Quantifying DEM Uncertainty and its Effect on Topographic Parameters

Suzanne P. Wechsler and Charles N. Kroll

Abstract

Digital elevation models (DEMs) are representations of topography with inherent errors that constitute uncertainty. DEM data are often used in analyses without quantifying the effects of these errors. This paper describes a Monte Carlo methodology for evaluation of the effects of uncertainty on elevation and derived topographic parameters. Four methods for representing DEM uncertainty that utilize metadata and spatial characteristics of a DEM are presented. Seven statistics derived from simulation results were used to quantify the effect of DEM error. When uncertainty was quantified by the average relative absolute difference, elevation did not deviate. The range of deviation across the four methods for slope was 5 to 8 percent, 460 to 950 percent for derived catchment areas and 4 to 9 percent for the topographic index. This research demonstrates how application of this methodology can address DEM uncertainty, contributing to more responsible use of elevation and derived topographic parameters, and ultimately results obtained from their use.

Introduction

Digital elevation models (DEMs) are a valuable and useful data source for many natural resource related Geographic Information System (GIS) applications. Representation of elevation in a grid framework enables neighborhood computations of parameters, such as slope and flow direction, which are used to identify ridges and valleys and derive the topographic index, a statistic that indicates the propensity for surface saturation (Beven and Kirkby, 1979; Quinn *et al.*, 1991). Although a DEM is a *model* of the elevation surface, it is often not treated as a model, but is accepted as a true representation of the earth's surface. DEM data, like other spatial data sets, are subject to error (USGS, 1995; Monmonier, 1991; Wright, 1942). Many DEM users perceive that DEM uncertainty affects the outcomes of their applications. However, the effects of DEM error on elevation and derived parameters are often not evaluated by DEM users (Wechsler, 2003), and methods to address DEM error have not been systematically integrated with GIS software packages. The methodology presented here integrates DEM uncertainty simulation with GIS. The impact of uncertainty on elevation and three derived topographic parameters are investigated: slope, upslope contributing area and topographic index (TI). This methodology could also be applied to other topographic parameters. The intent of this approach is to enable DEM users to assess the uncertainty of elevation and derived parameters, in the absence of higher accuracy elevation data.

Background

Derived Topographic Parameters
The DEM provides a base data set from which topographic parameters are digitally generated. These surface derivatives provide the basis for characterization of landform (Evans, 1998) and are used extensively in environmental applications such as hydrology, geomorphology, and environmental modeling. Various algorithms exist for calculating topographic parameters from a DEM, and each method can produce different results (Ryder and Voyadgis, 1996; Carter, 1990; Skidmore, 1989). The topographic parameters investigated in this paper are dependent on the current capabilities of ArcView™ GIS. Slope (tan B) is derived using the Horn method (also referred to as the "Kings Case") (Horn, 1981; Burrough and McDonnell, 1998; ESRI, 1998). The upslope contributing area represents the total area draining into each grid cell and is derived from flow direction. Many algorithms exist for computing flow direction (ESRI, 1998; Tarboton, 1997; Wolock and McCabe, 1995; Costa-Cabral and Burges, 1994; Quinn *et al.*, 1991; Jenson and Domingue, 1988). ArcView™ incorporates a single flow path method that computes flow direction based on the direction of steepest descent in one of eight directions from a center cell of a 3 × 3 window. TI is computed as: $TI = A_c/\tan B$ where A_c is the drainage area per unit contour length (flow path width) of a cell, here taken as the cell width. Derivation of these parameters (particularly upslope area and TI which are based on flow direction) does not reflect the most recent methods proposed in the hydrologic literature (Tarboton, 1997; Wolock and McCabe, 1995; Quinn *et al.*, 1991). In fact, there is little consistency in the hydrologic literature regarding the most appropriate estimators of many topographic parameters. The method used to compute the parameters is not relevant to the present analysis since the uncertainty of other estimators could be examined in a similar manner. Of interest in this analysis is assessing the effect of DEM errors on the uncertainty of derived topographic parameter estimators.

DEM Error Constitutes Uncertainty
Error is the departure of a measurement from its true value. In geographic analyses of complex natural systems using spatial data we often do not know or do not have access

Suzanne P. Wechsler is with the California State University - Long Beach, Department of Geography, 1250 Bellflower Boulevard, Long Beach, CA 90840 (wechsler@csulb.edu).

Charles N. Kroll is with Environmental Resources and Forest Engineering, SUNY College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210.

to a true value. Our lack of knowledge about the reliability of a measurement in its representation of the true value is referred to as uncertainty and is a measure of what we don't know. Uncertainty exists in spatial data, and in DEMs. DEM product information concerning data collection, processing or error distribution is generally unavailable to DEM users (Holmes *et al.*, 2000). In the United States the U.S. Geological Survey (USGS) is one of the most common producers of DEMs with over 14.5 million 7.5-minute DEMs downloaded between March 1998 and July 2001 (the last year statistics were available from the USGS) (USGS, 2004). Approximately 30 percent of DEM users use the USGS 7.5-minute DEM product (Wechsler, 2003). Because of the prevalence of USGS 7.5-minute DEMs, a review of the accuracy of this data product is warranted. The USGS recognizes that their DEM products are subject to "blunders," systematic errors, and random errors (USGS, 1995; Caruso, 1987). Blunders are vertical errors associated with the data collection process and are identified and removed prior to release of the data. Systematic errors are the result of procedures or systems used in the DEM generation process and follow fixed patterns that can cause bias or artifacts in the final DEM product. When the cause is known, systematic errors are eliminated or reduced. Random errors remain in the data after blunders and systematic errors are removed. The processes responsible for introducing random errors into a DEM are not sufficiently understood thus the exact nature and location of random error cannot be precisely determined. The major purpose of this paper is to demonstrate a methodology to address uncertainty that is due to random DEM errors.

Quantifying Vertical Accuracy in DEMs

The nature and extent of DEM errors are not provided to DEM users. The USGS provides users of their products with the Root Mean Square Error (RMSE) statistic, an estimate of the DEM's vertical accuracy. The Root Mean Square Error (RMSE) is expressed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - yt_i)^2}{N}} \qquad (1)$$

where: y_i refers to the ith interpolated elevation, yt_i refers to the ith known or measured elevation, and N is the number of sample points. For its 7.5-minute DEM product the USGS calculates the RMSE from $N = 28$ sample points (yt_i) distributed throughout the specified area and presumed to be representative of the terrain. However, these values do not always reflect actual elevations, rather they are the most probable elevation (USGS, 1995). While a valuable quality-control statistic, the RMSE does not provide the DEM user with an accurate assessment of how well each cell in the DEM represents the true elevation; it provides only an assessment of how well the DEM corresponds to the data from which it was generated. Researchers have used error measured at discrete points (such as from GPS surveys or data of higher resolution) to generate DEM error estimates and investigated the spatial structure of DEM error (Holmes *et al.*, 2000; Ehlschlaeger, 1998; Ehlschlaeger and Shortridge, 1996). In the absence of definitive information regarding the spatial structure of DEM error, the RMSE is the only statistic available. The purpose of our approach is to quantify uncertainty that is due to random errors in a DEM and derived parameters using information available to most DEM users, i.e., the RMSE. This methodology uses the RMSE as a mechanism for quantifying DEM uncertainty using Monte Carlo simulation. When additional information about the structure of DEM error becomes available, error fields could be adjusted accordingly.

