prediction map of a primary variable given secondary variable data available at all the cells to be estimated, such as TM images. The methods introduce the auxiliary information into the interpolation for improving estimation. For example, a simple collocated cokriging estimator is (Goovaerts, 1997)

$$z_p^{sck}(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{sck}(u)[z_p(u_\alpha) - m_{z_p}] + \lambda_y^{sck}(u)[y_q(u) - m_{y_q}] + m_{z_p} \quad (4)$$

where u is a cell to be estimated, $z_p^{sck}(u)$ is an estimate of a primary variable, $n(u)$ is the number of ground plots used for the estimation given a neighborhood size, for instance, within 500 m, m_{z_p} and m_{y_q} are the average values of the primary and secondary variable, and $\lambda_\alpha^{sck}(u)$ and $\lambda_y^{sck}(u)$ are the weights to a primary and auxiliary datum. For the weights, a system consisting of $n(u) + 1$ linear equations containing the auto- and cross-semivariograms must be solved. The auto-semivariograms of the primary variables are inferred directly from the sample data and the cross-semivariograms between a primary and secondary variable are approximated by a Markov model, that is, obtained by the auto-semivariogram of the primary variable, divided by its variance, then multiplying the covariance of two variables. Equation 4 can be directly extended for more than one secondary variable (Almeida and Journal, 1994).

Joint Sequential Co-Simulation for Multiple Variables

The joint sequential simulation algorithm is based on Bayes' conditional probability axiom to generate L joint realizations of P variables from their Cumulative Distribution Functions (CDFs) (Almeida, 1993). A realization implies that each cell of the grid is provided with an estimation vector of the variables, derived from the vectors of neighboring field plots and the vector of the secondary variables at the location to be estimated. Theoretically, a joint P-variable CDF characterizing the P random events can be decomposed into a product of $(P - 1)$ univariate conditional CDFs and a marginal CDF according to Bayes' axiom for conditional probability. From the decomposition, a general sequential simulation algorithm can be developed to jointly simulate the P dependent variables by drawing from the sequence of univariate conditional CDFs.

The conditional CDFs can be derived in the simplest way in which a multi-Gaussian model is assumed for the multivariate distribution. The assumption implies univariate normality. When the data of the variables are not normally distributed, a normal score transform should be performed so that the transformed data have means of zero with unit variances. Adding secondary variables such as remote sensing data into the joint simulation leads to joint co-simulation.

Before the co-simulation, a simple collocated cokriging estimator described above is selected to derive the statistical parameters (means and variances) of the conditional CDFs. The estimator can provide the best estimates of the

statistical parameters for the conditional CDFs. In a joint sequential co-simulation, first define a hierarchy of the primary variables and start with the most important one with the highest correlation. Then set a random path to visit each cell of grid once. At the uth cell to be visited, determine the mean and variance of the conditional CDF for the first primary variable given the $n(u)$ neighboring primary data and the collocated secondary data using the collocated cokriging. From that distribution, draw a value which will become the conditional datum for all subsequent drawings.

At the same cell u, determine the distribution and draw a value for the second primary variable; at that time the simulated value of the first primary variable is added into the conditional data. Continue the determination and drawing until all primary variables have been estimated. A joint sequential co-simulation is completed when all the cells are visited and provided with simulated values. Repeating the joint sequential co-simulation process many times with probable different visiting paths leads to a set of estimates at each location for each variable. Finally, an expected vector and covariance matrix for P variables at each location is calculated.

As the number of simulations increases, the variances decrease rapidly at the beginning, and then slowly and gradually become stable. The number at which the estimation variances tend to become stable can be chosen as the final number of simulation runs. In this study, the estimated variances stabilized at 500 runs.

Results

The coefficients of correlation of topographical factor *LS* with vegetation cover factor *C* and soil erodibility factor *K* was −0.2105 and −0.1204, and the coefficient between the *C* and *K* factors was 0.2255. The coefficients of correlation were low. The correlations of the factors with slope and image data are listed in Table 1. The *LS* factor had higher correlation with slope than with the images. The *C* and *K* factors were most correlated with the TM Ratio 5, that is, (TM7 + TM3)/TM4 and TM7, respectively. Overall, TM7 had the largest average coefficient of correlation with all the factors. TM7 and slope was thus selected as the two secondary variables to jointly map the factors by stratification, collocated cokriging, and joint sequential co-simulation.

