Article

Using Monte Carlo Simulation to Improve the Performance of Semivariograms for Choosing the Remote Sensing Imagery Resolution for Natural Resource Surveys: Case Study on Three Counties in East, Central, and West China

Juanle Wang [1,2,]* [ID], **Junxiang Zhu [3,]* and Xuehua Han [1,4]**

[1] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; hanxh@lreis.ac.cn
[2] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
[3] Australasian Joint Research Centre for Building Information Modelling, School of Built Environment, Curtin University, Bentley, WA 6102, Australia
[4] University of Chinese Academy of Sciences, Beijing 100049, China
*   Correspondence: wangjl@igsnrr.ac.cn (J.W.); junxiang.zhu@postgrad.curtin.edu.au (J.Z.); Tel.: +86-010-6488-8016 (J.W.)

Abstract: Semivariograms have been widely used in research to obtain optimal resolutions for ground features. To obtain the semivariogram curve and its attributes (*range* and *sill*), parameters including *sample size* (*SS*), *maximum distance* (*MD*), and *group number* (*GN*) have to be defined, as well as a mathematic model for fitting the curve. However, a clear guide on parameter setting and model selection is currently not available. In this study, a Monte Carlo simulation-based approach (MCS) is proposed to enhance the performance of semivariograms by optimizing the parameters, and case studies in three regions are conducted to determine the optimal resolution for natural resource surveys. Those parameters are optimized one by one through several rounds of MCS. The result shows that exponential model is better than sphere model; *sample size* has a positive relationship with R^2, while the *group number* has a negative one; increasing the simulation number could improve the accuracy of estimation; and eventually the optimized parameters improved the performance of semivariogram. In case study, the average sizes for three general ground features (grassland, farmland, and forest) of three counties (Ansai, Changdu, and Taihe) in different geophysical locations of China were acquired and compared, and imagery with an appropriate resolution is recommended. The results show that the ground feature sizes acquired by means of MCS and optimized parameters in this study match well with real land cover patterns.

Keywords: optimal resolution; Monte Carlo simulation; semivariogram; natural resource survey; remotely sensed image interpretation

1. Introduction

Remote sensing technology is developing rapidly as an efficient method for acquiring data from a distance, usually from outer space by a satellite. It is now becoming a popular tool for natural resource surveys, which determine the distribution and quantity of various natural resources over a specific area, such as soil, grassland, forest, and farmland. Remote sensing is making this process quicker, more efficient, and more economical, by providing a wide range of imageries varying in resolution and spectral band, and with the help of land cover classification and spatial analysis techniques [1–3].

The first task to be completed before conducting a remote sensing-based survey is choosing an appropriate remote sensing data source to determine the imagery resolution and spectral bands, which have a significant effect on the accuracy of land cover classifications [4]. The influence of spectral bands on the extraction and visualization of ground features, which are physical objects on the ground such as forest and water, is well known. For example, red and near-infrared bands are used to calculate the Normalized Difference Vegetation Index (NDVI) for vegetation detection or different band combinations of Landsat Thematic Mapper (TM) imagery are employed to highlight different ground features. However, the selection of the resolution, or the size of pixels that constituting an imagery, remains an unsolved problem. There is no "one size fits all" solution [5], because the resolution should be determined with respect to the size of ground features, and there is one optimal resolution for each ground feature. If a unique resolution is adopted for all the ground features, details of objects smaller than the resolution would be lost.

Currently, several methods are used for determining the size of ground features such as semivariograms, local variance (LV), wavelet method, and spatial autocorrelation [6–8]. Among them, semivariograms are widely adopted because of their mathematical simplicity and ease of interpretation. The *range* of a semivariogram is related to the size of the ground feature. It provides a measure for the size of the elements in the image and has been suggested to be a useful indicator in selecting the optimal spatial resolution [7,9]. Woodcock and Strahler analyzed and validated the feasibility of semivariogram in determining the size of ground features in remote sensing using both simulated and real satellite imagery [7,10]. Subsequently, semivariograms were applied in many remote sensing based studies, for example, to obtain the structure of forests from high-resolution remote sensing images [11–13]. Song and Liu studied the performance of semivariograms with respect to obtaining the canopy size via IKONOS or Quickbird imagery [14,15]. Guardiola developed a new methodology for modeling spatial variations of the relative wood density using variograms for XRCT images [16].

One limitation of the application of semivariograms in remote sensing is that they can only be applied to simple scenes, which merely contain one element and one background [7,10], while real satellite imagery tends consist of complex scenes that comprise multiple elements. The studies mentioned above mainly focused on merely one specific ground feature, such as grassland, forest, or canopy, and to meet simple scene requirements. Two main approaches below are used to solve this. Subimageries containing simple scenes were carefully extracted from the original imagery. This approach is called sub-area method (SAM) in this paper. Another way to solve this issue is separating those elements from one another using land cover classification techniques [17], which is called direct-analysis method (DAM) in this paper. Both methodologies have advantages and disadvantages. A more adaptive and reliable approach is needed for natural resource surveys.

Another limitation is the massive sample size introduced using imagery. In traditional domains, such as mining and geology, the sample size for calculating a semivariogram is limited, and usually remains under 1000 [18]. However, in the case of remote sensing, the area of interest is entirely sampled, and each pixel in the imagery serves as a sample. As a result, the sample size can easily reach one million (for a common 1000 by 1000 pixels image), which is overlarge and makes building semivariogram computationally unrealistic. Apparently, the more samples included in the calculation, the more stable and accurate the estimate will be. However, more samples also mean higher requirements on the computing and memory capacities; therefore, not all pixels can be incorporated and the reliability of the analysis results might thus be impaired. Finding the balance between the sample size, computing capacity, and accuracy would facilitate the application of semivariogram in remote sensing. Unfortunately, there is no clear guide on it.

This study proposes a Monte Carlo simulation-based approach (MCS) to enhance the performance of semivariograms in obtaining the optimal resolution for general ground features in conjunction with RapidEye imagery. First, the approach is compared with two other commonly used semivariogram-based methods for optimal resolution acquirement. The MCS is then applied to

optimize the parameters (*sample size*, *maximum distance*, *and number of distance groups*), as well as the model and number of simulation. Finally, the average size and the optimal resolution remote sensing images for three general ground features (grassland, farmland, and forest) in three counties in different geographic locations of China are obtained using the optimized parameters. The aim of this study is to improve the performance of semivariogram by reveling the influence of those parameters on the estimate accuracy and providing optimized parameters, to make *range* estimates more accurate and precise for the purpose of selecting appropriate remote sensing image for natural resource survey, in spite of the size of imagery used and the capacity of the computer.

2. Materials and Methods

2.1. Study Area

Three study areas in different geographic locations of China were chosen for this study: Taihe County in the Jiangxi Province in the mountainous area of eastern China; Ansai County in the Shaanxi Province in the Loess Plateau in central China; and Changdu County in Tibet, which is located in the Qinhai–Tibet Plateau of western China. The locations of these study sites are shown in Figure 1.
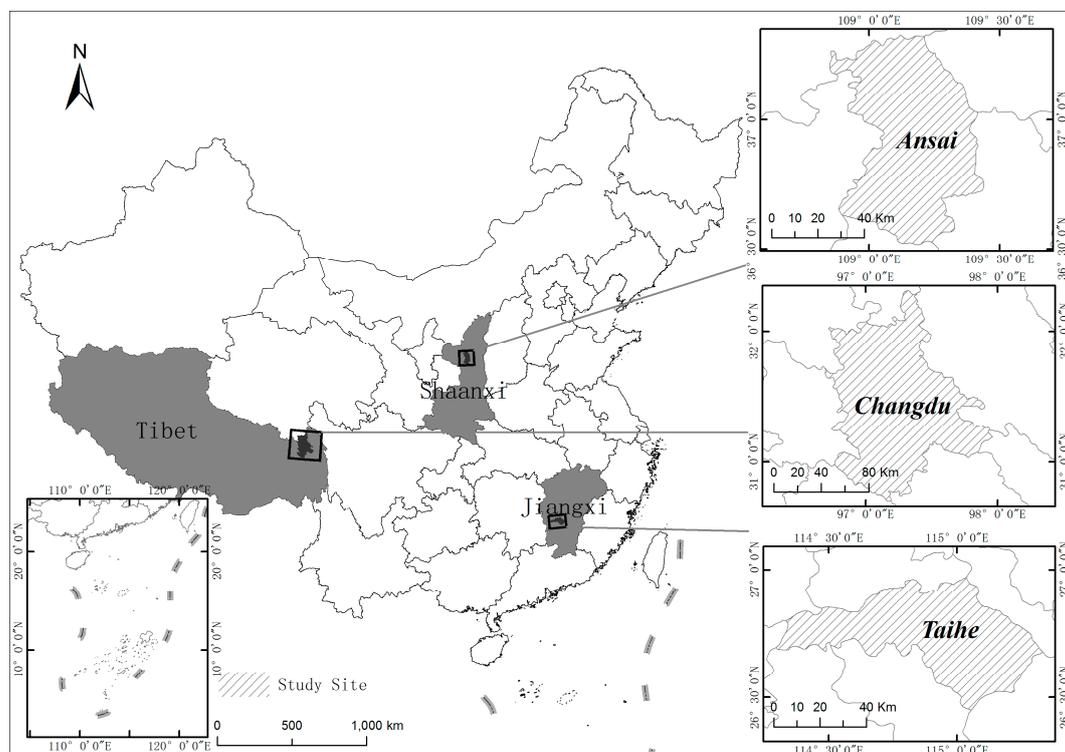
Figure 1. Locations of study sites.