Monte Carlo Simulation

In a Monte Carlo simulation of DEM uncertainty, the DEM is recognized as only one possible realization of the true elevation surface. Multiple simulations can be used to quantify DEM uncertainty through evaluation of statistics associated with a distribution of realizations. To create this distribution a number (N) of random error fields are generated where each cell represents the possible error at a co-located elevation. Each random field is added to the DEM generating a new realization of the elevation surface. Monte Carlo simulation methods have been used by many researchers to evaluate error in GIS data, including Lanter and Veregin (1992), Openshaw *et al.* (1991), and Heuvelink *et al.* (1989) and have been applied to specifically address DEM uncertainty. For example, Fisher (1991) applied Monte Carlo simulation techniques to evaluate the impact of DEM error on viewshed analyses. Lee *et al.* (1992) determined that small DEM errors significantly affected floodplain locations. Liu and Herrington (1993) investigated the impact of DEM error on a forest harvesting model. Ehlschlaeger and Shortridge (1996) evaluated the impact of DEM uncertainty on a least-cost-path application. Hunter and Goodchild (1997) investigated the effect of simulated changes in elevation at different levels of spatial autocorrelation on slope and aspect calculations. Holmes *et al.* (2000) simulated error in DEMs to evaluate slope failure prediction.

Monckton (1994) recognized that error values such as the RMSE ignore the spatial structure of error. To understand the effects of such error, researchers have developed various methods for representing DEM error through spatially autocorrelated random fields. Some of these approaches require prior knowledge of the spatial structure of DEM error obtained from a higher accuracy data source such as spot heights or check points (Holmes *et al.*, 2000, Monkton, 1994; Östman, 1987) or use of a DEM that is deemed of "higher accuracy" (Ehlschlaeger, 1998). Other methods use a swapping algorithm (Ehlschlaeger, 2002; Fisher, 1991), spatially autoregressive random fields (Hunter and Goodchild, 1997) and covariance functions (Oliver, 1995) to represent spatially autocorrelated error in random fields.

While progress has been made in the development of error propagation techniques, a consensus regarding appropriate technique(s) does not exist. The intent of this research is not to add new elements in the simulation process. As the literature attests, there are many ways of representing DEM error in the simulation process. The approach undertaken here provides a mechanism for representing error in the absence of any knowledge of its spatial structure and provides statistical methods for quantifying and visualizing the effect of DEM uncertainty not only for elevation, but also for slope, upslope area, and the topographic index. The methodology was successfully integrated within a GIS and provides a comprehensive toolkit for general application and quantification of DEM uncertainty.

Methodology

The methodology presented in this paper is intended for use in applications where a ground truth field survey or other higher accuracy data source is not available and all that is available to the DEM user is the RMSE. It is assumed that on average the true elevation is equal to the stated elevation value, but it could be any value within a specific probability distribution. Here, it is assumed these errors are normally distributed with a standard deviation equal to the RMSE. This provides the theoretical basis for using unfiltered random fields in the simulation. However, error is not likely to manifest itself in the pattern represented by unfiltered random fields, and therefore an error model ought to be based on an

assumption of spatial dependence of error (Hunter and Goodchild, 1997). Although the assumption is that error is random, elevation is spatially autocorrelated; hence, elevation error is spatially autocorrelated (Ehlschlaeger, 1998; Hunter and Goodchild, 1997; Lopez, 1997; Veregin, 1997; Fisher, 1991). Here spatial autocorrelation in the error field is created by smoothing or filtering the random error fields with an averaging filter. Filtering error fields increases the spatial autocorrelation of the random error fields. Different levels of spatial autocorrelation (or spatial dependence) are achieved through this filtering. This research applies a stochastic approach to representing DEM error through random fields and Monte Carlo simulation that enables the methodology to be easily integrated with a commonly used GIS package, making these simulation techniques and resulting uncertainty estimators accessible to DEM users. The approach could easily be extended to other GIS software programs.

Random Field Methods

The results of four representations of DEM error are presented. Two of these methods assume no, or only slight spatial auto-correlation. The other two methods depend on the distance of the spatial autocorrelation of the elevation error, referred to as the distance of spatial dependence (D). One method for obtaining a D value is through analysis of the sill of a semi-variogram of DEM error. This determination can be subjective and cannot be determined without extensive ground truth. Holmes *et al.* (2000) computed a semivariogram of error in a 30 m USGS DEM. A sill appears to occur at 200 to 300 meters. More thorough analyses of the structure of DEM error in various types of topography and physiographic regions are warranted. In the absence of such empirical information, results from Holmes *et al.* (2000) were used to guide the present analyses. A D of 270 m (9×9 grid cells) was selected as the filter neighborhood. Four methods to represent random errors in DEMs using random fields were used. The methods differ in how spatial autocorrelation of random error is treated; these methods are outlined in Table 1.

Unfiltered (U)

Liu and Herrington (1993), Lopez (1997) and Hunter and Goodchild (1997) investigated the spatial dependence of elevation and determined that an error model based on independent random error does not reflect the spatial structure of error. Due to spatial autocorrelation, positive correlations exist between adjacent errors. Fisher (1991) stated that independent errors are likely to contribute to only a small portion of DEM error. The problem, however, is that little is known about the nature and extent of random DEM errors. Until DEM producers and vendors provide this information, assumptions must be made about the spatial relationships of

random errors. Here the first method considered is based on completely random unfiltered error fields (U) and assumes no spatial autocorrelation. These random fields were generated with a mean of zero and a standard deviation equal to the RMSE for the DEM (Equation 1). The RMSE statistic is based on an assumption of independent errors (Liu and Herrington, 1993; Caruso, 1987). This simulation provides the largest errors in the parameters evaluated in this experiment, and could be considered a worst-case scenario of DEM uncertainty effects.

Neighborhood Autocorrelation (N)

The *neighborhood autocorrelation* filter method (N) incorporates spatial autocorrelation of error by passing a mean-low-pass 3×3 filter (Eastman, 1992) over the surface. Each cell in the random field is replaced by the mean of the value of a centered nine-cell window. This method increases the spatial autocorrelation of each random surface and decreases the standard deviation of the values. As long as the distance of spatial dependence (D) is greater than the cell resolution, this procedure will underestimate D.

Mean Spatial Dependence (S)

The *mean spatial dependence* filter method (S) incorporates characteristics inherent in the selected DEM and requires prior analysis of the DEM to quantify D. The mean spatial dependence filter passes a $D \times D$ window over each cell in a grid and replaces the center cell with the mean of all cells in the centered $D \times D$ neighborhood.

Weighted Spatial Dependence (W)

The *weighted spatial dependence* filter (W) further incorporates spatial autocorrelation specific to a selected DEM. Like the mean spatial dependence filter, the dimension of the filter kernel are based on the established D. Unlike the mean spatial dependence filter, the impact of error from neighboring cells decreases with distance, which is similar to semi-variogram trends. The filter passes a $D \times D$ kernel over a grid and calculates a weighted mean of all the cells within the kernel. Cells farther away from the center cell are assigned less weight. The filter can be expressed as follows:

$$\text{Weighted Grid} = \sum_{i=1}^{TL} \left(\mu L_i * \frac{TL - i + 1}{\sum_{i=1}^{TL} i} \right) \quad (3)$$

where: μL_i is the mean of the values in ring i, and TL is the total number of rings.

