

Article

An Efficient Vector-Raster Overlay Algorithm for High-Accuracy and High-Efficiency Surface Area Calculations of Irregularly Shaped Land Use Patches

Peng Xie ¹, Yaolin Liu ^{1,2,3,*}, Qingsong He ^{1,*}, Xiang Zhao ¹ and Jun Yang ⁴

¹ School of Resource and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; 2014102050008@whu.edu.cn (P.X.); zhaoxiang@whu.edu.cn (X.Z.)

² Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

³ Collaborative Innovation Center for Geospatial Information Technology, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

⁴ College of Urban and Environmental Science, Liaoning Normal University, Dalian 116021, China; yangjun@lnnu.edu.cn

* Correspondence: liuyaolin@whu.edu.cn (Y.L.); qingsong@ad.unc.edu (Q.H.); Tel.: +86-159-2762-5945 (Q.H.)

Academic Editor: Wolfgang Kainz

Received: 22 March 2017; Accepted: 24 May 2017; Published: 27 May 2017

Abstract: The Earth's surface is uneven, and conventional area calculation methods are based on the assumption that the projection plane area can be obtained without considering the actual undulation of the Earth's surface and by simplifying the Earth's shape to be a standard ellipsoid. However, the true surface area is important for investigating and evaluating land resources. In this study, the authors propose a new method based on an efficient vector-raster overlay algorithm (VROA-based method) to calculate the surface areas of irregularly shaped land use patches. In this method, a surface area raster file is first generated based on the raster-based digital elevation model (raster-based DEM). Then, a vector-raster overlay algorithm (VROA) is used that considers the precise clipping of raster cells using the vector polygon boundary. Xiantao City, Luotian County, and the Shennongjia Forestry District, which are representative of a plain landform, a hilly topography, and a mountain landscape, respectively, are selected to calculate the surface area. Compared with a traditional method based on triangulated irregular networks (TIN-based method), our method significantly reduces the processing time. In addition, our method effectively improves the accuracy compared with another traditional method based on raster-based DEM (raster-based method). Therefore, the method satisfies the requirements of large-scale engineering applications.

Keywords: surface area; landforms; vector-raster overlay; edge-effect; DEM; pixel boundary

1. Introduction

The surface area refers to the actual area of the Earth's surface and considers topographical conditions. In the field of geographic information systems (GIS), the Earth is usually simplified as a standard ellipsoid during modeling, and its actual topography is not considered. In resource surveys, an area calculation based on an ellipsoidal area rather than the actual surface area is often used to evaluate biological resources, the scale of land resources, and the value of ecological services [1–5]. In bio-geographical studies, the biodiversity is significantly increased by heterogeneity and the separation of terrain [6,7]. However, these methods do not distinguish whether humans or a variety of other creatures in nature live within any particular topographic relief surface [6,8]. Given the high fluctuation of the Earth's surface, its true surface area will be quite different from an ideal ellipsoid in

mountainous and hilly areas. Therefore, the use of ellipsoidal areas alone does not satisfy the needs of actual resource survey applications, such as land and forestry surveys [9].

With the development of GIS, several algorithms have been developed based on digital elevation models (DEM) to calculate the surface area of terrains [9–13]. A DEM is a generic model of elevations, which can be expressed either as triangulated irregular networks (TINs) or by raster-based DEMs. Therefore, surface area calculation methods can be divided into two major types: TIN-based methods and raster-based methods.

TINs express continuous terrain surfaces using 3-D elevation vertices and edge-connecting elevation vertices. Since TIN data is stored in a vector format, it can be used for accurate measurements and cutting. Therefore, TIN-based surface area calculations are generally considered to have a high accuracy [9,14]. However, due to the different parameters used in the generation process (such as the number of vertices and the elevation threshold), different TIN datasets will be generated. The uncertainty of this conversion is particularly evident during generation using a raster-based DEM, which is the most popular elevation data format. TINs usually need less storage space than raster-based DEMs and have a lower computation time for the surface area calculation. However, as high resolution elevation data is becoming more popular, the number of vertices in TIN data is increasing, which significantly increases the storage space and the computation time of the surface area [9,12].

Raster-based surface area calculation methods are based on raster-based DEM data, using a series of regular elevation points to represent the terrain surface. Raster-based DEM data is usually stored in a raster data format. Raster-based surface area calculation methods can be divided into the following four types. (1) The surface area is derived from the slope and aspect based on raster-based DEM; however, this method is known to be less accurate. [10,11]. (2) The area of the curved surface composed of the four adjacent cells is calculated by employed an integral formula using the four neighborhood cells of the raster-based DEM. The areas of the curved surfaces that fall into a polygon based on the relationships between the spatial locations are then summed. A commonly used numerical integration method is the parabolic quadrature method [15]. (3) A 3×3 cell window is created and moved across the DEM. Eight spatial triangles are generated in each window, and the surface area of the central pixel of each 3×3 cell window is calculated [9]. (4) The triangular regular network (TRN) is directly constructed based on the raster-based DEM. The TRN is overlaid with a vector polygon, and the surface area of the polygon is obtained by summing the area of the triangles that fall into the polygon based on the relationship between the spatial locations [16]. The first three methods are used to generate the surface area raster. For the calculation of the vector polygon (land use patch) surface area, a vector-raster conversion is required. The accuracy loss caused by the edge-effect problem cannot be ignored; that is whether or not the center point of a cell is within the polygon to decide whether the cell value belongs to the polygon [9]. The fourth method can handle the edge-effect problem effectively, but requires a complex spatial intersection. Therefore, it has a low calculation efficiency when the land use polygons are small and the number of land use polygons is particularly large.

China has a vast territory and an area of approximately 9.6 million km². It is a mountainous country and its mountainous areas account for more than two-thirds of the country's total area [12]. The State Council of the First Geographical National Census Leading Group Office stipulates that the county-level land surface area statistics for each land-use type are a key statistical indicator for the basic statistics of national conditions. For a national census with high resolution data computation tasks, the above methods are not sufficient. Therefore, it is a significant challenge for researchers to find a method to quickly and accurately calculate the surface area of irregularly shaped polygons, such as land use patches. In this study, the authors propose and experiment with an efficient vector-raster overlay algorithm (VROA) to solve this problem. The VROA-based method is a straightforward method for calculating the surface area of an irregularly shaped polygon. The study mainly aims to solve the following two problems: (1) calculation of the surface area of irregular polygons with raster-based DEM data and (2) delivery of a method with a high efficiency to meet the implementation of computing tasks of high resolution data within a national scope.