Taihe County is in the middle southern part of the Jiangxi Province, within $26°27'$–$26°59'$ N and $114°57'$–$115°20''$ E, and has subtropical monsoon climate. Its annual average temperature is 18.6 °C and the annual average precipitation is 1726 mm. The total area is 2667 km^2; high hills account for 5.92% of the total area, low hills make up 54.52%, and plains comprise 27.6%.

Ansai County is located in the north of Shaanxi Province, within $36°31'$–$37°19$ N and $108°6'$–$109°26'$ E, and a total area of 2950 km^2. It has a semiarid monsoon climate, annual average temperature of 8.8 °C, annual average precipitation of 505.3 mm. The vegetation species in this region gradually change northward, from broadleaved deciduous forests to shrub and grassland. This county suffers from severe water loss and soil erosion [19,20].

Changdu County is in the east of Tibet, within 30°44′–32°19′ N and 96°42′–97°58′ E, with a total area of 10,652 km². Its average altitude is more than 3500 m. Affected by the altitude and terrain, it has a plateau subtemperate sub-humid climate. Its annual average temperature varies between 3 °C and 8 °C and the annual average precipitation is 477.7 mm.

2.2. Data

The remote sensing data used in this study are RapidEye L1B imagery with a spatial resolution of 6.5 m. The details of the RapidEye image can be found in Table 1. It contains blue, green, red, red edge, and infrared bands [21]. Both red band and near infra-red band are sensitive to vegetation. Because red band is common in both old and latest remote sensing sensors, and some scholars also used it for semivariogram study [10,22], the red band is utilized for analysis in this paper.

Table 1. Image bands and sampling cell size.

Items	Detail
Blue	440–510 nm
Green	520–590 nm
Red	630–685 nm
Red Edge	690–730 nm
Infrared	760–850 nm
Sampling cell	6.5 m

Table 2 shows some extra information on the images used. There are four raw images for Ansai County, three of which were obtained in September 2010 and one was obtained in October of the same year. Taihe County is also covered by four raw images. They were captured on different dates; one was taken in November 2011, two in September 2012, and the last one in October 2010. For Changdu County, five raw images are available, but only the two images that cover the majority of this region with good quality were used in the end, which were captured in September 2010. Due to the low quality of the remaning images, the marginal area in the northern and southeast of Changdu County is not covered in this study. All raw images were orthorectificated and mosaicked into one image before interpretation, which was later cut by the boundary of their corresponding county.

Table 2. Image number and capture date.

Region	Number	Capture Date
Ansai County	4	3 [1] September 2010, 1 October 2010
Taihe County	4	1 November 2011, 2 September 2012, 1 October 2010
Changdu County	5	2 September 2010

[1] The values in parentheses indicate the number of images captured on that date.

2.3. Semivariogram

There are several types of calculations of semi-variograms in remote sensing; Equation (1) is mostly used [23].

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{N(h)} \{Z(s_i) - Z(s_i + h)\}^2, \tag{1}$$

where *N(h)* refers for the total number of pairs of points within a distance *h* and $Z(s_i)$ and $Z(s_i + h)$ are the values at Point s_i and Point $s_i + h$, separated by a distance of *h* pixels [24]. By changing *h*, an ordered set of semivariances is obtained, which constitutes the experimental semivariogram [25].

An ideal semivariogram curve is shown in Figure 2. The assumption in semivariogram theory is that closer objects are more related than distant objects. When the distance between two objects is small, those objects are probably similar or of low variance. In an extreme scenario, when the distance is 0, those two objects are the same. Thus, the variance is zero. As the distance increases, the variance

starts to increase. However, when the distance exceeds a specific threshold, the variance reaches its maximum and will no longer change. This threshold is known as *range*, indicating the influence scale of an event. In remote sensing, objects are pixels and the *range* represents the size of the ground feature. The maximum variance is called *sill*, which is an estimator for the true variance of those pixels [7,12]. Note that, in this study, *nugget* is not considered, and its value is set 0. Forest, grassland, and farmland are assumed to be isotropic.
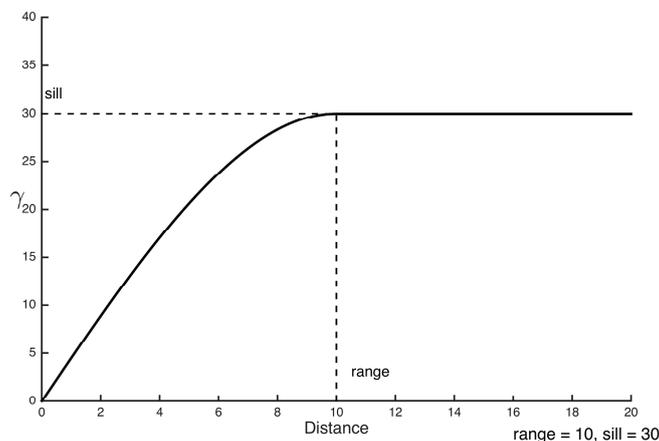
Figure 2. An ideal curve of a semivariogram with a range of 10 and a sill of 30.

As the *range* was measured in pixel, the real size of the ground feature can be calculated using Equation (2):

$$Size = Range \times Resolution, \tag{2}$$

where the *range* is the value estimated for the *range* and *resolution* is the size of the pixel, which is 6.5 in this study.

In practice, a serial of parameters should be defined before establishing a semivariogram. Firstly, to build the semivariogram, three parameters are needed, namely, the *sample size* (*SS*), *maximum distance* (*MD*), and *group number* (*GN*). The *sample size* refers to the number of randomly selected pixels within an image; all distances between pixels will be calculated. The *maximum distance* defines the radius of a circular area; all pixels outside of this area will be excluded from the calculation. The *group number* is the number of groups which the distances (from zero to *maximum distance*) are divided into. Then, to obtain its attributes, *range* and *sill*, a model is needed to fit the semivariogram curve. The most widely used models are the linear, spherical, and exponential models.

2.4. Data Processing

The workflow in this study is shown in Figure 3. It includes three data processing sessions: (1) image interpretation; (2) method comparison; and (3) parameter optimization.

1. Image interpretation: In this session, an object-based automatic classification method, as well as the manual visual interpretation (manually identifying ground features according to prior knowledge on them, such as feature's shape, size, pattern, tone, and texture), was used to identify farmland, built area, water, barren, grassland, and forest from remotely sensed images in eCognition software with its multiresolution segmentation algorithm [26]. First, some ground features such as farmland, built area, and water were visually interpreted, as they either mix easily with other adjacent ground features, such as farmland, or are hard to maintain their shape in the image segmentation, such as river (included in water). By doing this before the image segmentation, the shapes of those ground features could be maintained and high classification accuracy could be ensured. Image was segmented with a scale parameter (SP) of 200 into

unclassified objects, and then those objects were classified based on different indexes, such as normalized difference soil index (NDSI), digital number (DN), normalized difference vegetation index (NDVI), brightness, and normalized difference water index (NDWI). The indexes and their value used for a specific ground feature could be found in Table 3. The value for those indexes were optimized through incremental adjustment to get the best result. Note that, the ground features in those counties were identified in the specific order listed in Table 3, and the order should not be changed, otherwise, different results may be derived. For accuracy assessment, a number of random points were generated, and Google Earth was adopted to check the points one by one manually.

2. Method comparison: The proposed MCS-based approach was compared with the sub-area method (SAM) and direct-analysis method (DAM) by applying them to the same area. A pilot area in Ansai County was selected to conduct the method comparison.

   - The SAM approach randomly selects 30 sample points for each ground feature (e.g., forest and grassland) using a stratified random sampling method based on land cover map. A rectangular area with a size of 500 by 500 pixels around each sample point was then created, which was called *sub-area* and used to extract *subimage* from the raw image. Those subimages in SAM are square and only cover a portion of a ground feature, they are then analyzed and the average of their estimated ranges will be calculated. Note that land cover map is not a must for SAM, subimages may be manually extracted from the raw image without land cover map and just based on individual's judgement. However, this study used land cover map to improve efficiency of this step by automating this process.
   - The raw image was divided into five subimages based on the land cover map in the DAM. Therefore, there is one image for each ground feature. Different from the regular square subimages in SAM, the subimages in DAM were much larger, covering the entire area of a ground feature, and of irregular shapes. The images of different ground features are analyzed individually.
   - The MCS approach is built on the DAM. Instead of one-time analysis, MCS runs the analysis multi times to obtain a number of parallel estimations of the range. Monte Carlo methods are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results [27]. One significant feature of this method is the large quantity of repetitions. Here, each simulation outputs a value for the estimated *range* and the average of all fitted ranges is considered to be the estimator of the true range. MATLAB equipped with a parallel-computing toolbox [28,29] was used to shorten the processing time by running several analyses simultaneously.

3. Parameter optimization: The semivariogram parameters, namely, the *sample size (SS)*, *maximum distance (MD)*, and *group number (GN)*, as well as the *model* and the simulation number, are to be optimized one by one through four assessment rounds.

Table 3. Image interpretation, indexes used and values.