The ring refers to a distance measure from the center cell. The center cell is ring 1. The rows and columns immediately surrounding the center cell are considered ring 2, and so on. The weight for each ring is the rank of the ring (where the ring farthest from the center cell has the lowest rank) divided by the sum of the rings. While other weighting schemes may also be plausible, here we wish to show the impact of a distance weighting scheme that produces autocorrelated errors similar to those observed in practice (Holmes *et al.*, 2000). Figure 1 depicts semivariograms of the different filter methods.

Procedure

The methodology presented is based on the assumption that the DEM on hand is the best representation of the true elevation available. Therefore, the DEM and derived parameters are used as "truth" surfaces. Errors are added to the original DEM because, although a DEM is the best representation

Random Field Method	Description
(U) Unfiltered	Random fields with M = 0, SD = DEM RMSE.
(N) Neighborhood Autocorrelation	Mean 3×3 low-pass filter applied to (U), and rescaled to M = 0, SD = DEM RMSE.
(S) Mean Spatial Dependence	Mean $D \times D$ filter applied to (U); cells replaced with the mean of surrounding $D \times D$ cells and rescaled to M = 0, SD = DEM RMSE.
(W) Weighted Spatial Dependence	Filter (U) with a weighted $D \times D$ kernel; cells farther away are assigned less weight and rescaled to M = 0, SD = DEM RMSE.
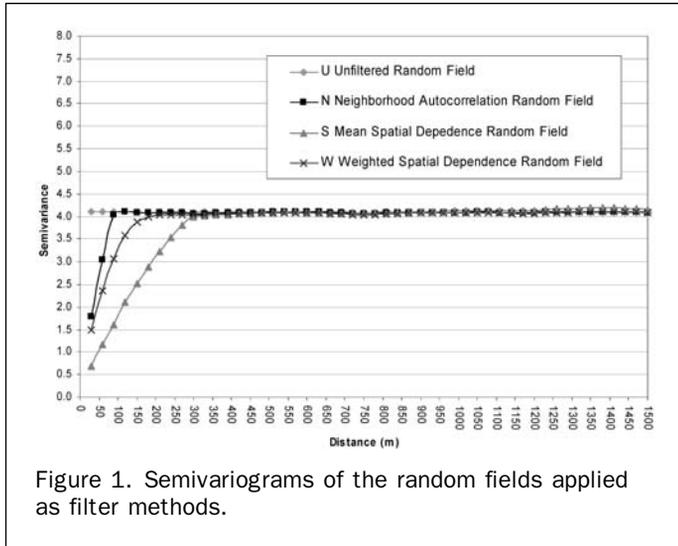
Figure 1. Semivariograms of the random fields applied as filter methods.
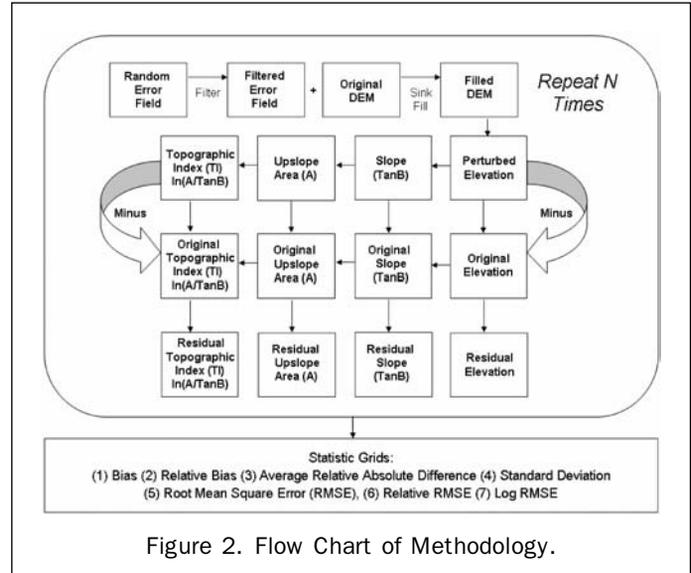
Figure 2. Flow Chart of Methodology.

of "truth" available, it contains errors and under uncertain conditions is assumed to be one of an infinite number of elevation realizations. Error simulation enables us to quantify uncertainty associated with elevation and derived topographic parameters. DEM uncertainty can be quantified by applying one of these filter techniques within a Monte Carlo simulation. A DEM is selected and a random field (created by the selected method) is added to the DEM, creating a new realization of the DEM. Depressions in the DEM (sinks) disrupt calculation of the upslope contributing area, requiring each realization to undergo a sink-filling procedure (ESRI, 1998; Jenson and Domingue, 1988). Four topographic parameter grids are computed from the new realization: elevation, slope, upslope contributing area, and the topographic index. This procedure is repeated N times, resulting in N realizations of each topographic parameter for each cell in the grid.

Quantification of DEM uncertainty through this methodology is based on the residuals of each parameter (i.e., the difference between each perturbed parameter and the original undisturbed parameter). Comparisons of the N realizations with the *true* values (from the original undisturbed grids) allow for an assessment of topographic uncertainty. The residual grids are produced by subtracting the original parameter values from each realization to quantify uncertainty by providing a measure of "forecast error" (Lettenmaier and Wood, 1993). Figure 2 provides a flow chart of the methodology employed in this analysis.

The methodology presented here allows the user to examine seven statistics for assessing the uncertainty in topographic parameters. These statistics are summarized in Table 2. *Bias* and *Relative Bias* (R-Bias) quantify systematic error. These statistics measure the degree to which predicted parameters are, on average, above or below the true parameter at a cell location. R-Bias standardizes bias measurements to the original true parameter value, providing the average percent deviation of the predicted value from the true value. Similar to R-Bias, the *Average Relative Absolute Difference* (ARAD) reports the percent absolute deviation of the predicted value from the true value. ARAD results are reported only as positive values and are hence more intuitive by indicating the average percent difference from the true value. Random errors can be quantified by the *Standard Deviation* (STD), which is a measure of the variability of the forecast about its mean value. The *Root Mean Square Error* (RMSE), *Relative Root Mean Square Error* (R-RMSE) and *Log Root Mean Square Error* (L-RMSE) are measures that incorporate both random error (variance) and systematic error (bias). The RMSE is a

TABLE 2. UNCERTAINTY ESTIMATOR STATISTICS FOR QUANTIFYING DEM UNCERTAINITY

Estimator	Formula
Bias	$\text{Bias} = \dfrac{\sum\limits_{i=1}^{N}(\hat{Y}_i - Y_i)}{N}$
Relative Bias (RBias)	$\text{RBias} = \dfrac{\sum\limits_{i=1}^{N}\left[\dfrac{(\hat{Y}_i - Y_i)}{Y_i}\right]}{N}$
Average Relative Absolute Difference (ARAD)	$\text{ARAD} = \dfrac{\sum\limits_{i=1}^{N}\left(\dfrac{|\hat{Y}_i - Y_i|}{Y_i}\right)}{N}$
Standard Deviation (STD)	$\text{STD} = \sqrt{\dfrac{\sum\limits_{i=1}^{N}(\hat{Y}_i - \overline{Y})^2}{N}}$
Root Mean Square Error (RMSE)	$\text{RMSE} = \sqrt{\dfrac{\sum\limits_{i=1}^{N}(\hat{Y}_i - Y_i)^2}{N}}$
Relative Root Mean Square Error (R-RMSE)	$\text{RRMSE} = \sqrt{\dfrac{\sum\limits_{i=1}^{N}\left(\dfrac{\hat{Y}_i - Y_i}{Y_i}\right)^2}{N}}$
Log Root Mean Square Error (L-RMSE)	$\text{LRMSE} = \sqrt{\dfrac{\sum\limits_{i=1}^{N}\left(\ln\left(\dfrac{\hat{Y}_i}{Y_i}\right)\right)^2}{N}}$

