An Assessment of the Effects of Cell Size on AGNPS Modeling of Watershed Runoff

Shuo-sheng Wu, E. Lynn Usery, Michael P. Finn, and David D. Bosch

ABSTRACT: This study investigates the changes in simulated watershed runoff from the Agricultural NonPoint Source (AGNPS) pollution model as a function of model input cell size resolution for eight different cell sizes (30 m, 60 m, 120 m, 210 m, 240 m, 480 m, 960 m, and 1920 m) for the Little River Watershed (Georgia, USA). Overland cell runoff (area-weighted cell runoff), total runoff volume, clustering statistics, and hot spot patterns were examined for the different cell sizes and trends identified. Total runoff volumes decreased with increasing cell size. Using data sets of 210-m cell size or smaller in conjunction with a representative watershed boundary allows one to model the runoff volumes within 0.2 percent accuracy. The runoff clustering statistics decrease with increasing cell size; a cell size of 960 m or smaller is necessary to indicate significant high-runoff clustering. Runoff hot spot areas have a decreasing trend with increasing cell size; a cell size of 240 m or smaller is required to detect important hot spots. Conclusions regarding cell size effects on runoff estimation cannot be applied to local watershed areas due to the inconsistent changes of runoff volume with cell size; but, optimal cells sizes for clustering and hot spot analyses are applicable to local watershed areas due to the consistent trends.

KEYWORDS: Spatial statistics, AGNPS, cell size, resolution effects, watershed modeling

Introduction

The Agricultural NonPoint Source (AGNPS) pollution model was developed by the U.S. Department of Agriculture in response to the need to quantitatively examine the influence of non-point source pollution on surface water and groundwater quality in agricultural watersheds. The model is designed to estimate sediment and nutrient yields from agricultural activities, and it is used most appropriately to compare the impacts of alternative land management strategies on surface water quality (Young et al. 1989; Panuska et al. 1991). This model is a distributed, cell-by-cell, event-based watershed model that predicts surface runoff (including total runoff volume and peak runoff rate), sediment yield, and nutrient loading of phosphorus, nitrogen, and chemical oxygen demand (COD) on the basis of 22 input parameters derived from soil, land cover, and elevation data (Young et al. 1994; Young et al. 1995).

The purpose of this study is to investigate how spatial statistics of AGNPS-simulated runoff change when model inputs of different cell sizes are used. Specifically, we are interested in the changes of the overall runoff clustering level and the location and size of runoff clusters. In addition, we investigate the changes of overland cell runoff and total runoff volumes, as well as whether the effects of cell size are consistent throughout the watershed. We calculated spatial statistics of the global Moran's index and the global G statistic (for the entire watershed) to indicate the overall clustering degree and the degree of high- or low-runoff clustering, respectively. We computed spatial statistics of the local Moran's index and the local G statistic (by cells) to locate clusters of similar runoff values as well as high- and low-runoff clusters, respectively. The runoff clustering degree and cluster locations

Shuo-sheng Wu, Department of Geography, Geology and Planning, Missouri State University 901 South National Avenue, Springfield, Missouri 65897. Phone: 417-836-5281. E-mail: <sw1020@gmail.com>. **E. Lynn Usery**, Center of Excellence for Geospatial Information Science, U.S. Geological Survey, 1400 Independence Road, Rolla, Missouri 65401. Phone: 573-308-3837. E-mail: <usery@usgs.gov>. **Michael P. Finn**, Center of Excellence for Geospatial Information Science, U.S. Geological Survey, 1400 Independence Road, Rolla, Missouri 65401. Phone: 573-308-3931. E-mail: <mfinn@usgs.gov>. **David D. Bosch**, Southeast Watershed Research Laboratory, Agricultural Research Service, U.S. Department of Agriculture, 2375 Rainwater Road, Tifton, Georgia 31794. Phone: 229-386-3899. E-mail: <David.Bosch@ars.usda.gov>.

are of great concern to watershed researchers and managers because high runoff contributes to soil erosion and deteriorates water quality; studying clustering patterns of modeled runoff helps decision making in optimal land management strategies. The results of this study have implications for watershed quality management and water resources planning.

In the following sections, we first briefly review past studies of cell size effects (or resolution effects) on watershed modeling. Then we describe our study area, data source, analysis procedures, and results. The calculation of spatial statistics is explained in detail. Lastly, we discuss limitations of our analyses and future research directions before concluding with major findings of this study.

Literature Review

Previous research has explored the effects of input cell size (or resolution) on watershed runoff modeling. For example, Zhang and Montgomery (1994) and Wolock and Price (1994) found that the predicted peak runoff rate from the TOPMODEL hydrological model increases with increasing DEM cell size; Bruneau et al. (1995) reported the same finding while noticing that the TOPMODEL-simulated catchment area has an opposite decreasing trend with increasing DEM cell size.

Using the KINEROS hydrological model, Thieken et al. (1999) found that the simulated peak runoff rate and total runoff volume both increase with increasing cell size. Canfield and Goodrich (2006) also observed an overall increasing trend of KINEROS2-simulated runoff volume and peak discharge, though the increase was found of no statistically significance.

In contrast to the increasing trend of runoff volume, Vieux and Needham (1993) found that the total runoff volume predicted from the AGNPS pollution model decreases with increasing cell size. Bhuyan et al. (2001) also reported a decreasing trend of AGNPS-simulated total runoff volume, but they observed that the peak runoff rate increases with increasing cell size. Davenport et al. (2003) found that both AGNPS-simulated total runoff volume and peak runoff rate increase with increasing cell size, noting the corresponding increasing trend of the drainage area contributing to the total flow.

In comparing results from past studies of cell size effects on runoff modeling, it is important to understand that different hydrological models have different input requirements and, therefore, modeling assumptions, and the results from different models may not be directly compared. For example, the TOPMODEL and KINEROS hydrological models generally assume spatially uniform soil- and land-use parameters while the AGNPS pollution model allows for spatially varied soil- and land-use input parameters. Nevertheless, a consistent finding from past literature is the increase of peak runoff rate with increasing cell size. Regarding trends of total runoff volume, studies using TOPMODEL, KINEROS, and KINEROS2 hydrological models tend to find increasing runoff volume with increasing cell size, while the AGNPS pollution model has produced mixed results.

It is of particular interest to understand why past studies have derived opposite conclusions regarding cell size effects on AGNPS-predicted runoff volume. This study investigates the cause by testing the hypothesis that, if the comparison of runoff volume modeled at different cell sizes is based on a common watershed boundaries, the runoff volume tends to decrease. On the other hand, if the comparison of runoff volume is based on individual watershed boundaries that generally increase with increasing cell size, the runoff volume tends to increase. Although past literature does not provide sufficient details of the data processing and analysis procedures to verify the hypothesis, the empirical results of this study have confirmed our hypothesis, as described later in this article.

In addition to investigating the different conclusions from past studies, this study is the one, if not the first, to use spatial statistics to (quantitatively) examine clustering level and hot spot patterns of watershed runoff and also to study cell size effects on the analyses; the reason being the intensive computation of spatial statistics applied to grid-based runoff data. Spatial statistics allow one to determine the overall clustering degree of cell runoff in the watershed and to locate statistically significant clusters of similar runoff values, which have particular implications for comparing alternative land management strategies in the context of AGNPS water quality modeling. High-runoff clustering contributes to erosion and pollution of the water quality; the degree of high-runoff clustering can be used as an indicator of the overall water quality in the watershed. Furthermore, visually examining statistically significant high-runoff clusters provides significant insight regarding critical watershed areas to mitigate the impact of human activities on surface water quality.

