

Spatial prediction and uncertainty assessment of topographic factor for revised universal soil loss equation using digital elevation models

Guangxing Wang^{a,*}, George Gertner^a, Pablo Parysow^b, Alan Anderson^c

^a Department of Natural Resources and Environmental Sciences, NRES, University of Illinois at Urbana-Champaign, W503 Turner Hall, 1102 S. Goodwin Avenue, Urbana, IL 61801, USA

^b School of Forestry, Northern Arizona University, Flagstaff, AZ, USA

^c USACERL, P.O. Box 9005, Champaign, IL, USA

Received 9 November 1999; accepted 15 May 2001

Abstract

Revised Universal Soil Loss Equation (RUSLE) is a model to predict longtime average annual soil loss, related to rainfall-runoff, soil erodibility, slope length and steepness, cover management, and support practice. The product of slope length L and steepness S is called topographic factor LS , implying the topographic effect on soil loss. This study focuses on (a) spatially predicting the topographic factor LS for RUSLE using a Digital Elevation Model (DEM), (b) selecting the appropriate DEM spacing for predicting the LS factor, and (c) modeling the loss of spatial variability of the predicted LS factor due to DEM resampling. The results show that using the physically based topographical factor LS equation and DEMs led to a higher correlation of predicted LS values with topographical features, compared to a spatial simulation method based on LS empirical models and sample data. The appropriate DEM spacing required to achieve prediction precision and detailed spatial variability of the LS factor was not identical for both requirements and a compromise may be made depending on the application aims. By modeling the spatial variability of predicted LS values for different DEM spacing, a new method to directly measure loss of spatial variability due to data resampling was developed. Compared to measures of entropy and global variance, the new method can reveal the different losses of spatial variability in different directions when the spatial variability is anisotropic. Published by Elsevier Science B.V.

Keywords: DEM spacing; Digital elevation model; Geostatistics; Soil loss; Spatial variability; Topographic factor; Uncertainty

1. Introduction

Using the Revised Universal Soil Loss Equation (RUSLE), the longtime average annual soil loss can be predicted as a product of rainfall-runoff erosivity

factor R , soil erodibility factor K , slope length factor L , slope steepness factor S , cover management factor C , and support practice factor P (Renard et al., 1997). The slope steepness factor (S) is a function of slope angles measured in degrees and reflects the influence of slope gradient on erosion. The slope length factor (L) is a function of slope length measured in meters. Soil erosion increases as slope length and steepness increase, and it is more sensi-

* Corresponding author. Tel.: +1-217-244-4603; fax: +1-217-244-3219.

E-mail address: wang12@uiuc.edu (G. Wang).

tive to slope steepness than to slope length. Their product, called the topographic factor *LS*, represents the topographic effect as the ratio of soil loss on a given slope length and steepness to soil loss from a slope that has a length of 22.13 m, and a steepness of 9% where all other conditions are the same. The topographic factor is the most sensitive in the prediction of soil loss (Renard and Ferreira, 1993; Risse et al., 1993).

In RUSLE, *L* and *S* are calculated using a set of empirical models (Renard et al., 1997). With a field sample, Wang et al. (2000) carried out spatial prediction and uncertainty analysis of the *LS* factor derived from the empirical models using geostatistical methods. When soil loss is estimated for large areas as part of a geographic information system (GIS) for converging and diverging terrain, the empirical models did not differentiate net erosion and those areas experiencing net deposition. A physically based topographical factor *LS* equation has thus been developed based on a digital elevation model (DEM) (Moore and Burch, 1986; Moore and Wilson, 1992). However, the precision for predicting the *LS* factor is related to the DEM accuracy and spacing, and the methods to derive topographical variables are related to *LS*. For example, Mitášová et al. (1996) investigated this approach by interpolating DEMs to finer spacing, and suggested that the commonly used 30-m spacing USGS DEMs are insufficient.

In addition to prediction precision, capturing the spatial variability with an appropriate DEM spacing is necessary to accurately represent the spatial characteristics of the *LS* factor. Recently, DEMs with different spacings have become readily available and lead to the problem of choosing an appropriate DEM spacing for a given task. Furthermore, the appropriate DEM spacing may be a function of other variables such as the complexity of the terrain, required precision, desired information, etc. Choosing an appropriate DEM spacing thus becomes very important.

Variance-based techniques have been widely used to determine appropriate cell spacing in remote sensing for mapping in natural resource inventory and environmental monitoring (Townshend and Justice, 1988; Marceau et al., 1994a,b). A local variance method to choose an appropriate cell spacing for mapping was developed by Woodcock and Strahler

(1987) based on the relationship between cell spacing and spatial dependence. In the studies by Atkinson and Danson (1988), Atkinson and Curran (1995, 1997) and Atkinson (1997), the appropriate cell spacing was determined based on the relationship between the spatial dependence of a variable and the cell spacing using the semivariance at a lag of one cell. This method can lead to the regularised semivariogram and the maximum semivariance at a lag of one cell for any size of support.

Spatial prediction of a variable implies that estimates of this variable are derived at any locations or sub-areas. An accurate spatial prediction provides not only an unbiased population estimate (mean global estimate) but also reliable local estimates. The unbiased and reliable estimates are related to requirements for prediction precision and detailed spatial information (spatial variability). If the objective is to derive reliable local estimates, the cell spacing of the dataset used to derive the estimates should be optimised for obtaining detailed spatial information. If the aim is to derive unbiased mean global estimates, precision of prediction becomes very important. The precision depends on the spatial variation of the dataset and measurement errors of the variable. If both global and local estimates are sought for, prediction precision and detailed spatial information of the estimates should be considered simultaneously. The desired prediction precision and detailed spatial information are thus the criteria for determining appropriate cell spacing for spatial prediction. These criteria can be applied to determine the appropriate spacing of the DEM used for spatial prediction of topographical factor *LS* related to soil loss.

Additionally, in geostatistics, a semivariogram function measures spatial variability of a variable. The spatial variability implies the average dissimilarity between data separated by a vector or distance given a direction. When a sample semivariogram is fitted using one of the three models commonly used (spherical, Gaussian and exponential), the nugget parameter of the models, called the nugget variance, indicates the measurement error and spatial variability within a support (Curran and Dungan, 1989). The larger the support, the smaller the measurement error and the smaller the spatial variability within the support. The support is related here to the DEM spacing. Thus, nugget variance of the semivariogram

estimates the random noise, consisting of the measurement error, and within-cell variability. Data resampling can reduce the nugget variance. Therefore, the semivariogram method may be used to investigate the appropriate DEM spacing for prediction of the *LS* factor.

Moreover, DEM variability affects slope steepness, slope length, and watershed area. Data resampling from finer DEM spacing to coarser may lead to loss of spatial information in these variables, and thus causes errors that can propagate to the *LS* factor. Vieux (1995) used entropy theory to measure the loss of spatial information. De Cola (1997) studied multi-resolution covariance and modeled error propagation using the variance difference between two sizes of cell spacing. However, the uncertainty and error propagation may vary over space due to anisotropy of spatial variability and should be modeled in terms of semivariograms in different directions.