Where $\hat{Y}_i$ refers to the estimator of the parameter Y_i and N is the number of simulations. $\overline{Y}$ is the average of the N $\hat{Y}_i$ values.

common measure of uncertainty employed in numerous applications. R-RMSE, like R-Bias and ARAD, standardizes the RMSE to the true parameter value and represents the percent variation of the estimator. Relative statistics, such as R-Bias and R-RMSE are useful when a wide range of true values is present. The L-RMSE transforms the RMSE into log space, and assigns more weight to underestimation than overestimation of the parameter. This statistic is useful for bounded estimators such as slope. L-RMSE and R-RMSE are first-order

equivalent for precise estimators (Kroll and Stedinger, 1996). The effects of DEM uncertainty on elevation and derived topographic parameters can be assessed by examining all or some of these seven statistical estimators.

Application of Methodology: Coweeta Hydrologic Watershed

The following analysis investigates how DEM error affects uncertainty of topographic parameters that are frequently employed in hydrologic investigations. Simulations were run for a portion of the Prentiss, North Carolina level-2 DEM that contains the Coweeta Hydrologic Watershed (Figure 3). This DEM has a 30 m grid cell resolution with 183 columns and 207 rows. The Coweeta watershed comprises a 2,169 ha (24,107 grid cells) experimental facility dedicated to forest hydrology research since its establishment in 1933 and is administrated by the USDA Forest Service (LTER, 2001). Numerous hydrologic studies have been conducted at the Coweeta Hydrologic Watershed (e.g., Bolstad and Swank, 1997; Day and Monk, 1974; Black, 1959).

The reported RMSE for this DEM was 2 m. As mentioned previously, a **D** of 270 m was selected for this analysis. A standard deviation analysis was applied to determine the appropriate number of simulations, as follows. A number (e.g., $N = 350$) of random simulations were performed, and across all N grids the average and standard deviation of each cell was computed. The process was repeated for $N = 349$,

$N = 348$ etc. The minimum number of needed simulations (N) was determined when the percent difference in standard deviation between subsequent simulations fell below 5 percent, which occurred at $N = 135$. The four different methods of random field generation were examined. The effect of error on topographic parameter estimators for each random field method was assessed. The methodology was programmed using the Avenue language as an extension for ArcView™ Spatial Analyst® GIS software (ESRI, 1998). This approach could be incorporated into other GIS software packages.

Uncertainty Statistics for the Coweeta Hydrologic Watershed

For each topographic parameter, statistical grids representing the seven statistics were computed. Statistics for the 24,107 grid cells in the Coweeta watershed summarize and quantify uncertainty in the topographic parameter estimators (Table 3). These summary values are more broadly descriptive than the RMSE provided with the DEM since they incorporate the spatial distribution of simulated uncertainty for every cell location in the watershed.

Figure 4 demonstrates the effect of filter methods on topographic parameters using simulation results for the RMSE uncertainty estimator. Elevation was unbiased and the estimator distribution remained similar between the filter methods. For elevation, only the U and N filters were not significantly different from each other ($p = 0.64$). Slope

Figure 3. Study area location: Coweeta Hydrologic Watershed, North Carolina.
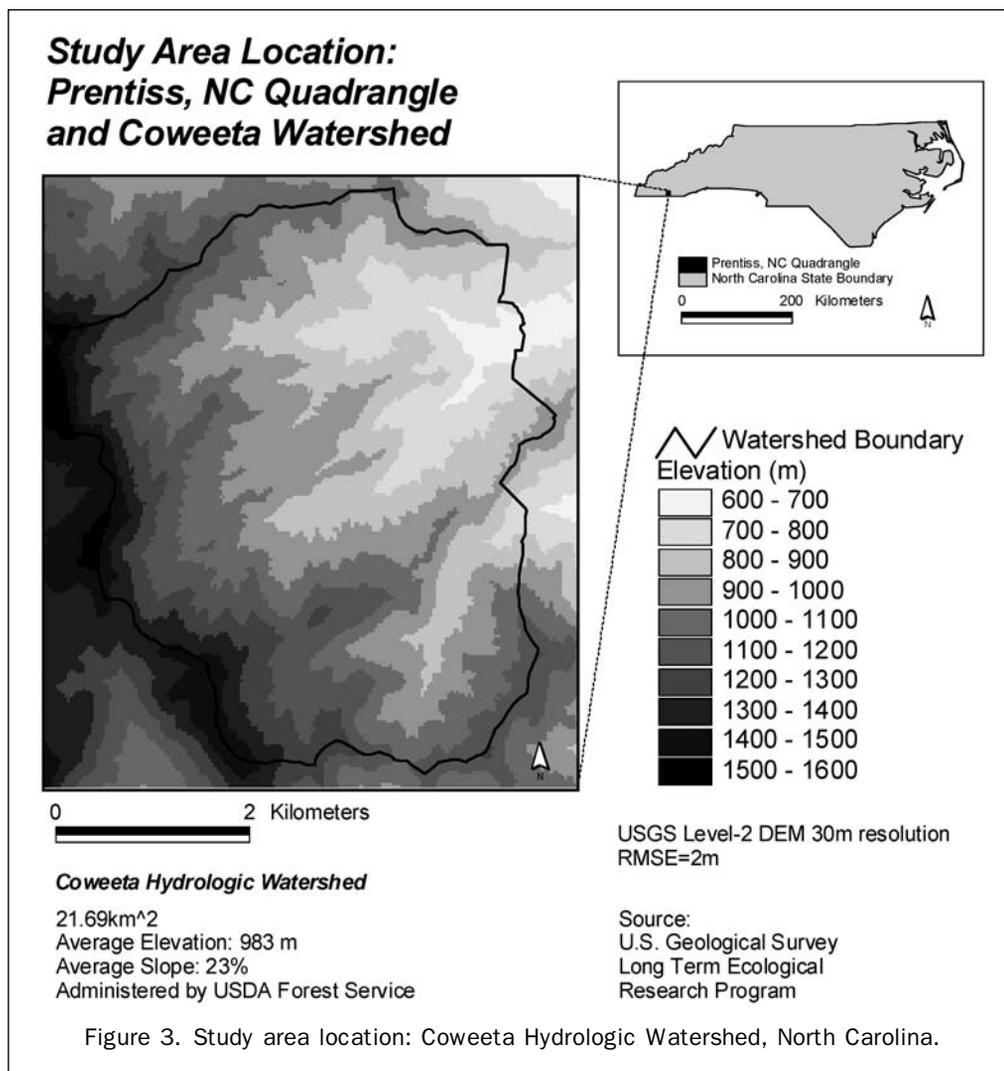
TABLE 3. UNCERTAINTY ESTIMATORS FOR THE COWEETA HYDROLOGIC WATERSHED