The structure of this paper is as follows. Section 2 introduces the study areas, data, and the following two steps of the VROA-based method: the generation of the surface area raster file with a high efficiency and high precision based on raster-based DEM and the calculation of the surface area of a polygon with a high efficiency when executing raster and vector overlay statistics, considering the vector polygon boundary. A comparative experiment is performed in Section 3. The advantages and disadvantages of VROA-based methods are discussed in Section 4. The conclusions are given in Section 5.

2. Data and Methods

2.1. Study Area and Data

Xiantao City (Figure 1a) is located in the south central area of Hubei Province and is the central city in the Jiangnan Plain. The topography of Xiantao is flat with small fluctuations, and it has an area of 2538 km².

Luotian County (Figure 1b) is located in the northeast of Hubei Province. The Dabie Mountains in the south lie at 30°35'N~31°16'N latitude and 115°06'E~115°46'E longitude. There are many hills in Luotian, but most of these hills are below 200 m a.s.l. The total area of Luotian County is 2145 km².

Shennongjia Forest District (Figure 1c) is the only designated “forestry district” in China. This area is located in the west of Hubei Province and is one of the Global Geoparks. The territory in Shennongjia is mostly mountainous. Its highest elevation of 3105.4 m is the highest point in central China and shows typical mountain landforms. The total area of Shennongjia Forest District is 3253 km².

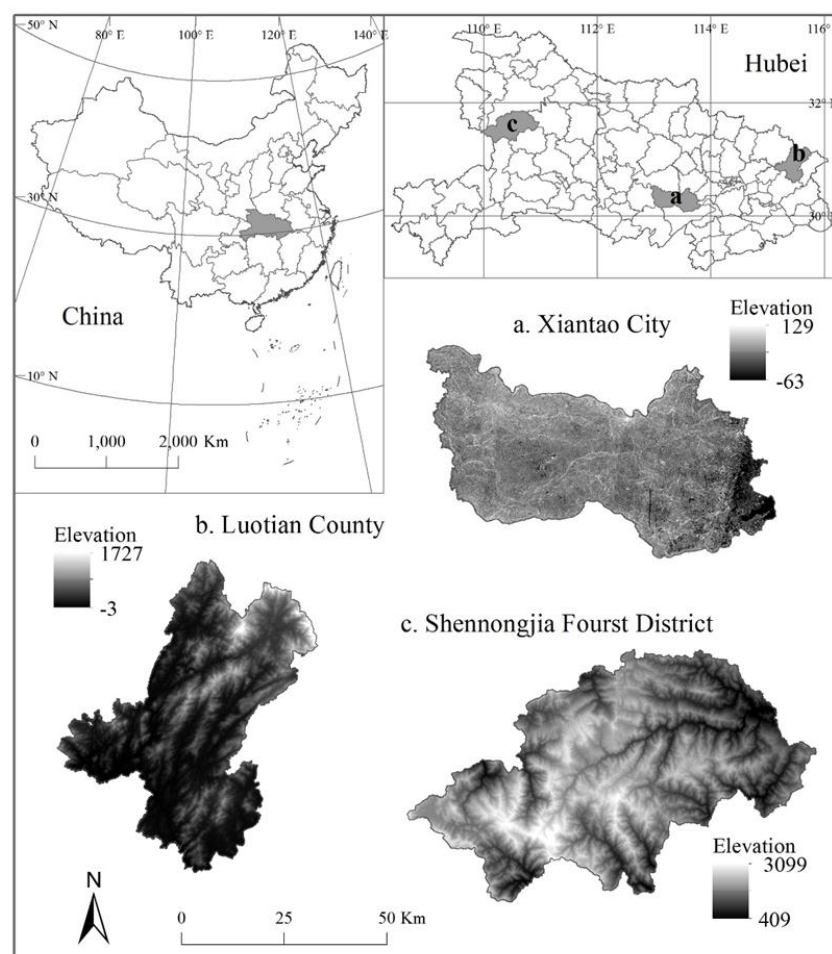


Figure 1. Study area.

The terrain data was represented by a raster-based DEM with a resolution of approximately 30 m and provided by the Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences. Land use was represented by a vector geodatabase generated from the second national land use investigation in 2009. The land use data were provided by the Wuhan Land Resource and Planning Bureau.

2.2. Methods

2.2.1. Generation of Raster File of Surface Area Based on the Raster-Based DEM

(1) Coordinate Transformation

A specific coordinate system is usually required to accurately describe the location of a point on the Earth's surface in three-dimensional (3-D) space. The most commonly used coordinate systems can be divided into the following two types: (1) the geodetic coordinate systems and (2) the Cartesian coordinate systems [17]. In a geodetic coordinate system, the geometric position of a point on the Earth's surface is defined by geodetic longitude (L), geodetic latitude (B), and geodetic height (H). A Cartesian coordinate system uses axes perpendicular to each other and the coordinates are defined as distances from the origin along the X , Y , and Z axes. For the calculation of a triangle area, it is necessary to transform the coordinates from a geodetic coordinate system to a Cartesian coordinate system. Any points (B , L , and H) in a geodetic coordinate system can be converted into a point (X , Y , and Z) in a Cartesian coordinate system using the following formulas:

$$X = (N + H)\cos(B)\cos(L) \quad (1)$$

$$Y = (N + H)\cos(B)\sin(L) \quad (2)$$

$$Z = \left[N(1 - e^2) + H \right] \sin(B) \quad (3)$$

In the above formulas B , L , and H represent the latitude, longitude, and elevation, respectively; X , Y , and Z denote the coordinate values in the Cartesian coordinate system; and e is the first oblateness of the Earth's ellipsoid. N can be calculated using Formula (4):

$$N = \frac{a}{\sqrt{1 - e^2 \sin^2 B}} \quad (4)$$

where a is the semi-major axle. a and e are both known constants that can be obtained from the definition of the Earth's reference ellipsoid.

(2) Process of Creating the Raster File of the Surface Area

Raster-based DEMs typically use raster images to represent the elevation of the Earth's surface. The Earth's surface is simulated using discrete and regular cells [9]. However, the surface of the Earth is a continuous surface, which cannot be obtained on the basis of a raster-based DEM. Therefore, the calculation of the surface area is usually based on using discrete cell center points in raster-based DEMs to construct the triangles that are connected to each other. In a Cartesian coordinate system, any three points can be guaranteed to be in the same plane, whereas this cannot be guaranteed for more than three points. In addition, the area cannot be calculated. The actual continuous surface of the Earth can be approximated by a triangular network consisting of interconnected triangles. Given vertexes A , B , and C of a triangle and their coordinates (X_A , Y_A , Z_A), (X_B , Y_B , Z_B), and (X_C , Y_C , Z_C), respectively, the area S of this triangle can be calculated using Heron's formula:

$$S = \sqrt{p(p-a)(p-b)(p-c)} \quad (5)$$

In Heron's formula, $p = (a + b + c)/2$ and a , b , and c represent the length of the triangle's three sides. The side length l_{ij} of the spatial triangle can be calculated using Formula (6):

$$l_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (6)$$

where l_{ij} represents the distance between point i and j in 3-D space, and (x_i, y_i, z_i) and (x_j, y_j, z_j) are the 3-D coordinates of the two points, respectively.