The objective of this study is to develop an approach to spatially predict the topographical factor *LS* using DEM, to choose the appropriate DEM spacing and to assess the loss of the spatial variability due to data resampling. The spatial prediction is first made using the physically based topographical factor *LS* equation by Moore and Burch (1986) and Moore and Wilson (1992) for different DEM spacing. The uncertainty of the spatial prediction (Chilès and Delfiner, 1999; Myers, 1997), including the variance, spatial variability, entropy and relationship of estimates with DEM spacing, is then investigated. The spatial variability of the predicted *LS* values is modeled. The effect of DEM spacing on the uncertainty of estimates is studied. The appropriate DEM spacing is analysed. A new method to model the loss of spatial variability caused by data resampling is presented and compared to existing loss measures using entropy and global variance.

2. Case study area and datasets

The study area is located in Central Texas in Bell and Coryell Counties approximately 256 km southwest of Dallas, TX. Its area is approximately 87,890 ha (Wang et al., 2000). The landscape exhibits a

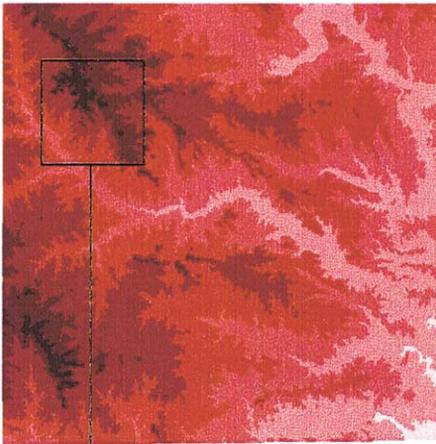
stair-step topography consisting of a gently rolling to rolling dissected remnant plateau. Higher elevations occur at the northwest portions of the study area and the lowest ones at the southeast (Fig. 1). A strip of territory with steeper slopes goes from the northwest to the southeast. In the central lower area, there is a flat sub-region. Belton Reservoir and Stillhouse Hollow Reservoir are close to the southeast border and the southeast corner, respectively. Surface water drains mostly in an easterly direction.

A 7-min DEM with 30-m spacing and vertical resolution of 1 m for this area was acquired from the US Geological Survey (Fig. 1). This DEM was classified as Level-2. The root mean square error in elevation was 5.13 m. Slope ranges from 0° to 38° with an average of 2.94° and steep slopes mainly occur as bluffs along the flood plain and at the slopes of the mesa-hills. Using nearest neighbour resampling, new datasets with 50-, 100-, 200- and 400-m spacings were derived from the original DEM. Moreover, a small area of 10 by 10 km was selected at the northwest where there is high variation of the predicted *LS* values (see Fig. 1). For the small area, datasets with 20-, 10- and 5-m spacings were created from the original DEM using an interpolation method called regularised spline with tension (Mitášová and Mitáš, 1993).

3. Methodology

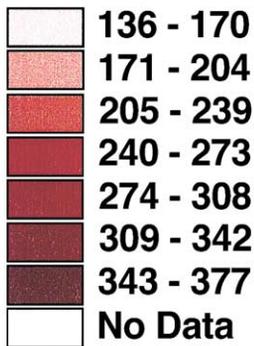
In this study, the five DEMs mentioned in Section 2 were used to derive estimates of slope, aspect, flow-line length, flow-line density, and up-slope contributing area using the Geographic Resources Analysis Support System (GRASS, 1993) GIS. The *LS* factor was calculated according to a physically based topographical factor *LS* equation (Moore and Burch, 1986; Moore and Wilson, 1992). The spatial variability of the *LS* grids was then presented by plotting semivariogram values with data separation distances in different directions using the Geostatistical Software Library (GSLIB) (Deutsch and Journel, 1998). The spatial variability structures were modeled using spherical, exponential and Gaussian models for different DEM spacing using S-Plus (Kaluzny et al., 1998).

DEM



A small area

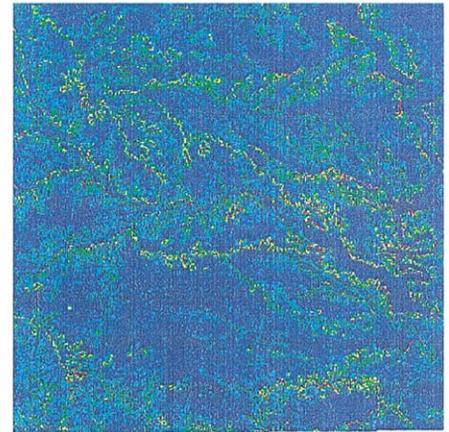
DEM (m)



0 7000 Meters



Slope



Slope (degree)

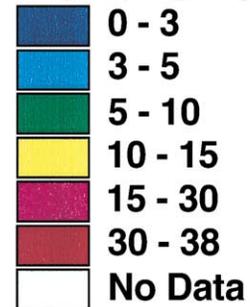


Fig. 1. The original DEM with 30-m spacing and its slope map. A small area at the northwest, used in other tests below, is outlined in the DEM map.

Furthermore, for each DEM spacing, global variance, mean local variance (in a 3×3 neighbourhood) and mean semivariance at a lag of one cell were derived from the predicted *LS* grids. The results were used together with the modeled semivariograms to choose the appropriate DEM spacing. In addition to entropy and global variance of estimates, a method to model the loss of spatial variability due to data resampling was derived. The relationships between losses of global variance, spatial variability and entropy were discussed. A program developed by the authors was used to calculate the relative

losses. ArcView GIS (Hutchinson and Daniel, 1997) was employed to display the raster data.

3.1. Calculation of topographic factor *LS*

The physically based topographical factor *LS* equation to estimate the *LS* factor is:

$$LS = \left[\frac{A}{22.13} \right]^m \left[\frac{\sin \beta}{0.0896} \right]^n \quad (1)$$

where m and n are constants equal to 0.6 and 1.3, respectively; β is the land surface slope in degrees, A is the up-slope contributing area per unit width of cell spacing [$\text{m}^2 \text{m}^{-1}$] from which the water flows into a given grid cell. The area A for a given grid cell is calculated from the sum of the grid cells from which the water flows into the cell, as follows (Mitášová et al., 1996):

$$A = \frac{n\mu a}{b} \quad (2)$$

where a is the area of a grid cell; n is the number of cells draining into the cell; μ is the weight depending on the runoff generation mechanism and infiltration rates; and b is the cell spacing. With uniform rainfall and infiltration in this study area, the weight μ was assumed to be one (Mitášová et al., 1996). Because a is constant for a specific cell spacing, $a = bb$. Thus, $A = nb$. In practice, A can be approximated by multiplying the down-slope flow-line density with the DEM spacing.

For each DEM spacing, the LS factor was predicted using Eq. (1) for each cell of the study area and the variance of the LS factor was also derived for each cell using the predicted LS values within a window. The cell for calculating variance was located at the center of the window and the size of window was related to the separation distance by which the cell values tend to be independent. Additionally, the grids of predicted LS values and their variances derived using this method above were compared to those created using LS empirical models and sample data with a spatial simulation geostatistical method. For the empirical models, the sample data and the geostatistical method, readers may refer to Wang et al. (2000).

