


## Article

# Improving the Accuracy of Remote Sensing Land Cover Classification by GEO-ECO Zoning Coupled with Geostatistical Simulation

Ling Zhu <sup>1,\*</sup> , Jing Li <sup>1</sup>, Yixuan La <sup>1</sup> and Tao Jia <sup>2</sup>

<sup>1</sup> School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture, Beijing 100044, China; 2108521519014@stu.bucea.edu.cn (J.L.); 2108521517042@stu.bucea.edu.cn (Y.L.)

<sup>2</sup> School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430070, China; tao.jia@whu.edu.cn

\* Correspondence: zhuling@bucea.edu.cn

**Abstract:** Land cover products obtained from remote sensing image classification inevitably contain a large number of false classification or uncertain pixels because of spectral confusion, image resolution limitation, and ground object complexity. The confusion matrix used to evaluate the classification accuracy cannot reflect the spatial variation. The information provided to users of land cover products is incomplete and uncertain. In this study, a method is presented to evaluate and improve the accuracy of land cover classification products by coupling Geo-Eco zoning and Markov chain geoscience statistical simulation. Validation points collected from various sources are used in the model calculation and accuracy verification of results. The pre-classified image that needs to be improved and Geo-Eco zoning attribute data are used as auxiliary data for co-simulation. Results show that the accuracy of Globeland30 data can be improved by more than 10% by coupling Geo-Eco zoning and Markov chain geostatistical simulation.

**Keywords:** Geo-Eco zoning; geostatistical simulation; Co-MCSS; accuracy improvement



**Citation:** Zhu, L.; Li, J.; La, Y.; Jia, T. Improving the Accuracy of Remote Sensing Land Cover Classification by GEO-ECO Zoning Coupled with Geostatistical Simulation. *Appl. Sci.* **2021**, *11*, 553. <https://doi.org/10.3390/app11020553>

Received: 10 December 2020

Accepted: 5 January 2021

Published: 8 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Land cover is a concept emerging with the development of remote sensing technology, and remote sensing is the only effective means for large-scale land cover mapping [1]. Since the 1980s, the international scientific community has been highly concerned about remote sensing mapping of global land cover [2–4]. A variety of 1000, 300, 30, and 10 m resolution global, regional, or national land cover products have been developed [5–12]. Classical supervised and unsupervised classification technologies are commonly used, but the accuracy is not high, which is about 60% to 70% [13,14]. A deep learning algorithm has not been applied to the classification of land cover products on a large scale [15]. The second-level land cover product classification is more difficult to obtain reliable results via automatic classification. Classification accuracy varies spatially, and spatial variation in accuracy should be quantitatively evaluated. Evaluating and improving the accuracy of land cover product classification (hereinafter referred to as pre-classified image) obtained by conventional classification algorithm and quantifying the uncertainty of classification are necessary to study global change, geographical condition census, social environment planning, and ecological resource management.

Two kinds of methods are used to improve the accuracy of remote sensing classification. One of the methods is the application of Geoscience Knowledge Rules. Since the 1980s, experts and scholars have introduced expert systems and knowledge engineering to solve remote sensing classification problems [16–20]. In previous land cover mapping, auxiliary data, such as digital elevation model (DEM) data, ecological region data, vegetation data of countries or regions [5], global mangrove atlas, global human settlements, and regional data of global urban coverage (<http://maps.elie.ucl.ac.be/CCI/viewer/index.php>),

MODIS (moderate-resolution imaging spectroradiometer, NASA, USA) NDVI (normalized different vegetation index) data, global geographic information data, global DEM data, various thematic data, and online high-resolution images are employed to improve product accuracy [21]. However, the expert knowledge and reference auxiliary data used in land cover mapping are sporadic and unsystematic, and no global integral system is available to manage expert knowledge and auxiliary data for reuse.

Eco geographical regions are areas where ecosystems (and the type, quality, and quantity of environmental resources) are generally similar. They are relatively large units of land containing a distinct assemblage of natural communities and species, with boundaries that approximate the original extent of natural communities prior to major land-use change [22]. Eco geographical regions can be applied as a frame to construct a global expert knowledge base and assist in remote sensing image classification. Zhu et al. [23] adopted the “world terrestrial ecological region” established by the World Wildlife Fund for natural protection as the basic framework of the global Geo-Eco zoning knowledge base [22]. An object-oriented method is used to construct a rule base and help identify spurious changes in remote sensing image detection. Five kinds of attributes, namely, DEM, slope, NDVI, temperature, and moisture, of each Geo-eco zone, are collected to identify spurious change. The accuracy of change detection is improved to a certain extent [23].

Another method to improve the accuracy of remote sensing classification is the application of geostatistics. Geostatistics is used to analyze and predict the values associated with spatial or spatiotemporal phenomena. Geostatistics has been applied in remote sensing since the 1980s but not popular. Meer [24] made a detailed review. Among them, studies on using geostatistics to improve the accuracy of land cover classification products are limited. Bruin [25] used the sequential indicator simulation algorithm and the cokriging method to predict the area range of olive trees by using the classified image as the soft data and the interpretation sample from the aerial image as the hard index. Tsendbazar et al. [14] used indicator Kriging to estimate the spatial variation in the accuracy of the source land cover products and use integration methods, which consider the local accuracy of each source product to obtain Africa’s land cover products. Carvalho et al. [26] improved the accuracy of land cover classification by using a direct sequential co-simulation algorithm combined with field observations and remote sensing images classified by maximum likelihood classification. Tang et al. [27] utilized a multipoint geostatistic calculation method to take the maximum likelihood classification result as the training image, the maximum likelihood probability as the soft condition, and the training sample as the hard condition and improve the accuracy of classification products. Li and Zhang [28] proposed to use the Markov chain random field (MCRF) method to evaluate the uncertainty of the land cover classification. They carried out conditional random simulation at the sample pixels interpreted by experts referring to high-resolution images and other auxiliary information. Unlabeled pixels are regarded as uncertain regions, and the class is obtained by using the MCRF algorithm [29]. However, the disadvantage of this method is that a large number of expert-interpreted samples are needed to ensure the reliability of results; consequently, the method becomes time-consuming and laborious. Li et al. [30] further improved the algorithm by introducing a pre-classified image to the simulation process. The co-simulation algorithm of the MCRF (Co-MCSS) is used to improve the classification accuracy by integrating sample pixels and pre-classified images.