County	Visual Interpretation	Object-Based Automatic Interpretation
Ansai County	Farmland, built area, and river	Forest (NDVI > 0.3), grassland (NDVI > 0.05), lake (NDWI > 0.3), barren (Brightness > 5600), others (all the rest)
Taihe County	Built area and river	Lake (NDWI > 0.25), farmland (NDSI > −0.17), forest (NDVI > 0.25 and Brightness < 4260), barren (Brightness > 5500), grassland (NDVI > 0), other (all the rest)
Changdu County	Farmland, built area and water	Forest (DN < 1950), grassland (NDVI > 0.07), others (all the rest)

To make the data processing procedure more efficient, Python was introduced as a script language for ArcGIS to conduct batch processes such as image separation and random points generation.

In addition, a customized function was developed to generate rectangular areas with assigned sizes (width and height), which were used to extract subimages from the raw image.
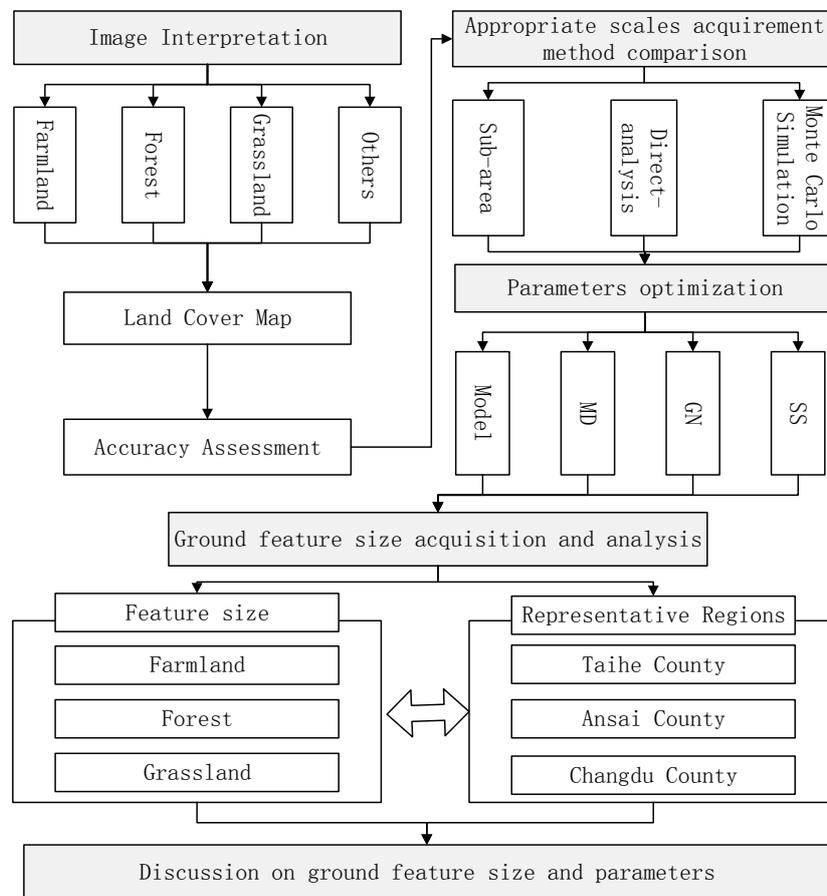
Figure 3. Workflow of this study.

3. Experiment

3.1. Comparison of Semivariogram-Based Methods

3.1.1. Sub-Area Method (SAM)

Figure 4 shows the square root semivariogram curves created via the SAM approach for the pilot area. The square root of $\gamma(h)$ is used in much the same way that a standard deviation is more easily interpreted than a variance [7]. Each subplot contains 30 curves and the distances of each curve are divided into 50 groups. Overall, most of the curves first increase sharply and then more or less remain flat. The curves for the built area, forest, farmland, and grassland show clear sills, while those for water seem to keep slowly increasing. The drawback of SAM is that, if a random point happens to be generated close enough to or on the boundaries between feature classes, the subimage derived from it would probably cross a boundary and contain pixels of multiple feature classes, which would contrast the simple scene assumption.

3.1.2. Direct-Analysis Method (DAM)

Using DAM, a total of six semivariograms were derived including a curve for the raw image, as shown in Figure 5. All curves evidently increase in the beginning and then remain stable, except for the built area, which undergoes fluctuations. The problem with DAM is that, if the image size is overly large, the sample size has to be controlled because of the limited computing capacity. Although the

random sampling method is unbiased, if only a small part of the pixels can be incorporated into the construction of the semivariogram, consequently, it may also lead to potential estimation bias.
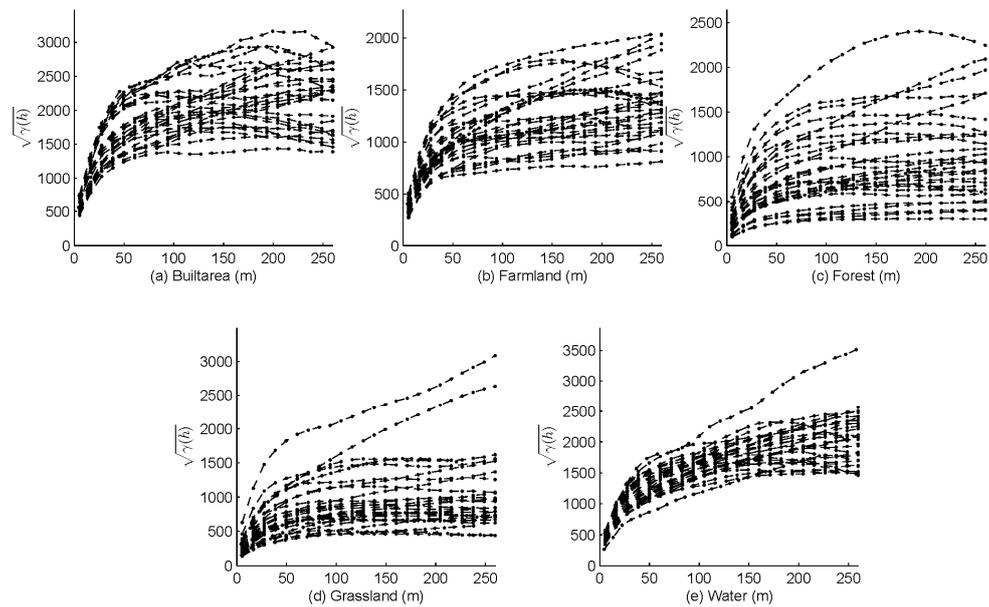
Figure 4. Semivariograms for: built area (**a**); farmland (**b**); forest (**c**); grassland (**d**); and water (**e**), via sub-area method (SAM).
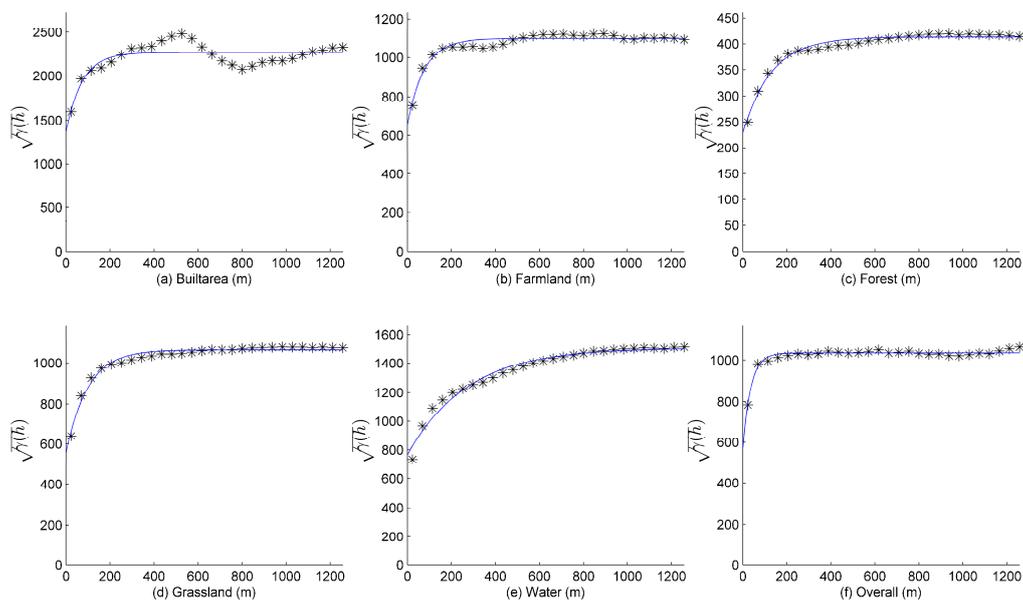
Figure 5. Semivariograms for: built area (**a**); farmland (**b**); forest (**c**); grassland (**d**); water (**e**); and overall (**f**) based on direct-analysis method (DAM).