Figure 4 shows the experimental and modeled standardization semivariograms of (a) *C* factor, (b) *LS* factor, and (c) *K* factor. Their sill parameters were forced to be unit one. A spherical model was used to fit the semivariograms for the three factors. As an example, the sample cross-semivariogram between factors *LS* and *C* was compared to its results based on a Markov model in Figure 4d. For these two factors, the Markov model depicted their cross-spatial variability over distance well, although the experimental cross-semivariogram fluctuated.

With the sample data set, modeled semi-variograms, and the auxiliary data, prediction maps of factors *LS*, *C*, and *K* using three methods were created. In Table 2, the statistical parameters from the prediction maps were com-

TABLE 1. THE COEFFICIENTS OF CORRELATION BETWEEN THE ESTIMATED VARIABLES (VEGETATION COVER FACTOR *C*, TOPOGRAPHICAL FACTOR *LS*, AND SOIL ERODIBILITY FACTOR *K*) AND AUXILIARY VARIABLES (SLOPE, LANDSAT TM BANDS AND THEIR TRANSFORMATIONS). (* AV(ABS) IS THE AVERAGE OF ABSOLUTE VALUES OF THE COEFFICIENTS)

Variables	Slope	TM1	TM2	TM3	TM4	TM5	TM7
C	−0.14405	0.57200	0.57209	0.59197	0.41603	0.57463	0.60386
LS	**0.46017**	−0.10863	−0.10862	−0.11618	−0.05995	−0.1282	−0.13450
K	−0.19795	0.28998	0.29379	0.29522	0.27888	0.31324	**0.31531**
Av(abs)*	0.26738	0.32353	0.32483	0.33445	0.25162	0.33867	**0.35122**
Variables	NDVI	Ratio1	Ratio2	Ratio3	Ratio4	Ratio5	Ratio6
C	0.39567	0.61784	0.60861	0.61754	0.61037	**0.62096**	0.61978
LS	−0.07416	−0.13556	−0.15822	−0.13546	−0.13494	−0.15034	−0.14822
K	0.09782	0.24954	0.27466	0.29557	0.29548	0.26727	0.25902
Av(abs)	0.18922	0.33431	0.34716	0.34952	0.34693	0.34619	0.34234

TABLE 2. STATISTICAL PARAMETERS OF SAMPLE DATA AND PREDICTION MAPS USING STRATIFICATION, COLLOCATED (COL.) COKRIGING, AND CO-SIMULATION WITH LANDSAT TM7 AND SLOPE. THE PARAMETERS INCLUDED MINIMUM AND MAXIMUM VALUE, AVERAGE, LOWER AND UPPER LIMITS OF CONFIDENCE INTERVAL AT PROBABILITY OF 95 PERCENT, AND COEFFICIENT OF CORRELATION BETWEEN THE ESTIMATED AND OBSERVED VALUES

				Confidence Interval		Coefficient of
	Min	Max	Average	Lower	Upper	Correlation
			LS factor (dimensionless)			
Sample data	0.0762	15.8393	0.7014	0.6107	0.7921	
Stratification	0.1774	6.2219	0.8845			0.1542
Col. Cokriging	−0.3292	9.3886	0.8722			0.8879
Co-simulation	0.1133	9.1463	0.7237			0.4589
			C factor (dimensionless)			
Sample data	0.009	0.17091	0.05112	0.0495	0.0528	
Stratification	0.0274	0.0857	0.0502			0.6018
Col. cokriging	−0.0073	0.1331	0.0505			0.8628
Co-simulation	0.0102	0.2552	0.0512			0.7093
			K factor (t · ha · h/(ha · MJ · mm))			
Sample data	0.0125	0.0589	0.0357	0.0351	0.0363	
Stratification	0.0245	0.0408	0.0351			0.1621
Col. cokriging	0.0222	0.0547	0.0351			0.5928
Co-simulation	0.0180	0.0709	0.0354			0.4053