Using the above theoretical basis, the raster file of the surface area can be generated through the following steps.

M denotes the number of rows of a DEM raster, N denotes the number of columns, and C denotes the resolution of the DEM raster. An empty raster file of $M-1$ rows and $N-1$ columns is first created, and all cells in the raster file are set to a value of 0 by default. The raster values are a double precision type and the new raster file has the same coordinate system as the original DEM raster. If the starting point of the latitude and longitude coordinates of the original DEM is (B_1, L_1) , the new raster file starting point is (B', L') .

$$B' = B_1 + C \cdot 0.5 \quad (7)$$

$$L' = L_1 + C \cdot 0.5 \quad (8)$$

The DEM raster is then iterated row by row, and a TRN is built to calculate the surface area of the cell in the surface area raster using the following steps.

- i. Let i be the current traversal row number, let j be the column number, and let the initial values of i and j be 2. Then, iterate the DEM raster from the second row and the second column. Label the current location of the cell as (i, j) . Obtain the raster cells of locations $(i-1, j-1)$, $(i, j-1)$, and $(i-1, j)$. Obtain their center points and label these points as 1, 2, 3, and 4 (Figure 2).

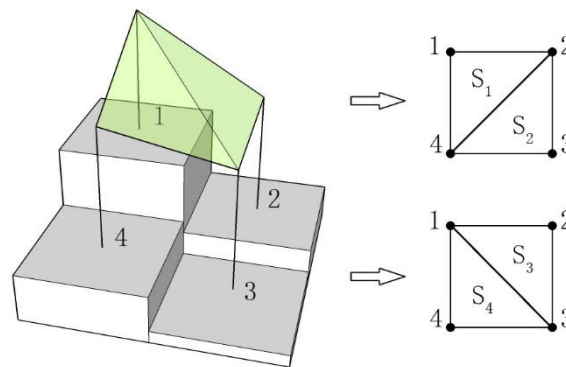


Figure 2. Schematic diagram of a surface area cell.

- ii. Construct four triangles S_1 , S_2 , S_3 , and S_4 by connecting the diagonal points 1–3 and 2–4.
- iii. Calculate the area of the four triangles as $(S_1, S_2, S_3, \text{ and } S_4)$ using Heron's formula (Formula (5)).
- iv. The surface area of zones 1-2-3-4 can be expressed by $(S_1 + S_2 + S_3 + S_4)/2$. The surface area can be written to its corresponding location, which can be identified by the row and column number of the surface area raster file. Repeat steps (1) to (4) until all rows and columns have been traversed. Each value assigned to the cells of the new raster represents the surface area of the particular cell.

2.2.2. Vector-Raster Overlay Statistics

For the raster and vector overlay statistics, when the center of the cell falls within a polygon, the cell is considered to belong to that polygon; otherwise, it is not part of that polygon. Using the center point of a cell in raster-based spatial analyses, rather than the entire area of the cell, simplifies the calculation to improve the computational efficiency. This method is the statistical method used for vector-raster overlay in ArcGIS, which is the most commonly used GIS software around the world. As shown in Figure 3, each square represents a cell and the dots represent the center point of that cell. The area of the polygon in Figure 3a is the total sum of the area of the gray cells in Figure 3b. This value does not match the real area of the polygon; however, this method of vector-raster overlay statistics is useful for learning the approximate area. The accuracy of this method is related to the spatial resolution of the DEM raster; the lower the DEM raster resolution is, the lower the statistical accuracy of the area. The results will also be significantly affected when the polygon patch is relatively small.

The surface area of a polygon can be accurately obtained by calculating the area of the intersection of all other polygons after raster-vectorization and intersection between polygons. The attribute table of each intersecting polygon contains the location information of the original raster cells. The intersecting polygon can then be calculated and assigned from the cell value (surface area) based on the ratio between the intersecting polygon area and the cell area. The assigned value of all the intersection polygons (the information relating to the polygon that is being calculated also exists in the intersection polygon) is summed to obtain the area of the polygon. However, this method was only used to calculate the area of Xiantao City. The projection plane area is 2538 km², covering 3266 rows and 1763 columns for a 30 m resolution raster-based DEM. The land use map contains 51,732 vector-polygon elements. These results are unsatisfactory because the data processing is time-consuming. It would take the raster-to-vector process 8 min and 19 s to achieve these results. In contrast, the intersection with the land use map takes 4 h, 35 min, and 23 s. Therefore, this method is unsuitable for engineering applications.

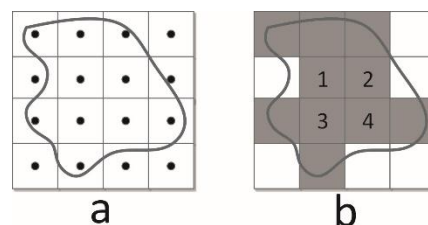


Figure 3. Schematic of vector-raster overlay statistics.

As shown in Figure 3, only raster cells 1, 2, 3, and 4 are completely covered by polygons, and the remaining pixels are partially covered. The GIS software frequently used to manage the partial coverage of raster and vector overlay statistics has only two possible options: either the cell value is completely counted in the polygon or else it is completely excluded. For a partially covered cell, it is difficult to accurately calculate the surface area of the intersecting section between the polygon and the cell directly. However, a viable option is to use the ratio of the intersecting polygon area and the cell area to decide how many intersecting polygons will be assigned from the raster values. This method ensures the highest level of accuracy of the area. Thus, at this point, the problem becomes how to quickly and accurately calculate the area of the intersecting polygons generated by the overlap between the polygon and the cell. The vector polygon is composed of arcs and each arc consists of two points. The vector polygon can be regarded as the connection between the points. The area of any vector polygon can be derived from the sum of the vector triangle areas, which is composed of two node points in sequence between the polygon boundary and any point. The origin point is usually chosen with coordinates (0, 0). The vector triangle area equals the cross-product of the two sides from the origin point [18]. If it is assumed that the point coordinates of each polygon are $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$ (Figure 4), then the area of the polygon S_p can be calculated using

Formula (9) [18]. It should be noted that the order of the node points will result in a positive or negative area value. The area value will be negative when the order of node points is clockwise, and it will be positive when the order of node points is counterclockwise.

$$Sp = \sum_{k=1}^{+\infty} S_{\Delta P_0 P_k P_{k+1}} = \frac{1}{2} \sum_{k=1}^{+\infty} (x_k y_{k+1} - x_{k+1} y_k) \quad (9)$$

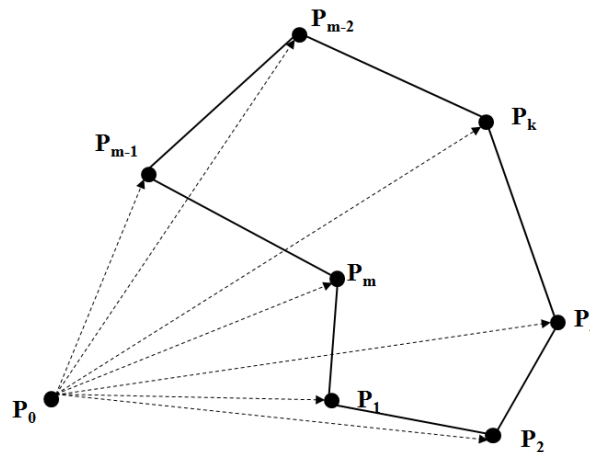


Figure 4. Schematic of the area vector method.