In this study, the boundary of geostatistics is confined in each Geo-Eco zone, and the Co-MCSS method [30] is used to quantify the classification local accuracy by integrating the pre-classified image, the verification points, and the attribute data based on the Geo-Eco zoning. Three contributions are made in terms of determining, evaluating, and improving the accuracy of the land cover product.

- (1) Taking Geo-Eco zoning as the area of geostatistics. Generally the area of geostatistics is based on the selected image area, which often contains boundary effect, that is, the fact that a class may have statistically biased smaller frequencies of transitions if it has a higher chance of occurring at boundaries of the study area because boundary polygons

are incomplete and have no transition to other classes beyond the boundary [28]. Generally, the boundary of Geo-Eco zoning does not cross two different types of land cover to avoid the boundary effect.

- (2) In previous research, the large number of sample pixels were interpreted by experts to obtain a transiogram model [28], but this process is time-consuming and laborious. In this study, reliable verification data published by some websites or institutions related to land cover research are reused to reduce the work on visual interpretation.
- (3) The Co-MCSS method not only combines remote sensing pre-classified images but also takes various attributes of Geo-Eco zoning (such as DEM, slope, temperature, and humidity) as auxiliary data to participate in the simulation and calculation of cross field transition probability. With the combination of additional attribute information, the simulation algorithm becomes more robust.

## 2. Method

### 2.1. Geo-Eco Zoning Rule Base

In this study, the global eco-regions established by the World Wild Fund for natural protection were selected. With these eco-regions, the terrestrial world was subdivided into 14 biomes and 8 biogeographic realms and nested 867 Geo-Eco zones [22]. The framework of Zhu et al. [23] Geo-Eco zoning rule base was adopted to collect and sort out the natural attribute data sets, including the DEM, slope, NDVI, temperature, and moisture. This kind of knowledge was highly correlated with land cover. Each attribute was expressed as a layer, which was used as the auxiliary data of co-simulation with the pre-classified image.

### 2.2. Markov Chain Co-Simulation

#### 2.2.1. Process

The technical process is shown in Figure 1. In this study, a set of transiogram models of each Geo-Eco zone were estimated by using the sample pixels. The cross field transition probability from the sample data to the auxiliary data set was estimated. The Co-MCSS algorithm was used to generate the optimal prediction map and occurrence probability map.

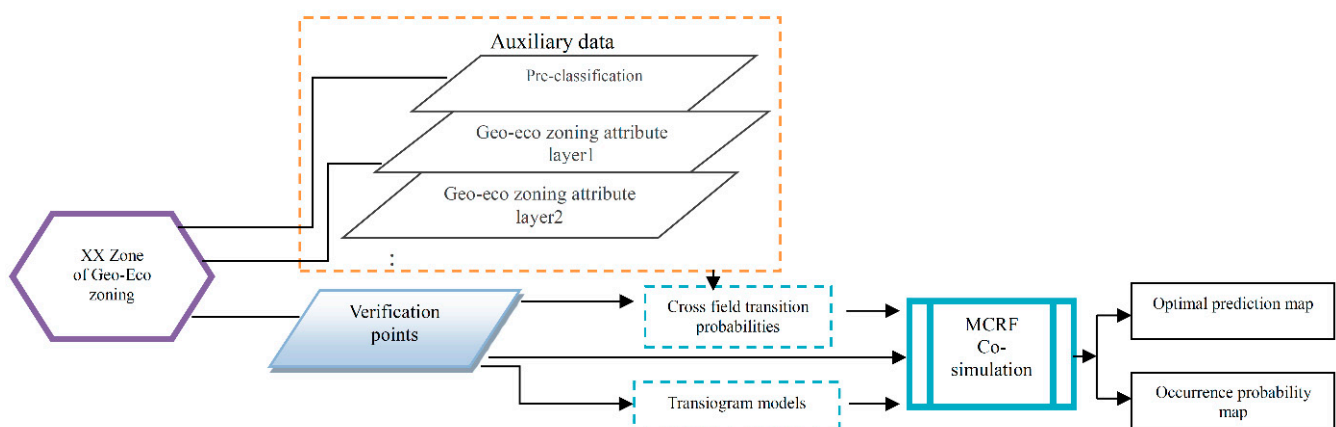


Figure 1. Flow chart.

The main steps were as follows:

- (1) The land cover verification points from networks were collected from the relevant websites (citations are provided below). If the amount of verification points from networks was not enough for the transiogram estimation, then visual interpretation of sample points was added as a supplement to form the sample data set.
- (2) Traditional methods, such as the maximum likelihood method, were used to obtain pre-classified images. The natural attribute layers (such as DEM, slope, and aspect) of Geo-Eco zoning constituted the auxiliary data set for co-simulation.

- (3) A set of transiograms were estimated by using the sample data set.
- (4) The cross field transition probabilities were estimated by the sample and auxiliary data set.
- (5) Co-MCSS algorithm was carried out under the condition of sample data and auxiliary data.
- (6) The optimal prediction map and occurrence probability map were obtained.

In the above steps, except that some of the sample points needed to be interpreted manually, other works were coded by Matlab and realized automatically. These steps will be further presented in detail in the Results part.