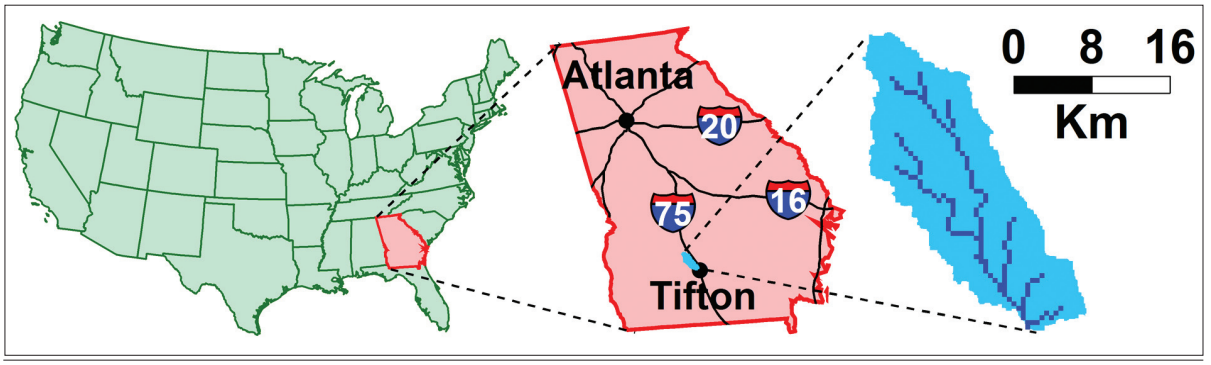


Figure 1. Little River Watershed, Georgia, USA.

Designed to analyze the distribution of geographic features, spatial statistics differ from traditional statistics in that space and spatial relationships are an integral and implicit component of analysis (ESRI 2005). Spatial statistics are commonly used to analyze vector-based socioeconomic data. In this study, we used spatial statistics to examine cell-based watershed modeling outputs by converting runoff raster data to vector point data. Watershed surface runoff varies geographically; therefore, it is appropriate to use spatial statistics to examine spatial autocorrelation properties and reveal geographic patterns of watershed runoff. Results from this spatial statistics analysis will have implications for water resources planning and water quality management, particularly in the context of investigating optimal cell size for effective and efficient watershed runoff modeling.

Study Area and Data Source

The study area is the Little River Watershed in Georgia, USA, which is approximately 12 km wide and 35 km long, and covers 334 square kilometers (Figure 1). The Little River Watershed is one of the experimental watersheds for the U. S. Department of Agriculture (USDA) Agricultural Research Service (ARS) Southeast Watershed Research Laboratory (SEWRL). The primary goal of the SEWRL is to "develop an improved understanding of basic hydrologic and water quality processes on Coastal Plain watersheds and evaluate the effects of agricultural management practices on the region's natural resources and environment" (Bosch et al. 2007).

This study used the runoff datasets of eight different cell sizes that have been previously simulated using the AGNPS pollution model by Usery et al. (2004). The modeling was conducted based on a simulated precipitation event of 17.78 cm, with input data at eight cell sizes of 30-m, 60-m, 120-m, 210-m, 240-m, 480-m, 960-m, and 1,920-m. The input data of elevation, land cover, and soil at different cell sizes were resampled from 30-m resolution baseline datasets. The 210-m cell datasets were generated because a 210-m cell, approximately 10 acres in size, is commonly used by USDA scientists for watershed modeling. The 30-m elevation data set comes from the USGS National Elevation Dataset (NED) (USGS 2006). The 30-m land cover data set is based on the USGS National Land Cover Database (NLCD) (USGS 2007). The 30-m soil data set originates from the USDA Digital Soil Geographic Databases (USDA 2006); this data set was originally in a vector polygon format and was converted to a 30-m raster to match the land cover and elevation datasets for watershed modeling purpose. The land cover and soil data sets are categorical data and were resampled using the nearest neighbor method, while the elevation data set is numerical data and was resampled using the bilinear interpolation method.

Examining Overland Cell Runoff and Watershed Runoff Volumes

Overland cell runoff and the watershed total runoff volume for the eight data sets of different cell sizes were examined first. When visually comparing the overland runoff at different cell sizes, runoff levels are high at tributary sources (Figure 2). For all cell sizes, the overland runoff ranges from 0.13 cm (0.05 inch) to 18.54 cm (7.3 inches). High runoff patches are at approximately the same locations and of the same sizes. It appears that there is no cell size effect on the overland runoff pattern except the generalization effect due to larger cells.

One consequence of resampling is that the boundaries and areas of the watershed become larger at

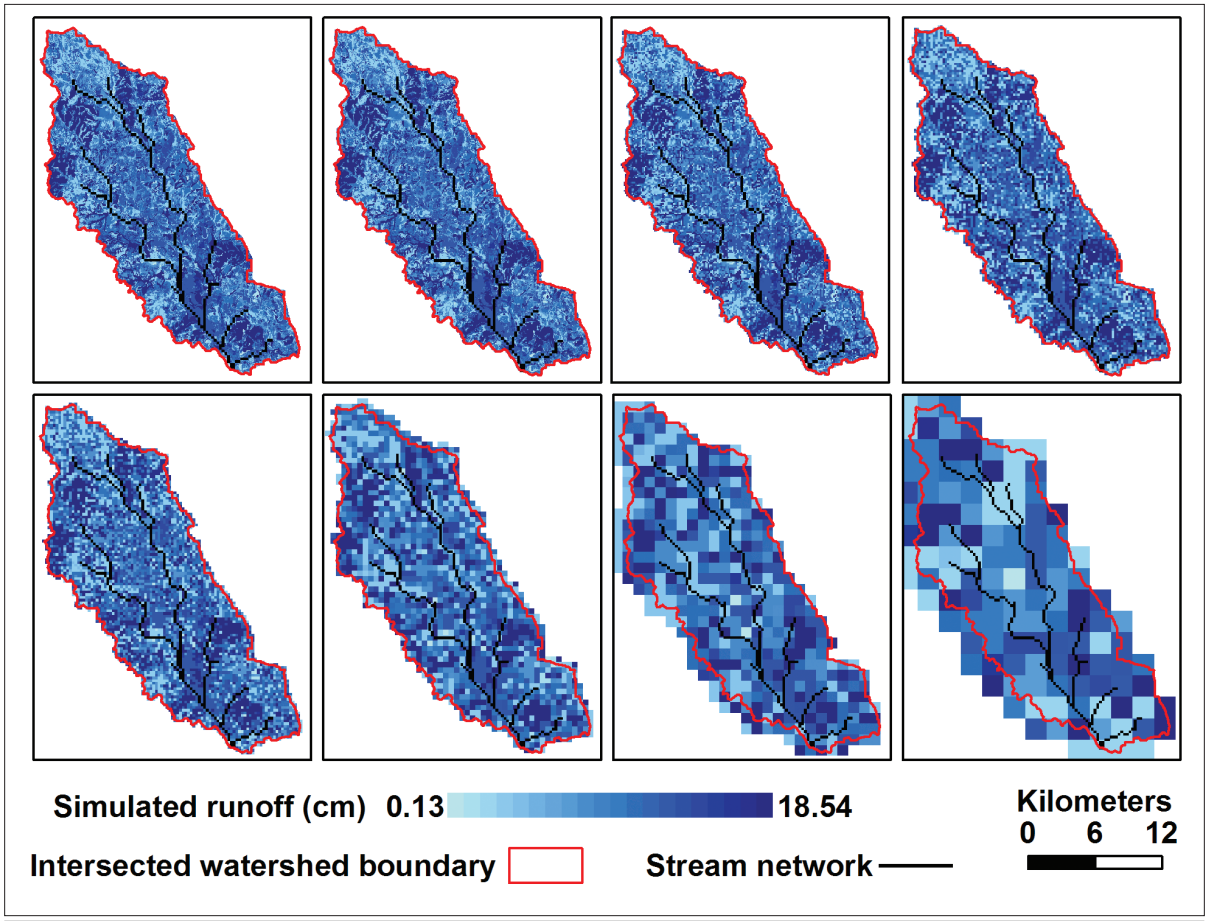


Figure 2. AGNPS simulated runoff maps at different cell sizes, Little River, Georgia, USA.

larger cell sizes (Figure 2 and Figure 3). Because the large watershed areas at large cell sizes are artificial, total runoff volumes at intersected watershed boundaries were compared, in addition to those based on individual watershed boundaries. The intersected watershed boundaries were calculated as the geometric intersection of common watershed areas of different cell sizes because watershed boundaries at large cell sizes do not necessarily completely contain watershed boundaries at small cell sizes. To compare the total runoff volumes at different cell sizes, the products of cell area and overland runoff for the entire watershed were summed. The total runoff volume based on individual watershed boundaries increases with increasing cell size (Figure 4A), which is consistent with the trend of total watershed areas. However, the total runoff volume based on the intersected watershed boundaries has an overall decreasing trend with increasing cell size (Figure 4B), which represents the real effect of cell size on runoff volumes.