3.1.3. Monte Carlo Simulation (MCS)

The MCS approach is based on the DAM approach, which is repeated 999 times for each ground feature. It took 36 h to complete the MCS on a computer with a i5-2400 CPU and 16 GB memory. In total, 4995 curves and their corresponding ranges were derived. Those curves were not displayed in the same way as those of the SAM and DAM because of the extremely large amount. Instead, they are presented as a customized plot for the probability distribution of ranges, with which more information

regarding *range* could be revealed. As shown in Figure 6, the *x*-axis is the type of ground feature, the *y*-axis is the range, and the *z*-axis is the probability.
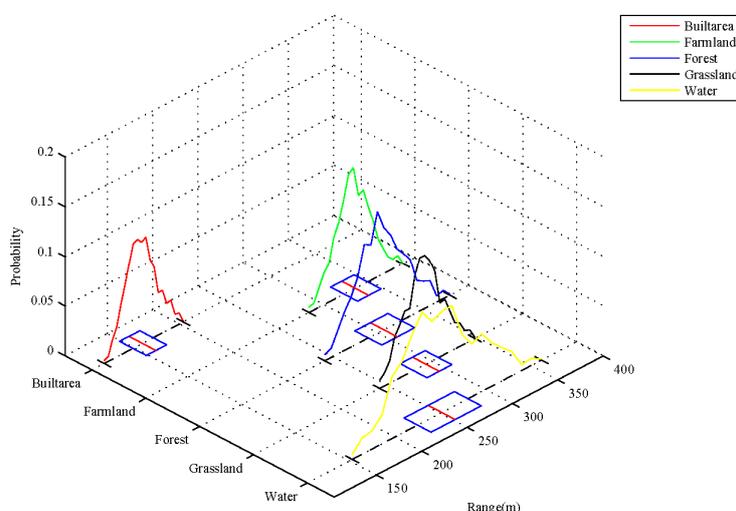
Figure 6. Probability distributions for the built area (**red**), farmland (**green**), forest (**blue**), grassland (**black**), and water (**yellow**).

The MCS approach is an improvement of the DAM, as the result of DAM is essentially just one possibility of results acquired by the MCS approach. It could improve the accuracy by doing more parallel experiments, which is a usual method used in sampling to eliminate unexpected error. Since it is based on DAM, it inherits the advantages of DAM that overcome the disadvantages of SAM by separating the classes as well. Therefore, in general, the MCS approach is suitable for the determination of the ground feature size in remote sensing scenarios against large datasets.

3.2. Parameter Optimization

In this section, the parameters for the calculation and fitting of the semivariogram, namely, the *maximum distance* (*MD*), *group number* (*GN*), *sample size* (*SS*), and *model*, as well as the number of MCSs, are optimized or selected based on several rounds of experiments. In each round, the estimated range and corresponding R^2 are derived. The R^2 is used to indicate the fitting accuracy. The parameters optimized in one round will be used in the next round of the test.

3.2.1. Model and Maximum Distance (MD)

Based on a pre-test, initial values were assigned to the parameters in the beginning of experiments, that is, 500 pixels for *MD*, 50 for *GN*, 10,000 for *SS*, and 200 for number of simulations. The semivariogram curves were fitted using both spherical and exponential models. The value for *MD* was set to three times the average ranges in the next round of the experiment.

The value of average R^2 for the exponential model is generally higher than that for the spherical model, as shown in Table 4. This indicated that the exponential model has a higher accuracy.

Table 4. R^2 values for the exponential and spherical models.

	Ansai County			Taihe County			Changdu County		
	Forest	**Grass**	**Farmland**	**Forest**	**Grass**	**Farmland**	**Forest**	**Grass**	**Farmland**
Exponential	0.93	0.94	0.79	0.92	0.90	0.81	0.87	0.56	0.89
Spherical	0.56	0.61	0.49	0.36	0.38	0.50	0.38	0.27	0.39

3.2.2. Group Number (GN) and Sample Size (SS)

The value for *GN* increases from 10 to 200 at a step size of 5. For each *GN*, 200 simulations were conducted and the average R^2 was calculated. Figure 7 shows the scatter plot of *GN* versus R^2 for each ground feature in each region. Each point inside the subplots stands for a *GN*–average R^2 pair.
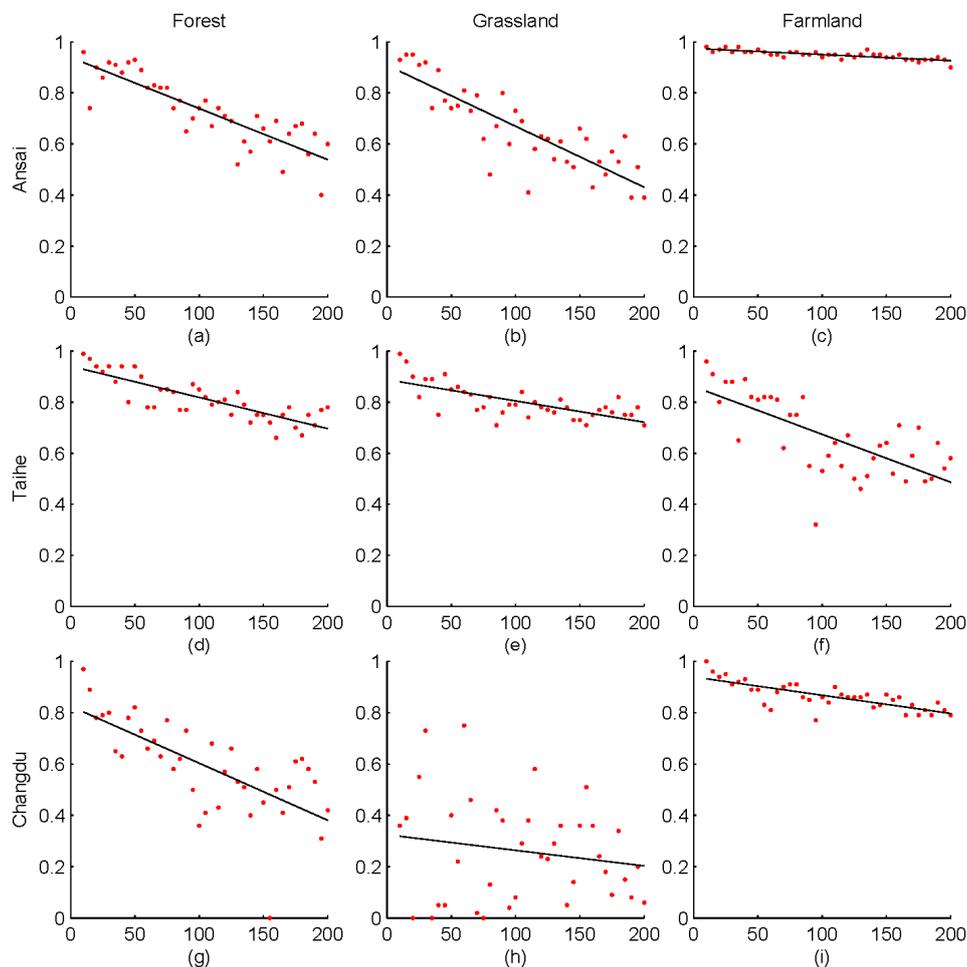
Figure 7. Relationship between the group number (*GN*) and fitting R^2 for Ansai's forest (**a**), grassland (**b**), farmland (**c**), Taihe's forest (**d**), grassland (**e**), farmland (**f**), and Changdu's forest (**g**), grassland (**h**), farmland (**i**); the *x*-axis is *GN*, and the *y*-axis is R^2.

The R^2 value of all subplots has a negative relationship with *GN*. As the *GN* increases, the R^2 decreases. This means that a larger *group number* decreases the accuracy of the estimation. Since no "turning point" could be observed, a specific optimized value could not be identified for *GN*. Nevertheless, based on the plots trends in Figure 7, it could be concluded that the initial value for *GN*, 50, is a little overlarge, and needs to be reduced; thus, 35 is used in this study, which is about two-thirds of the original one.

In the third round, the value of *SS* increases from 1000 to 10,000 at a step size of 250 to reveal the influence of *SS* on the estimation accuracy; for each *SS*, 200 simulations were conducted. The result is shown in Figure 8.