### 2.2.2. Transiogram

Traditionally, variograms are commonly used to describe the correlation between variables. However, variograms cannot describe directional asymmetry when land cover categories appear sequentially, thus they cannot effectively express the parallel relationship between categories. Markov chain conditional simulation provides a general framework for a non-linear non-kriging geostatistical approach. It needs a powerful index of spatial heterogeneity to realize multi-type simulation. Li [31] proposed the concept of transiogram, which is a one-dimensional transition probability function (conditional probability between two points) model at distance  $h$ :

$$p_{ij}(h) = Pr[z(x+h) = j | z(x) = i] \quad (1)$$

where  $x$  is a specific location, and  $P_{ij}(h)$  is the transition probability of a random variable  $z$  from class  $i$  to class  $j$ . When  $h$  increases gradually,  $P_{ij}(h)$  forms a graph, i.e., a transiogram.  $P_{ii}(h)$  denotes an auto-transiogram, which describes the dependency of a type itself; a cross-transiogram describes the dependency between types (including cross correlation, parallel relationship, and directional asymmetry).

Transiograms can estimate multistep transition probabilities from sparse point samples for two-dimensional Markov chain simulations, which involves the rich spatial heterogeneity characteristics of land cover types. The transition probabilities of different spatial steps (or lag) can form a one-dimensional continuous transition probability graph. The role of transiograms on Markov chain geostatistics is similar to that of variograms for Kriging geostatistics.

Transiograms can only rely on a large number of reliable and soundly distributed sample points, which need an amount of visual interpretation. When the number of sample points is insufficient, transiograms may show false fluctuations and cannot convey reliable information [31].

In recent years, some organizations participating in land cover mapping around the world have released their validation data to the public, thereby providing a reference for subsequent research. Geoweb-based tagging system also enables users to tag geographical information for land cover data acquisition [32]. We collected these reference sample data, and they were mainly from the following websites: (1) GOF-C-GOLD Land Cover Project Office in coordination with reference data producers ([http://www.gofcgold.wur.nl/sites/gofcgold\\_refdataportal.php](http://www.gofcgold.wur.nl/sites/gofcgold_refdataportal.php)). The GOF-C-GOLD includes the consolidated GLC 2000 reference (GLC200ref) [33], the consolidated GlobCover 2005 reference (GlobCover2005ref) [34], the System for Terrestrial Ecosystem Parameterization (STEP) reference [35], the Visible Infrared Imaging Radiometer Suite Surface-Type reference [36] and the GLCNMO 2008 datasets [37]; (2) Geo-Wiki crowdsourced data (<https://www.geo-wiki.org/>); (3) DCP volunteers (<http://www.confluence.org>); (4) other research institutions, such as the global validation sample set developed by Tsinghua University [38]; [http://data.ess.tsinghua.edu.cn/data/temp/GlobalLandCoverValidationSampleSet\\_v1.xlsx](http://data.ess.tsinghua.edu.cn/data/temp/GlobalLandCoverValidationSampleSet_v1.xlsx)); (5) Flickr photo-sharing website ([www.flickr.com](http://www.flickr.com)); and (6) LACO-Wiki open access online portal for land cover validation (<http://www.laco-wiki.net>). Existing reference sample datasets built for calibrating and validating global land cover maps have high reliability and can be reused. However, the density and distribution of these verification points was not balanced because of the

scattered collection and may not meet the density requirements of model estimation. Nevertheless, the workload of the visual interpretation of sample points can be appropriately reduced. As more and more organizations release their validation data sets, reusing these data to reduce visual interpretation will become feasible.

### 2.2.3. Cross Field Transition Probability Matrix

The transition probability from a primary variable to an auxiliary variable can be called a cross field transition probability matrix. Each auxiliary variable is considered independent of each other. The cross field transfer probability is calculated as follows:

$$\hat{q}_{ik} = \frac{f_{ik}}{\sum_{j=1}^n f_{ij}} \tag{2}$$

where  $f_{ik}$  is the frequency of class  $i$  being transformed into class  $k$  in the space of auxiliary variables, and  $n$  is the number of auxiliary variables.

### 2.2.4. MCRF Co-Simulation (Co-MCSS)

Similar to the cokriging model in classical geostatistics, Co-MCSS can be built by extending Markov chain random fields. According to Bayesian reasoning theory, Co-MCSS can be regarded as the Bayesian updating of the Markov chain random field model based on new evidence on auxiliary data.

If  $X$  is the target classification variable to be estimated, and  $E$  is the auxiliary data set, then the Bayesian inference formula can be written as follows:

$$P(X|E) = \frac{P(E|X)P(X)}{P(E)} = \frac{P(E|X)P(X)}{\sum_X P(E|X)P(X)} = C^{-1}P(E|X)P(X) \tag{3}$$

where  $C = \sum_X P(E|X)P(X)$  is a constant. Based on Bayes principle, the simplified Equation (3) is extended to the Co-MCSS model, and the auxiliary data are combined via co-simulation. The contribution of auxiliary variables can be realized in different ways. In this study, auxiliary variable data are considered the nearest neighbor of an unknown location in other variable spaces. The Co-MCSS model with  $k$  auxiliary variables can be expressed as follows:

$$p\left[i_0(u_0)|i_1(u_1), \dots, i_m(u_m); r_0^{(1)}(u_0^{(1)}), \dots, r_0^{(k)}(u_0^{(k)})\right] = \frac{p_{i_1 i_0}(h_{10}) \prod_{g=2}^m p_{i_0 i_g}(h_{0g}) \prod_{l=1}^k q_{i_0 r_0^{(l)}}}{\sum_{f_0=1}^n \left[ p_{i_1 f_0}(h_{10}) \prod_{g=2}^m p_{f_0 i_g}(h_{0g}) \prod_{l=1}^k q_{f_0 r_0^{(l)}} \right]} \tag{4}$$

where  $r_0^{(k)}$  is the class of the  $k$ th auxiliary variable at  $u_0^{(k)}$ , and  $q_{i_0 r_0^{(k)}}$  is the transition field probability matrix between  $i_0$  in the space of the primary variable at position  $u_0$  and  $r_0$  in the space of the auxiliary variable. In this study, two auxiliary variables are selected: (1) Pre-classified image and (2) DEM. Therefore,  $k = 2$  in Equation (4).