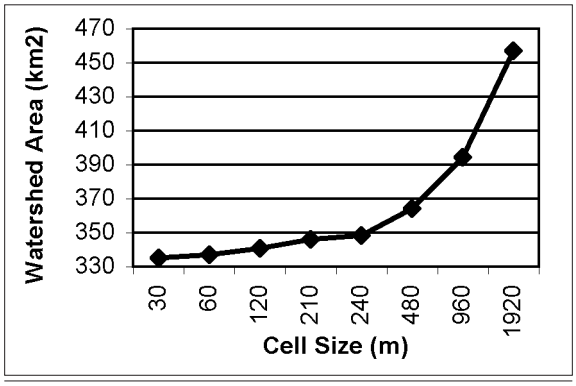


Figure 3. Watershed areas at different cell sizes.

The total runoff volume at 30-m cell size was used as a benchmark in comparisons involving runoff volumes at larger cell sizes. The total runoff volume based on individual watershed boundaries has a substantial change rate of up to 30 percent at larger cell sizes (Figure 4C); the change corresponds with the dramatic increase of watershed areas at large cell sizes (Figure 3). In contrast, the total runoff volume based on intersected watershed boundaries has a maximum change of 2.5 percent as cell sizes increase (Figure 4D). Total runoff volumes in the range are therefore accepted as representative estimates. The implication for watershed scientists is that they can use data sets at 1920-m cell size to estimate the total runoff volume within 2.5 percent accuracy if the calcula-

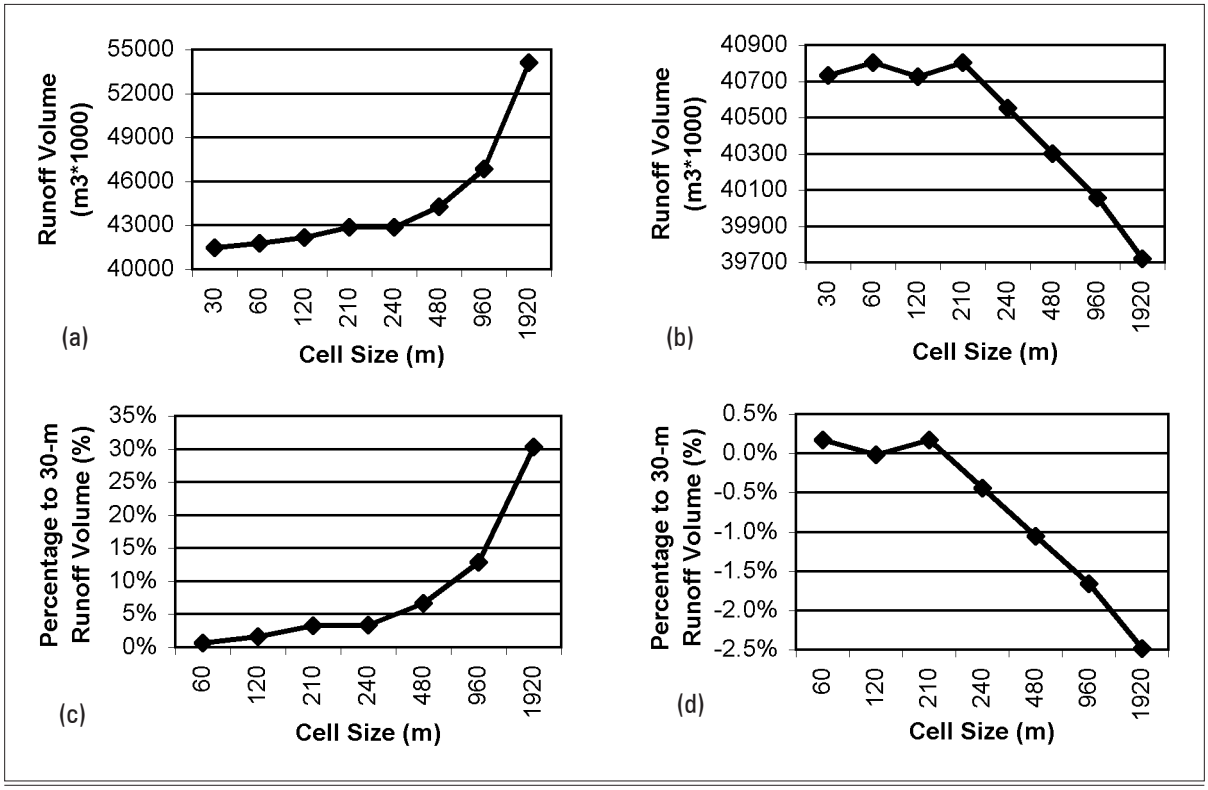


Figure 4. Total runoff volumes based on individual watershed boundaries (a) and intersected watershed boundaries (b) at different cell sizes; percentages to 30-m runoff volumes based on individual watershed boundaries (c) and intersected watershed boundaries (d).

tion of the total runoff volume is based on relatively accurate watershed boundaries. Furthermore, the change of runoff volumes from 30-m to 210-m cell sizes fluctuates within a 0.2 percent range (Figure 4D), indicating that a relatively accurate estimate of total runoff volume can be obtained using relatively coarse data sets of up to 210-m cell size.

Analyzing Overall Clustering Level of Runoff

To investigate the effect of cell size on the overall clustering level of watershed runoff, the global Moran's index was used to represent the clustering degree (Wong and Lee 2005). The global Moran's index (I) is calculated as:

$$I = \frac{n\sum_i \sum_j W_{ij}\left(X_i - \bar{X}\right)\left(X_j - \bar{X}\right)}{\sum_i \sum_j W_{ij} \sum_i \left(X_i - \bar{X}\right)^2} \quad (1)$$

where:

X_i = the runoff value for the ith cell;
$\bar{X}$ = the mean runoff of the watershed;
X_j = the runoff value for the jth cell;
W_{ij} = weight parameter for the pair of cells i and j to represent proximity relations; and
n = the number of runoff cells in the watershed.

The equation basically tests if near cells have more similar values than distant cells. The sum of cross-products in the equation will be positive if more neighboring cells have similar values; the value of I will be greater than 0, indicating similar values are clustered. The sum of the cross-products in the equation will be negative if there are more neighboring cells with dissimilar values; the value of I will be less than 0, indicating similar values are dispersed. If some neighboring cells have similar values and some do not and there are roughly as many cell pairs with positive cross-products as there are with negative cross-products, the result of summing the cross-products will be close to 0, indicating a random distribution.

A cutoff neighborhood distance was used to define the weight function in calculating the global Moran's index. Cell pairs within this distance were assigned a weight of 1 and all other pairs were assigned a weight of 0. This neighborhood distance can be regarded as the scope of spatial interaction and can be determined from the variogram analysis, using the variogram range as the neighborhood

distance. The variogram for watershed runoff was calculated, and the variogram range was determined on the basis of the first semivariance peak.

The Z score of the Moran's index was computed to compare the clustering degrees at different cell sizes. The Z score indicates whether similar runoff values are closer than would occur by chance. It is, therefore, a normalized statistic that can be appropriately used to compare data sets of different numbers of cells. The Z score of the Moran's index (Z_I) is calculated as:

$$Z_I = \frac{I_O - I_E}{S(I_E)} \tag{2}$$

where:

I_O = the observed I;

I_E = the expected I for a random distribution; and

$S(I_E)$ = the standard deviation of the expected I.

The expected value of I for a random distribution depends on the number of cells in the data set, calculated as -1/(n-1), where n is the number of runoff cells in the watershed. This is a very small number that is close to 0; an observed I greater than 0 would result in a positive Z score, indicating a clustering pattern, and an observed I less than 0 would result in a negative Z score, indicating a dispersed pattern.