3.2. Spatial variability and appropriate DEM spacing

For spatial prediction and uncertainty assessment of the LS factor using DEM, it is very important to choose an appropriate DEM spacing and to capture spatial variability. Ideally, the DEM spacing should be determined such that the desired precision of estimates and detailed spatial information can be simultaneously achieved using the least data. Since

data resampling leads to reduction of data noise, the global variance of predicted values generally decreases as DEM spacing becomes coarser. The decrease is rapid at the beginning, then slows, and finally the variance stabilizes. Given the precision requirements, the global variance and the appropriate DEM spacing for the desired precision can be determined. Data resampling, on the other hand, results in loss of spatial information. The appropriate DEM spacing for the desired precision may be too coarse to capture and allow interpretation of spatial variability. Thus, the appropriate DEM spacing for each of these two requirements may differ. In this study, the global variance and mean local variance of the whole grid and semivariance at a lag of one cell are integrated to search for an appropriate compromise in DEM spacing.

A local variance (Woodcock and Strahler, 1987) is defined as the variance within a 3×3 window. The mean local variance for a grid is taken to be the mean of the local variances through the entire grid with the exception of the border grid cells. Given a method of data resampling for spatial prediction of the LS factor, if the DEM spacing is considerably finer than the appropriate DEM spacing to capture the spatial variability of LS , the local variances (or mean local variance) will generally be low because of high correlation between the cell values in the 3×3 neighbourhood. If the DEM spacing approximates the appropriate DEM spacing, the local variances tend to be higher because of lower correlation between the cell values. When the DEM spacing is further increased, the local variances start to decrease again because of similarity between the neighbours. The maximum in the mean local variance as a function of DEM spacing is an indication of the appropriate DEM spacing to capture spatial variability of the LS factor. It can thus be used to choose an appropriate DEM spacing for spatial prediction of the LS factor.

On the other hand, spatial variability of a variable can be modeled as a realization of a random function (Atkinson and Curran, 1997). Let $Z(x)$ be a random function defined for position x in two-dimensional space as $Z(x) = m_v + e(x)$ with m_v as the local mean of Z in a neighbourhood V , and $e(x)$ as the random function with a zero mean. With the so-called intrinsic hypothesis of stationarity, a semivariogram

$\gamma(h)$ measures the average dissimilarity between data separated by a lag distance h given a direction:

$$\begin{aligned}\gamma(h) &= \frac{1}{2} \text{var}[Z(x) - Z(x+h)] \\ &= \frac{1}{2} E[Z(x) - Z(x+h)]^2\end{aligned}\quad (3)$$

The semivariogram is usually unknown and can be approximated with an experimental semivariogram obtained by sampling. Assume that the variable Z is sampled in the study area and n data $z(\mathbf{u}_\alpha)$ are obtained, where $\alpha = 1, 2, \dots, n$ and \mathbf{u}_α is the vector of spatial coordinates of the α th sample. The experimental semivariogram is calculated as half the average squared difference between the components of every data pair:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} (z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + h))^2 \quad (4)$$

where $N(h)$ is the number of data pairs used, h is the distance separating two sample values, $z(\mathbf{u}_\alpha)$ and $z(\mathbf{u}_\alpha + h)$ are two samples at locations \mathbf{u}_α and $\mathbf{u}_\alpha + h$ separated by a lag distance of h .

When the lag distance is equal to the cell spacing, the value obtained by Eq. (4) is the semivariance at a lag of one cell (Atkinson and Danson, 1988). The relationship between the cell spacing and the semivariance at a lag of one cell is similar to that between the cell spacing and local variance mentioned above. At the beginning, the semivariance increases with increased cell spacing. At some cell spacing, the spatial variability of the variable is captured best and the semivariance at a lag of one cell reaches a maximum value, then it decreases again and becomes stable. The semivariance at a lag of one cell is thus an indication of the appropriate DEM spacing to capture the desired spatial variability of the LS factor.

3.3. Modeling spatial variability and uncertainty

The experimental semivariograms should be checked for detecting anisotropy in different directions such as azimuth 0° , 45° , 90° , and 135° . If they are isotropic, an omnidirectional experimental semivariogram can be calculated and is often fitted using

spherical, exponential or Gaussian models. Otherwise, the experimental semivariograms should be fitted separately in different directions. Eqs. (5), (6) and (7) listed below are the spherical, exponential and Gaussian models, respectively, with nugget effects:

$$\begin{aligned}\hat{\gamma}(h) &= \begin{cases} c_0 + c_1 \left[1.5 \left(\frac{h}{a_0} \right) - 0.5 \left(\frac{h}{a_0} \right)^3 \right] & 0 \leq h < a_0 \\ c_0 + c_1 & h \geq a_0 \end{cases}\end{aligned}\quad (5)$$

$$\hat{\gamma}(h) = \begin{cases} c_0 + c_1 \left[1 - \exp\left(-\frac{3h}{a_0}\right) \right] & 0 \leq h < a_0 \\ c_0 + c_1 & h \geq a_0 \end{cases}\quad (6)$$

$$\hat{\gamma}(h) = \begin{cases} c_0 + c_1 \left[1 - \exp\left(-\frac{(3h)^2}{a_0^2}\right) \right] & 0 \leq h < a_0 \\ c_0 + c_1 & h \geq a_0 \end{cases}\quad (7)$$

where c_0 and c_1 are the nugget variance and structure variance, respectively, and $c = c_0 + c_1$ is the sill variance; a_0 is the actual range parameter for the spherical model. For the exponential and Gaussian models, a_0 is the effective range parameter. The effective range is defined as the distance at which $\hat{\gamma}(a_0) = 0.95c$. When $c_0 = 0$, the equations above represent pure spherical, exponential and Gaussian models.

The nugget variance c_0 of a semivariogram can be inferred by the y -intercept of the fitted model and arises from measurement error variance and micro-scale variance (Atkinson, 1997; Goovaerts, 1997). When the experimental semivariograms are calculated using raster data, the nugget variance implies a

noise term, that is, measurement error variance and within-cell variability. The sill variance c implies the variance of the raster data. The nugget variance and sill variance generally decrease as the cell spacing increases. The range parameter a_0 implies the distance when spatial dependence disappears, and it tends to increase as the cell spacing increases. In this study, these parameters for the grids of predicted LS values were modeled and used together to investigate the relationship of spatial variability of predicted LS values with the DEM spacing.

One could resample a DEM to a coarser spacing and then compute all these quantities related to the LS factor for another spacing, or one could compute LS for the original DEM and then resample LS to a coarser spacing. The objective of this study is, however, to investigate the appropriate DEM spacing from which the LS factor can be derived. The second method may not directly result in the appropriate DEM spacing; thus, the first method was used in this study. The loss of spatial variability of elevation due to data resampling is propagated to the calculation of slope and up-slope contributing area, and thus to prediction of the LS factor. The loss of spatial variability was modeled using the measures of entropy, global variance and semivariogram.