In Figure 8, most of the plots show an increasing trend even after 10,000 such as the one in Figure 8g for forest in Changdu County, indicating that increasing the number of sample could still increase the accuracy, but for some plots such as the one in Figure 8c, a turning point around 2500 could be observed, after which, the accuracy varies slightly. Similar situations could be found in Figure 8e,i.
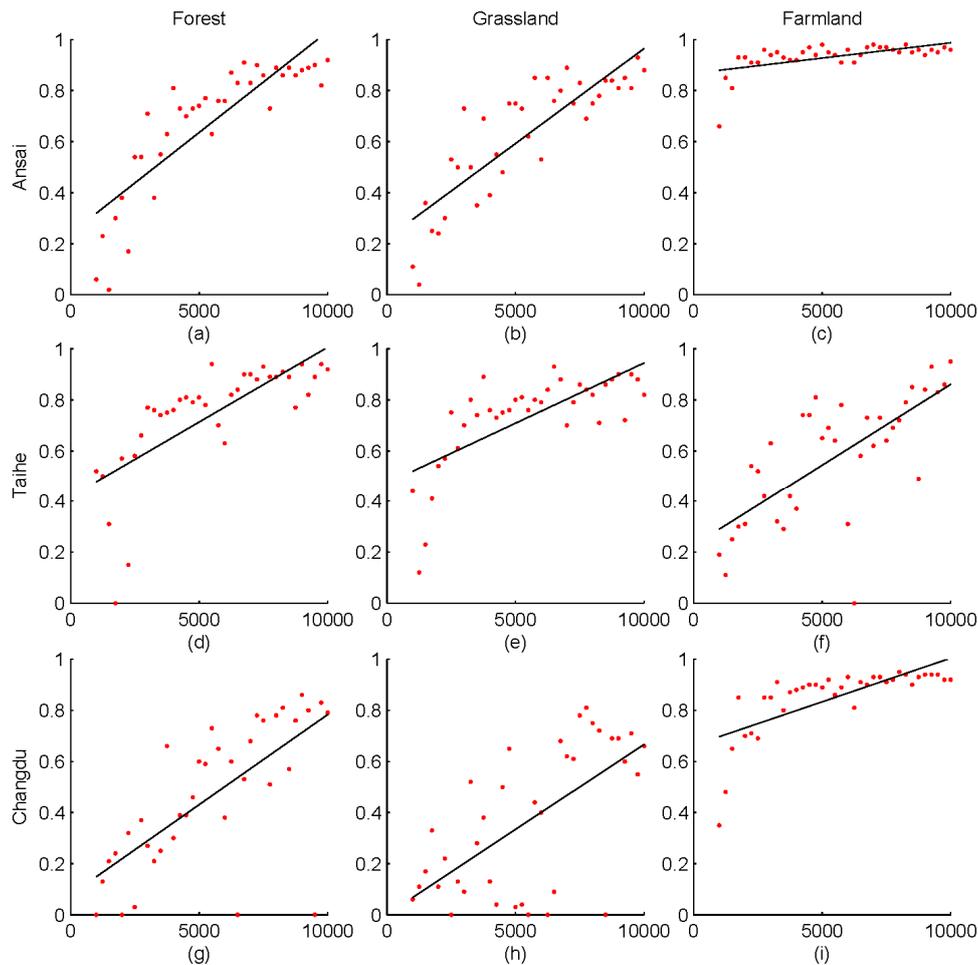
Figure 8. Relationship between sample size (*SS*) and R^2 for Ansai's forest (**a**), grassland (**b**), farmland (**c**), Taihe's forest (**d**), grassland (**e**), farmland (**f**), and Changdu's forest (**g**), grassland (**h**), farmland (**i**); the *x*-axis is *SS* and the *y*-axis is R^2.

It can be inferred that the *SS* and fitting R^2 have a positive relationship. When the *SS* value increases, the R^2 increases simultaneously. However, a threshold is assumed to exist for *SS*, after which the R^2 value would remain stable. By far the selection of *SS* has been greatly limited by the computing capacity, especially the memory capacity. In this study, since most of the plots do not have a "turning point" (threshold), 10,000, the maximum number of samples that the microcomputer could handle, is recommended as the optimized *SS* value.

3.2.3. Comparison between the Results before and after Parameter Optimization

Another 200 simulations were run using the new, optimized parameters. The standard deviations of the 200 ranges derived from the simulation were calculated and compared in Table 5.

The numbers in Table 5 indicate that, after optimization, the standard deviations of the fitting ranges are generally reduced, which indicates a higher precision of the estimation, especially for the farmland in Ansai County. It changed the most, decreasing from 92.9 to 5.8, i.e. 93.8%, followed by farmland in Taihe County, which fell from 167.9 to 22.4 (86.7%). Compared with others, however, the behavior for grassland in Changdu County is rather abnormal: it increased from 418.2 to >7000. The cause needs to be further studied.

Table 5. Standard deviation of the ranges.

	Ansai County			Taihe County			Changdu County		
	Forest	**Grassland**	**Farmland**	**Forest**	**Grassland**	**Farmland**	**Forest**	**Grassland**	**Farmland**
Before (B)	24.5	29.8	92.9	142.4	110.5	167.9	81.9	418.2	127.5
After (A)	10.9	13.4	5.8	137.9	74.9	22.4	62.4	7053.7	34.3
A−B	−13.6	−16.4	−87.1	−4.5	−35.6	−145.5	−19.5	6635.5	−93.2
Percentage	55.5%	55.0%	93.8%	3.2%	32.2%	86.7%	23.8%	——	73.1%

3.2.4. Number of Simulations

Figure 9 shows the lines representing the relationship between the number of simulations and standard deviation of average ranges in Ansai County. The number of simulations increases from 20 to 5000 at a step of 20. It can be observed that the standard deviation of the average fitting ranges first drops obviously as the number of simulation increases. However, once the number of simulations exceeds ~1000, the curves remain stable. In a nutshell, the number of simulations has an influence on the precision and accuracy of the estimation; a high number is recommended.
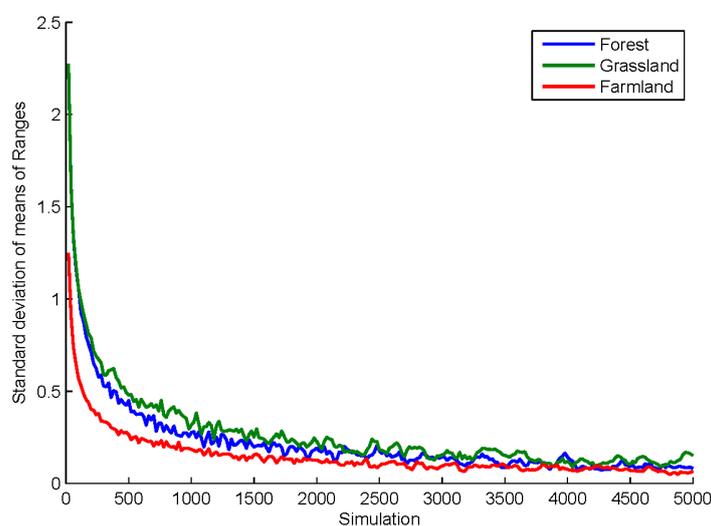
Figure 9. Relationship between the simulation times and standard deviation of average fitting ranges for forest (**blue**), grassland (**green**), and farmland (**red**) in Ansai County.

4. Results

4.1. Image Interpretation

Figure 10a–c shows the land cover maps of Ansai County, Taihe County, and Changdu County, respectively.

The classification accuracy was assessed using Google Earth images with higher spatial resolution. In total, 256 samples were randomly selected for each region and manually checked one by one to validate their accuracy. The overall classification accuracies and Kappa coefficients were then calculated (Table 6). Due to the capability of object-oriented classification methods in dealing with high-resolution images, the overall accuracy and Kappa coefficient are relatively high (almost the same).

Table 6. Accuracy assessment of the classification results for Ansai, Taihe, and Changdu.

	Overall Accuracy	Kappa Coefficient
Ansai County	0.89	0.87
Taihe County	0.88	0.86
Changdu County	0.89	0.87
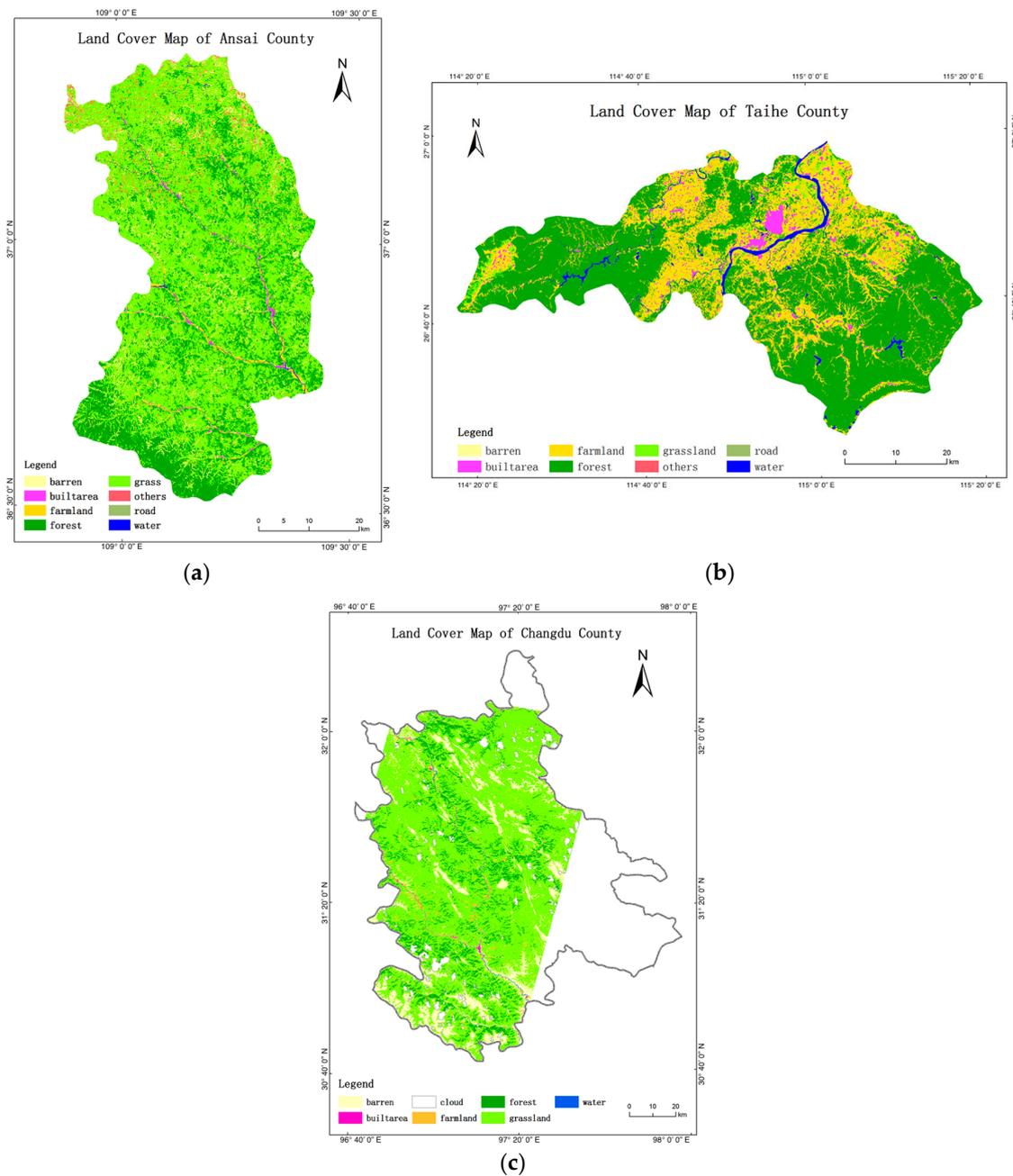
Figure 10. Land cover maps for: (**a**) Ansai County; (**b**) Taihe County; and (**c**) Changdu County.