The graph of the global Moran's index Z score against cell size (Figure 5a) shows that the clustering degree (represented by the Moran's index Z score) decreases with increasing cell size and that watershed runoff has statistically significant clustering patterns at all cell sizes except 1920 m. The implication for watershed scientists is that they can use modeled runoff data of a large cell size of up to 960 m (for data collection convenience and model computational efficiency) to detect whether surface runoff across the watershed is clustered with statistical significance.

Furthermore, to determine whether the overall clustering pattern is contributed by high- or low-runoff clustering, the global G statistic (Getis 1991) was calculated as:

$$G = \frac{\sum_i \sum_j W_{ij} \left(X_i \cdot X_j \right)}{\sum_i \sum_j \left(X_i \cdot X_j \right)} \tag{3}$$

where:

X_i = the runoff value for the ith cell;

X_j = the runoff value for the jth cell; and

W_{ij} = a weight parameter for the pair of cells i and j to represent proximity relations.

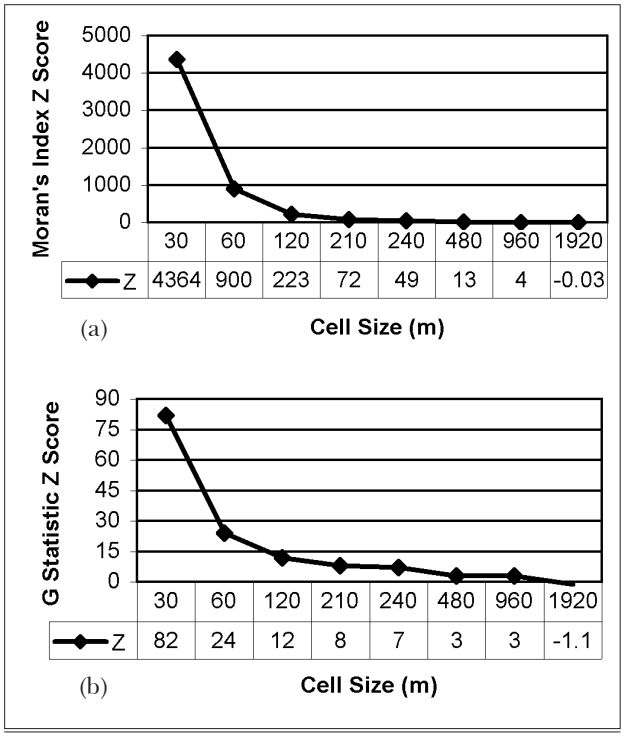


Figure 5. Global Moran's index Z scores (a) and global G statistic Z scores (b) at different cell sizes.

A large value of G indicates that high-runoff cells are clustered, while a small value of G indicates that low-runoff cells are clustered. We used the same cutoff neighborhood distance parameter to calculate the global Moran's index and then the global G statistic. The Z score of the global G statistic was computed to indicate the level at which the clustering pattern is significantly different from a random distribution of surface runoff. The Z score of G (Z_G) is calculated as:

$$Z_G = \frac{G_O - G_E}{S(G_E)} \tag{4}$$

where:

G_O = the observed G;

G_E = the expected G were there are no particular concentration of high- or low-runoff at the specified neighborhood distance parameter; and

$S(G_E)$ = the standard deviation of the expected G.

If the observed G is larger than the expected G, the Z score is positive, indicating high-runoff cells are clustered. The expected G (G_E) at a specified neighborhood distance is:

$$G_E = \frac{W}{n(n-1)} \tag{5}$$

where:

W = the sum of the weights for all cell pairs; and

n = the number of runoff cells in the watershed.

The graph of the global G statistic Z score against cell size (Figure 5b) shows that the degree of high-runoff clustering decreases with increasing cell size. Except for the 1920-m cell size, high-runoff patches in the watershed have an overall clustering pattern with statistical significance (a Z score of 2.33 indicates a 98 percent confidence interval; a Z score of 1.96 indicates a 95 percent confidence interval; and a Z score of 1.65 indicates a 90 percent confidence interval). At the 1,920-m cell size, the Z score becomes negative, indicating more low-runoff patches are clustered. The implication for watershed scientists is that they can use runoff data at a large cell size of up to 960 m to derive reliable results regarding whether high- or low-runoff patches are more clustered across the watershed.

Analyzing Neighborhood Distance Effect on Global Clustering Statistics

The neighborhood distance parameter specified in calculating the global clustering statistics will affect the results of the statistics and the interpretation of the clustering level (Getis and Ord 1992). To determine the effect of neighborhood distance on clustering statistics, a comparison was conducted of the global Moran's index Z score and the global G statistic Z score at neighborhood distances ranging from 210 m to 9600 m, using data sets at four cell sizes (210 m, 240 m, 480 m, and 960 m). The global Moran's index Z score becomes smaller as neighborhood distances increase (Figure 6a). At a neighborhood distance of 9600 m, the clustering degree becomes relatively low for the 960-m cell size data set. The implication is that large neighborhood distances can make the actual clustering patterns appear to be of no statistical significance. Considering that it is more computationally efficient to calculate clustering statistics using a large neighborhood distance, a neighborhood distance equal to, or smaller than, 8640 m is preferable.

The global Moran's index Z score reaches a maximum at a neighborhood distance of 1920 m. This indicates that cells separated at a distance of 1920 m have values more similar to each other than they do at other distances, suggesting that large runoff patches in the watershed are of about the same size as a 1,920-m radius neighborhood (Figure 2).

The global G statistic Z score has a consistent decreasing trend with increasing neighborhood distance (Figure 6b), indicating that the degree of high-runoff clustering decreases with increasing neighborhood distance. At a neighborhood distance of 5760 m, the confidence level of high-runoff clustering drops below 90 percent (with a Z score of 1.5) for the 960-m cell size data set. Therefore, a neighborhood distance equal to, or smaller than, 4800 m is preferred to calculate the high-runoff clustering statistic.

Analyzing Location and Size of Runoff Cluster

While the global Moran's index Z scores confirm an overall clustering pattern of watershed runoff, areas with similar runoff and areas where similar and dissimilar runoffs are interspersed are of particular interest when examining the source of clustering. The local Moran's index was calculated for each runoff cell so as to locate spatial clusters of similar runoff values (Anselin 1995). The local Moran's index (I_i) for the ith cell is calculated as:

$$I_i = \frac{\left(X_i - \bar{X}\right) \cdot \sum_j W_{ij}\left(X_j - \bar{X}\right)}{S^2} \tag{6}$$

where:

Xi = the runoff value for the ith cell;

$\bar{X}$ = the mean runoff of the watershed;

Xj = the runoff value for the jth cell;

Wij = a weight parameter for the pair of cells i and j to represent proximity relations; and

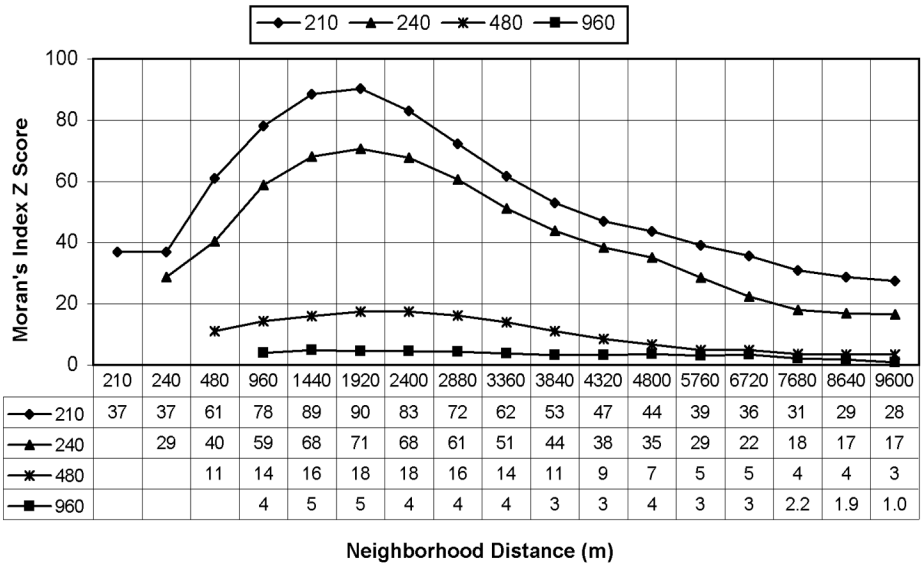
S = the standard deviation of runoff in the watershed.