In practical applications, considering many nearest known neighbors in different directions is unnecessary and difficult. For the pixel data of a remote sensing image, the four main directions are easily considered. Therefore, Equation (4) of the Co-MCSS considering two auxiliary variables and four main directions is as follows:

$$p\left[i_0(u_0)|i_1(u_1), \dots, i_m(u_m); r_0^{(1)}(u_0^{(1)}), \dots, r_0^{(4)}(u_0^{(k_2)})\right] = \frac{p_{i_1 i_0}(h_{10}) \prod_{g=2}^4 p_{i_0 i_g}(h_{0g}) \prod_{l=1}^2 q_{i_0 r_0^{(l)}}}{\sum_{f_0=1}^n \left[ p_{i_1 f_0}(h_{10}) \prod_{g=2}^4 p_{f_0 i_g}(h_{0g}) \prod_{l=1}^2 q_{f_0 r_0^{(l)}} \right]} \tag{5}$$

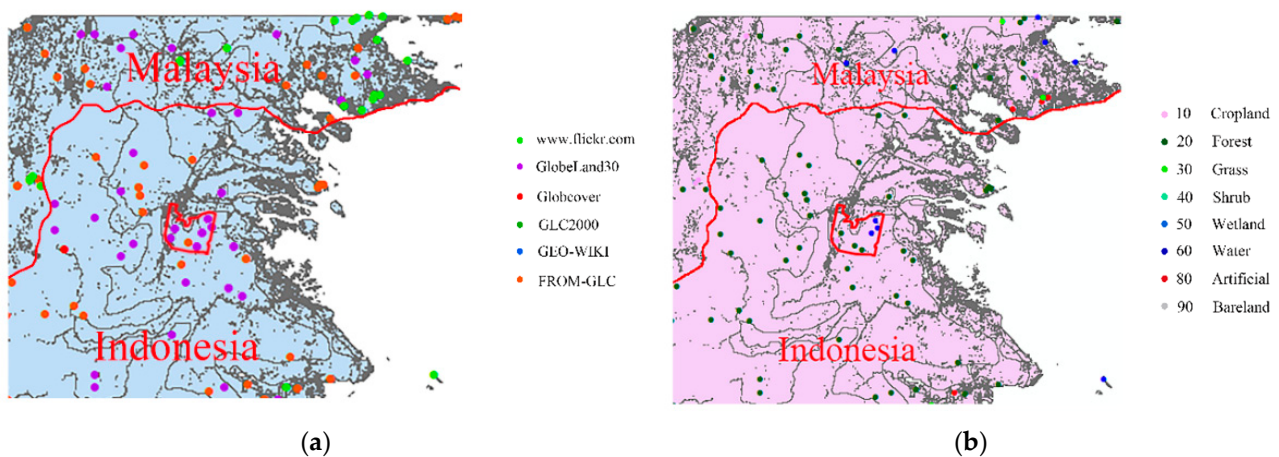
Equation (5) is the Co-MCSS model used in this study.

### 3. Results

#### 3.1. Study Area

The selected study area was located in Indonesia, Southeast Asia. Indonesia is extremely rich in biological species, and the forest coverage rate has reached 67.8% (according to Globeland30 land cover product 2010, [www.globeland30.com](http://www.globeland30.com)).

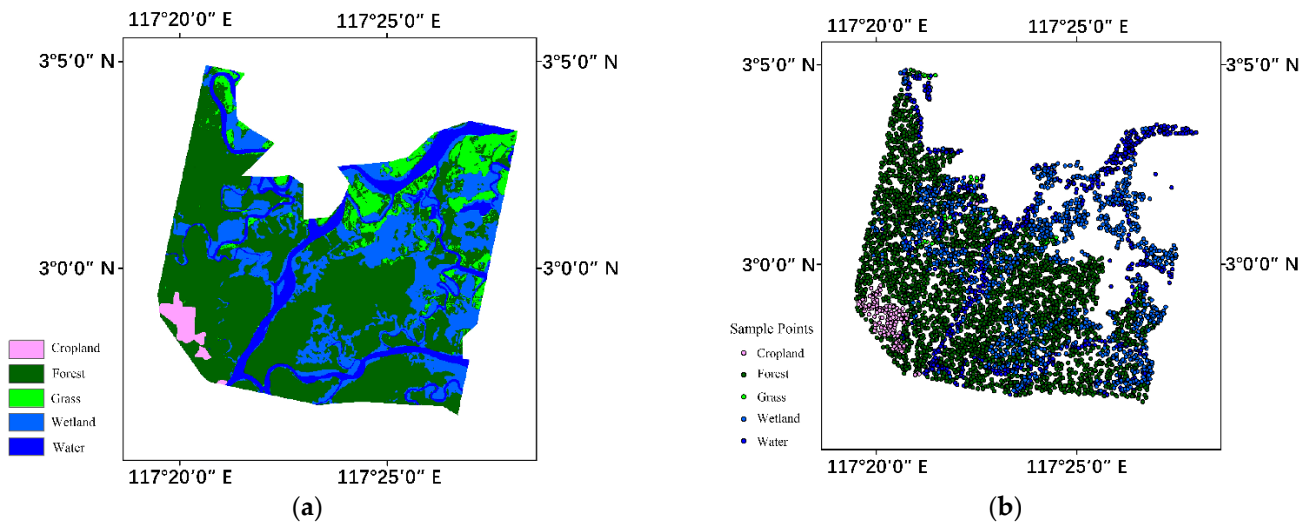
The collected verification points whose resolution and time met the requirements include reference data of GLC 2000 (GLC2000ref), reference data of Globcover 2005 (Globcover2005ref), reference data of system for Territorial Ecosystem parameter (STEP), reference data of visible infrared imaging radiometer Suite (VIIRS), GLCNMO 2008 reference data, Tsinghua University Global validation sample set, and Globeland30 verification points provided by NGCC (National Geomatics Center of China). The collected verification points were filtered because of the inconsistency semantics and resolution between the verification points and the pre-classified image. Figure 2a shows the distribution of reserved verification points according to the source of verification points; Figure 2b illustrates the distribution of reserved verification points in the study area according to the land cover class of Globeland30-2015.