4.2. Ground Feature Size

The optimized parameters, namely, 35 for *GN* and 10,000 for *SS*, and the exponential model, were used in the analysis of three ground features (farmland, grassland, and forest) in the Ansai, Taihe, and Changdu counties, respectively. The simulation was repeated 10,000 times for each ground feature in each study area. The simulation generated a total of 90,000 range values. These ranges were statistically analyzed and displayed in a customized 3D plot (Figure 11). There is a chart for each area, at the bottom of which is a conventional box plot, upon it is the corresponding probability distribution of range. The values for the maximum, minimum, median, first, and third quartiles are shown.

Figure 11a shows the result for Ansai County. Overall, the probability distributions of the three general ground features first increase and peak at some point before they decrease. The curves are not completely separated; instead, they intersect, indicating that different ground features may

have similar sizes. To be specific, the ranges for grassland vary between 10 and 65 m, while most of the ranges are ~35 m. The ranges for forest are between 18 and 72 m, while the majority is close to 40 m. The ranges for farmland are between 48 and 78 m; most of them are ~64 m. The 99% confidence intervals for forest, grassland, and farmland in Ansai County are (44.8, 45.3), (38.4, 39.0), and (63.3, 63.6), respectively.

(**a**)

(**b**)

(**c**)
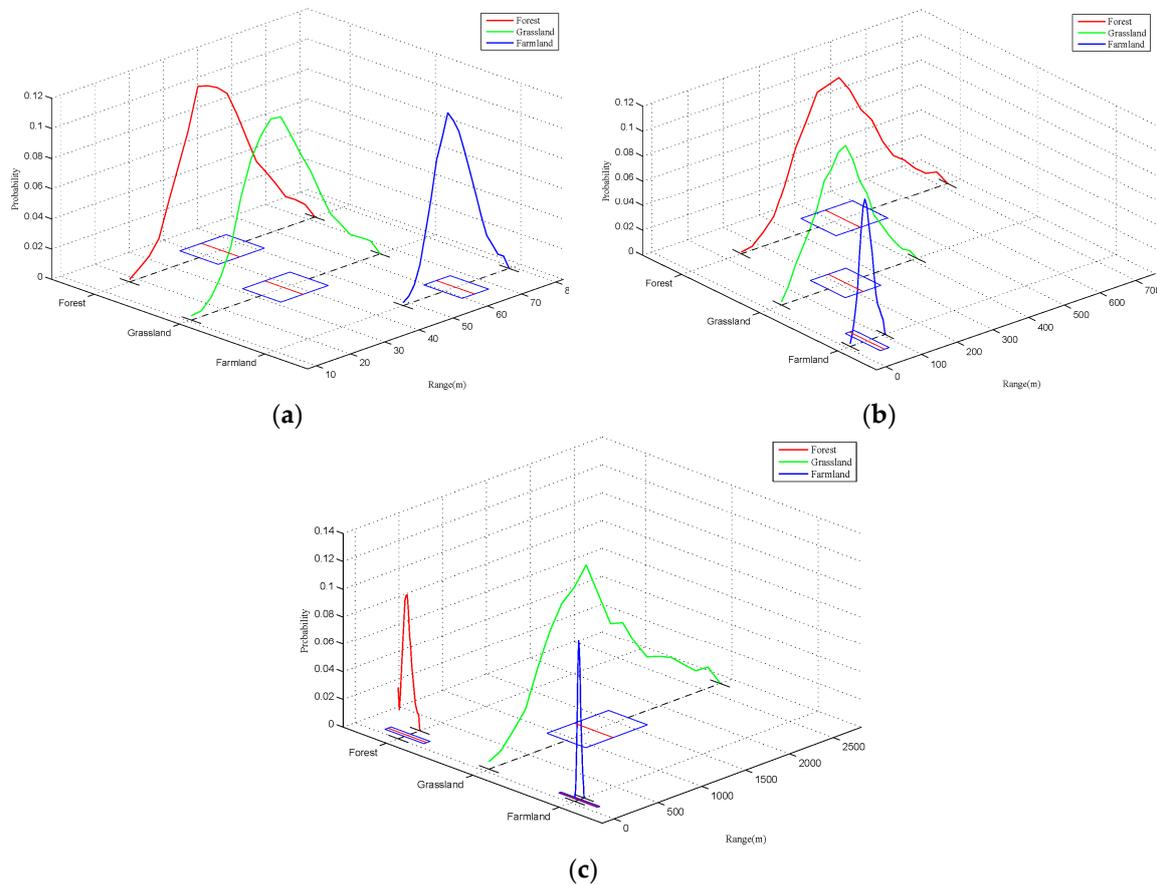
Figure 11. Probability distributions for the ranges for: Ansai County (**a**); Taihe County (**b**); and Changdu County (**c**).

Figure 11b shows the ground feature sizes for Taihe County. The probability distributions of the three ground features are normally distributed, and the curves are intersecting. The ranges for forest are between 130 and 720 m; the ranges for the other two ground features are relatively smaller. The 99% confidence intervals for forest, grassland, and farmland in Taihe County are (430.1, 435.6), (214.7, 218.5), and (57.1, 58.1), respectively.

Figure 11c shows the range distribution in Changdu County. The grassland extent of the range is extraordinarily large, from near 0 to 2700 m, while the range for farmland is between 50 and 150 m, and forest is between 0 and 250 m. The 99% confidence intervals for the forest, grassland, and farmland in Changdu County are (106.6, 109.4), (1293.0, 1323.7), and (97.1, 98.1), respectively.

5. Discussion

5.1. Comparison of Semi-Variogram-Based Methods

1. The SAM is a development of the inefficient manual process of choosing sub-area images. By controlling the amount and size of subimages, the estimation accuracy of the ranges can be guaranteed and the data size can be minimized. The advantage of SAM is that the estimation

is unbiased and reliable because all pixels of a subimage are used to construct the semivariogram. The drawback is that if a random point generated by the stratified random sampling method happens to be close to or on the boundary, the subimage derived from it will probably cross the boundary and contain pixels from other feature classes, which lead to an estimation bias.

2. The DAM deals with the disadvantages of SAM. In DAM, feature classes are completely separated from each other and semivariogram analysis is conducted. Thus, each subimage can be treated as simple scene. However, if the image size is overly large due to a large survey area, this method is limited by the computing and memory capacities of the computer and the number of pixels needs to be controlled. As a result, only a part of the pixels can be used to construct the semivariogram, which may lead to an estimation bias.

3. The MCS is a DAM upgrade. Basically, it increases the number of simulations, from DAM's mere one simulation to many simulations. This helps to overcome the drawback of DAM and avoid the potential biased result. Meanwhile, it utilizes the advantage of DAM with respect to eliminating or minimizing the influence of neighboring feature classes. Furthermore, this method does not depend as much on the capacity of the computer. By adjusting the sample size and amount of simulations, MCS can obtain a reliable estimation of the range. Due to its flexibility, the MCS approach is suitable for this study.

5.2. Impact of Parameters on the Semivariogram

The parameters (i.e., *GN*, *SS* and *MD*), model, and number of simulation have notable influences on the outputs. The preliminary comparison between the outputs before and after parameter optimization shows that the ranges of all three ground features have a lower variance after optimization. Therefore, parameter optimization is necessary. (1) For *GN*, a larger value is not recommended because it might incorporate too many details, which would consequently reduce the fitting quality. This study suggests that the *GN* should be controlled at <50. (2) *SS* is the actual number of pixels used for the analysis. Based on more samples, the output is supposed to be more reliable. This study reveals that there may exist a threshold for *SS*, after which, the accuracy would only vary slightly, such as farmland in Ansai County, grassland in Taihe County, and farmland in Changdu County. These thresholds could be used as optimized values for *SS*. However, in most cases, as restricted by the capacity of the computer, the sample size cannot be extended large enough in experiment to identify the threshold. In this scenario, the maximum sample size the computer can handle should be identified and applied to subsequent research. (3) The model selected affects the fitting accuracy. The general criterion for an ideal model is a relatively high R^2. In this study, the exponential model outperforms the spherical model. (4) The *MD* is the most difficult parameter to control. Theoretically, the *MD* should be large enough to reflect the true range. To obtain a qualified value for *MD*, two factors need to be considered, namely, the true range and to what extent the *MD* should exceed the true range. While the true range can be roughly estimated using initial tests, the second factor is the real problem. In this study, three times the average range derived in the parameter optimization session is used as the recommended value to make sure it is large enough to reflect the true range. This is a rough estimate and needs to be further studied.