A large positive value for I_i indicates that the cell is surrounded by cells with similar values. A negative value for I_i indicates that the cell is surrounded by cells with dissimilar values. A cutoff neighborhood distance of 1920 m, at which the overall clustering degree is maximized (Figure 6A), was used to calculate the local Moran's index. The Z score of the local Moran's index was computed to determine at which level the local clustering is significantly different from a random distribution. The Z score of the local Moran's index, $Z(I_i)$, is calculated as:

$$Z(I_i) = \frac{O(I_i) - E(I_i)}{S\left(E(I_i)\right)} \tag{7}$$

where:

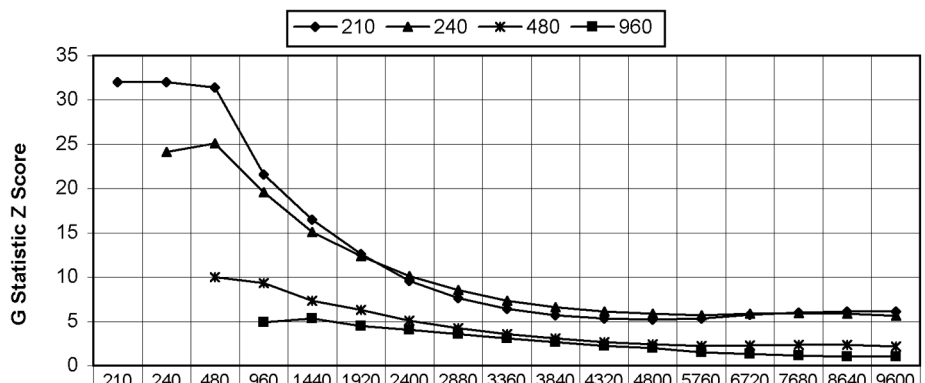
$O(I_i)$ = the observed I_i;



	210	240	480	960	1440	1920	2400	2880	3360	3840	4320	4800	5760	6720	7680	8640	9600
210	37	37	61	78	89	90	83	72	62	53	47	44	39	36	31	29	28
240		29	40	59	68	71	68	61	51	44	38	35	29	22	18	17	17
480			11	14	16	18	18	16	14	11	9	7	5	5	4	4	3
960				4	5	5	4	4	4	3	3	4	3	3	2.2	1.9	1.0

Neighborhood Distance (m)

(a)



	210	240	480	960	1440	1920	2400	2880	3360	3840	4320	4800	5760	6720	7680	8640	9600
210	32	32	31	22	16	13	10	8	6	6	5	5	5	6	6	6	6
240		24	25	20	15	12	10	9	7	7	6	6	6	6	6	6	6
480			10	9	7	6	5	4	4	3	3	2.4	2.2	2.3	2.4	2.4	2.2
960				5	5	4	4	4	3	3	2.2	2.0	1.5	1.4	1.2	1.0	1.0

Neighborhood Distance (m)

(b)

Figure 6. Global Moran's index Z scores (a) and global G statistic Z scores (b) at seventeen neighborhood-distance parameters using datasets at four cell sizes.

$E(I_i)$ = the expected I_i for a random distribution; and

$S(E(I_i))$ = the standard deviation of the expected I_i.

The expected I_i for a random distribution is calculated as:

$$E(I_i) = \frac{-\sum_j W_{ij}}{n-1} \tag{8}$$

where:

W_{ij} = the weight parameter for the pair of cells i and j to represent proximity relations; and

n = the total number of cell pairs in the watershed.

A high positive Z score for a cell indicates that the neighboring cells have similar runoff values. A very negative Z score for a cell indicates that the cell is surrounded by cells of dissimilar runoff values. The Z scores of the I_i values for the 60-m, 120-m, 210-m, 240-m, 480-m, and 960-m cell sizes were mapped to show where statistically significant clusters of similar runoff values are (Figure 7). It is apparent that statistically significant cluster patches correspond with runoff patches in the runoff maps (Figure 2). With increasing cell size, the locations of statistically significant clusters remain approximately the same, while their sizes decrease. The shrinking of statistically

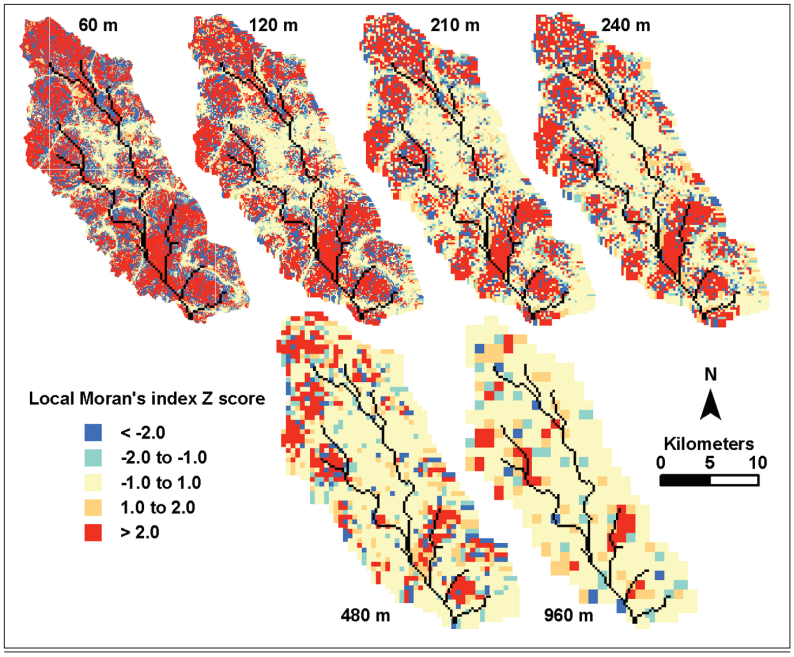


Figure 7. Runoff clusters, i.e., concentrations of similar runoff values, at different cell sizes. The red areas are statistically significant runoff clusters.

significant clusters is consistent with the decreasing of the overall clustering degree (Figure 5A). As cell sizes become larger, statistically significant clusters become less definable and small clusters disappear, particularly at a cell size beyond 480 m, which is regarded as the limit of cell size for mapping runoff clusters.

Spatial clusters of similar runoff values may be from high- or low-runoff clustering. High-runoff clusters are of particular interest to researchers due to their implication for erosion-induced pollution. We computed the local G statistic for each cell in order to locate statistically significant high-runoff clusters (Ord and Getis 1995). The local G statistic (G_i) for the ith cell is calculated as:

$$G_i = \frac{\sum_j W_{ij} X_j}{\sum_j X_j} \tag{9}$$

where:

X_j = the runoff value for the jth cell; and
W_{ij} = a weight parameter for the pair of cells i and j to represent proximity relations.

A cutoff neighborhood distance of 1920 m (at which the overall clustering level is maximized) was used to compute the statistic. A group of cells with high G_i indicates a concentration of cells with high-runoff values. Conversely, a group of cells with low G_i indicates a concentration of cells with low-runoff values. A G_i near 0 indicates there is no concentration of either high or low runoff surrounding the target cell. This occurs when the surrounding cell values are near the mean, or when the target cell is surrounded by a mix of high- and low-runoff values. The Z score of the G_i was computed (in the same way as for the local Moran's index) to determine the level at which the local clustering is significantly different from a random distribution. As with the G_i value itself, a high Z score for a cell indicates its neighbors have high runoff values, and *vice versa*; a Z score near 0 indicates no apparent concentration of similar values. We mapped the Z scores of the G_i for the 60-m, 120-m, 210-m, 240-m, 480-m, and 960-m cell sizes to show where statistically significant high- and low-runoff clusters (termed hot spots and cold spots, respectively) are (Figure 8). Observe that hot spots correspond to high-runoff patches and cold spots correspond to low-runoff patches in the original runoff map (Figure 2). As the sizes of hot and cold spots decrease with increasing cell size, small hot and cold spots disappear, particularly at a cell size beyond 480 m. This limit of cell size for mapping hot and cold spots corresponds to that for mapping runoff clusters (Figure 7).