The theory of entropy was used to measure loss of information content (Vieux, 1995). Let I denote the entropy, B the number of discrete intervals of the LS factor, and P_i the probability of the LS factor occurring within the interval i . The entropy of grid data at a specific spacing is (Shannon and Weaver, 1964):

$$I = - \sum_{i=1}^B P_i \log(P_i) \quad (8)$$

Measuring the entropy of the LS factor for different DEM spacing provides an estimate of the rate of information content loss due to the data resampling. The entropy is also a measure of spatial variability when applied to raster data. The entropy is additionally consistent with the variance of raster data. When the topographic surface is a plane, the elevation is a constant and the probability P_i is 1.0. This means zero entropy, zero information content, and zero uncertainty. The higher entropy is associated with surfaces of highly variable elevation or of uniform slope. The loss of entropy from the original DEM

spacing v_0 to a DEM spacing v_k due to data resampling can be represented as:

$$\Delta I(v_0 \rightarrow v_k) = I_{v_0} - I_{v_k} \quad (9)$$

Moreover, the variance change after data resampling is:

$$\Delta \sigma^2(v_0 \rightarrow v_k) = \sigma^2(v_0) - \sigma^2(v_k) \quad (10)$$

where $\sigma^2(v_0)$ and $\sigma^2(v_k)$ are the variance of LS raster data for DEM spacing v_0 and v_k , respectively.

If the spatial variability of the predicted LS factor for two different sizes of DEM spacing is modeled using either Eq. (5), Eq. (6) or Eq. (7), the difference of the modeled semivariograms between two sizes of DEM spacing at any lag indicates implicitly the loss of spatial variability. The integral of the differences over the grid is thus the total loss of the spatial variability for the whole grid, and can be denoted as:

$$\Delta \gamma_{v_0 \rightarrow v_k} = \iint [\gamma_{v_0}(h) - \gamma_{v_k}(h)] d(h) d(az) \quad (11)$$

where az is an azimuth at which the semivariogram is derived. Given a specific direction, the integral can be evaluated over distance h :

$$\Delta \gamma_{v_0 \rightarrow v_k} = \int_0^H [\gamma_{v_0}(h) - \gamma_{v_k}(h)] dh \quad (12)$$

where H is the maximum distance in a specific direction. When a semivariogram model is selected, the equation for loss of spatial variability is specific. Using the spherical model, for example, the loss of spatial variability can be represented as:

$$\begin{aligned} \Delta \gamma_{v_0 \rightarrow v_k} = & \left\{ \Delta C_0 h + \Delta C_1 h^2 - \Delta C_2 h^4 \right\} \Big|_0^{\min(a_{00}, a_{k0})} \\ & + \left\{ \Delta C_0 h + c_{01} h - \Delta C_3 h^2 \right. \\ & \left. + \Delta C_4 h^4 \right\} \Big|_{\min(a_{00}, a_{k0})}^{\max(a_{00}, a_{k0})} \\ & + \left\{ \Delta C_5 h \right\} \Big|_{\max(a_{00}, a_{k0})}^H \end{aligned} \quad (13)$$

where:

$$\begin{aligned} \Delta C_0 &= c_{00} - c_{k0} & \Delta C_1 &= \frac{1.5(c_{01} a_{k0} - c_{k1} a_{00})}{2 a_{00} a_{k0}} \\ \Delta C_2 &= \frac{0.5(c_{01} a_{k0}^3 - c_{k1} a_{00}^3)}{4 a_{00}^3 a_{k0}^3} & \Delta C_3 &= \frac{1.5 c_{k1}}{2 a_{k0}} \\ \Delta C_4 &= \frac{0.5 c_{k1}}{4 a_{k0}^3} & \Delta C_5 &= c_{00} - c_{k0} + c_{01} - c_{k1} \end{aligned}$$

and c_{00} , c_{01} , and a_{00} are the nugget variance, structure variance and actual range of influence of the

semivariogram for the original DEM spacing (v_0), respectively; and c_{k0} , c_{k1} , and a_{k0} the same quantities for the new DEM spacing (v_k).

Eqs. (9) (10) and (11) generally indicate the entropy loss, global variance loss, and spatial variability loss from the original DEM spacing v_0 to a new spacing v_k . If the losses are divided by the corresponding entropy, global variance and semivariogram for the original DEM spacing v_0 , the relative losses are obtained. Furthermore, the relative losses of spatial variability and global variance can be plotted against the relative entropy loss.

4. Results

The grids of the predicted LS factor obtained using the physically-based topographical factor LS equation and different DEM spacing are shown in Fig. 2. For comparison, a predicted LS grid derived using LS empirical models with sample data and a spatial simulation method are also presented in the lower right of Fig. 2. The outline indicates the boundary of the area where slope and slope length data in 219 field sample areas were used.

When the physically based topographical factor LS equation and DEMs were used (Fig. 2), at the areas from northwest to southeast and at the corner of southwest, high LS factor values were derived because of the hilly terrain. The central-south area is flat, thus, low LS estimates were predicted. Furthermore, the lakes at the southeast were filled with estimates of zero. The features above are similar when the DEM spacing is less than 100 m, slightly different between 30- and 200-m DEM spacings, and very different between 30- and 400-m DEM spacings. The finer the DEM spacing, the more detailed the spatial variability.

A geostatistical spatial simulation algorithm was used to estimate the LS values at the unknown locations using the empirical models and sample data (lower right of Fig. 2). The resulting LS grid had a 100-m spacing. Within this grid, high and low LS values were obtained respectively in the hilly and flat areas. However, most of the areas had no sample data and were filled with the predicted values close to the average of the sample data. The topographical

features affecting LS values cannot obviously be observed in this grid. For example, this grid does not show the areas filled with zero, indicating flat areas and lakes, but the other grids in Fig. 2 do show these features.

The variance grids of the predicted LS values for different DEM spacing were computed by the window method mentioned in Section 3.1 and are shown in Fig. 3. The window size used was 800 by 800 m, i.e. the size in cells varied according to the DEM spacing. For comparison, the variance grid of predicted LS values created by LS empirical models with sample data and a spatial simulation is also given in the lower right of Fig. 3. The grid spacing was 100 m.

When the physically based topographical factor LS equation and DEMs were used (Fig. 3), large variances are spatially distributed along the hilly areas from northwest to southeast, and at the southwest corner. Small variances are mainly found at the central-south of the flat area and the lake areas. The features of variance spatial distribution correspond to topographic features of the area and the spatial variability of the predicted LS factor. The relationship of the DEM spacing to the predicted LS values described in Fig. 2 also applies to the variances. For the variance grid of the empirical models in the lower right of Fig. 3, the relationship of the variances with the topography was not distinct. In this case, the variances depended on the sample values and data configuration, not on topography. In the areas with lower LS values and denser samples, the variances are smaller, and vice versa. Outside of the sampling area, the variances are high.