Computational efficiency is an important issue that should be addressed in vector-raster overlay statistics. As mentioned above, the statistical speed of the vectorized surface area raster and polygon intersection is extremely slow due to the high requirement for spatial operations. When there are too many cells or polygons, the time-consumption of the algorithm is not suitable for large-scale engineering applications. Therefore, an efficient way to improve the statistical speed is to reduce the number of spatial operations as much as possible. In this study, the VROA is presented, which considers the precise clipping of the raster cells by the vector polygon boundary. VROA builds a rectangle through the border extraction of the surface raster row by row, using a rectangle to cut the polygon, and then obtaining the intersecting area of the polygon-rectangle. The vector polygon is composed of points. The specific location of a point in the original surface area raster can be determined based on the vector polygon coordinates. A polygon is enclosed by all of the points in the vector and its plane area can be calculated using Formula (9). This area divides the plane area of the raster cell to obtain the ratio. This ratio can be regarded as the proportion of the intersection ellipsoid area to cell ellipsoid area. The diagram (Figure 5) and flowchart (Figure 6) for VROA are shown as follows, and the pseudocode is presented in the appendix section.

As described in the algorithm above, the key to VROA is determined by the intersections of the arc and the cell border. The coordinates of each cell border are extrapolated based on the two endpoints of the arc. Each intersecting point can be regarded as the end point of the previous intersecting polygon and the start point of the next intersecting polygon. If no intersection exists, then a closed polygon can be formed directly.

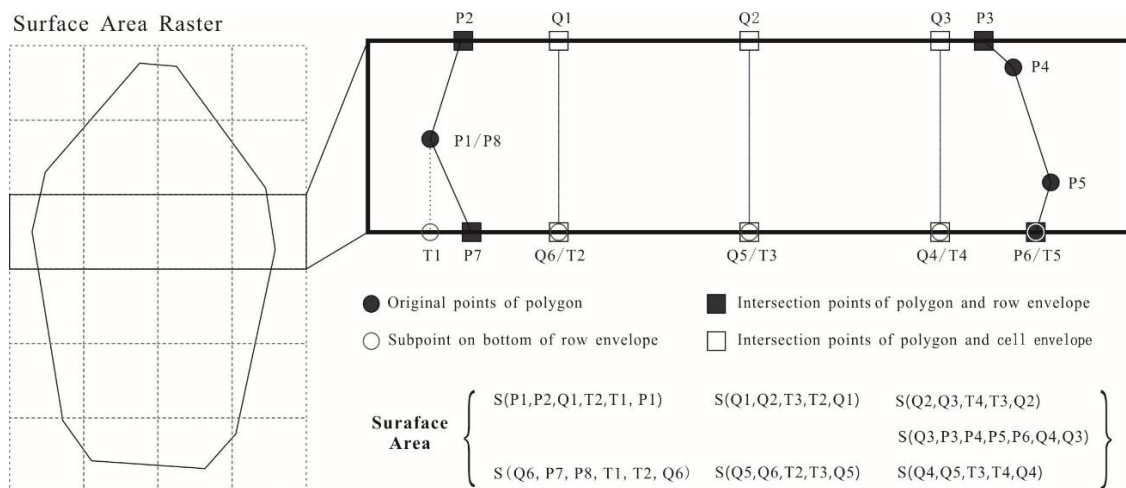


Figure 5. Schematic of the calculation of an irregular polygon.

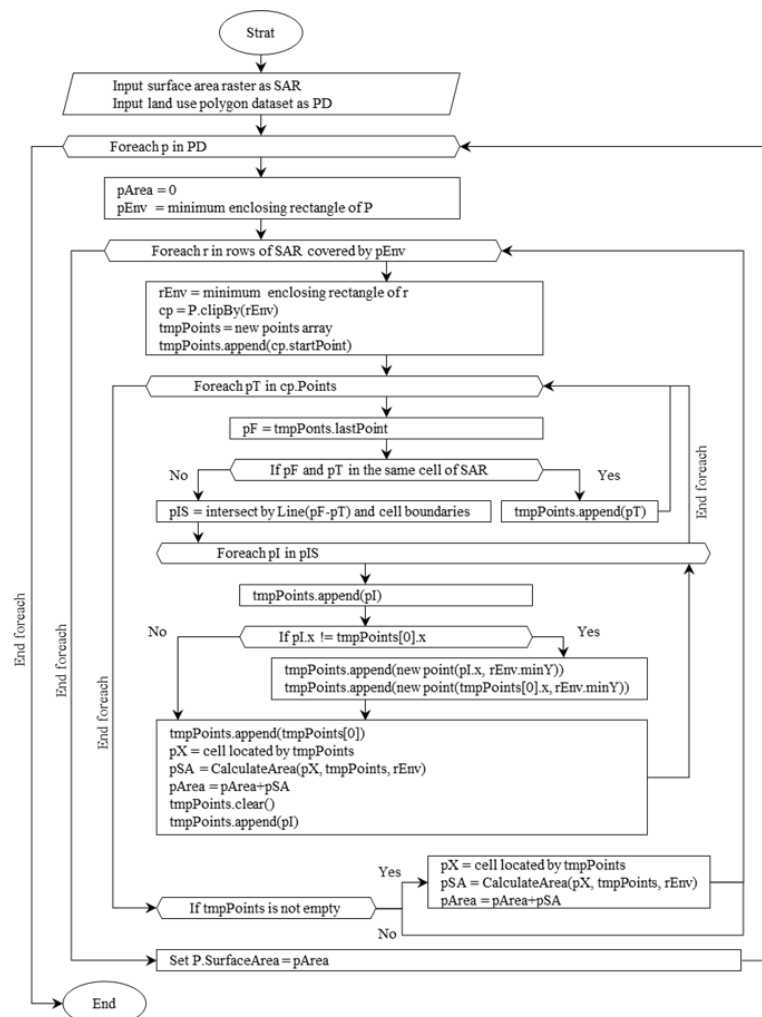


Figure 6. Flowchart of the VROA algorithm.