Analyzing Neighborhood Distance Effect on Hot Spot Patterns

The neighborhood distance parameter in mapping hot spots may have an effect on the size and location of hot spots (Getis and Aldstadt 2004) and the subsequent interpretation of potentially high-erosion areas in the watershed. To study the neighborhood distance effect on hot spot patterns and to investigate optimal neighborhood distances for identifying hot spots, hot-spot maps were generated at neighborhood distances of 1440 m, 960 m, and 480 m, using data sets at five cell sizes (Figure 9). These hot-spot maps and previous hot-spot maps at a neighborhood distance of 1920 m (Figure 8) show that the smaller the neighborhood distance, the more fragmented the hot spots are. The fragmentation at a neighborhood distance of 480 m makes

the cluster pattern unrealistic (when referenced with the original runoff map) and undesirable; therefore a neighborhood distance of 960 m or larger is preferred for hot spot analysis. The upper limit of neighborhood distance parameter can be determined using the previous analysis of neighborhood distance effect on high-runoff clustering degree (Figure 6b). It was concluded that a neighborhood distance of 4800 m or smaller is needed to have a reliable estimate of whether high runoff is clustered with the statistical significance across the watershed.

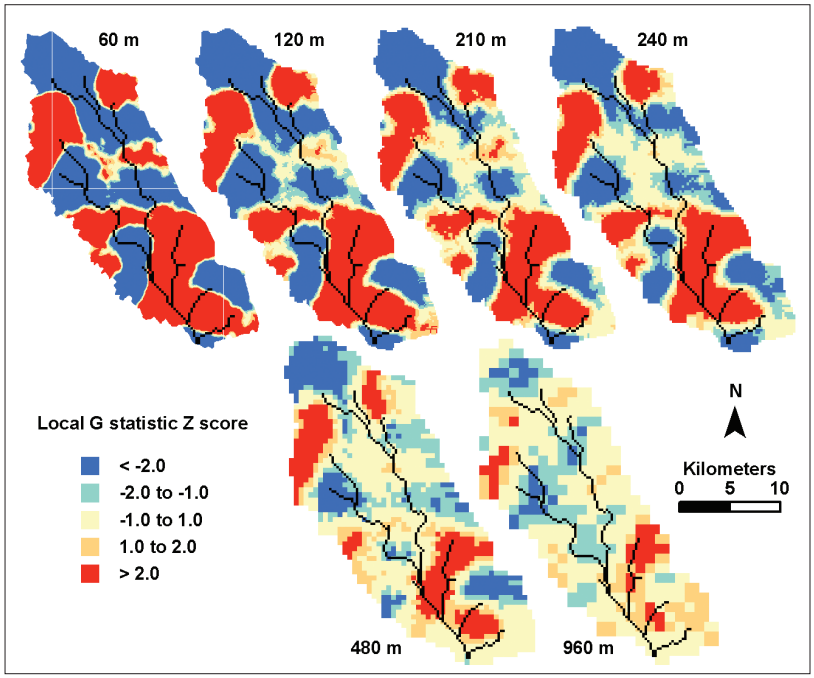


Figure 8. Runoff hot spots (red) and cold (blue) spots at different cell sizes (at a neighborhood distance of 1920 m). The red hot spots are statistically significant high-runoff clusters; the blue cold spots are statistically significant low-runoff clusters.

Analyzing Runoff Statistics at Watershed Level vs. Local Level

Previous analyses of cell size effects on runoff statistics are based on the assumption that these effects are systematic throughout the watershed. To investigate local variability of cell size effects, the Little River Watershed was partitioned into 30 areas and their individual statistics were examined (Figure 10). For the purpose of this analysis, the partition of watershed areas was based on visually identified, relatively homogeneous runoff patches. Average overland runoff, the total runoff volume, and the overall clustering statistics for individual watershed areas at different cell sizes were calculated. The average runoff for individual areas does not have consistent trends with increasing cell size (Figure 11), which corresponds with the constant runoff range with changing cell size for the entire watershed area (Figure 2). We then investigated whether the total runoff volume within a local watershed area has a consistent trend with changing cell size. The change ratios of runoff volume between successive cell sizes were mapped for the thirty watershed areas (Figure 12). If an area stays blue throughout successive cell sizes, the area has a consistent decreasing runoff volume. In contrast, if an area stays red throughout successive cell sizes, the area has a consistent increasing runoff volume. The change ratio maps show that no single area has a consistent runoff change, while three areas have a consistent decreasing runoff from 210-m to 1,920-m cell sizes. The runoff volume change for local areas, then, is not as consistent and significant as that for the entire watershed area. The implication is that the optimal cell sizes (of 210 m or smaller) for estimating the total runoff volume for the entire watershed area may not be applicable to local watershed areas.

Next, spatial statistics at the local level were compared with those at the watershed level. The global Moran's index and the global G statistic were calculated for individual watershed areas at different cell sizes (Figure 13). The global Moran's index Z scores for individual watershed areas were all positive (indicating clustering of similar runoff values) and all decreased as cell size increased (Figure 13a), which corresponds with the trends for the entire watershed area (Figure 5a). As for the global G statistic, most watershed areas had positive G statistic Z scores (indicating high-runoff clustering) and a few had negative G statistic Z scores (indicating low-runoff clustering) (Figure 13b), which corresponds with the overall positive Z scores at the watershed level (Figure 5b).

Regardless of whether a local area had positive or negative G statistic Z scores, the statistical significance of clustering decreased as cell size