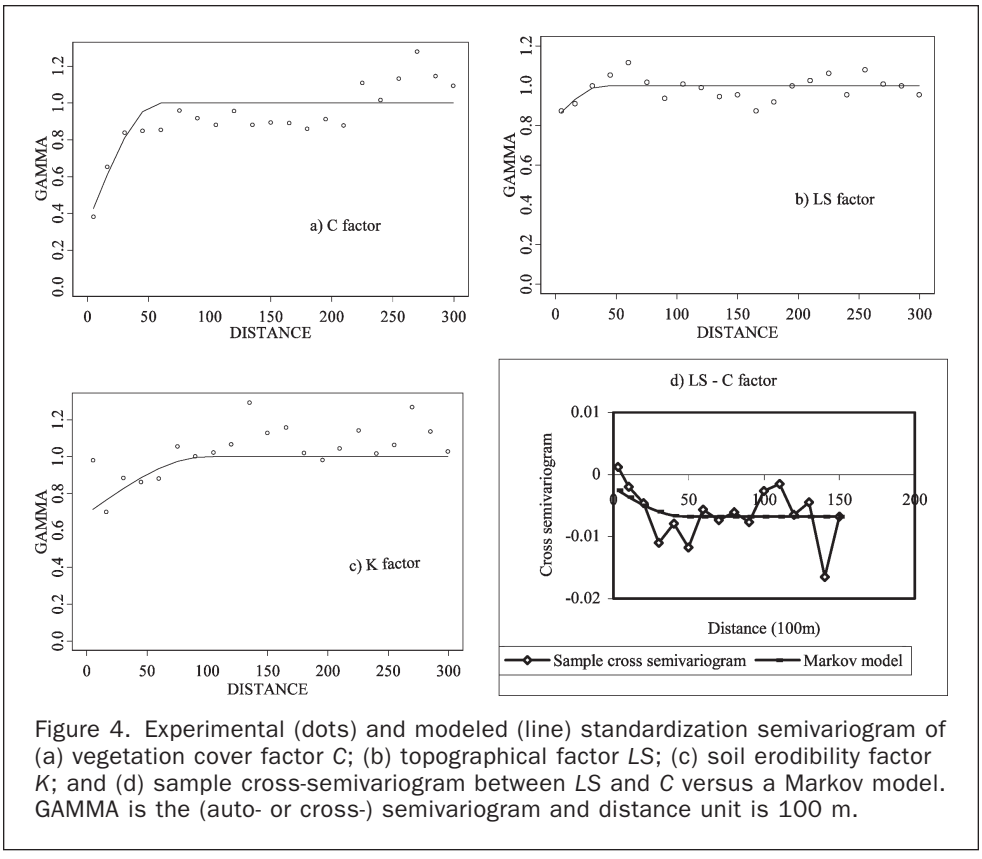

Figure 4. Experimental (dots) and modeled (line) standardization semivariogram of (a) vegetation cover factor *C*; (b) topographical factor *LS*; (c) soil erodibility factor *K*; and (d) sample cross-semivariogram between *LS* and *C* versus a Markov model. GAMMA is the (auto- or cross-) semivariogram and distance unit is 100 m.

pared to those of the sample data. All the map mean estimates for the three factors fell within their corresponding 95 percent confidence intervals, except that the stratification and cokriging led to the map mean estimates of the *LS* factor that fell out of their confidence interval. The coefficients of correlation between the estimated and observed values using two geostatistical methods varied from 0.4053 to 0.8879 and were significantly larger than those using the stratification, from 0.1542 to 0.6018. The collocated cokriging resulted in the best correlation, from 0.5928 to 0.8879.

Figure 5 shows the field plot values of factor *LS* calculated according to the RUSLE empirical equations given by Renard *et al.* (1997), and the *LS* prediction maps using the three methods. Both stratification and co-simulation methods led to similar spatial distribution of estimates to that of the observed *LS* values. Higher spatial variability of observed and estimated values for the *LS* factor was mainly located along the areas from northwest to southeast. The spatial variability was reproduced well. In addition, the spatial patterns of the *LS* values were visually consistent with those of slope values in Figure 3, which might imply the existence of spatial correlation between the slope map and *LS* factor. The predicted *LS* values varied depending on topographical characteristics, and this feature was more obviously seen on the map by the co-simulation.

The comparison of the results using the three methods and the observed values are further illustrated in Figure 6 for factor *C*. Factor *C* had large values and high spatial variability at the southwest areas, and small values and low spatial variability at the east and northeast areas. The similar spatial distribution of the *C* factor values was visually found for the three prediction maps. The spatial variability of the factor was highly related to that of the TM7 image in Figure 3. Furthermore, the cokriging and co-simulation resulted in smaller estimates than the stratification at the east and northeast areas, and the result was reverse at the southwest areas. The estimates by the cokriging and co-simulation were spatially more correlated with the observed values than were those by the stratification. The description for the comparison of three methods based on factor *C* in Figure 6 was similar for factor *K* in Figure 7.