**Figure 2.** Verification points of the study area (a) Verification points according to the origin; (b) Verification points according to the classification of Globeland 30-2015.

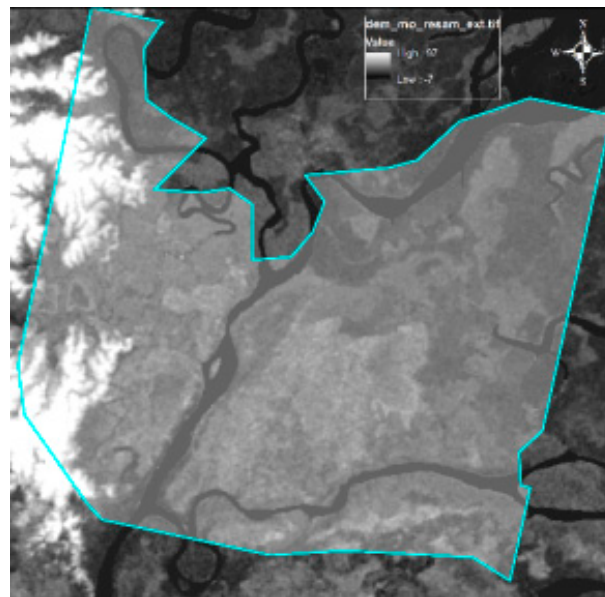
Because of the limitation of the operation speed and data volume of the algorithm, a small block, as shown in the red frame of Figure 2a,b, was selected as the demonstration area based on the richness of land cover type. The demonstration area was located in the IM0104 of the Geo-Eco zone [22]. Only one zone was involved, thus the subsequent processing was carried out uniformly within this zone. Figure 3a presents the land cover classification map of Globeland30-2015, which was used as the pre-classified image in this study. There were five kinds of land cover in this area according to GlobeLand30 classification schema [11]. Globeland30-2015 had a resolution of 30 m and was based on the classification of remote sensing images within year 2015. The demonstration area had more than 70,000 pixels. The number of available verification points was insufficient to meet the requirements of model calculation. Therefore, some sample pixels were interpreted as supplements to form a sample data set as shown in Figure 3b.

Using the “create random points” function in ArcGIS, the random points were generated and uniformly distributed of the density of 1 point per 4000 m<sup>2</sup>. If there was no network collected verification point at or near the position of the random point, then visually interpreted on Google Earth high resolution images was needed. A total of 15,826 sample points were divided into two parts. Of these points, 7913 were used for model simulation (about 10% of the pixels in the demonstration area), and the remaining points were utilized for the final accuracy verification.



**Figure 3.** Demonstration area of (a) Globeland30-2015 and (b) interpreted sample points distribution map.

In addition to the sample point data, Geo-Eco zoning related attribute data should also be collected. For the demonstration area, the final data source was GDEM 30 m resolution digital elevation data, considering the availability of data, the independence of data, resolution, and year close to 2015 (Figure 4).



**Figure 4.** IM0104 Geo-eco zone digital evaluation model (DEM).

### 3.2. Transfer Probability Diagram

Figure 5 is the result of the self- and cross-transiograms of each land cover type. The type is indicated by a code, and the specific meaning is shown in Table 1.

For the cultivated land, the curve was more tortuous as shown in Figure 5a. The self-transfer (cultivated land with itself) and cross-transfer (with the other four categories of forest, grassland, wetland, and water body) probabilities of the forest were the highest, wetland were lower than forest, water body were lower than wetland, and grassland were basically 0. This finding was due to the scattered distribution of grassland. The transfer probability map results were consistent with the results of expert interpretation.