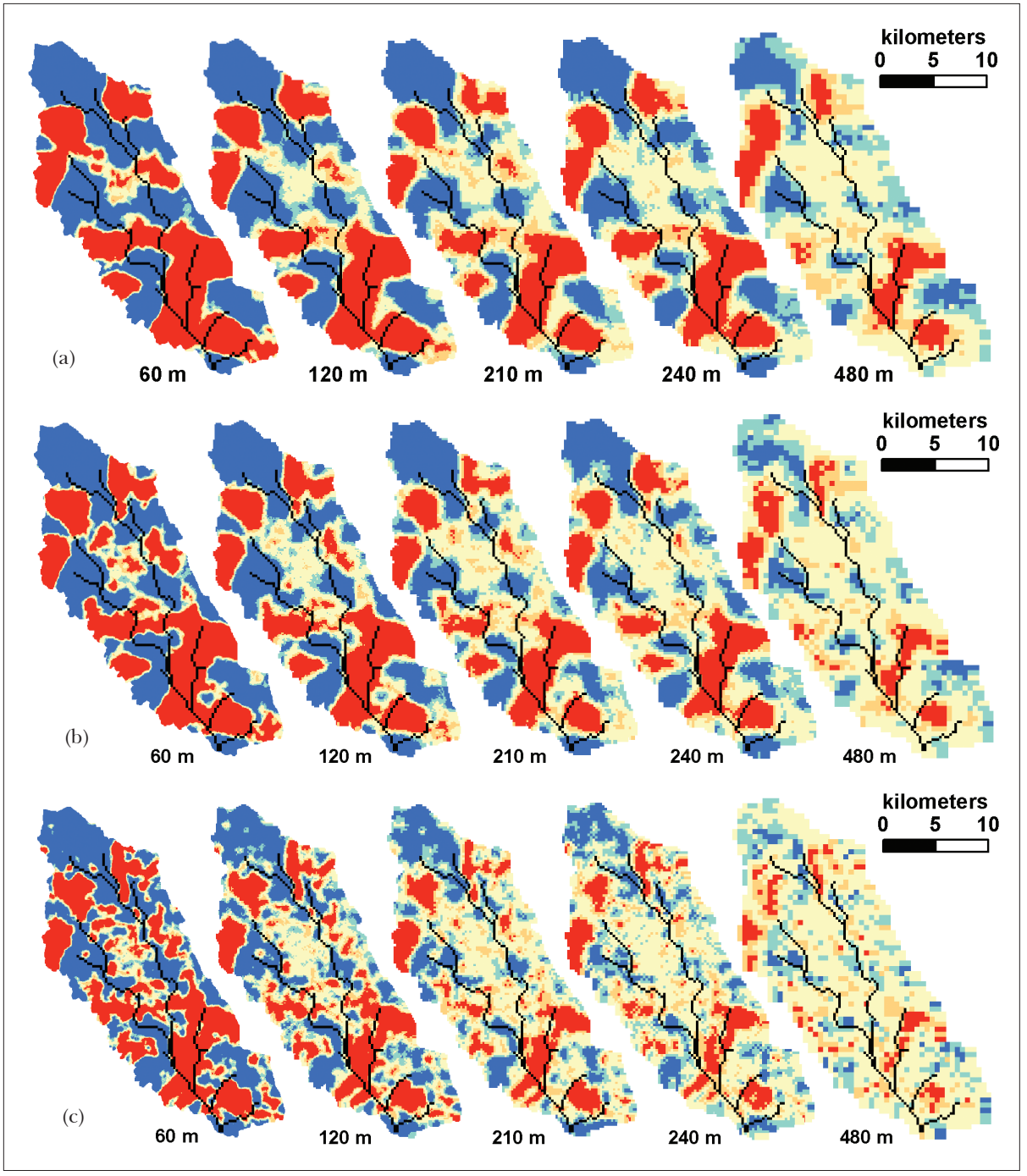


Figure 9. Runoff hot spots and cold spots at five different cell sizes at a neighborhood distance of 1440 m (a), 960 m (b), and 480 m (c).

increased, which corresponds with the trend for the entire watershed area. Because the effects of cell size on spatial statistics are consistent throughout the watershed, the optimal cell sizes for calculating spatial statistics (960 m or smaller for global clustering statistics and 480 m or smaller for local clustering statistics by cells) for the entire watershed area are considered applicable to local watershed areas.

The uniform clustering statistics of different local watershed areas at large cell sizes (Figure 13) indicate that natural runoff variability and extremes are lost as cell size increases. In the context of natural resource modeling, the extremes may convey important information about the watershed that watershed scientists can use to study local factors of hydrological processes. For that reason, watershed runoff modeling using

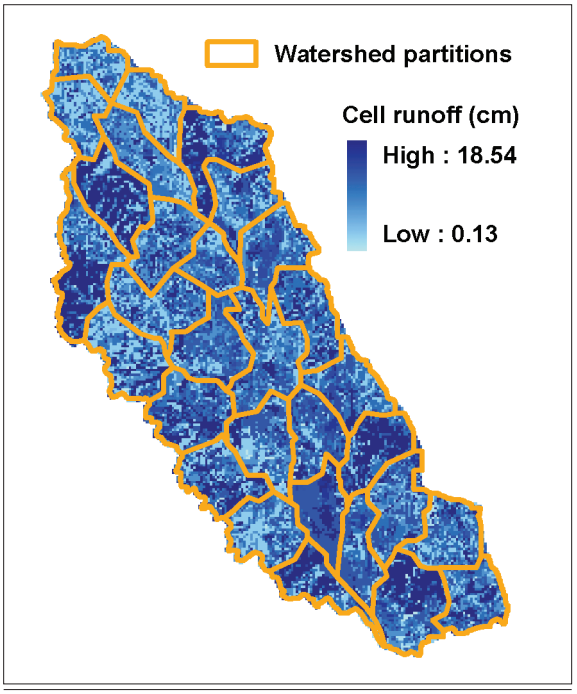


Figure 10. Thirty partitioned watershed areas.

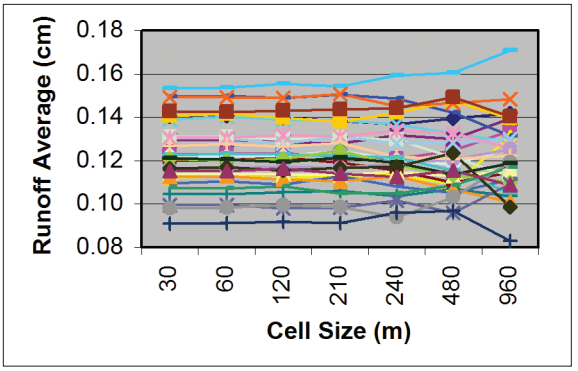


Figure 11. Runoff averages for the thirty watershed areas at different cell sizes.

data sets of cell sizes smaller than or equal to 210 m is preferred.

regarding appropriate cell sizes for calculating runoff volumes and spatial statistics. The eight cell sizes are not an exhaustive continuity and, therefore, the suggestions of optimal cell sizes are relatively general. Future studies may test additional, intermediate cell sizes to draw conclusions that are more specific.

In this study, 17 successive neighborhood distances were tested to calculate the overall clustering statistics and four neighborhood distances to derive hot-spot maps; recommendations regarding

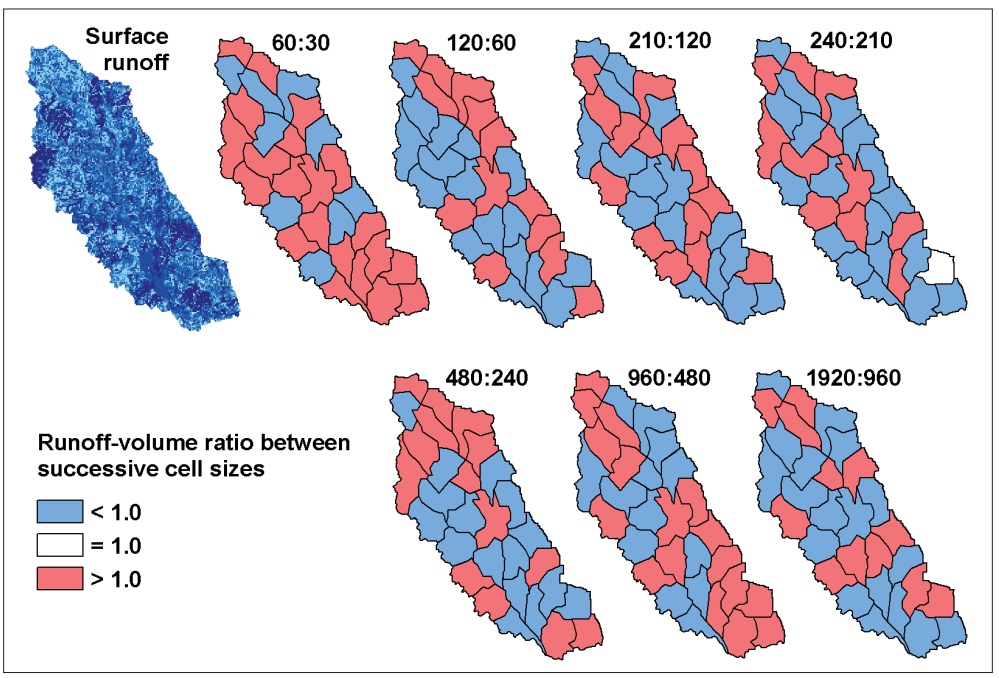


Figure 12. Runoff-volume ratios between successive cell sizes for 30 watershed areas referenced with a surface runoff map.

Discussion

Previously modeled runoff data sets at eight successive cell sizes were used to draw conclusions appropriate neighborhood distances were subsequently provided. The conclusions of optimal neighborhood distances for calculating the overall clustering statistics are more specific compared to

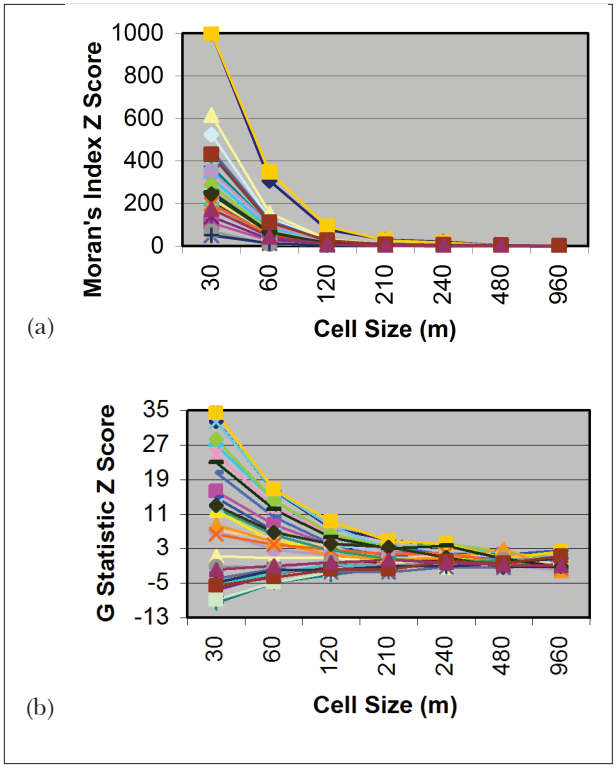


Figure 13. Global Moran's index Z scores (A) and global G statistic Z scores (B) for the 30 watershed areas at different cell sizes.

those for deriving hot-spot maps. Future studies may test additional, intermediate neighborhood distances in generating hot-spot maps to refine the neighborhood distance suggestions.