Filter	Stat	Elevation				Slope				TI				Upslope			
		Avg.	Std.	5th %	95th %	Avg.	Std.	5th %	95th %	Avg.	Std.	5th %	95th %	Avg.	Std.	5th %	95th %
U	ARAD	0.2%	0.03%	0.1%	0.2%	7%	9%	3%	14%	9%	7%	1.2%	22%	949%	11393%	0%	623%
N	ARAD	0.2%	0.03%	0.1%	0.2%	8%	10%	3%	16%	6%	6%	1.0%	17%	720%	9877%	0%	275%
W	ARAD	0.2%	0.03%	0.1%	0.2%	7%	9%	3%	14%	6%	6%	0.9%	17%	678%	9277%	0%	273%
S	ARAD	0.2%	0.03%	0.1%	0.2%	5%	6%	2%	10%	4%	5%	0.5%	13%	461%	7916%	0%	119%
U	Bias	0.01	0.2	−0.3	0.3	0.1	0.3	−0.4	0.5	−0.04	0.6	−0.9	0.7	3,469	335,407	−10,564	19,364
N	Bias	0.02	0.2	−0.3	0.3	0.1	0.4	−0.4	0.6	0.00	0.5	−0.6	0.6	2,341	327,921	−5,577	8,852
W	Bias	0.01	0.2	−0.3	0.3	0.1	0.3	−0.3	0.5	−0.01	0.5	−0.4	0.6	1,887	316,988	−5,912	8,893
S	Bias	0.02	0.2	−0.3	0.3	0.0	0.2	−0.3	0.4	0.00	0.4	−0.4	0.4	1,975	267,848	−2,813	3,511
U	R-Bias	0.002%	0.02%	−0.03%	0.03%	0.6%	7%	−0.9%	2%	0.5%	9%	−13%	16%	923%	11394%	−35%	593%
N	R-Bias	0.002%	0.02%	−0.03%	0.03%	0.8%	8%	−1.0%	2%	0.6%	7%	−9%	12%	703%	9877%	−29%	264%
W	R-Bias	0.002%	0.02%	−0.03%	0.03%	0.6%	7%	−0.9%	2%	0.5%	7%	−9%	12%	661%	9277%	−29%	261%
S	R-Bias	0.002%	0.02%	−0.03%	0.03%	0.3%	4%	−0.7%	1%	0.4%	6%	−7%	8%	451%	7916%	−22%	113%
U	Std	2.0	0.1	1.8	2.2	2.8	0.2	2.5	3.1	0.7	0.6	0.1	1.9	74,261	442,841	0	176,582
N	Std	2.0	0.1	1.8	2.2	3.3	0.2	2.9	3.6	0.5	0.5	0.1	1.4	60,672	419,243	0	79,687
W	Std	2.0	0.1	1.8	2.2	2.8	0.2	2.5	3.1	0.5	0.5	0.1	1.4	58,522	409,885	0	77,967
S	Std	2.0	0.1	1.8	2.1	2.1	0.1	1.9	2.3	0.3	0.5	0.0	1.0	42,887	365,293	0	25,079
U	RMSE	2.0	0.1	1.8	2.2	2.9	0.2	2.6	3.2	0.8	0.8	0.1	2.2	88,018	553,521	0	195,019
N	RMSE	2.0	0.1	1.8	2.2	3.3	0.2	3.0	3.7	0.6	0.7	0.1	1.7	72,411	530,792	0	89,031
W	RMSE	2.0	0.1	1.8	2.2	2.8	0.2	2.5	3.1	0.6	0.7	0.1	1.7	69,784	516,765	0	86,547
S	RMSE	2.0	0.1	1.8	2.2	2.1	0.1	1.9	2.3	0.4	0.6	0.0	1.2	50,803	452,155	0	27,187
U	R-RMSE	0.2%	0.04%	0.1%	0.3%	9%	11%	4%	18%	13%	10%	2%	32%	2170%	22680%	0%	1649%
N	R-RMSE	0.2%	0.04%	0.1%	0.3%	10%	13%	4%	20%	9%	9%	1%	25%	1716%	20715%	0%	690%
W	R-RMSE	0.2%	0.04%	0.1%	0.3%	8%	11%	4%	17%	9%	9%	1%	26%	1618%	19654%	0%	659%
S	R-RMSE	0.2%	0.04%	0.1%	0.3%	6%	8%	3%	13%	6%	8%	1%	20%	1089%	16472%	0%	257%
U	L-RMSE	0.002	0.0004	0.001	0.003	0.1	0.1	0.04	0.2	0.1	0.1	0.02	0.3	0.8	0.8	0.00	2.2
N	L-RMSE	0.002	0.0004	0.001	0.003	0.1	0.1	0.04	0.2	0.1	0.1	0.01	0.2	0.5	0.7	0.00	1.7
W	L-RMSE	0.002	0.0004	0.001	0.003	0.1	0.1	0.04	0.2	0.1	0.1	0.01	0.3	0.5	0.7	0.00	1.7
S	L-RMSE	0.002	0.0004	0.001	0.003	0.1	0.1	0.03	0.1	0.1	0.1	0.01	0.2	0.3	0.6	0.00	1.2

F = Filter: U = Unfiltered simulation, N = Neighborhood 3 × 3 filter, W = Weighted $D \times D$ filter, S = Mean $D \times D$ filter. D = 270 meters. N = 24,107; the number of grid cells in the Coweeta Watershed.
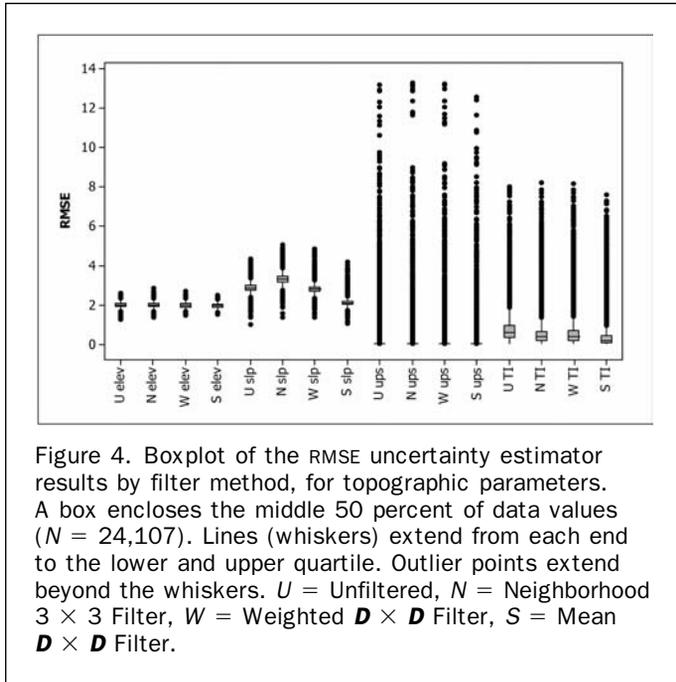
Figure 4. Boxplot of the RMSE uncertainty estimator results by filter method, for topographic parameters. A box encloses the middle 50 percent of data values (*N* = 24,107). Lines (whiskers) extend from each end to the lower and upper quartile. Outlier points extend beyond the whiskers. *U* = Unfiltered, *N* = Neighborhood 3 × 3 Filter, *W* = Weighted *D* × *D* Filter, *S* = Mean *D* × *D* Filter.

uncertainty increased under the *N* filter, indicating that the *U* filter does not represent a worst-case scenario when error is propagated to this derivative. RMSE slope result grids for the four filter methods were significantly different from each other ($p = 0.00$). The positive skew for upslope area reflects the impact that uncertainty can have on this parameter. For upslope area, only the *N* and *W* filters were not significantly different from each other ($p = 0.58$). The effect of filter method on TI is also evident. For TI, the *N* and *W* filters were also not significantly different from each other ($p = 0.58$). It is interesting to note the asymmetry of the boxplots for upslope area in Figure 4. This indicates that for some simulations, large errors in these parameters are found. One would expect this, since flow accumulation for a particular cell may change dramatically if the topographic valley, where flow accumulation is the highest, moves slightly. This error is also observed for TI, which is derived from upslope area.