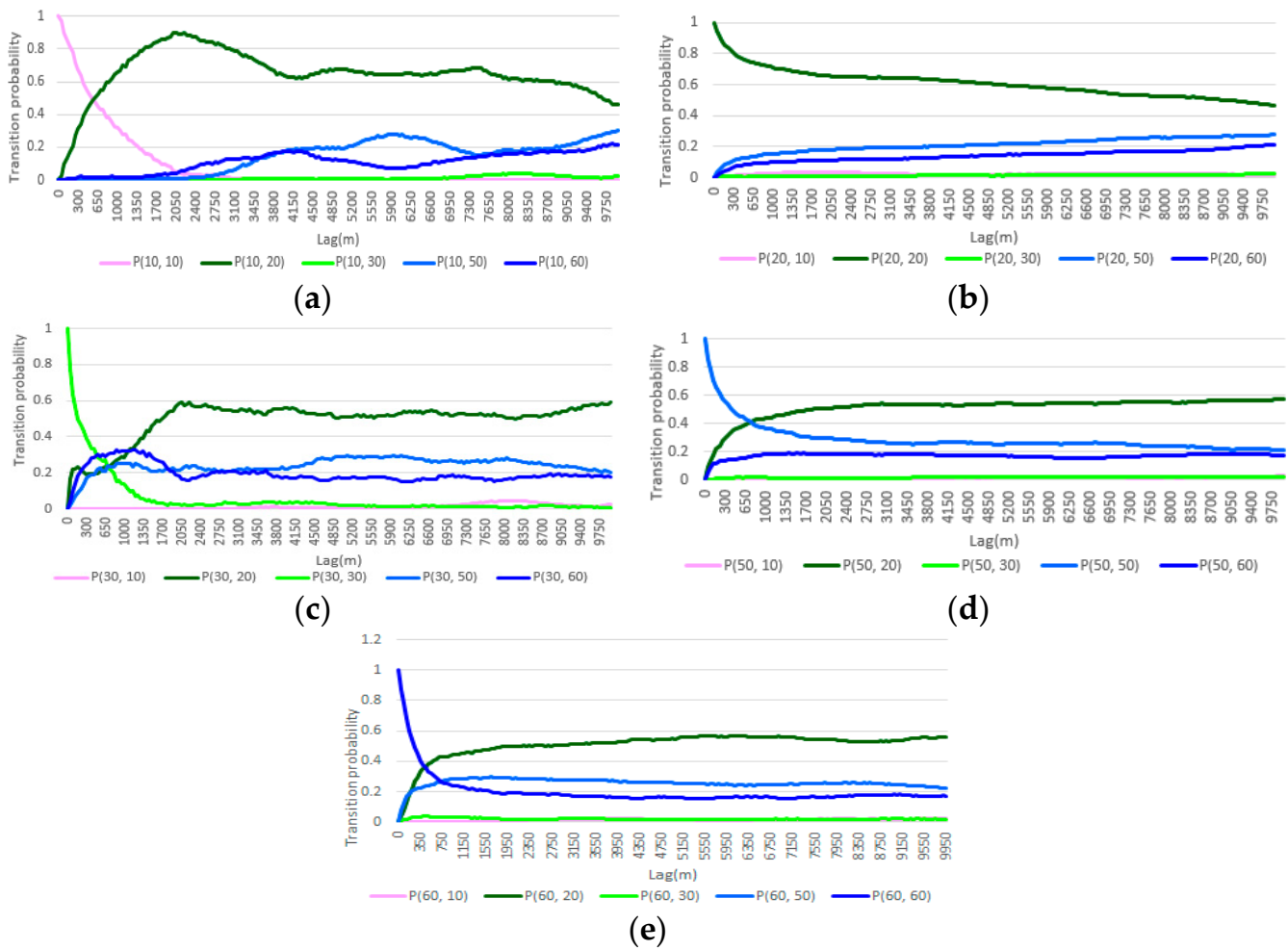


Figure 5. Transiogram models (a) cultivated land-others; (b) forest-others; (c) grassland-others; (d) wetland-others; (e) water-others.

Table 1. Land cover types and codes.

Color	Code	Land Cover Type
	10	Cropland
	20	Forest
	30	Grass
	40	Shrub
	50	Wetland
	60	Water
	80	Artificial
	90	Bareland

The forest had the largest proportion, thus its self-transfer probability was higher than those of the four other categories as shown in Figure 5b. The transfer probabilities of the wetland and the water body were lower than the autocorrelation transfer probability of the forest. The conversion relationship from the cultivated land to the grassland and the forest almost did not exist, because the area of these two categories was relatively small. This finding was consistent with the results of expert interpretation.

In Figure 5c, the distribution of the grassland was scattered, thus the curve of the transiogram was very tortuous and not as smooth as those of the other categories. Generally, the cross-transfer probability of the grassland and the cultivated land was almost 0, and the short-distance cross-transfer probabilities of the forest land, the water body, and the



wetland were similar. The long-distance cross-transfer probability of the forest was higher than that of the other categories because the proportion of the forest was the largest.

As shown in Figure 5d, the curve of the wetland was relatively smooth. As the distance increased, the probability of the conversion of the wetland to the forest gradually increased. The probability of the conversion of the wetland to the water body was lower than that of the forest. Almost no conversion relationship was observed in the cultivated land and the grassland. In addition, the wetlands and water bodies always appeared together, and this finding was consistent with our common sense that wetlands always form around water bodies.

The shapes of the autocorrelation transfer probability curve and the cross-correlation transfer probability curve of the water body were similar to those of the wetland by comparing Figure 5d,e. As distance increased, the probability of the conversion of the water body to the forest increased gradually. The probability of the conversion of the water body to the wetland was less than that of the forest. Almost no conversion relationship was observed in the cultivated land and the grassland. In addition, wetlands and water bodies always appeared together, wetlands always formed around water bodies, thus their transfer probability curves were very similar.

In general, each category had the highest probability of conversion to forest, because the proportion of forest was relatively large. In addition, the forest, wetland, and water sample points were rich, the transfer probability curve was stable, and the credibility was high. Although the curve of cultivated land was not smooth enough, the transiogram model was reliable because of its concentrated distribution. There were few and scattered samples in grassland, and the transfer probability map was not smooth enough, and the confidence level was low.

### 3.3. Simulation Results

For comparison, two kinds of simulation methods were adopted: One was using MCRF, which involves sample pixels only, and the other applies the Co-MCSS method combined with auxiliary data (including pre-classified image, i.e., Globeland30-2015 data and DEM layer). The simulation results include the occurrence probability map of each category and the optimal prediction map.

#### 3.3.1. Simulation Results of MCRF

The occurrence probability map of the simulation results obtained from the sample pixels is shown in Figure 6. The deeper the hue is, the higher the likelihood of a category to be correct, and vice versa. The black part indicates that the confidence level of the type is nearly 1, i.e., it can be considered almost 100% certain. The white part implies that the confidence level of the type is approximately 0, which can be considered impossible. The color of the forest is the deepest, and the range is the widest, suggesting that the reliability of the forest land is highest. Although the scope of the cultivated land is small, the color is deep. As such, the simulation of the cultivated land is reliable. For grassland, few dark places and many gray areas are found, implying that the simulation of the grassland is insufficiently reliable. The light tone part can be considered the warning area for further survey. Wetlands and water bodies are always accompanied by each other, thus their hues are almost complementary, and the range of gray colors is large. Therefore, wetlands and water bodies are easily misclassified.