Soil loss values for the sample plots and all the cells of the grid at a spatial resolution of 90 m by 90 m were calculated using Equation 1. The plot soil loss values were derived using the observed values of factors *LS*, *C*, and *K* for the plots. Three prediction maps of soil loss were obtained using the prediction maps of factors *LS*, *C*, and *K* for the three different methods. The *R* factor map used to calculate soil loss was the same (Figure 2).

Figure 8 shows the soil loss values based on field sample plots and the corresponding prediction maps based on three methods. The soil loss values of the sample plots was small in the east and northeast areas because of small values of factors *C* and *K*, although the values of factor *LS* were large. On the other hand, the soil erosion was high in the west areas because of high values of factors *C* and *K*, although the *LS* observed values were low. The spatial distribution of soil loss values was obviously coded into the prediction maps of soil loss using the two geostatistical methods. Relative to the geostatistical methods, the spatial pattern of soil loss values based on stratification was not so well defined. On the other hand, the stratification might result in underestimation at the west areas and overestimation at the east areas. If compared to the DEM and slope map used and shown in Figure 3, moreover, the spatial distribution of soil loss values in the prediction maps were also consistent with the topographical features. The flat, narrow, and long valleys with very small values of soil loss could be found on the maps of Figure 8. In addition, the cokriging led to

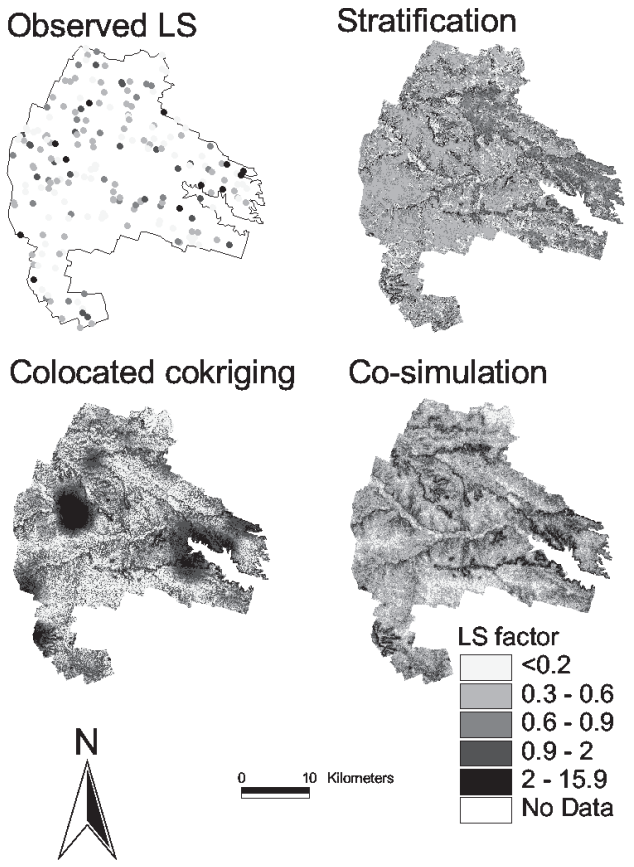

Figure 5. Comparison of prediction maps of topographical factor *LS* using stratification, collocated cokriging, and co-simulation.

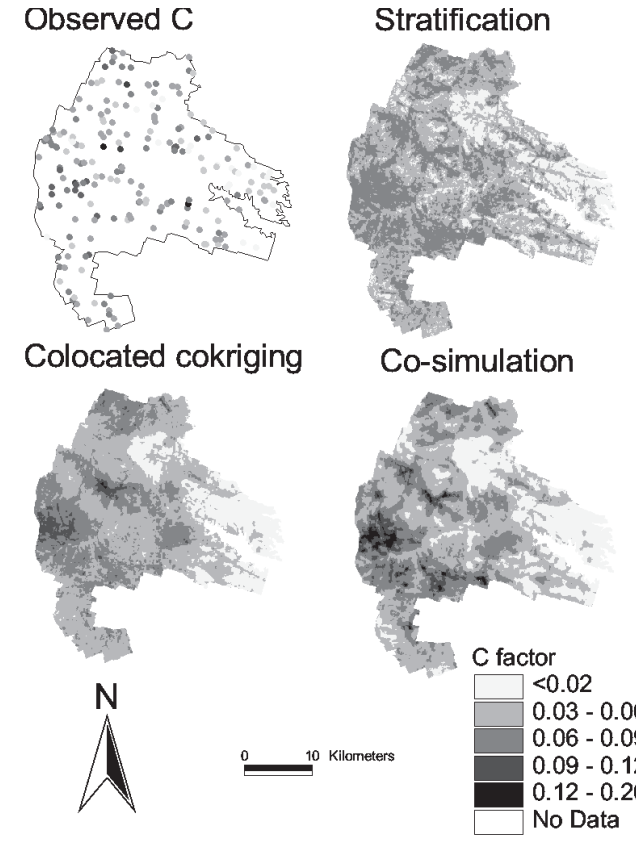

Figure 6. Comparison of prediction maps of vegetation cover factor *C* using stratification, collocated cokriging, and co-simulation.