3. Results

The software and hardware configuration of the computing device used to execute the algorithm are as follows: (1) CPU: Core (TM) i5-3470, two kernels four threads, 3.2 GHz; (2) Memory: 4 GB;

(3) Operating system: Windows 10, 64 bit version; and (4) other: Net Framework 4.5. To verify the performance and accuracy of the algorithm, a comparative experiment was performed against the TIN-based method and the raster-based method. The comparison was achieved using ArcGIS Desktop 10.2.2 software produced by ESRI. The TIN-based method was implemented using the tool “Add Surface Information”, and the TIN dataset was built using elevation points which were converted from the raster-based DEM. The raster-based method was implemented using the tool “Zonal Statistics as Table” with surface area raster data and a land use map. The surface area raster was first generated using the method described in Section 2.2. Since the TIN-based method is considered to have a relatively high accuracy [9], and it is impossible to obtain the true value of the surface area for the current data and computational environment, this paper used the results obtained by the TIN-based method as the standard value. A ratio of the raster-based value to the TIN-based value (RatioR), and a ratio of the VROA-based value to the TIN-based value (RatioV), were employed to analyze the improvement when using the VROA-based method.

The statistical results of the total surface area for the three study areas are shown in Table 1. In this table, TIN_SA, Raster_SA, and VROA_SA represent the surface areas of the TIN-based method, the Raster-based method, and the VROA-based method, respectively. The box plots of RatioR and RatioV are shown in Figure 7.

Table 1. Descriptive statistics of the patch surface area in the three study zones.

	Xiantao	Luotian	Shennongjia
Number of land use polygons	51,233	65,722	22,988
Total surface area of TIN-based method	2,268,766,822.94	2,268,766,822.94	3,902,985,501.19
Total surface area of raster-based method	2,270,267,464.98	2,270,267,464.98	3,905,941,441.71
Total surface area of VROA-based method	2,269,227,094.11	2,269,227,094.11	3,905,701,021.62
RatioR of total surface area	1.000433	1.000661	1.000757
RatioV of total surface area	1.000162	1.000203	1.000696
Mean RatioR of patch surface areas	1.011051	1.023171	1.013790
Mean RatioV of patch surface areas	1.000468	1.001099	1.003095
Standard deviation of RatioR	0.183528	0.217023	0.209578
Standard deviation of RatioV	0.001055	0.002853	0.005538

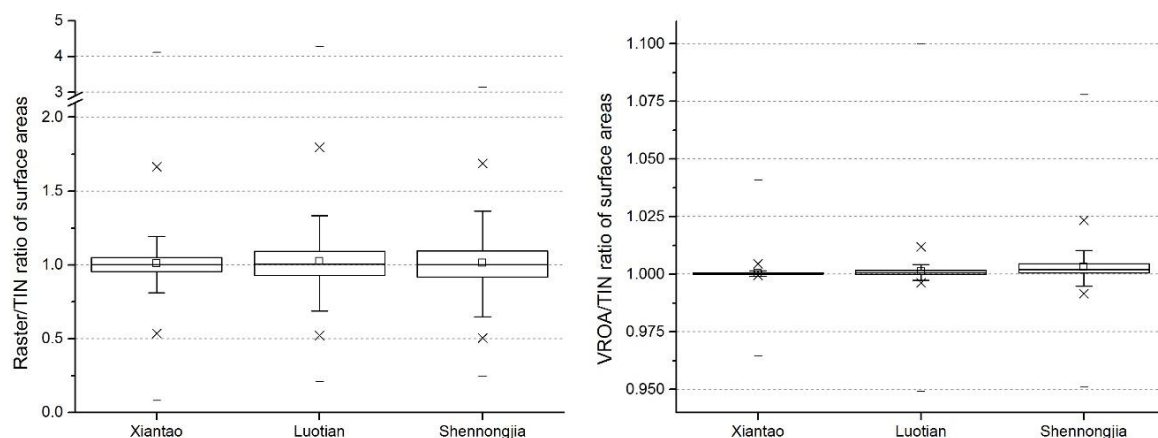


Figure 7. Box plots of RatioR and RatioV in the three study zones.

RatioR and RatioV are close to 1 for the total area (Table 1) of each of the three study areas. However, the standard deviation of RatioR (between 0.183528 and 0.209578) is larger than that of RatioV (between 0.001055 and 0.005538). A larger standard deviation indicates that the raster-based method is more unstable than the VROA-based method. It can also be seen that between the plain area (Xiantao) and the mountainous area (Shennongjia), the mean values of both RatioR and RatioV show

a tendency to be slightly, but consistently, far away from 1 and the standard deviation of RatioR and RatioV tends to be bigger.

Figure 8 shows the range of values for land-use classed RatioR and RatioV. The results show that for each land use class, the mean value of RatioV is closer to 1 than that of RatioR and the variance of RatioV is much smaller than that of RatioR.

The Land Road class and the Tour Land class in Shennongjia achieved the maximum difference in mean values of RatioR and RatioV from 1, with values of 0.834451 and 1.006352, respectively. The Tour Land class and the Road Land class in Luotian achieved the maximum standard deviations for RatioR and RatioV, with values of 0.420764 and 0.013960, respectively.