Table 3 contains simulation results for each method. Results for simulations based on the *N* filter are explained here as an example. On average, elevation in the watershed is unbiased (Bias: and R-Bias are not significantly different than zero). There is a slight upward bias in elevation, although it is not significantly different from zero. This is due to the sink filling procedure of DEMs required for drainage area calculations. Elevation can be expected to deviate from the true value by 0.2 percent (ARAD) and can be expected to vary by approximately 2 meters (STD and RMSE) or 0.2 percent (R-RMSE). Errors are likely to be small since there was no significant difference between R-RMSE and L-RMSE ($p = 0.98$).

DEM error appeared to exert a greater influence on percent slope. Slope values, on average, had a positive bias (0.13) and deviated from the expected value by 0.8 percent (R-Bias) with an average absolute deviation of 8 percent (ARAD). Slope can be expected to vary on average by 3 percent (RMSE and STD) with a relative variation of 10 percent (R-RMSE). R-RMSE and L-RMSE for slope were not significantly different from each other ($p = 0.35$).

Upslope contributing area estimators were highly variable as exhibited by the standard deviations of uncertainty estimators. Although bias values are large, they were not

significantly different from zero and thus upslope area was unbiased. On average, upslope area can be expected to deviate from the true value by approximately 700 percent (R-Bias) with an absolute deviation of 720 percent (ARAD). Values can be expected to vary by approximately 61,000 m^2 (STD) and 72,000 m^2 (RMSE). Upslope area can be expected to exhibit a relative variation of 1,700 percent (R-RMSE). R-RMSE and L-RMSE were significantly different from each other ($p = 0.00$), further indicating a biased estimator. The low L-RMSE indicates that the upslope area was overestimated in this simulation.

The average topographic index was unbiased and can be expected to deviate from the true value by 0.6 percent (R-Bias) with an absolute deviation of 6 percent (ARAD). STD and RMSE were significantly different from each other ($p = 0.00$) indicating some bias; Mean R-RMSE and L-RMSE were also significantly different ($p = 0.01$). TI values for this example can be expected to vary by 0.5 (STD) or 0.6 (RMSE) with an average variation of 9 percent (R-RMSE).

Although the uncertainty estimators reported in this section were averaged for the entire watershed, they provide a comprehensive representation of uncertainty and a significant improvement in our ability to represent uncertainty for elevation and derived parameters. These values are more descriptive than the non-spatial RMSE provided with the DEM.

Visualization

Statistical grids produced as a result of the simulations quantify uncertainty on a watershed or on a cell-by-cell basis and enable visualization of the impacts of random DEM error. An investigator can obtain estimator values for a specific location interactively by clicking on a cell in a GIS. Plate 1 provides an example of slope uncertainty estimators. Slope estimators are quite sensitive to DEM errors. The lower elevation valley areas are especially sensitive.

The impact of topographic relationships can be discerned through visual inspection of the result grids. From the RMSE result grids (Plate 2), it is evident that higher RMSE estimates occur in valley areas for upslope area and TI. The upslope area values are inherently larger in valleys and these locations incorporate the influence of upslope uncertainty. RMSE is lower in valley areas for elevation and slope. These findings could have implications for studies located in areas of lower elevation (such as agricultural watersheds) that incorporate upslope contributing area and topographic index parameters generated from a DEM.

Conclusions

Error occurs in spatial data. The traditional response to error has been to ignore it because easily employed methods to address error do not exist (Openshaw, 1989). Openshaw provided the following reasons for the GIS community's neglect of errors in spatial data: (a) the severity of the problem is unknown due to a lack of empirical research; (b) the effects of uncertainty are application and data specific, so there is no consensus as to whether it really is a problem; and (c) there are no established procedures for dealing with error in many GIS functions or for tracking error propagation (Openshaw, 1989, p. 264). In more than a decade since these issues were acknowledged, little progress has been made in providing methods for GIS users to address DEM errors.

The methodology and its application presented here demonstrate an easily employed method to assess DEM error and its impact on derived topographic parameter estimators. Previous research used techniques that require higher accuracy data sources such as GPS surveys or higher resolution DEMs. The reality is that most DEM users do not have such data. In the absence of this data, the only

Plate 1. Uncertainty estimators of slope percent: (a) ARAD, (b) Bias, (c) R-Bias, (d) STD, (e) RMSE, (f) R-RMSE, and (g) L-RMSE. Values represent the standard deviation from the mean value for each uncertainty estimator resulting from the Unfiltered (*U*) uncertainty simulation, *N* = 135.
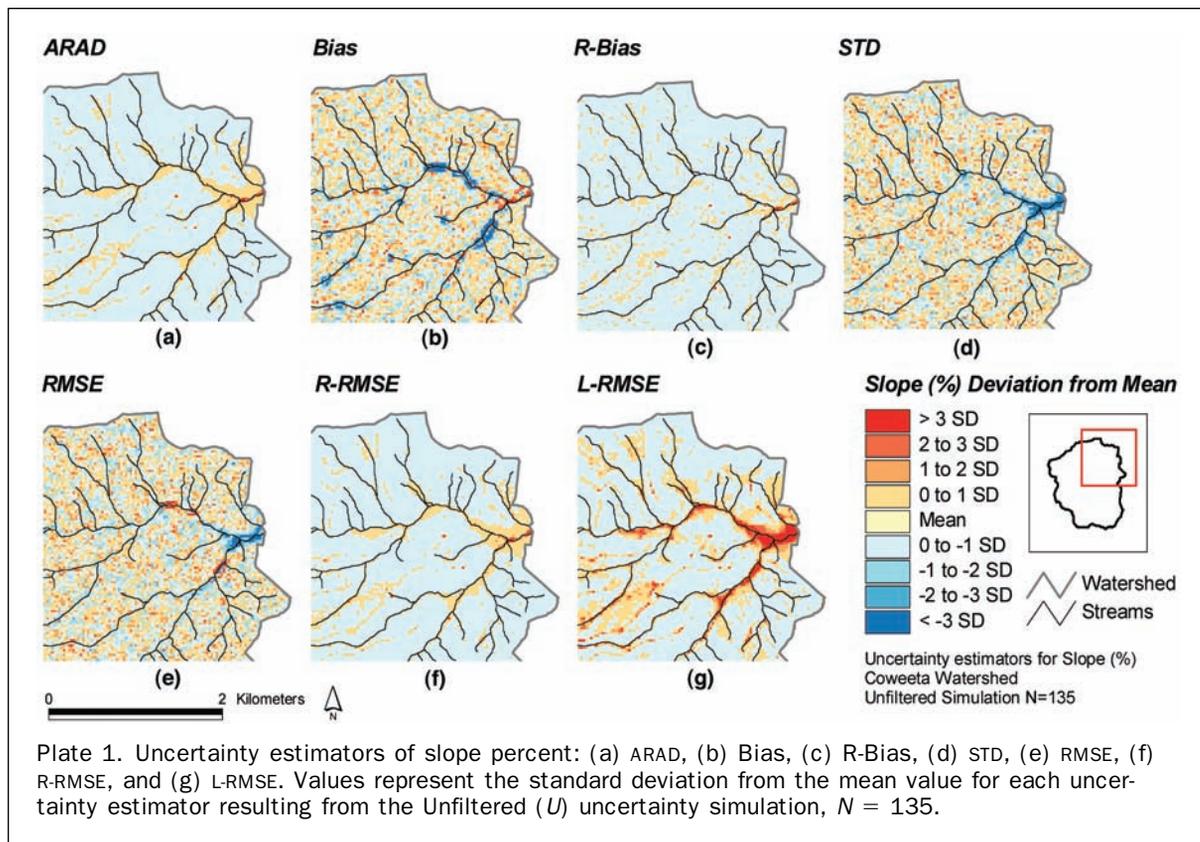
information regarding vertical accuracy is the RMSE provided by the DEM provider (such as the USGS). The methodology described is one that most users could integrate using the basic functions of most GIS packages. According to Veregin (1997), ". . . commercial GIS packages offer few capabilities for documenting and tracking error through sequences of data processing steps, such that users are often unable to ascertain the reliability of derived data." (p. 67). The methodology presented here was designed to remedy this issue by providing a methodology that was successfully implemented within a commercial GIS.