smoothing and high estimates of soil erosion at several small areas because of the significantly large *LS* values predicted in Figure 5.

The three methods resulted in the variance maps of the estimates for factors *LS*, *C*, *K*, and soil loss. As an example, the variance maps of *LS* estimates are presented in Figure 9. The stratification and co-simulation generated a similar spatial distribution of the variances. Comparing the prediction and variance maps in Figure 5 and Figure 9, it was found that, at the areas with high spatial variability of *LS* estimates, the variances of estimates by the stratification and co-simulation were large, and at the areas with low spatial variability, the variances were small. That is, the estimation variances reproduced the spatial variability of the *LS* factor well. However, the variances of the estimates by the collocated cokriging were obviously smoothed. The variances based on collocated cokriging depended on the spatial configuration of the ground sample plots, not on the sample data values themselves. The closer the cells to be estimated were to the ground sample plots, the smaller the variances. As the cells got farther away from the ground sample plots, the variances of the cell estimates increased. The variances by the cokriging did not reflect spatial variability of the plot *LS* values and topographical features of the slope map.

The three methods were compared based on the percent root-mean-square error (Figure 10). Based on the sample data, the coefficient of variation for the *LS* factor was 187 percent. Due to this large coefficient of variation, the relative errors for the prediction of *LS* were large, varying from 100 percent to 226 percent. The corresponding errors for the *C* and *K* factors were small, from 24 percent to 38 percent and from 1 percent to 25 percent, respectively. This was due to the small coefficient variations of the factors based on the sample data. For each factor, the cokriging led to the smallest relative errors, from 1 percent to 100 percent, then co-simulation, from 24 percent to 185 percent, and stratification, from 25 percent to 226 percent.

Discussion and Conclusions

Accurately predicting annual soil erosion is very difficult because soil loss spatially varies considerably, and depends on the input factors, including rainfall-runoff *R*, soil erodibility *K*, topographical factor *LS*, vegetation cover *C*, and their interactions. This is especially true when reproducing the auto- and cross-spatial variability and correlation of the factors. Two geostatistical methods, including collocated cokriging and joint sequential co-simulation, and a traditional stratification, were compared in this study. The comparison was done in terms of the estimates of the input factors and their uncertainties, further of the estimates and uncertainties of soil loss at all locations using a sample data set, TM images, and a slope map.

The results showed that the collocated cokriging and joint sequential co-simulation were much better than the traditional stratification in terms of estimation accuracy. A reason might be because the geostatistical methods use the spatial correlation and similarity between sample data of

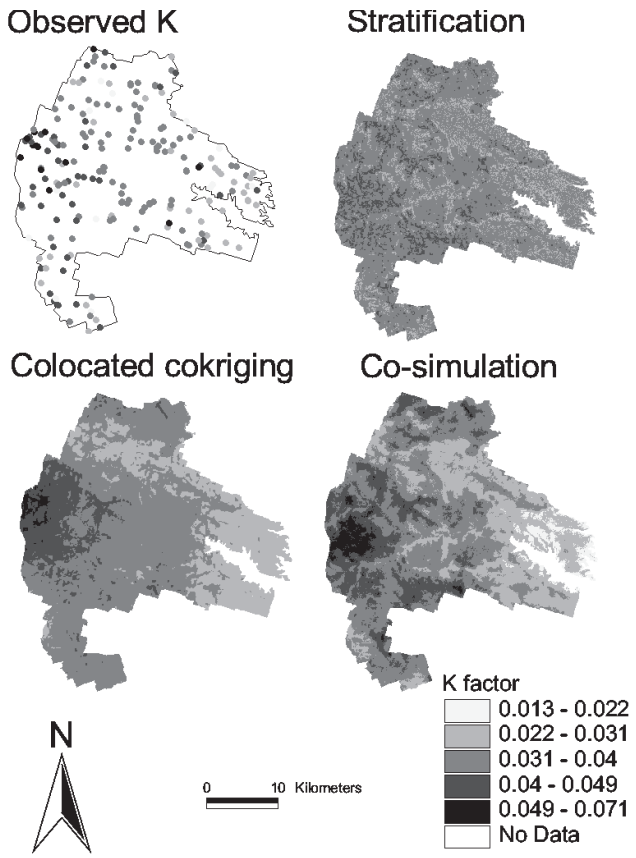

Figure 7. Comparison of prediction maps of soil erodibility factor *K* using stratification, collocated cokriging, and co-simulation. The unit for factor *K* is t · ha · h/(ha · MJ · mm).