The authors made those recommendations mainly based on the tremendous amount of calculations and the consequent relationship between parameters and fitting R^2. Even though some of the parameters, such as *GN*, cannot be clearly defined as optimized, the influence of those parameters on the accuracy of semivariogram has be well revealed, based on which, better decisions could be made on setting the parameter for semivariogram than before.

5.3. Differences in the Ground Feature Size in Different Regions

Figure 12 shows the average ranges for general ground features in each region.

Based on the land cover map of each region, the area for each ground feature was calculated; the percentage is shown in Figure 13.

Grassland is dominant in Ansai County, accounting for 64.7% of the total area. Forest accounts for 23.4% and farmland for only 0.6%, as shown in Figure 13a. Grassland comprises the largest area in this region, but its estimated range is the smallest among the three regions, only 45 m, indicating that the grassland in Ansai is fragmentized. This may be explained by severe soil loss in that region [20], which results in the grassland being cut into pieces. Farmland covers the smallest area but the largest estimated range in this region, reaching 63 m, which means that the farmland is more concentrated than the other two ground features in Ansai. The reason is mainly that most farmland has to be concentrated and distributed along the river for easier irrigation because of lacking water in this area [20].
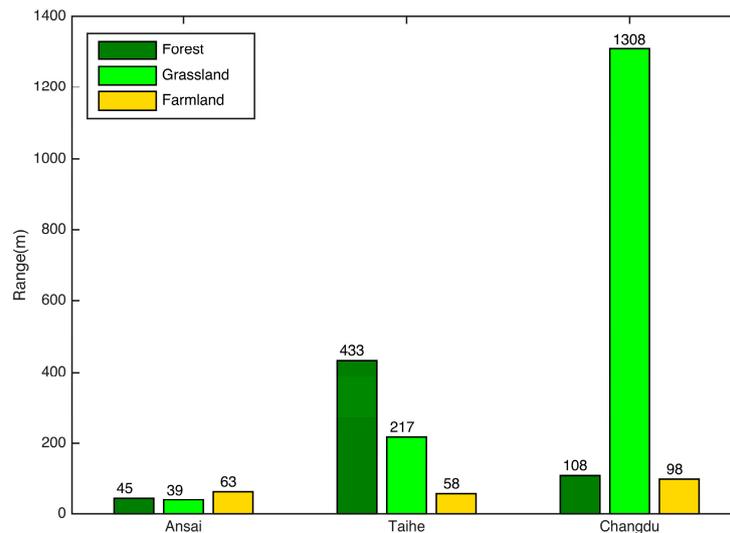
Figure 12. Appropriate scales for forest, grassland, and farmland at the study sites.
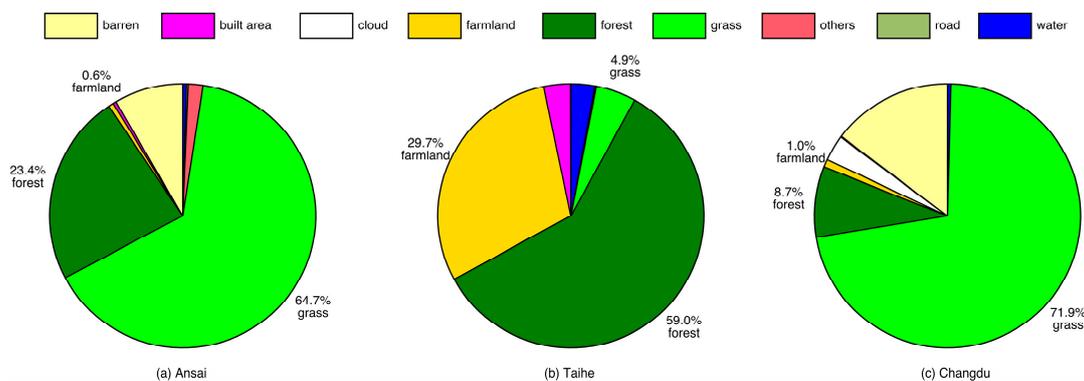
Figure 13. Proportion of the area of the ground features in: (**a**) Ansai; (**b**) Taihe; and (**c**) Changdu.

Forest is the primary ground feature in Taihe County, accounting for up to 59.0%. Grassland comprises only 4.9%, while farmland covers 29.7% of the total area, as shown in Figure 13b. The forest occupies the largest area in this region and the largest average range of 433 m among those regions, indicating that forest is widely and continuously distributed in Taihe County. This matches the real situation well. Taihe County is located in the southeast of China where there is abundant water and large forests [30]. The relatively small average range for farmland shows that the farmland is dispersed. East China has a large population, which leads to a high demand on food supply. Much of the land has been reclaimed for food production, which is why the farmland has a higher proportion than the other two counties. However, the farmland is separated by hills, which makes the farmland relatively small [31].

Changdu County is located in the Qinghai–Tibet Plateau with extensive land and a small population. The landscape there is simple. Grassland accounts for 71.9%, forest occupies 8.7%, and farmland contributes to only 1.0% (Figure 13c). The grassland area in Changdu County is large and continuous [32]; therefore, the estimated range is also large, reaching 1308 m. The ranges for forest and farmland are similar; both are ~100 m, which is close to the real situation.

The scale features of ground features acquired using the proposed method for the three representative counties of China could match well with the real local landscape characteristics, which indicates that the proposed method could achieve more than other relevant quantitative studies. In terms of study in other regions, Wen obtained the size of farmland, 98.5 m, for the Sanjiang Plain in Northeast China [33]. Yang reported that 30 m is an appropriate scale for depicting built area, while 90 m for water in Baoji City in the Loess Plateau [34]. However, these studies lack description and explanation regarding simulation accuracy, and only focus on one single region, i.e. no comparison was conducted between different regions, which limits its universal application. This study demonstrates the advantage of MCS on directly revealing the size characteristics of ground features in different geographical regions, and has the potential to receive wide popularity.

5.4. Ground Feature Size and Optimal Resolution

The ground feature size is mainly confined within two dimensions in this study and can be considered as an abstract representation of its real size. It is not real because the size of an object in the real world should be characterized based on at least two aspects, length and width, while the "size" calculated from the estimated range of the semivariogram is just a measure of the length. The optimal resolution of an image for the ground feature should be smaller than that size to retain information about that object. The finer the resolution is, the more detail is included in the imagery.

According to Woodcock, variation at a scale finer than the scale of regularization (resolution) cannot be detected and variations less than two to three times the scale of regularization (resolution) cannot be reliably characterized [7]. Therefore, the optimal resolution for a ground feature should be, at most, one-third the size of that feature, for the variation within it to be reliably detected.

Table 7 recommends the maximum applicable satellite imagery for survey of general ground features in those three counties, based on the maximum resolution that is one-third of average size of ground feature. For example, it could be inferred that the Sentinel-2A imagery with a resolution of 10 m could be used to survey all three areas, and Landsat 8 imagery (15 m) could meet the requirement of almost all areas, except the grassland in Ansai County, while CBERS-2 with a resolution of 20 m could be used to survey most of the areas. Apart from those listed in Table 7, many commercial satellite imageries provide a resolution less than 10 m that could also be leveraged, such as WorldView-4 (0.31 m), GeoEye-1 (0.46 m), Pleiades-1A (0.5 m), KOMPSAT-3A (0.55 m), QuickBird (0.65 m), Gaofen-2 (0.8 m), TripleSat (0.8 m), IKONOS (0.82 m), SPOT-6 (1.5 m), SPOT-7 (1.5 m), ALOS (2.5 m), CARTOSAT-1 (2.5 m), SPOT-5 (2.5–5 m), and Dove Satellite (3 m) [35].

Table 7. Applicable satellite imagery for general ground features in pilot areas.

	Ground Features	Average Size (m)	Maximum Resolution (m)	Applicable Satellite Imagery
Ansai County	Forest	45	15	e.g., Landsat 8 (15 m)
	Grassland	39	13	e.g., Sentinel-2A (10 m)
	Farmland	63	21	e.g., CBERS-2 (20 m)
Taihe County	Forest	433	144	e.g., CBERS-2 (20 m)
	Grassland	217	72	e.g., CBERS-2 (20 m)
	Farmland	58	19	e.g., Landsat 8 (15 m)
Changdu County	Forest	108	36	e.g., CBERS-2 (20 m)
	Grassland	1308	436	e.g., CBERS-2 (20 m)
	Farmland	98	32	e.g., CBERS-2 (20 m)

6. Conclusions

This study uses MCS to improve the performance of semivariograms by optimizing its parameters, for selecting remote sensing imagery with an appropriate resolution for natural resource surveys. The main contribution of this paper is revealing the relationship between semivariogram parameters and simulation accuracy, as well as the suggestions given on selecting appropriate values for parameters. Apart from that, the sizes of those common ground features in three counties of China are acquired, and satellite image for survey is recommended, which is more objective than the usual subjective way of selecting datasets by experience. The main findings include:

(1) The proposed MCS performs better than SAM and DAM in determining the ground feature size from large remote sensing imagery by providing more accurate and precise estimations.