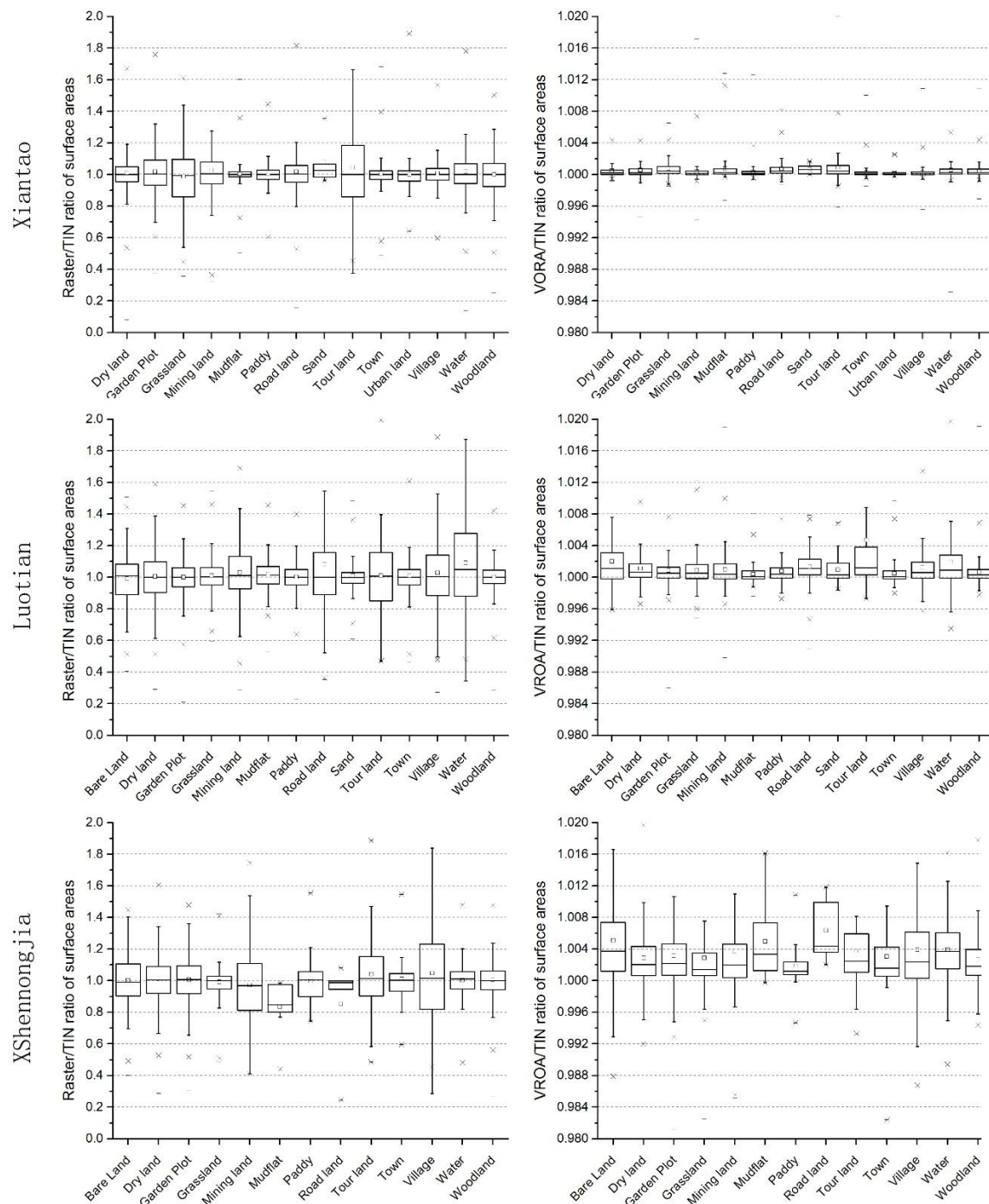


Figure 8. Box plots of land use classed by RatioR and RatioV in the three study zones.

Figure 9 shows the value ranges of the size-classed RatioR and RatioV. The breaks in size classes were designed by the surface areas of the TIN-based method (Table 2). The mean values of RatioV were closer to 1 than those of RatioR for all size classes, and the standard deviations of RatioV were closer to 0 than those of RatioR for all size classes. As the size increased, the mean values of RatioV and RatioR both tended to be closer to 1, and the standard deviations of RatioV and RatioR both tended to be closer to 0. For all size classes, the mean values of RatioV and RatioR were slightly higher than 1. Between the plain area (Xiantao) and the mountainous area (Shennongjia), the mean value for both RatioR and RatioV tended to be away from 1, and the standard deviations of RatioR and RatioV tended to be bigger.

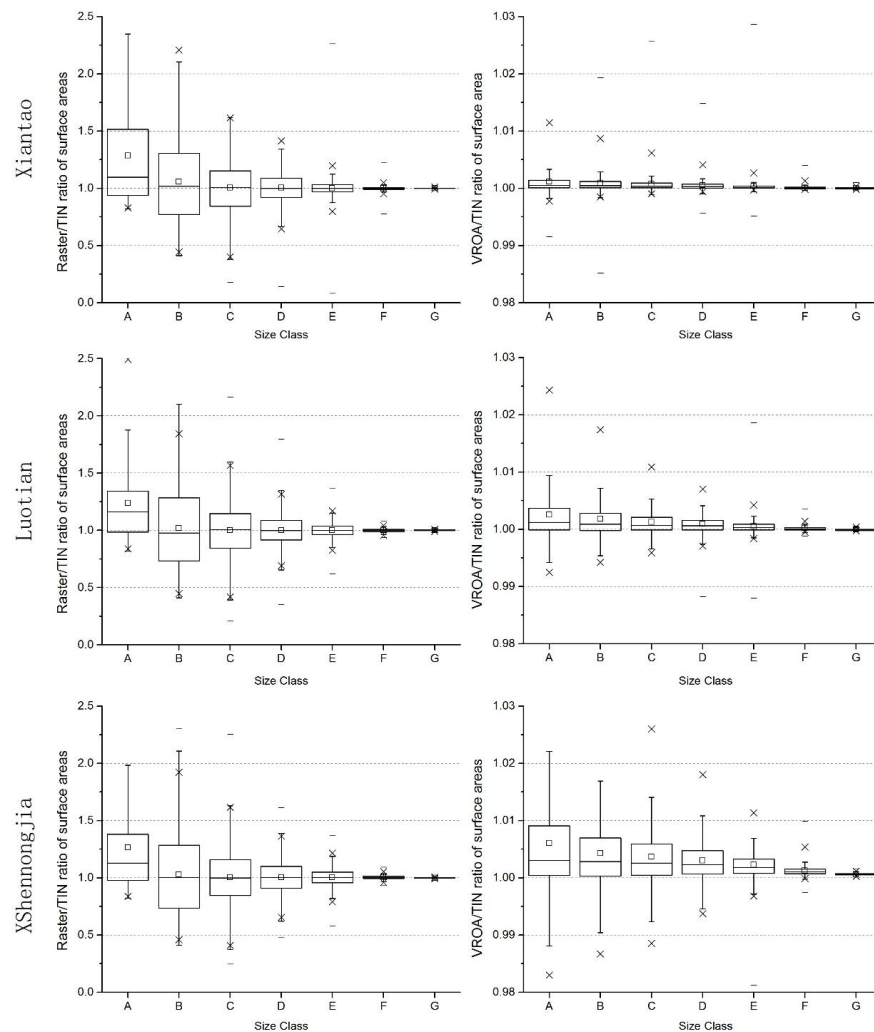


Figure 9. Box plots of size-classed RatioR and RatioV in the three study zones.

Table 2. Size class by TIN-based surface areas.