The experimental semivariograms of the predicted LS grid were derived in four directions using the original 30-m DEM spacing (Fig. 4). The semivariograms have the similar development trend over distance in all four directions, and they increase rapidly from 0 to 400 m, then slowly from 400 to about 1200 m, and stabilize after 1200 m. The nugget variances are different in different directions. The experimental semivariograms for 0° azimuth and different DEM spacing are presented in Fig. 5. These semivariograms are similar in shape, and may be modeled using the same spherical model. For the same lag distance, the semivariogram values decrease from finer to coarser spacing. The semivari-

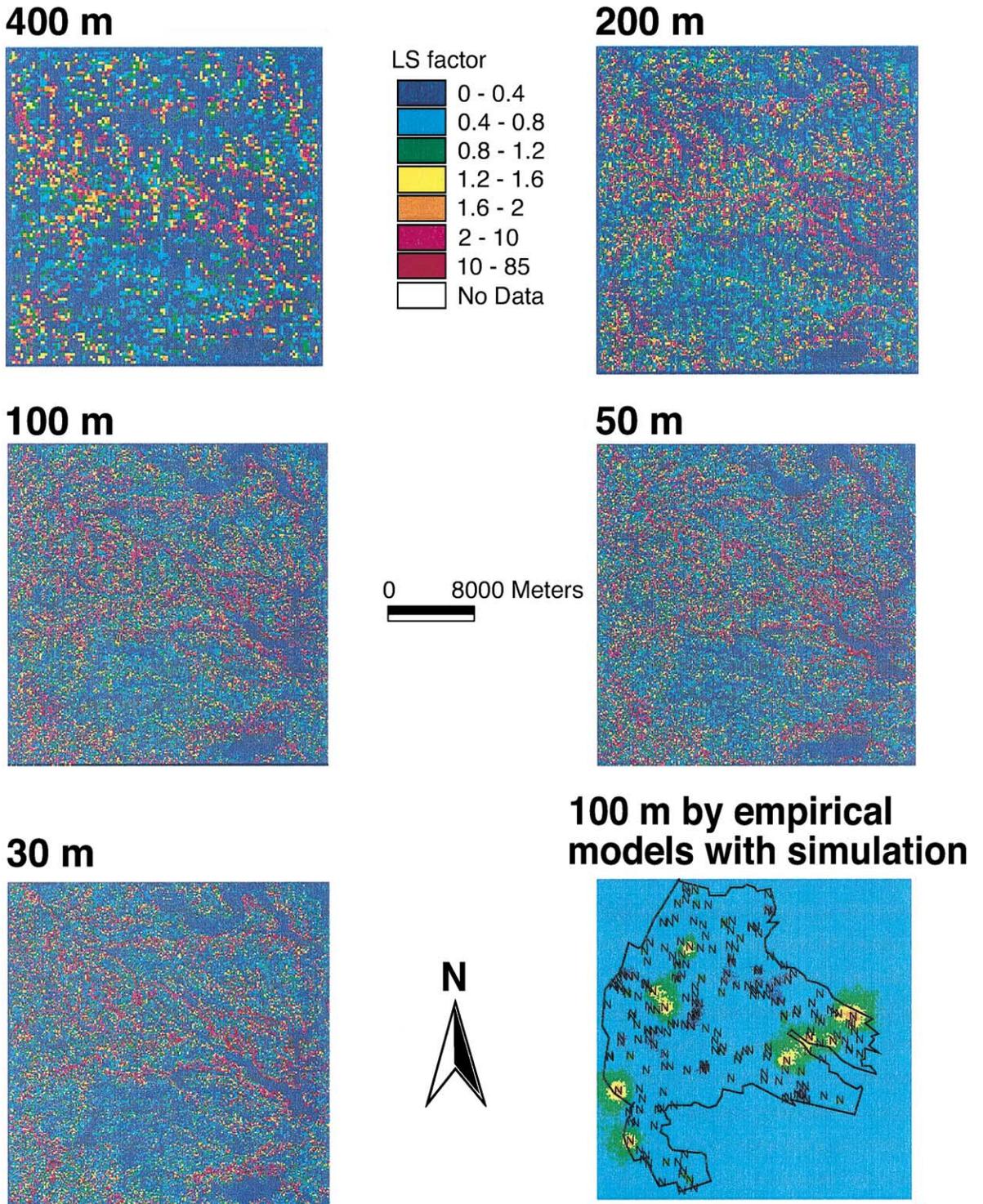
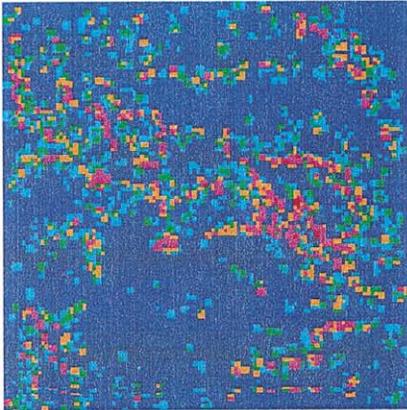
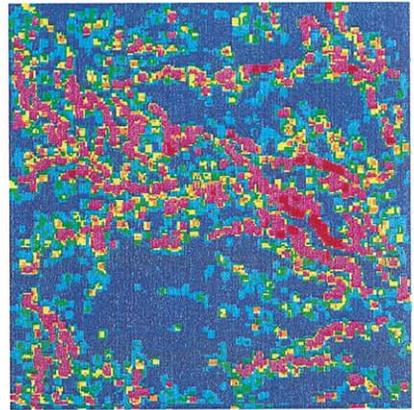


Fig. 2. Comparison of the predicted *LS* grids for different DEM spacing. Additionally, a predicted *LS* grid (100-m spacing) created by *LS* empirical models with sample data (with marked field sample areas) and a spatial simulation method is shown in the lower right.