The optimal prediction map of the simulation results is shown in Figure 7, which was the assembly of the most likely type of each location. The coincidence degree of the forest, cultivated land, and part of grassland, wetland, and water bodies was higher than that of the original classification products (Globeland30-2015), and the other places with deviation were likely to be the ones with false classifications.

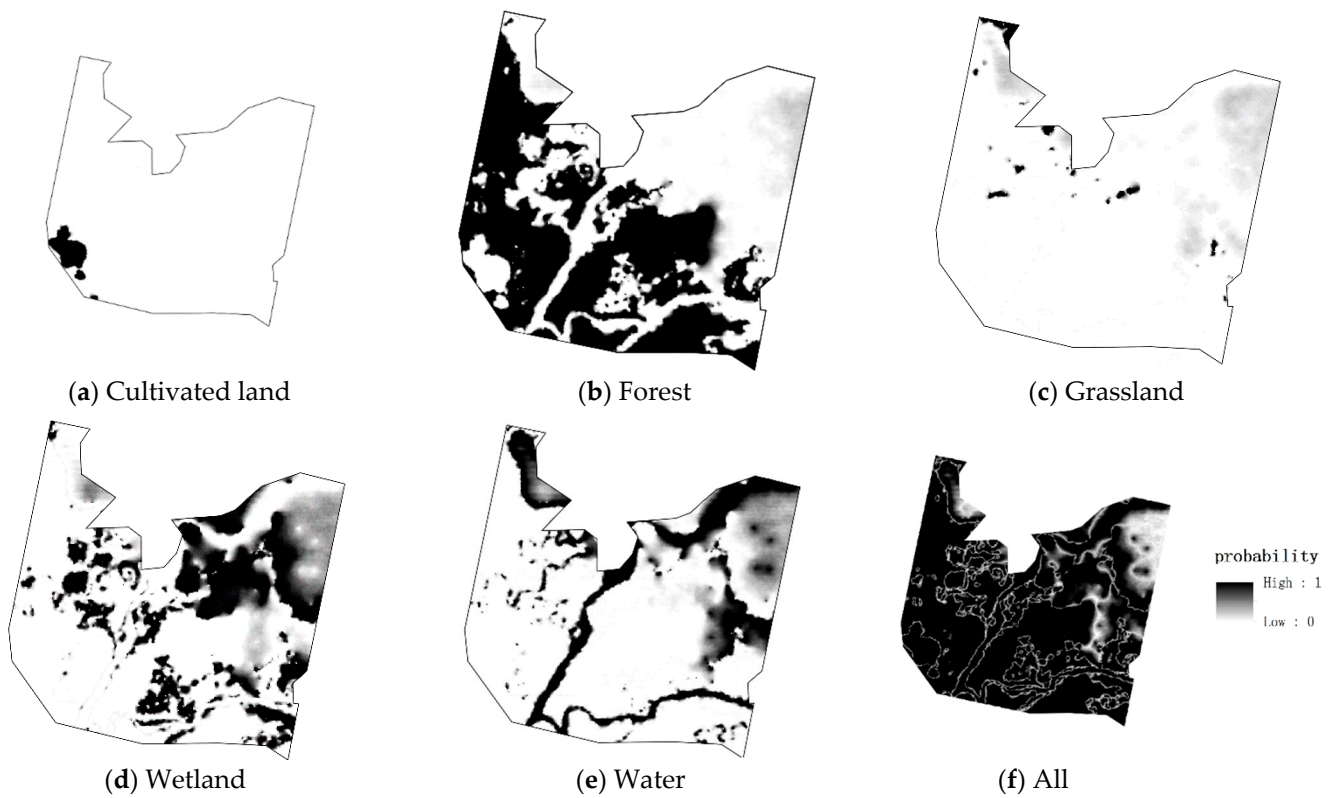


Figure 6. Occurrence probability map results of using sample points.