The trends of watershed statistics with changing cell size may be similar to those obtained in other watershed studies using the same approach, but conclusions regarding appropriate cell sizes and neighborhood-distance parameter in calculating spatial statistics are likely to be different, depending on the spatial scales of individual watersheds.

Results and Conclusions

This study examines the effects of cell size on spatial properties and statistics of simulated watershed runoff. Specifically, the change of overland cell runoff, total cell runoff volumes, overall runoff clustering degree, and runoff cluster patterns were examined. The trends of these runoff properties and statistics are observed in maps and graphs, and appropriate cell sizes for the analyses are suggested. In addition, the appropriate neighborhood-distance parameter for calculating spatial statistics was evaluated, and the cell size effects for the entire watershed area versus the local watershed areas were compared.

The results show that the range of overland runoff within the watershed does not change with cell size, while the total runoff volume of the watershed has an overall decreasing trend. Researchers can estimate the total runoff volume from the modeling results within 2.5 percent accuracy using data set at a cell size of 1920 m or smaller, and within 0.2 percent accuracy using data set at a cell size of 210 m or smaller.

The overall clustering degrees (indicated by the Z scores of the global Moran's index and the global G statistic) have a decreasing trend with increasing cell size. The appropriate cell size for calculating both clustering statistics is 960 m or smaller. The upper limitation of the neighborhood-distance parameter for calculating the clustering statistics is 8640 m for the global Moran's index and 4800 m for the global G statistic.

Runoff clusters become smaller at larger cell size, while the locations of runoff clusters do not change. Runoff clusters can be barely defined at a cell size beyond 480 m, which is therefore regarded as the limit of cell size for mapping runoff clusters.

Hot spots (high-runoff clusters) become more fragmented at smaller neighborhood-distance parameters. A neighborhood distance equal to or larger than 960 m is suggested to generate hot spot maps, while the upper limit of neighborhood distance is determined as 4800 m based on the analysis of the high-runoff clustering degree for the watershed.

Runoff properties and statistics at the local level were studied by partitioning the watershed into thirty areas. Analysis results show that the runoff volume change at local areas is not consistent/significant as that for the entire watershed area. On the other hand, the decreased clustering with increasing cell size is comprehensive at every local watershed area. As local variability of spatial statistics is reduced with increasing cell size, a cell size smaller than or equal to 210 m is suggested for retaining information of extreme conditions within local watershed areas.

The results of this study have implications for how watershed models are applied and results are interpreted. While the results should be verified across other regions, the results presented here establish several guidelines for watershed modeling, specifically in the Coastal Plain area. Future research may study other watersheds of different sizes and characteristics and compare the results. Other watershed modeling outputs, such as sediment yield and nutrient loadings of phosphorous

and nitrogen, may also be studied (using the same approach) and the results be compared.

REFERENCES

Anselin, L. 1995. Local indicators of spatial association-LISA. *Geographical Analysis* 27(2): 93-115.

Bhuyan, S.J., K.R. Mankin, J.K. Koelliker, L. Marzen, and J.A. Harrington. 2001. Effect of cell size on AGNPS predictions. Paper presented at ASAE Annual International Meeting (Paper Number: 01-2002), American Association of Agricultural Engineering, Sacramento, California, 30 July-1 August.

Bosch, D.D., J.M. Sheridan, R.R. Lowrance, R.K. Hubbard, T.C. Strickland, G.W. Feyereisen, and D.G. Sullivan. 2007. Little River Experimental Watershed Database. *Water Resources Research* 43(9): W09470.

Bruneau, P., C. Gascuel-Odoux, P. Robin, P. Merot, and K. Beven. 1995. Sensitivity to space and time resolution of a hydrological model using digital elevation data. *Hydrological Processes* 9: 69-81.

Canfield, H.E., and D.C. Goodrich. 2006. The impact of parameter lumping and geometric simplification in modelling runoff and erosion in the shrublands of southeast Arizona. *Hydrological Processes* 20: 17-35.

Davenport, I.J., M. Silgram, J.S. Robinson, A. Lamb, J.J. Settle, and A. Willig. 2003. The use of earth observation techniques to improve catchment-scale pollution predictions. *Physics and Chemistry of The Earth* 28(33-36): 1365-76.

ESRI. 2005. *The ESRI guide to GIS analysis*, vol. 2: *Spatial measurements and statistics*. Redlands, California: ESRI Press.

Getis, A. 1991. Spatial interaction and spatial autocorrelation: A cross-product approach. *Environment and Planning A* 23(9): 1269-77.

Getis, A., and J. Aldstadt. 2004. Constructing the spatial weights matrix using a local statistic. *Geographical Analysis* 36(2): 90-104.

Getis, A., and J.K. Ord. 1992. The analysis of spatial autocorrelation by use of distance statistics. *Geographical Analysis* 24(3): 189-206.

Ord, J.K., and A. Getis. 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis* 27(4): 287-306.

Panuska, J.C., I.D. Moore, and L.A. Kramer. 1991. Terrain analysis—integration into the agricultural nonpoint source (AGNPS) pollution model. *Journal of Soil and Water Conservation* 46(1): 59-64.

Thieken, A.H., A. Lucke, B. Diekkruger, and O. Richter. 1999. Scaling input data by GIS for hydrological modeling. *Hydrological Processes* 13: 611-30.

USDA. 2006. Digital Soil Geographic Databases. [http://soils.usda.gov/survey/geography.]

USGS. 2006. National Elevation Dataset. [http://ned.usgs.gov.]

USGS. 2007. National Land Cover Database 2001 (NLCD 2001). [http://www.mrlc.gov/mrlc2k_nlcd.asp.]

Usery, E., M. Finn, D. Scheidt, S. Ruhl, T. Beard, and M. Bearden. 2004. Geospatial data resampling and resolution effects on watershed modeling: A case study using the agricultural non-point source pollution model. *Journal of Geographical Systems* 6: 289-306.

Vieux, B.E., and S. Needham. 1993. Nonpoint-pollution model sensitivity to grid-cell size. *Journal of Water Resources Planning and Management* 119(2): 141-57.

Wolock, D.M., and C.V. Price. 1994. Effects of digital elevation model map scale and data resolution on a topography-based watershed model. *Water Resources Research* 30 (11): 3041-3052.

Wong, D., and J. Lee. 2005. *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*. Wiley. 464p.

Young, R.A., C.A. Onstad, D.D. Bosch, and W.P. Anderson. 1989. AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds. *Journal of Soil and Water Conservation* 44(2): 168-73.

Young, R.A., C.A. Onstad, D.D. Bosch, and W.P. Anderson. 1994. *Agricultural non-point source pollution model, version 4.03, AGNPS user's guide*. North Central Soil Conservation Research Laboratory, Morris, Minnesota, USA.

Young, R.A., C.A. Onstad, D.D. Bosch, and W.P. Anderson. 1995. AGNPS: An agricultural nonpoint source model. In: Singh, V.P. (ed.), *Computer models of watershed hydrology*. Highlands Ranch, Colorado: Water Resources Publications. pp. 1010-20.

Zhang, W., and D.R. Montgomery. 1994. Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resources Research* 30(4): 1019-28.