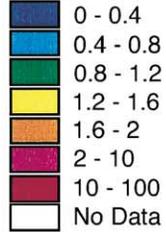
400 m



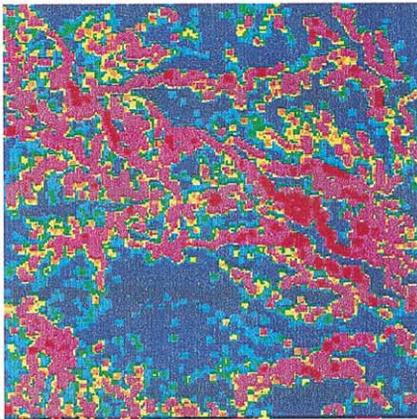
200 m



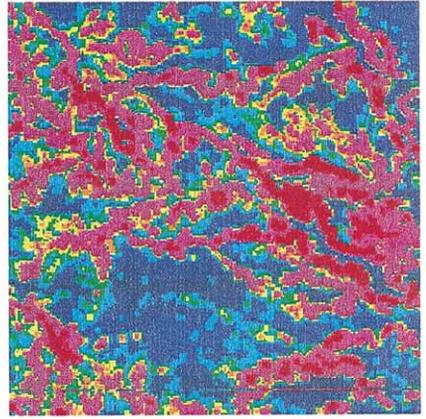
LS factor variance



100 m



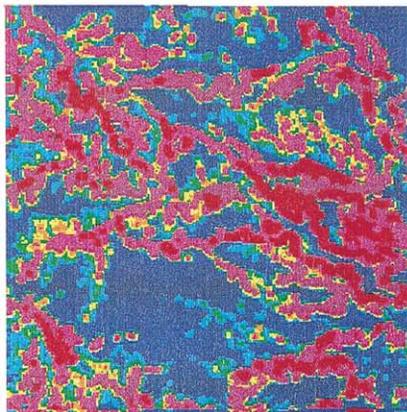
50 m



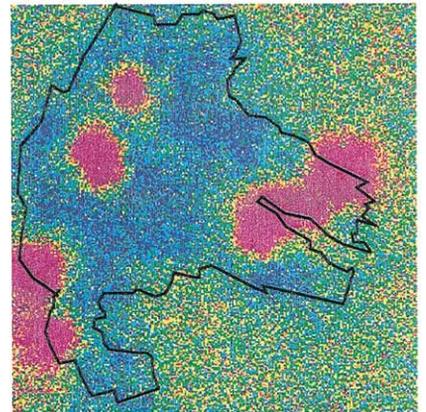
0 8000 Meters



30 m



100 m by empirical models with simulation



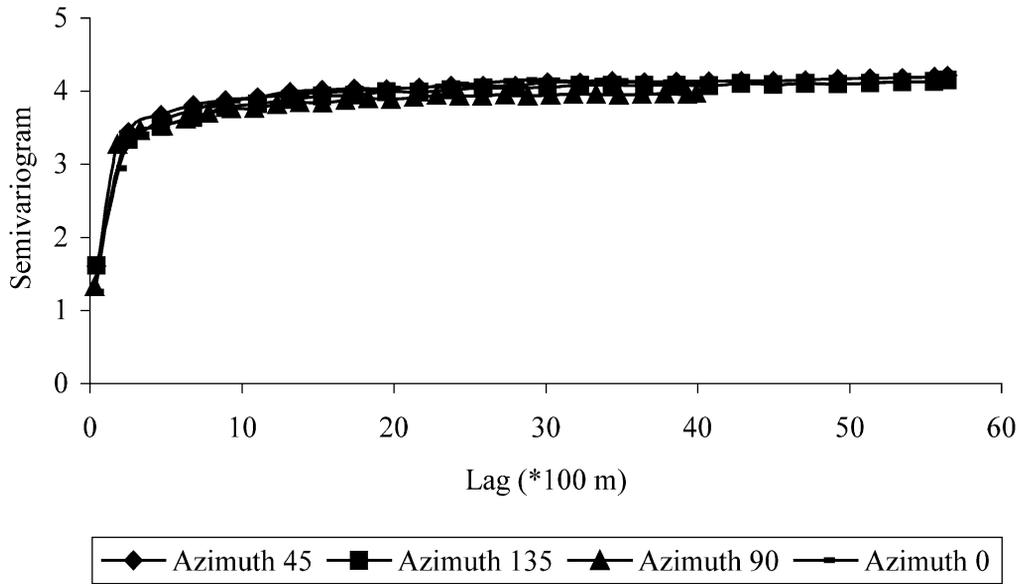


Fig. 4. Experimental semivariogram of predicted *LS* grid for 30-m DEM spacing and four directions.

ograms for 100- and 200-m DEM spacing look very similar and overlap, but have different nugget variances.

The semivariograms for different DEM spacing and four directions were modeled using the spherical model and the parameters are listed in Table 1. For each direction, the nugget and structure variance decrease and the range parameter increases as the DEM spacing increases except for the 100-m DEM spacing. The decrease in the nugget variance implies a reduction of noise and within-cell spatial variability due to data resampling. For the same DEM spacing, the nugget and structure variances and range parameter vary depending on the direction. However, the sill variances (nugget plus structure variance) are very similar in different directions. At the southeast, the nugget variance and range parameter are the largest because of the complex topography and the high spatial variability in elevation and slope. The exception mentioned above is that the 100-m DEM spacing led to the smallest nugget variances in all directions, and to smaller range parameters in three

directions compared to the 50-m DEM spacing. The reason has not been clarified yet and further investigation is needed.

The change of global variance, mean local variance and semivariance at a lag of one pixel over the DEM spacing for the whole area (denoted with W_{\cdot}) and a small area (denoted with S_{\cdot}) are shown in Fig. 6. For the small area, the global variance and mean local variance of the predicted *LS* values, and semivariance at a lag of one cell at first increase with DEM spacing. The largest global variance is found at the 30-m spacing, and it decreases slowly from 30 to 50 m, rapidly from 50 to 200 m, and after 200 m slowly again. The mean local variance and semivariance at a lag of one cell reach their largest values at the 50-m spacing, then decrease slowly and gradually stabilize.

For the entire area, the global variance of the predicted *LS* values decreases rapidly with the DEM spacing from 30 to 100 m, and then becomes stable after 100 m (Fig. 6). The mean local variance and semivariance at a lag of one cell increase from 30- to

Fig. 3. Comparison of variance grids of the predicted *LS* values for different DEM spacing. Additionally, a variance grid of predicted *LS* values (100-m spacing) created by *LS* empirical models with sample data and a spatial simulation method is shown in the lower right.

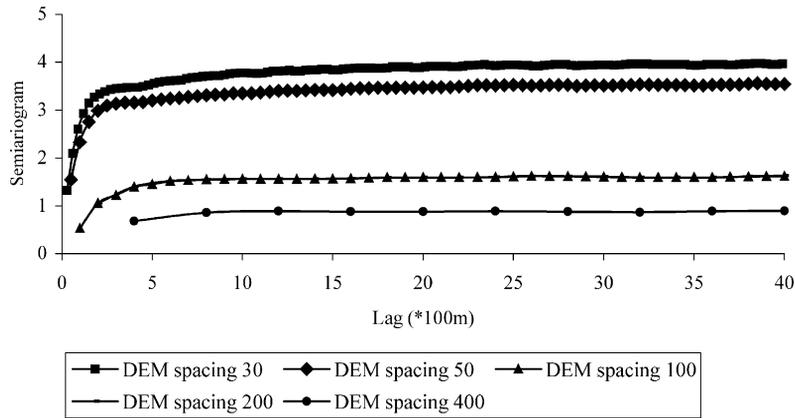


Fig. 5. Experimental semivariogram of predicted *LS* grids for different DEM spacing and 0° azimuth.

50-m spacing, and after 50 m decrease rapidly. There is a secondary wide peak at the 200-m spacing. The reason why the mean local variance and semivariance at a lag of one cell are smaller for the 100-m DEM spacing compared to the ones for 200 m is not clear, as mentioned previously for the nugget variances. The 50-m spacing at which the largest local variance and semivariance at a lag of one cell are observed may be appropriate for detailed spatial information (distribution and variability) for spatial prediction of the *LS* factor.

The similar behaviour of the mean local variance and semivariance at a lag of one cell with the DEM spacing implies that the semivariance at a lag of one cell can replace the local variance as a guide for choosing the appropriate DEM spacing, when the aim is to obtain detailed spatial information. Both the local variance and semivariance at a lag of one cell measure the mean local spatial variability, while the global variance measures the global variation of the estimates and shows a considerably different behaviour with an increase of DEM spacing. The global

Table 1

Spherical semivariogram model parameter estimates of topographical factor *LS* for different DEM spacing and spatial directions

DEM spacing (m)	c_0	c_1	c	a_0 (m)	DEM spacing (m)	c_0	c_1	c	a_0 (m)
<i>Azimuth 0°</i>					<i>Azimuth 45°</i>				
30	2.09	1.98	4.07	638	30	1.79	2.24	4.03	681
50	1.85	1.78	3.63	603	50	1.70	1.88	3.58	760
100	0.26	1.34	1.60	578	100	0.56	1.09	1.65	711
200	0.73	0.88	1.61	765	200	0.84	0.82	1.66	826
400	0.49	0.44	0.93	1602	400	0.57	0.37	0.94	1581
<i>Azimuth 90°</i>					<i>Azimuth 135°</i>				
30	1.39	2.43	3.82	326	30	2.51	1.53	4.04	1129
50	1.30	2.15	3.45	356	50	2.40	1.20	3.60	1275
100	0.38	1.27	1.65	746	100	0.60	1.03	1.63	927
200	0.74	0.92	1.66	968	200	1.16	0.49	1.65	1739
400	0.47	0.44	0.91	1179	400	0.62	0.30	0.92	1882

c_0 and c_1 are the nugget variance and structure variance, respectively.