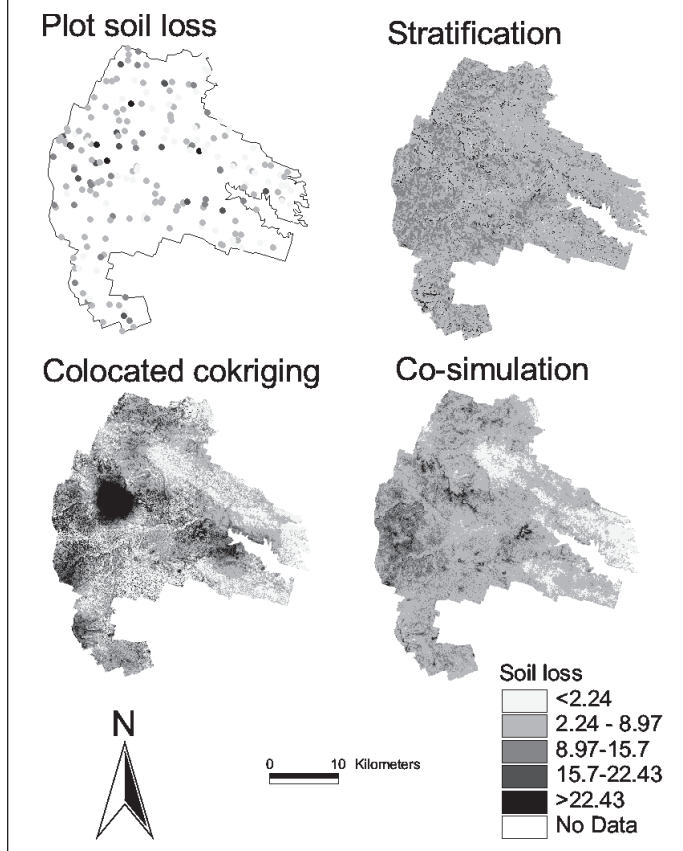

Figure 8. Comparison of prediction maps of soil loss using stratification, collocated cokriging, and co-simulation. The unit for soil loss is t/(ha · year).

variables in terms of auto and cross semi-variograms when mapping the input factors. This results in the significant improvement. The traditional stratification does not assume and apply the spatial correlation, but, at the same time, smoothes the estimates within each stratum. This leads to underestimation at the areas with large values and overestimation at the areas with small values.

Furthermore, according to the coefficient of correlation and relative root-mean-square error between the estimates and observations, the collocated cokriging was better than the co-simulation. This is because for the kriging algorithm a local estimate at any location is directly derived and is locally the best with minimum error variance. For the simulation algorithm, a number of local estimates at each location are obtained by randomly drawing from the conditional cumulative distribution function determined by the kriging above, and an expected estimate is then calculated and it is unbiased but might not be locally the best. The co-simulation is also more complicated in practice and computationally very intense. The computer time needed for the co-simulation is about 50 times that for the collocated cokriging.

However, the collocated cokriging led illogically to negative estimates at some locations. More important was that the cokriging smoothed the variance map, and the variances of local estimates depended only on the data configuration but not on the sample data themselves. That is, the variances cannot be used as uncertainty measures of estimates. The co-simulation reproduced well the spatial uncertainty of the local estimates in terms of their estimation variances. The variance map by the co-simulation realistically reflects the precision of local estimates. At the densely sampled areas with low variation of the sample data, the variances of local estimates are small, but otherwise are large. The variance map of local estimates is reliable and will be very useful in natural resources and in ecological and environmental management to assess risks in decision making.