TIN-Based Surface Areas							
From value	0	1000	2000	5000	10,000	100,000	1,000,000
To value	1000	2000	5000	10,000	100,000	1,000,000	
Size Class	A	B	C	D	E	F	G

The computation time for each of the three methods is shown in Table 3 and resulted in the following results: (1) The raster-based method had the lowest time consumption and spent most of the required time generating the surface area raster, (2) the TIN-based method had the highest

time consumption and spent most of the time calculating the polygon surface area with TINs, and (3) the VROA-based method had a time consumption that was between the other two methods and was approximately a tenth of that of the TIN-based method and ten times higher than the raster-based method.

Table 3. Time consumption of the three methods.

Region Name	Rows and Columns	Vector Polygon No.	Time Consumption (seconds)		
			Tin-Based	Raster-Based	Vroa-Based
Xiantao City	3524 × 6530	51,732	4151	38	491
Luotian County	5128 × 4998	67,145	4633	43	517
Shennongjia Forestry District	4332 × 7844	23,244	3034	45	328

4. Discussion

Accuracy and efficiency are two important indexes for practical surface area calculations. The raster-based method is very efficient and widely used for large area computing tasks. The TIN-based method is considered to be a highly accurate method of surface area calculation and is used for high precision computing tasks in small areas. As data observation technology has improved, higher resolutions and data quality are becoming more prevalent, leading to higher requirements for surface area calculations. The VROA design is suitable for achieving both a high accuracy and a high efficiency. The advantages and disadvantages of the VROA method are discussed in the following paragraphs.

4.1. Precision Analysis

Raster-based methods are the most widely used methods for surface area calculations in large areas. In terms of the total surface area, the VROA-based method only achieves an accuracy that is slightly higher than the raster-based method. However, for land use polygons, the VROA-based method shows smaller value ranges than the raster-based method for both land-use classed and size-classed methods, which means that the performance of the VROA-based method has a better stability than the raster-based method. This feature can ensure the accuracy for each land use polygon in a large surface area calculation task.

In this paper, the value obtained using the TIN-based method is considered to be the standard value. The reason for the differences between the TIN-based method and the VROA-based method could arise from either method. A raster-based DEM dataset can usually be converted to a TIN dataset based on a certain elevation threshold. Different elevation thresholds will lead to different choices of elevation points for TIN construction. Therefore, the same raster-based DEM data can generate different TIN datasets for different elevation thresholds, which leads to a difference in the surface area calculated based on TIN data [9]. There are several interpolation methods available to determine the surface information on triangulated surfaces. Different interpolation methods will result in different surface areas [19]. A further TIN dataset needs a projection coordinate system to calculate the surface area and deformation will also affect the calculation results. The first step of the VROA-based method is to generate a surface area raster. The surface area raster is built directly from the raster DEM data based on the geodetic coordinate system, thereby avoiding a loss of accuracy due to projection conversion. Using four adjacent cells in the raster-based DEM, two sets of triangles can be constructed to represent the cell surface area. In this paper, the average value of two sets of triangle areas was used to represent the surface area. For the polygon surface area calculation step, VROA was implemented based on the ratio between the intersecting polygon area and the cell area to determine the area of intersecting polygons that will be assigned from the raster values. In this way, the VROA-based method obtains values that are close to the TIN-based method.

4.2. Efficiency Analysis

Raster-based DEM is currently the most popular type of digital elevation model. The raster-based method with a raster-based DEM has a very high efficiency for calculating the surface area, which is an important reason for its wide-scale implementation. The raster-based method is highly applicable for large areas at a national scale, especially when there are low accuracy requirements. However, with the development of spatial data observation technologies, high resolution data is becoming more popular. It is difficult to meet the accuracy requirements for land use polygon surface area computing tasks using this method. The TIN-based method is considered to be a high precision surface area calculation method, but it is not suitable for large scale surface area computation due to its low computational efficiency for high precision TIN data. Although the computational efficiency can be improved by reducing the number of TIN vertices, this will inevitably introduce a loss in computational accuracy. The VROA method considers the characteristics of both methods and uses vector boundary precision overlay technology to obtain the surface areas of irregular polygons with a high efficiency. The computational efficiency is 8–10 times better than the TIN-based method, while ensuring the accuracy and stability for land use polygon surface area calculations in a large area.

4.3. Compatibility Analysis

The VROA-based method was designed based on both raster data and vector data, thus solving the edge-effect problem that occurs during raster-vector analysis. Vector land use data is converted to raster data while using the raster-based surface area calculation method. The vector-raster data conversion inevitably introduces a loss in precision. In the TIN-based method, the output data can be constructed differently if different parameters are used during the process of generating the TIN dataset, and calculating the surface area and this uncertainty are important sources of uncertainty of the surface area. The two outcomes of the VROA-based method are the surface area raster and the surface areas of each land use polygon. The surface area raster can be used for surface-area ratio calculations, which may be further used in the calculation of topographic roughness or ruggedness and cost rasters. The land use polygon surface area could be further used for statistical analysis and pattern index construction. This characteristic is very useful for comprehensive calculation tasks for large areas.

5. Conclusions

In this study, a method based on a vector-raster overlay analysis algorithm (the VROA-based method) was proposed and applied to obtain the surface area statistics of land-use polygons in different landforms, using the efficient and accurate clipping of raster cells by the vector polygon boundary. The algorithm was described in the following two parts: building the raster file of the surface area based on the raster-based DEM and calculating the surface area for the vector polygons by VROA. The following three conclusions were obtained.

- (1) In vector-raster overlay statistics analysis, the proposed algorithm maintains the actual boundary of the vector polygon to clip the raster cells. Compared to the raster-based method, the results generated by the VROA-based method are closer to the results generated by the TIN-based method. Since the TIN-based method is considered to have relatively high accuracy, the VROA-based method can perform with a higher accuracy than the raster-based method.
- (2) The algorithm also significantly reduces the number of spatial operations and computation time compared to the TIN-based method.
- (3) The VROA-based method was designed based on both raster data and vector data, which avoids a loss in precision during the conversion to different data formats. The outcomes could be used in further calculation tasks.

Thus, this method meets the accuracy and efficiency requirements for large-scale applications. Furthermore, the surface area has a wide range of application prospects for regional resource assessments, including forest resource assessments, rice yield assessments, and ecological geography

research, including the study of the carbon sequestration capacity and the calculation of the value of ecosystem services. All of these fields require the surface area as basic data, and the traditionally used projection plane may be inappropriate for these areas. These issues will be discussed in a further study.

Acknowledgments: This research was financially supported by the Special Fund of the Ministry of Land and Resources of China in the Public Interest (ID. 201511001).

Author Contributions: Peng Xie, Yaolin Liu, and Qingsong He conceived and designed the experiments; Peng Xie and Xiang Zhao performed the experiments; Jun Yang helped in data analysis; and Peng Xie wrote the paper.

Conflicts of Interest: The authors declare no conflicts of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix Appendix

The pseudo code of the **Algorithm A1** algorithm is given in the following table.