$c = c_0 + c_1$ is the sill variance and a_0 is the actual range parameter.

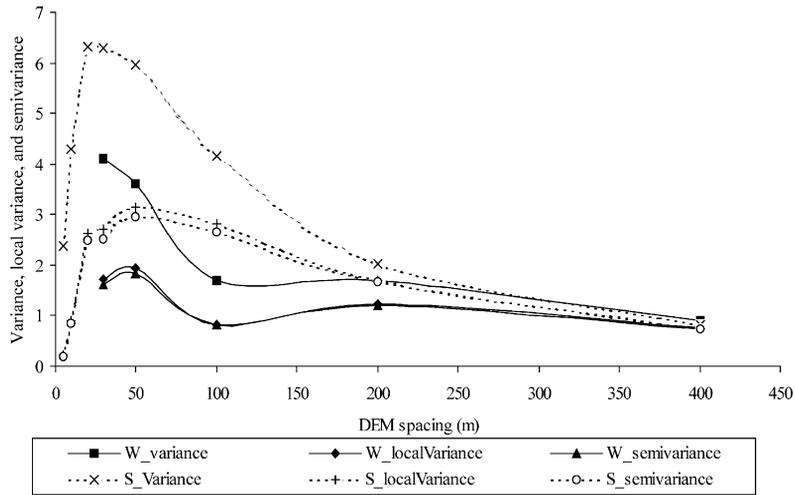


Fig. 6. Global variance and mean local variance of the predicted *LS* values and semivariance at a lag of one cell for different DEM spacing. W_ indicates the whole grid and S_ the small area shown in Fig. 1.

variance can be used to determine the DEM spacing when the objective is to provide an unbiased mean global estimate.

The loss of spatial information measured based on entropy is shown in Fig. 7. The grid of predicted *LS* values at the original 30-m DEM spacing has the largest entropy for both the entire area and the small area. Compared to the entropy of the original 30-m spacing, the entropy decreases, indicating the loss of information for both finer and coarser spacing. This

means that interpolating a DEM to new datasets of finer spacing may not be able to provide more information. The reason why the entropy drops more rapidly for the small area than for the whole area may be that there is a higher spatial variability of the *LS* factor within the small area. There is no significant difference in the relative loss of entropy from 30 to 100 and from 30 to 200 m for the entire area.

Fig. 8 shows the relationship between the relative losses due to resampling from 30-m spacing to a

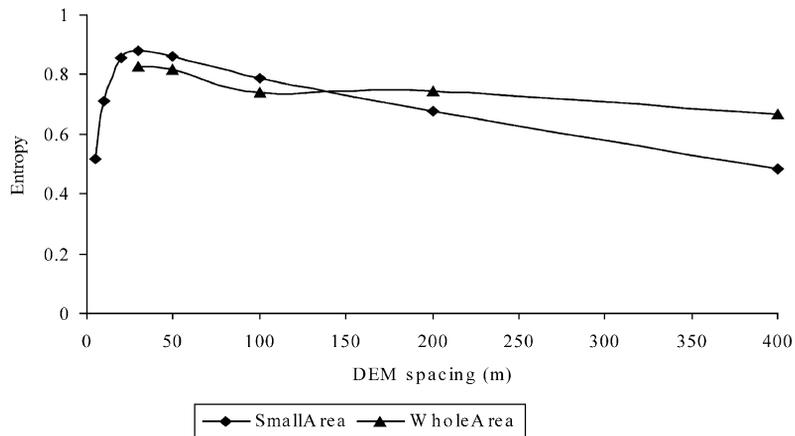


Fig. 7. Entropy varying with DEM spacing as measure of information loss caused by data resampling.

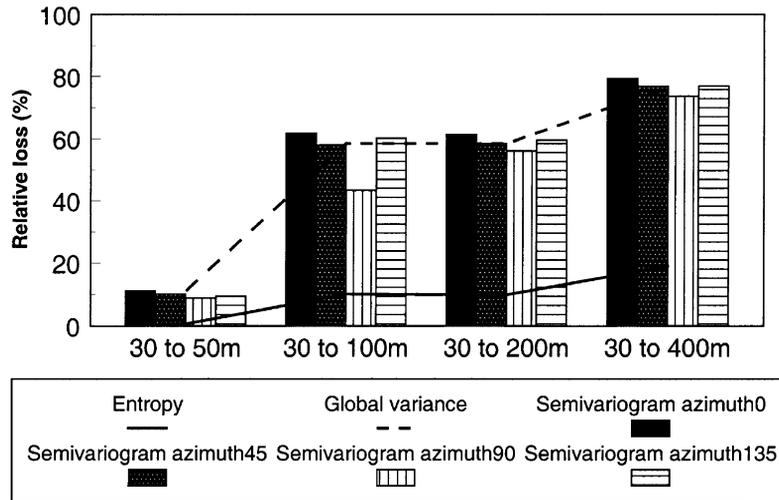


Fig. 8. Relative spatial variability loss due to data resampling, as measured by entropy (Eq. (9)), global variance (Eq. (10)), and spatial variability given by the integral of the differences of semivariogram functions (Eq. (11)) in different directions.

larger spacing. The measures of loss are based on entropy (Eq. (9)), global variance (Eq. (10)), and spatial variability as expressed by the integral of the differences of semivariogram functions at different directions (Eq. (11)). For the specific data resampling used here, the spatial variability and global variance loss were high compared to the relative loss of entropy. For example, when the data was resampled from the original 30-m DEM spacing to 50 m, a slight entropy loss of 0.71% occurred; however, the relative losses of spatial variability and global variance were larger than 8.9%.

The data resampling from 30-m spacing to 100 m led to a relative entropy loss of 10.10% and relative global variance loss of 58.5%. The relative loss of spatial variability varies from 43% to 62% depending on the directions, with the smallest loss occurring at an azimuth of 90°. This suggests that the spatial variability of the *LS* factor in different directions may be different in structure and amount (at least for certain DEM spacing), and this anisotropy leads to the above differences in the loss of spatial variability. In the process of further DEM resampling to 200-m spacing, the relative losses are similar to those of 100 m, except for the relative loss of spatial variability at the azimuth 90°. When the data were resampled to 400 m, an entropy loss of 19.04% and very large relative losses of spatial variability and

global variance occurred in all directions. These results indicate that the 50-m spacing may be an appropriate choice as it leads to little losses compared to 30 m (while the DEM data are reduced by about 2.8), while the next larger spacing of 100 m leads to significant losses.

5. Conclusion and discussion

This study deals with spatial prediction and uncertainty analysis of the topographic factor *LS* involved in RUSLE using a DEM in a case study area. The focus is on investigating the use of DEM and appropriate DEM spacing for spatial prediction of the *LS* factor, and modeling the loss of spatial variability due to data resampling. The predicted *LS* and its variance grids derived using the physically-based topographical factor *LS* equation and DEMs are spatially consistent and correlated with the topographical features. That is, in the hilly areas the predicted *LS* values and variances are high, and in flat areas *LS* values and variances are low. The lake areas are filled with *LS* values of zero. The improved correlation of the predicted *LS* values with the topography is obvious compared to the corresponding grids by a spatial simulation based on empirical models and sample data.