In fact, the co-simulation jointly created a set of estimates at each location for each variable. An expected vector and covariance matrix at each unknown location can be thus derived. A probability map of each variable for the expected estimates being larger than a critical soil loss tolerance value can also be calculated. The probability maps are applicable for making management plans such as delineating conservation areas of soil and water. The covariance maps together with the variance maps can be employed as the input information for spatial uncertainty analysis in which the relative variance contributions of each variable and its interactions with other variables to the total uncertainty of predicting soil loss can be determined. The uncertainty analysis thus provides decision-makers with useful information to assess the error sources of the decisions being made. This is deemed to be an advantage of the co-simulation compared to the collocated cokriging and traditional stratification.

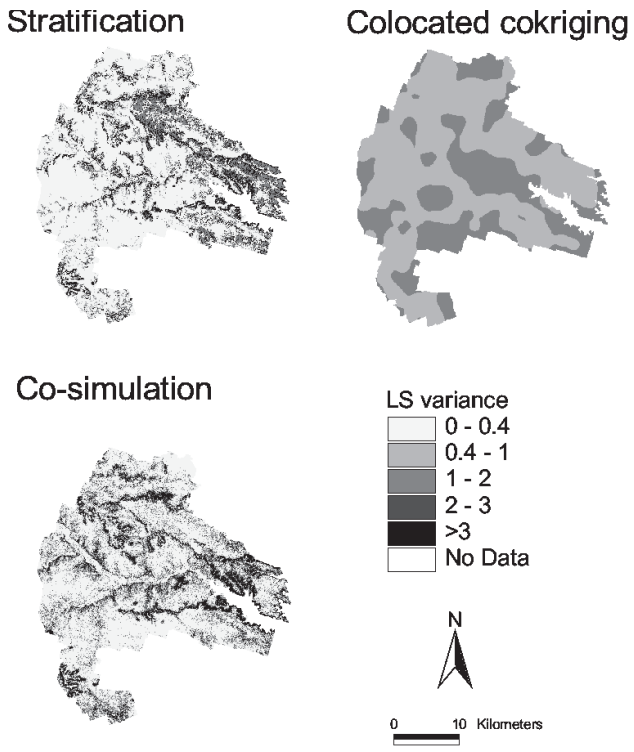

Figure 9. Comparison of variance maps of *LS* factor estimates using stratification, collocated cokriging, and co-simulation.

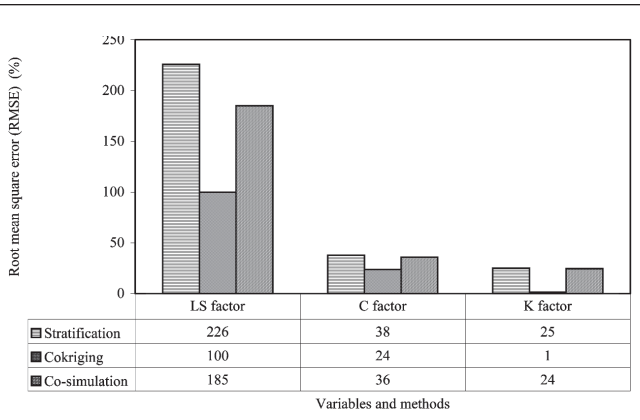

Figure 10. Comparison of stratification, collocated cokriging, and co-simulation based on percent root-mean-square error (RMSE).

Acknowledgment

We are grateful to SERDP (Strategic Environmental Research and Development Program) for providing support for this study, and to the U.S. Army Corps of Engineers, Construction Engineering Research Laboratory (USA-CERL) for the data sets.

References

Almeida, A.S., 1993. *Joint Simulation of Multiple Variables with a Markov-Type Coregionalization Model*, Ph.D. dissertation, Department of Applied Earth Sciences, Stanford University, Stanford, California, 199 p.

Almeida, A.S., and A. Journel, 1994. Joint simulation of multiple variables with a Markov-type coregionalization model, *Mathematical Geology*, 26:565–588.

Barata, M.T., M.C. Nunes, A.J. Sousa, F.H. Muge, and M.T. Albuquerque, 1996. Geostatistical estimation of forest cover areas using remote sensing data, *The Fifth Geostatistics Congress, Geostatistics Wollongong '96 2* (E.Y. Baafi and N.A. Schofield, editors), 22–27 September, Wollongong, Australia (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp. 1244–1257.

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(1):74–78.

Biesemans, J., M.V. Meivenne, and D. Gabriels, 2000. Extending the RUSLE with Monte Carlo error propagation technique to predict long-term average off-site sediment accumulation, *Journal of Soil and Water Conservation*, First Quarter:35–42.