The purpose of this methodology is to provide DEM users with a suite of tools by which they can evaluate the effect of uncertainty in DEMs and derived topographic parameters. DEM users often apply the DEM as a *truth* surface rather than as a model (as its name implies). Although the simulation techniques employed in the demonstrated methodology are well-established procedures in general, their application to DEM derivatives is novel. This application addresses the sensitivity of DEM derivatives to errors, which are rarely acknowledged and thus not adjusted for in subsequent applications.

The need for and use of accurate DEMs is evident (Mondello *et al.*, 2004). Numerous applications utilize DEMs. The effects of DEM error will vary by DEM product, application, and location. The demonstrated methodology did not assess uncertainty associated with specific DEM applications such as hydrologic modeling or hazard analyses. Researchers can, however, use the uncertainty estimates provided by the proposed simulation techniques to better assess uncertainty for projects that utilize DEMs and DEM-derived data. For example, input parameters for hydrologic models (such as those for non-point source pollutant prediction) might require elevation and slope values that are frequently obtained directly from a DEM. A modeler might refine model inputs based on the calculated uncertainty of these

parameters, and thus be better suited to understand the uncertainty in the output of their models. The methodology can be applied to investigate relationships between terrain and uncertainty in elevation and derived parameter. Subsequent research could also apply the methodology to investigate the scaling of uncertainty by evaluating impacts of DEM error on elevation and derived parameters at various grid scales.

It is the responsibility of the DEM user to determine whether uncertainty in the DEM will affect results from specific analyses that utilize data derived from a particular DEM. Use of the demonstrated methodology may result in a more veridical use of DEM data and reporting of results derived from their use. Random DEM errors that result in uncertainty may have little effect on data derived from DEMs and decisions based on DEM analyses; however, it is appropriate to make this judgment only after a thorough evaluation. This research provides tools to assist in such an evaluation.

Acknowledgments

References
Beven, K., and M. Kirkby, 1979. A physically based variable contributing area model of basin hydrology, *Hydrological Sciences Bulletin*, 24(1):43–69.

Black, P., 1959. Interception of rainfall by a hardwood canopy, *Dergisi* 9(2):218–224.
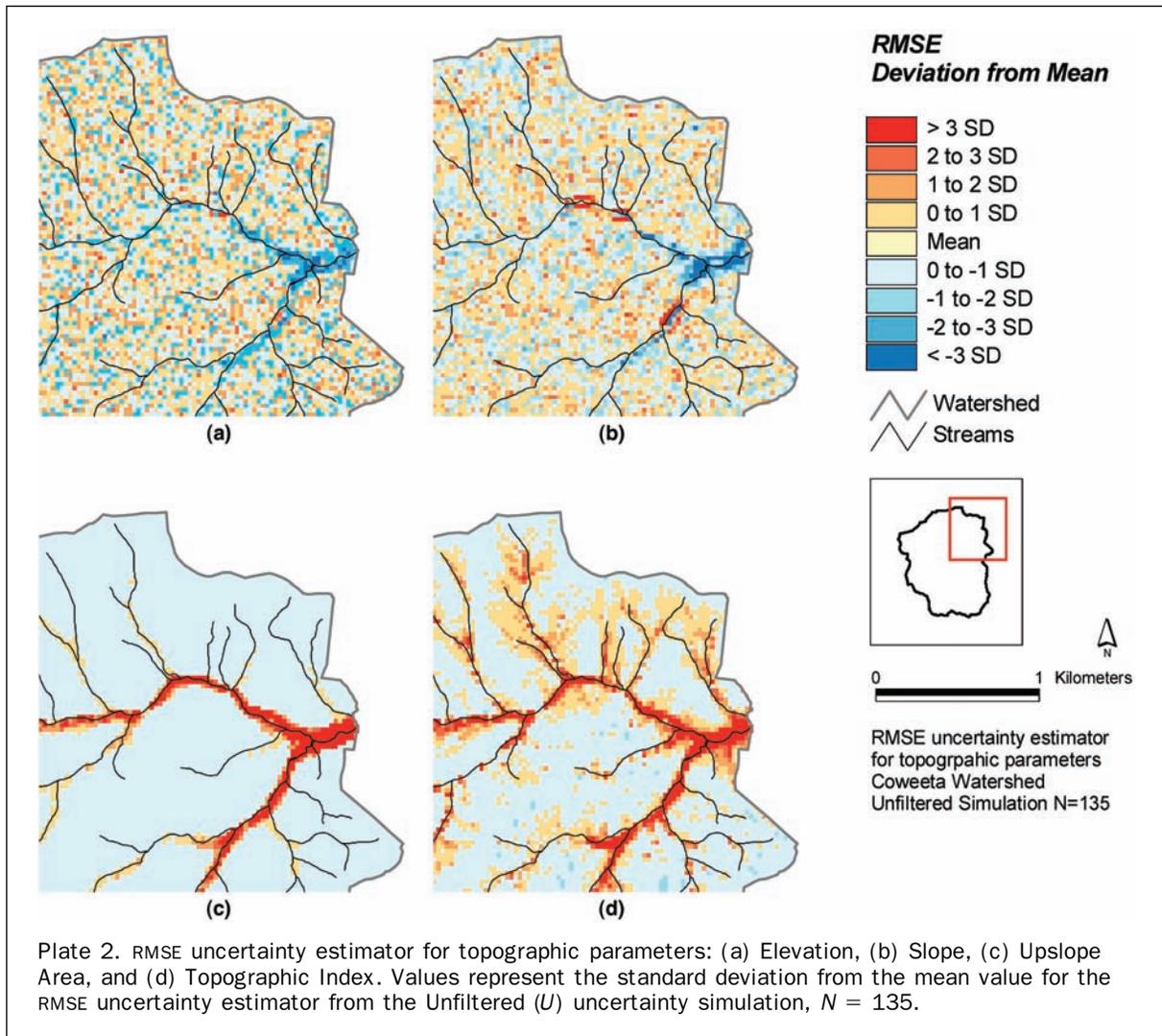
Plate 2. RMSE uncertainty estimator for topographic parameters: (a) Elevation, (b) Slope, (c) Upslope Area, and (d) Topographic Index. Values represent the standard deviation from the mean value for the RMSE uncertainty estimator from the Unfiltered (*U*) uncertainty simulation, *N* = 135.