Algorithm A1 VROA	
	Inputs: Surface Area Raster(SAR), Polygon Dataset(PD)
	Outputs: Surface Area for Each Polygon
	Function: VROA(SAR, PD)
1	For each polygon(p) in PD:
2	Define pArea=0 as the surface area of cp
3	Obtain minimum enclosing rectangle (pEnv) of p
4	For each row between start and end rows of SAR covered by pEnv
5	Obtain minimum enclosing rectangle (rEnv) of row
6	Clip p by rEnv and name the clipped polygon as cp
7	Get the start point in cp and save it in a temp points array named tmpPoints
8	For each point (pT) rest in cp:
9	Define pF as the last point in tmpPoints
10	If pF and pT locate in same pixel:
11	Append pT into tmpPoints and do next
12	Else:
13	Get the intersected points(pIS) of Line(pF-pT) and the boundaries of cells
14	For each pI in pIS:
15	Append pI into tmpPoints
16	If pI.x != tmpPoints[0].x:
17	Append new point (pI.x, rEnv.minY) into tmpPoints
18	Append new point (tmpPoints[0].x, rEnv.minY) into tmpPoints
19	Append tmpPoints[0] into tmpPoints
20	Define pX as the cell tmpPoints located
21	Define pSA = CalculateArea(pX, tmpPoints, rEnv)
22	Set pArea=pArea+pSA
23	Clear tmpPoints and append pI into tmpPoints
24	If tmpPoints is not empty:
25	Define pX as the cell tmpPoints located
26	Define pSA = CalculateArea(pX, tmpPoints,rEnv)
27	Set pArea=pArea+pSA
28	Outputs pArea
	Inputs: raster cell (pX), points array (tmpPoints), row minimum enclosing rectangle (rEnv)
	Outputs: surface area of the polygon presented by tmpPoints
	Function: CalculateArea(pX, tmpPoints, rEnv)
1	Define pTmp as the polygon presented by tmpPoints
2	Get the ellipsoid area of pTmp as pEA
3	Get the ellipsoid area of pX as pxEA
4	Get the surface area of pX as pxSA
5	Get the surface area of pTmp (pSA) by formula pEA*pxSA/pxEA
6	Return pSA

References

1. Li, J.C.; Wang, W.L.; Hu, G.Y.; Wei, Z.H. Changes in ecosystem service values in Zoige Plateau, China. *Agric. Ecosyst. Environ.* **2010**, *139*, 766–770. [CrossRef]
2. Li, H.; Wang, S.; Ji, G.; Zhang, L. Changes in land use and ecosystem service values in Jinan, China. *Energy Proced.* **2011**, *5*, 1109–1115.
3. Liu, Y.; Li, J.; Zhang, H. An ecosystem service valuation of land use change in Taiyuan City, China. *Ecol. Model.* **2012**, *225*, 127–132. [CrossRef]
4. Chen, H.; Yang, G.; Peng, C.; Zhang, Y.; Zhu, D.; Zhu, Q.; Hu, J.; Wang, M.; Zhan, W.; Zhu, E. The carbon stock of alpine peatlands on the Qinghai–Tibetan Plateau during the Holocene and their future fate. *Quat. Sci. Rev.* **2014**, *95*, 151–158. [CrossRef]
5. Cui, Q.; Wang, X.; Li, C.; Cai, Y.; Liu, Q.; Li, R. Ecosystem service value analysis of CO₂ management based on land use change of Zoige alpine peat wetland, Tibetan Plateau. *Ecol. Eng.* **2015**, *76*, 158–165. [CrossRef]
6. Zhou, T.; Chen, B.M.; Liu, G.; Huang, F.F.; Liu, J.G.; Liao, W.B.; Wang, Y.Y.; Ren, S.J.; Chen, C.Q.; Peng, S.L. Biodiversity of Jinggangshan Mountain: The importance of topography and geographical location in supporting higher biodiversity. *PLoS ONE* **2015**, *10*, e0120208. [CrossRef] [PubMed]
7. Adhikari, H.; Heiskanen, J.; Maeda, E.E.; Pellikka, P.K.E. The effect of topographic normalization on fractional tree cover mapping in tropical mountains: An assessment based on seasonal Landsat time series. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 20–31. [CrossRef]
8. Eisenlohr, P.V.; Alves, L.F.; Bernacci, L.C.; Mcg, P.; Torres, R.B.; Emb, P.; Fam, D.S.; Assis, M.A.; Ramos, E.; Alc, R. Disturbances, elevation, topography and spatial proximity drive vegetation patterns along an altitudinal gradient of a top biodiversity hotspot. *Biodivers. Conserv.* **2013**, *22*, 2767–2783. [CrossRef]
9. Jenness, J.S. Calculating landscape surface area from digital elevation models. *Wildl. Soc. Bull.* **2004**, *32*, 829–839. [CrossRef]
10. Hodgson, M.E. What cell size does the computerd slope/aspect angle represent? *Photogramm. Eng. Remote Sens.* **1995**, *61*, 513–517.
11. Lopez, M.; Berry, J.K. Use surface area for realistic calculations. *GeoWorld* **2002**, *15*, 25.
12. Zhang, W.; Ai-Nong, L.I. Study on calculating surface area in China based on SRTM DEM data. *Geogr. Geo-Inf. Sci.* **2014**, *30*, 51–55.
13. Zhang, Y.; Zhang, L.-N.; Yang, C.-D.; Bao, W.-D.; Yuan, X.-X. Surface area processing in GIS for different mountain regions. *For. Stud. China* **2011**, *13*, 311–314. [CrossRef]
14. Song, G.; Chen, Y.; Zhou, Y. Algorithm for Estimating Mountainous Surface Area Based on GPS Data. In Proceedings of the 2nd International Conference on Green Communications and Networks 2012 (GCN 2012), Chongqing, China, 12–14 December 2012; pp. 679–687.
15. Zhou, Q.; Liu, X. *Digital Terrain Analysis*; Science Press: Beijing, China, 2006; pp. 114–122.
16. Xie, C.L.; Zhao, R.; Liang, Y. Arithmetic research on computing surface area of statistic units in national geographic conditions monitoring based on geographic coordinate system. *Remote Sens. Inf.* **2014**, *29*, 47–51.
17. Lu, Z.; Qu, Y.; Qiao, S. *Geodesy: Introduction to Geodetic Datum and Geodetic Systems*; Springer: Berlin, Germany, 2014; pp. 176–182.
18. O'Rourke, J. *Computational Geometry in C*; Cambridge University Press: New York, NY, USA, 1994.
19. Esri. ArcGIS Help 10.1—Add Surface Information (3D Analyst). Available online: <http://resources.arcgis.com/en/help/main/10.1/index.html#/00q900000016000000> (accessed on 10 July 2016).