(2) Optimizing the parameters for building and fitting semivariograms is necessary and strongly recommended by the authors because they greatly affect the performance of semivariograms in terms of precision and accuracy. The optimized parameters could improve the precision of the estimations. A larger *SS* and larger number of simulations is recommended, while the *GN* should not be too large.

(3) The size and corresponding optimal resolution for specific ground feature could be acquired using the proposed MCS and the optimized parameters. In this study, three counties in China were used as case study, and appropriate remote sensing image was recommended.

(4) The sizes of the same ground feature in different regions vary; the sizes of different ground features in the same region also vary. That means the result is specific to a certain area. Nevertheless, the proposed method can also be used for choosing the remote sensing imagery resolution in other places.

There still exists some limitation which should be studied further. First, the influence of the classification accuracy on the estimation has not been studied in this research because the overall accuracies for the three regions are almost identical. The relation between the classification and estimation accuracies is still unclear and needs to be further investigated. In addition, high performance computing (HPC) could be introduced in the future study, and cost–benefit analysis would be incorporated to assess the cost and benefit of this method, to further improve its efficiency.

Author Contributions: Juanle Wang was responsible for the research design and analysis and designed and reviewed the manuscript. Junxiang Zhu drafted the manuscript and was responsible for data preparation, image interpretation, experiment, and analysis. Xuehua Han was responsible for the data processing and archiving. All authors contributed to editing and reviewing the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Saturno, W.; Sever, T.L.; Irwin, D.E.; Howell, B.F.; Garrison, T.G. Putting us on the map: Remote sensing investigation of the ancient maya landscape. In *Remote Sensing in Archaeology*; Springer: Berlin, Germany, 2007; pp. 137–160.

2. Shen, F.; Kuang, D. Remote sensing investigation and survey of qinghai lake in the past 25 years. *J. Lake Sci.* **2002**, *15*, 289–296.

3. Sankey, J.B.; Draut, A.E. Gully annealing by aeolian sediment: Field and remote-sensing investigation of aeolian-hillslope-fluvial interactions, colorado river corridor, arizona, USA. *Geomorphology* **2014**, *220*, 68–80. [CrossRef]

4.　Kim, J.; Grunwald, S.; Osborne, T.; Robbins, R.; Yamataki, H.; Rivero, R. Spatial resolution effects of remote sensing images on digital soil models in aquatic ecosystems. In *Computing Ethics: A Multicultural Approach*; CRC Press: Boca Raton, FL, USA, 2016; p. 121.

5.　De Klerk, H.M.; Gilbertson, J.; Lück-Vogel, M.; Kemp, J.; Munch, Z. Using remote sensing in support of environmental management: A framework for selecting products, algorithms and methods. *J. Environ. Manag.* **2016**, *182*, 564–573. [CrossRef] [PubMed]

6.　Qiu, B.; Sui, Y.; Chen, C.; Tu, X. Identification of optimal spatial resolution with local variance, semivariogram and wavelet method. In Proceedings of the Accuracy 2010 Symposium, Leicester, UK, 20–23 July 2010; pp. 353–356.

7.　Woodcock, C.E.; Strahler, A.H.; Jupp, D.L.B. The use of variograms in remote sensing: I. Scene models and simulated images. *Remote Sens. Environ.* **1988**, *25*, 323–348. [CrossRef]

8.　Ming, D.; Ci, T.; Cai, H.; Li, L.; Qiao, C.; Du, J. Semivariogram-based spatial bandwidth selection for remote sensing image segmentation with mean-shift algorithm. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 813–817. [CrossRef]

9.　Treitz, P.; Howarth, P. High spatial resolution remote sensing data for forest ecosystem classification: An examination of spatial scale. *Remote Sens. Environ.* **2000**, *72*, 268–289. [CrossRef]

10.　Woodcock, C.E.; Strahler, A.H.; Jupp, D.L.B. The use of variograms in remote sensing: II. Real digital images. *Remote Sens. Environ.* **1988**, *25*, 349–379. [CrossRef]

11.　Kayitakire, F.; Hamel, C.; Defourny, P. Retrieving forest structure variables based on image texture analysis and ikonos-2 imagery. *Remote Sens. Environ.* **2006**, *102*, 390–401. [CrossRef]

12.　Levesque, J.; King, D.J. Airborne digital camera image semivariance for evaluation of forest structural damage at an acid mine site. *Remote Sens. Environ.* **1999**, *68*, 112–124. [CrossRef]

13.　Wolter, P.T.; Townsend, P.A.; Sturtevant, B.R. Estimation of forest structural parameters using 5 and 10 meter spot-5 satellite data. *Remote Sens. Environ.* **2009**, *113*, 2019–2036. [CrossRef]

14.　Song, C. Estimating tree crown size with spatial information of high resolution optical remotely sensed imagery. *Int. J. Remote Sens.* **2007**, *28*, 3305–3322. [CrossRef]

15.　Liu, Y.; Chen, D.; Zheng, C. Estimation on forest stand crown of the picea schrenkiana based on high spatial resolution imagery. *J. Xinjiang Agric. Univ.* **2013**, *36*, 153–157.

16.　Guardiola-Albert, C.; Ballesteros-Cánovas, J.; Stoffel, M.; Díez-Herrero, A. How to improve dendrogeomorphic sampling: Variogram analyses of wood density using X-ray computed tomography. *Tree Ring Res.* **2015**, *71*, 25–36. [CrossRef]

17.　Zhu, J.; Wang, J. An appropriate scale extraction method for complicated scene model based on semivariogram analysis. *Geogr. Geo-Inf. Sci.* **2015**, *31*, 33–37.

18.　Webster, R.; Oliver, M.A. How large a sample is needed to estimate the regional variogram adequately? In *Geostatistics Tróia'92*; Springer: Berlin, Germany, 1993; pp. 155–166.

19.　Han, X.; Yang, G.; Liu, Z. Vegetation changes of loess plateau and ansai county and the responding characteristics of climate. *Acta Bot. Boreali-Occident. Sin.* **2009**, *29*, 593–600.

20.　Zhou, D.; Zhao, S.; Zhu, C. Impacts of the sloping land conversion program on the land use/cover change in the loess plateau: A case study in Ansai county of Shaanxi province, China. *J. Nat. Resour.* **2011**, *26*, 1866–1878.

21.　Maxwell, A.; Strager, M.; Warner, T.; Zégre, N.; Yuill, C. Comparison of naip orthophotography and rapideye satellite imagery for mapping of mining and mine reclamation. *GISci. Remote Sens.* **2014**, *51*, 301–320. [CrossRef]

22.　Atkinson, P.M.; Curran, P.J. Choosing an appropriate spatial resolution for remote sensing investigations. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 1345–1351.

23.　Jupp, D.L.B.; Strahler, A.H.; Woodcock, C.E. Autocorrelation and regularization in digital images. I. Basic theory. *IEEE Trans. Geosci. Remote Sens.* **1988**, *26*, 463–473. [CrossRef]

24.　Miranda, F.; Fonseca, L.; Carr, J. Semivariogram textural classification of jers-1 (fuyo-1) sar data obtained over a flooded area of the amazon rainforest. *Int. J. Remote Sens.* **1998**, *19*, 549–556. [CrossRef]

25.　Atkinson, P.M.; Curran, P.J. Defining an optimal size of support for remote sensing investigations. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 768–776. [CrossRef]

26.　Baatz, M. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. *Angew. Geogr. Informationsverarbeitung* **2000**, *12*, 12–23.

27. Robert, C.; Casella, G. *Monte Carlo Statistical Methods*; Springer Science & Business Media: Berlin, Germany, 2013.

28. Choy, R.; Edelman, A. Parallel matlab: Doing it right. *Proc. IEEE* **2005**, *93*, 331–341. [CrossRef]

29. Sharma, G.; Martin, J. Matlab®: A language for parallel computing. *Int. J. Parallel Program.* **2009**, *37*, 3–36. [CrossRef]

30. Wu, Z.; Dai, E.; Ge, Q.; Xi, W.; Wang, X. Modelling the integrated effects of land use and climate change scenarios on forest ecosystem aboveground biomass, a case study in taihe county of china. *J. Geogr. Sci.* **2017**, *27*, 205–222. [CrossRef]

31. Wang, P. Volving Law of Land Use and Driving Force in Jiangxi Province. Master's Thesis, Jiangxi Agriculture University, Nanchang, China, 2013.

32. He, Z.-W.; Yang, Z.; Luo, J. Analysis on information entropy of land use structure in changdu area of Tibet. *Resour. Environ. Yangtze Basin* **2007**, *16*, 192–195.

33. Wen, Z.; Zhang, S.; Bai, J.; Ding, C.; Zhang, C. Agricultural landscape spatial heterogeneity analysis and optimal scale selection: An example applied to sanjiang plain. *Acta Geogr. Sin.* **2012**, *67*, 346–356.

34. Yang, X.; Wang, X.; Hu, T.; Cao, Y. Appropriate scale in typical objects feature extraction. *J. Mt. Sci.* **2012**, *30*, 607–615.

35. Satellite Imaging Corporation. Satellite Sensors. Available online: https://www.satimagingcorp.com/satellite-sensors/ (accessed on 28 December 2017).