Bolstad, P., and W. Swank, 1997. Cumulative impacts of landuse on water quality in a southern Appalachian watershed, *Journal of the American Water Resources Association*, 33(3):519–533.

Burrough, P., and R. McDonnell, 1998. *Principles of Geographic Information Systems*, Oxford University Press, New York, New York, 333 p.

Carter, J., 1990. Some effects of spatial resolution in the calculation of slope using the spatial derivative, *ASPRS-ACSM Annual Convention Technical Papers, Volume 1*, 18–23 March, Denver, Colorado, pp. 43–52.

Caruso, V., 1987. Standards for digital elevation models, *ASPRS-ACSM Annual Convention Technical Papers, Volume 4*, ASPRS, Falls Church, Virginia, pp. 159–166.

Costa-Cabral, M., and S. Burgess, 1994. Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas, *Water Resources Research*, 30(3):1681–1692.

Day, F., and C. Monk, 1974. Vegetation patterns on a southern Appalachian watershed, *Ecology*, 55:1064–1074.

Eastman, J.R., 1992. *IDRISI Technical Reference Version 4.0*, Clark University Graduate School of Geography, Worcester, Massachusetts.

Ehlschlaeger, C., 2002, Representing multiple spatial statistics in generalized elevation uncertainty models: moving beyond the variogram, *International Journal of Geographical Information Science*, 16(3):259–285.

Ehlschlaeger, C., 1998. *The Stochastic Simulation Approach: Tools for Representing Spatial Application Uncertainty*, Ph.D. Dissertation, University of California, Santa Barbara, California.

Ehlschlaeger, C., and A. Shortridge, 1996. Modeling elevation uncertainty in geographical analyses, *Proceedings of the International Symposium on Spatial Data Handling*, Delft, Netherlands, pp. 9B.15–9B.25.

Environmental Systems Research Institute (ESRI), 1998, *ArcView Spatial Analyst OnLine User Guide*, Redlands, California.

Evans, I., 1998. What do terrain statistics really mean?, *Landform Monitoring, Modelling and Analysis* (S.N. Lane, K.S. Richards, and J.H. Chandler, editors), John Wiley and Sons, pp. 119–137.

Fisher, P., 1991. First experiments in viewshed uncertainty: The accuracy of the viewshed area, *Photogrammetric Engineering & Remote Sensing*, 57(10):1321–1327.

Heuvelink, G., P. Burrough, and A. Stein, 1989. Propagation of errors in spatial modeling with GIS, *International Journal of Geographical Information Systems*, 3(4):303–322.

Holmes, K., O. Chadwick, and P. Kyriakidis, 2000. Error in a USGS 30-meter digital elevation model and its impact on terrain modeling, *Journal of Hydrology*, 233:154–173.

Horn, B., 1981. Hillshading and the reflectance map, *Proceedings of the IEEE*, 69(1):4–47.

Hunter, G., and M. Goodchild, 1997. Modeling the uncertainty of slope and aspect estimates derived from spatial databases, *Geographical Analysis*, 29(1):35–49.

Hunter, G., and M. Goodchild, 1995. Dealing with error in spatial databases: A simple case study, *Photogrammetric Engineering & Remote Sensing*, 61(5):529–537.

Jenson, S., and J. Domingue, 1988. Extracting topographic structure from digital elevation data for geographic information system

analysis, *Photogrammetric Engineering & Remote Sensing*, 54(11):1593–1600.

Kroll, C., and J. Stedinger, 1996. Estimation of moments and quantiles using censored data, *Water Resources Research*, 32(4):1005–1012.

Lanter, D., and H. Veregin, 1992. A research based paradigm for propagating error in layer-based GIS, *Photogrammetric Engineering & Remote Sensing*, 58(6):825–833.

Lee, J., P. Snyder, and P. Fisher, 1992. Modeling the effect of data errors on feature extraction from digital elevation models, *Photogrammetric Engineering & Remote Sensing*, 58(10): 1461–1467.

Lettenmaier, D., and E. Wood, 1993. Hydrologic forecasting, *Handbook of Hydrology* (D. Maidment, editor), McGraw Hill, New York, New York.

Liu, R., and L. Herrington, 1993. The Effects of Spatial Data Error on Forest Management Decisions, *Proceedings of the GIS 1993 Symposium,* February, Vancouver, British Columbia, pp. 1–8.

Long Term Ecological Research (LTER), 2001, Coweeta information site characteristics, URL: http://sparc.ecology.uga.edu/webdocs/coweeta/html/ronsite-chars.html, (last date accessed: 19 June 2006).

Lopez, C., 1997. Locating some types of random errors in digital terrain models, *International Journal of Geographical Information Science*, 11(7):677–698.

Mondello, C., G. Hepner, and R. Williamson, 2004. 10-year industry forecast phases I-III – Study documentation, *Photogrammetric Engineering & Remote Sensing*, 70(1):7–58.

Monmonier, M., 1991. *How to Lie with Maps*, The University of Chicago Press, Chicago, Illinois, 207 p.

Monkton, C., 1994. An investigation into the spatial structure of error in digital elevation data, *Innovations in GIS 1* (M.F. Worboys, editor), London, Taylor and Francis, pp. 201–211.

Oliver, D., 1995. Moving averages for Gaussian simulation in two and three dimensions, *Mathematical Geology*, 27(8):939–960.

Openshaw, S., M. Charlton, and S. Carver, 1991. Error propagation: A Monte Carlo simulation, *Handling Geographic Information: Methodology and Potential Applications* (I. Masser and M. Blackemore, editors), Longman Scientific and Technical, UK, pp. 78–102.

Östman, A., 1987. Accuracy estimation of digital elevation data banks, *Photogrammetric Engineering & Remote Sensing*, 53(4): 425–430.

Quinn, P., K. Beven, P. Chevallier, and O. Plancon, 1991. The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models, *Hydrological Processes*, 5:59–79.

Ryder, W., and D. Voyadgis, 1996. Measuring the performance of algorithms for generating ground slope, *Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium*, 21–23 May, Fort Collins, Colorado, pp. 208–216.

Skidmore, A., 1989. A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model, *International Journal of GIS*, 3(4):323–334.

Tarboton, D., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water Resources Research*, 33(2):309–319.

U.S. Geological Survey (USGS), 2004. USGS geodata download statistics, U.S. Geological Survey, Reston, Virginia, URL: http://edc2.usgs.gov/stats/geoData.asp?year=1998 (last date accessed: 19 June 2006).

U.S. Geological Survey (USGS), 1995. *National Mapping Program Technical Instructions, Part 2-Specifications*, Department of the Interior, Washington, D.C.

Veregin, H., 1997. The effects of vertical error in digital elevation models on the determination of flow path direction, *Cartography and Geographic Information Systems*, 24(2):67–79.

Wechsler, S., 2003. Perceptions of digital elevation model uncertainty by DEM users, *Urban and Regional Information Systems Association Journal*, 15(2):57–64.

Wolock, D., and G. McCabe, 1995. Comparison of single and multiple flow direction algorithms for computing topographic parameters, *Water Resources Research*, 31(5):1315–1324.

Wright, J., 1942. Map makers are human: Comments on the subjective in maps, *The Geographical Review*, 32(4):527–544.