The DEM spacing should be chosen considering simultaneously the required prediction precision and the detailed spatial information of the *LS* factor. In choosing a single DEM spacing optimally for both requirements, a compromise may be needed, depending on the users' emphasis on one of the requirements or both. Global variance and semivariance at a lag of one cell can be used in combination to achieve the above purpose. In addition, modeling the experimental semivariograms and using them to estimate spatial variability loss due to data resampling can help users determine the appropriate DEM spacing.

For the same spatial direction, the sill parameters and the nugget variances of the modeled semivariograms generally decrease from finer to coarser DEM spacing while the range parameters generally increase. The more complex the topographic features, the larger the nugget variances and range parameters. In addition to the within-cell spatial variability, the nugget variances may be considered as an estimate of noise caused by errors from elevation measurements, data resampling, models used, and calculation of the variables related to the *LS* factor. Developing a method to separate the noise from the within-cell spatial variability is important in order to determine an appropriate DEM spacing.

In addition to entropy and global variance as a general measure of information loss in Eqs. (9) and (10), a new method to directly measure the loss of spatial variability was presented. This method (Eq. (11)) is based on the modeled semivariograms and varies depending on the semivariogram model (e.g., spherical) chosen. Once a model is determined, the loss measure function of spatial variability can be easily derived and calculated by differentiation and integration. The results in Fig. 8 showed that the losses of spatial variability calculated by the new method are similar in three of the four directions, but different in one direction. This implies that the new method can reveal differences in spatial variability and spatial variability loss due to data resampling in different directions when anisotropy exists.

Acknowledgements

We are grateful to SERDP (Strategic Environmental Research and Development Program) for provid-

ing support for this study, to US Army Corps of Engineers, to the Construction Engineering Research Laboratory (USA-CERL) for providing the datasets, and to Mr. William Jackson for his help to convert raster data from GRASS to ArcView GIS format.

References

- Atkinson, P.M., 1997. On estimating measurement error in remotely-sensed images with the variogram. *International Journal of Remote Sensing* 18 (14), 3075–3084.
- Atkinson, P.M., Curran, P.J., 1995. Defining an optimal size of support for remote sensing investigations. *IEEE Transactions on Geoscience and Remote Sensing* 33 (3), 768–776.
- Atkinson, P.M., Curran, P.J., 1997. Choosing an appropriate spatial resolution for remote sensing investigations. *Photogrammetric Engineering and Remote Sensing* 63 (12), 1345–1351.
- Atkinson, P.M., Danson, F.M., 1988. Spatial resolution for remote sensing of forest plantations. *Proceedings of IGARSS '88 Symposium*, 13–16 September, Edinburgh, Scotland, ESA SP-284 (IEEE 88CH2497-7), ESA Publications Division, pp. 221–223.
- Chilès, J., Delfiner, P., 1999. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York.
- Curran, P.J., Dungan, J.L., 1989. Estimation of signal-to-noise: a new procedure applied to AVIRIS data. *IEEE Transactions on Geoscience and Remote Sensing* 27 (5), 620–628.
- De Cola, L., 1997. Multi-resolution covariation among Landsat and AVHRR vegetation indices. In: Quattrochi, D.A., Goodchild, M.F. (Eds.), *Scale in Remote Sensing and GIS*. Lewis Publishers, CRC Press, Boca Raton, FL, pp. 73–92.
- Deutsch, C.V., Journel, A.G., 1998. *Geostatistical Software Library and User's Guide*. Oxford Univ. Press, Oxford.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford Univ. Press, Oxford.
- GRASS, 1993. *Geographic Resources Analysis Support System (GRASS) version 4.1 user's reference manual*. USA Corps of Engineers, Construction Engineering Research Laboratories, Champaign, IL. <http://www.baylor.edu/grass/> (accessed May 6, 2001).
- Hutchinson, S., Daniel, L., 1997. *Inside ArcView GIS*. OnWord Press, Santa Fe, NM, USA.
- Kaluzny, S.P., Vega, S.C., Cardoso, T.P., Shelly, A.A., 1998. *S+ Spatialstats: User's Manual for Windows and Unix*. Mathsoft, Cambridge, MA, USA, See also <http://www.insightful.com/products/spatial/default.html> (accessed May 6, 2001).
- Marceau, D.J., Howarth, P.J., Gratton, D.J., 1994a. Remote sensing and the measurement of geographical entities in a forested environment. 1. The scale and spatial aggregation problem. *Remote Sensing of Environment* 49 (2), 93–104.
- Marceau, D.J., Gratton, D.J., Fournier, R.A., Fortin, J., 1994b. Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial resolution. *Remote Sensing of Environment* 49 (2), 105–117.

- Mitášová, H., Mitáš, L., 1993. Interpolation by regularized spline with tension: I. Theory and implementation. *Mathematical Geology* 25 (6), 641–655.
- Mitášová, H., Hofierka, J., Zlocha, M., Iverson, L.R., 1996. Modelling topographic potential for erosion and deposition using GIS. *Journal of Geographical Information Science* 10 (5), 629–641.
- Moore, I.D., Burch, G.J., 1986. Physical basis of the length-slope factor in the Universal Soil Loss Equation. *Soil Science Society of America Journal* 50 (5), 1294–1298.
- Moore, I.D., Wilson, J.P., 1992. Length-slope factors for Revised Universal Soil Loss Equation: simplified method of estimation. *Journal of Soil and Water Conservation* 47 (5), 423–428.
- Myers, J.C., 1997. *Geostatistical Error Management: Quantifying Uncertainty for Environmental Sampling and Mapping*. Van Nostrand-Reinhold, New York.
- Renard, K.G., Ferreira, V.A., 1993. RUSLE model description and database sensitivity. *Journal of Environmental Quality* 22 (3), 458–466.
- Renard, K.G., Foster, C.R., Weesies, G.A., McCool, D.K., Yoder, D.C., 1997. *Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE)*. US Department of Agriculture, Agriculture Handbook Number 703, Government Printing Office, Washington, DC, pp. 1–404.
- Risse, L.M., Nearing, M.A., Nicks, A.D., Laflen, J.M., 1993. Error assessment in the Universal Soil Loss Equation. *Soil Science Society of America Journal* 57 (3), 825–833.
- Shannon, C.E., Weaver, W., 1964. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL.
- Townshend, J.R.G., Justice, C.O., 1988. Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations. *International Journal of Remote Sensing* 9 (2), 187–236.
- Vieux, B.E., 1995. DEM aggregation and smoothing effects on surface runoff modeling. In: Lyon, J.G., McCarthy, J. (Eds.), *Wetland and Environmental Applications of GIS*. Lewis Publishers, CRC Press, Boca Raton, FL, pp. 205–229.
- Wang, G., Gertner, G.Z., Parysow, P., Anderson, A.B., 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.
- Woodcock, C.E., Strahler, A.H., 1987. The factor of scale in remote sensing. *Remote sensing of environment* 21 (3), 311–322.