Campbell, J.B., 1996. *Introduction to Remote Sensing*, A Division of Guilford Press, Inc., New York, N.Y., 622 p.

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*, Oxford University Press, Inc., New York, N.Y., 483 p.

Hunner, G., H.T. Mowrer, and R.M. Reich, 2000. An accuracy comparison of six spatial interpolation methods for modeling forest stand structure on the Fraser Experimental Forest, Colorado, *Accuracy 2000 Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences* (G.B.M. Heuvelink and M.J.P.M. Lemmens, editors), 12–14 July, Amsterdam, The Netherlands (Delft University Press, Delft, The Netherlands), pp. 305–312.

Mitášová, H., J. Hofierka, M. Zlocha, and L.R. Iverson, 1996. Modelling topographic potential for erosion and deposition using GIS, *International Journal of Geographical Information Science*, 10(5):629–641.

Poso, S., G. Wang, and S. Tuominen, 1999. Weighting alternative estimates when multisource auxiliary data for forest inventory, *Silva Fennica*, 33(1):41–50.

Price, D.L., A.B. Anderson, W.R. Whitworth, and P.J. Guertin, 1995. *Land Condition Trend Analysis Data Summaries: Preliminary Data Applications*, USACERL Technical Report 95/39, U.S. Army Corps of Engineers, Construction Engineering Research Laboratory, Champaign, Illinois, 86 p.

Renard, K.G., and V.A. Ferreira, 1993. RUSLE model description and database sensitivity, *Journal of Environmental Quality*, 22(3):458–466.

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),* USDA, Agriculture Handbook Number 703, Government Printing Office, Washington, D.C., 404 p.

Risse, L.M., M.A. Nearing, A.D. Nicks, and J.M. Laflen, 1993. Error assessment in the Universal Soil Loss Equation, *Soil Science Society of America Journal*, 57(3):825–833.

Schertz, D.L., 1983. The basis for soil loss tolerances, *Journal of Soil and Water Conservation*, 38:10–14.

Steele, B.M., J.C. Winne, and R.L. Redmond, 1998. Estimation and mapping of misclassification probabilities for thematic land cover maps, *Remote Sensing of Environment*, 66:192–202.

Tazik, D.J., S.D. Warren, V.E. Diersing, R.B. Shaw, R.J. Brozka, C.F. Bagley, and W.R. Whitworth, 1992. *U.S. Army Land Condition Trend Analysis (LCTA) Plot Inventory Field Methods*, USACERL Technical Report N-92/03, U.S. Army Corps of Engineers, Construction Engineering Research Laboratory, Champaign, Illinois, 62 p.

Tomppo, E., 1996. Multi-source national forest inventory of Finland, *New Trusts in Forest Inventory — Proceedings of the Subject Group S4.02–00 "Forest Resource Inventory and Monitoring' and Subject Group S4.12–00 'Remote Sensing Technology', Vol. I, IUFRO XX World Congress* (R. Paivinen, J. Vanclay, and S. Miina, editors), 06–12 August 1995, Tampere, Finland, pp. 27–42.

Wang, G., G.Z. Gertner, P. Parysow, and A.B. Anderson, 2000. Spatial prediction and uncertainty analysis of topographical

factors for the Revised Universal Soil Loss Equation (RUSLE), *Journal of Soil and Water Conservation*, 55(3):373–382.

Wang, G., G.Z. Gertner, X. Liu, and A.B. Anderson, 2001a. Uncertainty assessment of soil erodibility factor for revised universal soil loss equation, *CATENA*, 46:1–14.

Wang, G., G.Z. Gertner, S. Wente, and A.B. Anderson, 2001b. Vegetation classification and accuracy assessment using image-aided sequential indicator co-simulation, *Conference Proceedings of ASPRS 2001 — Gateway to the New Millennium*, 23–27 April, St. Louis, Missouri (American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland), unpaginated CD-ROM.

Wang, G., G.Z. Gertner, V. Singh, S. Shinkareva, P. Parysow, and A.B. Anderson, 2002. Spatial and temporal prediction and uncertainty of soil loss using revised universal soil loss equation: A case study in rainfall and runoff erosivity for soil loss, *Ecological Modeling*, 153:143–155.

Wischmeier, W.H., and D.D. Smith, 1978. *Predicting Rainfall Erosion Loss: A Guide to Conservation Planning*, USDA, Agriculture Handbook Number 537, Government Printing Office, Washington, D.C., 58 p.

(Received 20 August 2001; accepted 17 October 2002; revised 08 November 2002)