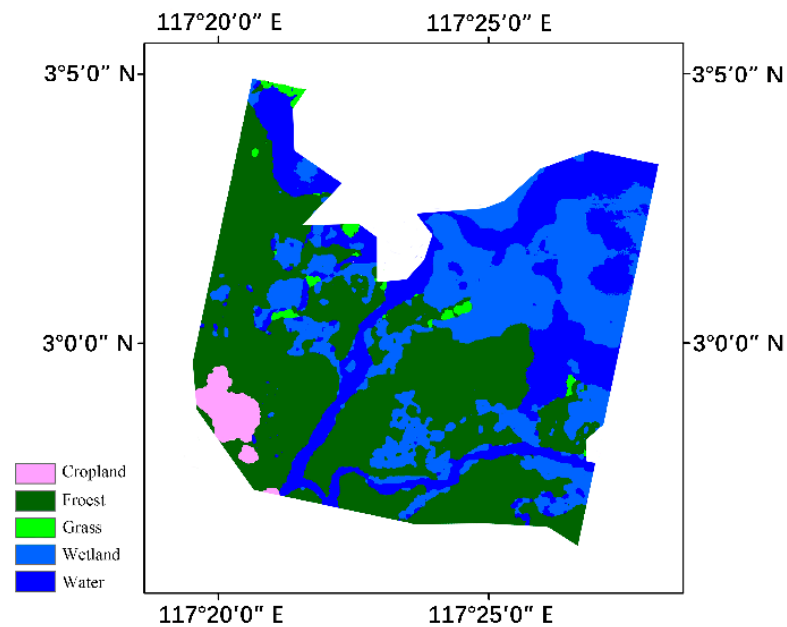
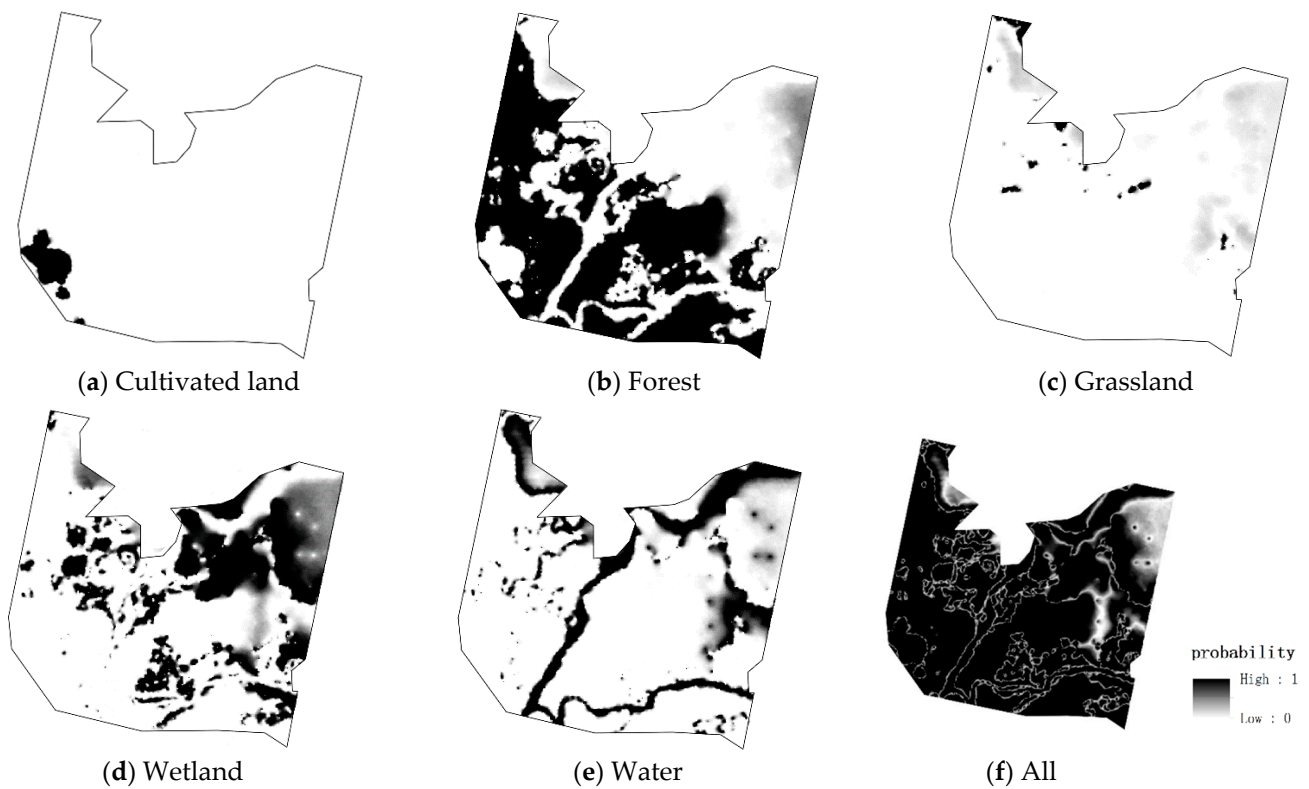


Figure 7. Optimal prediction map (no auxiliary data added).

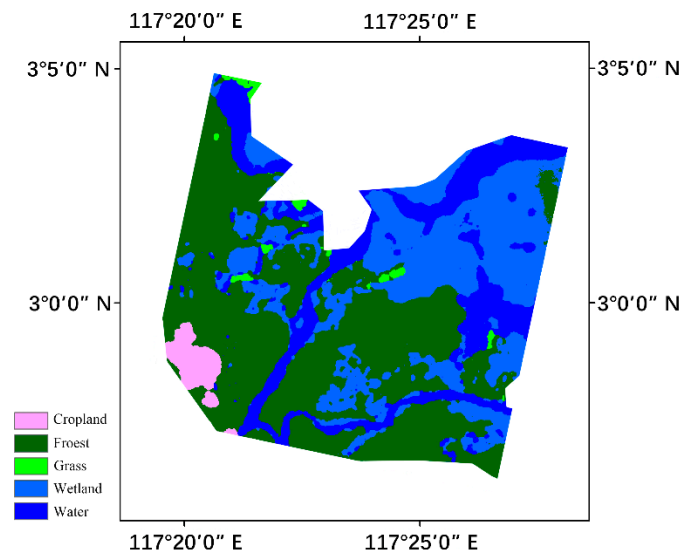
### 3.3.2. Simulation Results of Co-MCRF with Auxiliary Data

After the auxiliary data were added, the occurrence probability map of the co-simulation results is shown in Figure 8. The simulation results of the forest, cultivated land, grassland, wetland, and water body was similar to MCRF. The distribution of the wetland, water body, and grassland was slightly different from that involving the sample data only.



**Figure 8.** Occurrence probability map results of auxiliary data added.

The simulation result of the optimal prediction map with the auxiliary data are shown in Figure 9. The simulation results revealed that the cultivated land, forest, and the original classification products had slight variations, the distribution of the cultivated land was relatively concentrated, and the coverage of forest distribution was very wide. The main changes were found in the grassland, the wetland, and the water body. Wetlands and water bodies were always together, thus they were easily misclassified. For grassland, the simulation result considerably differed from the original pre-classified image because of two possible reasons. One was that more pixels were wrongly classified into grassland, and the other is that the transiograms obtained were insufficiently reliable.



**Figure 9.** Optimal prediction map (with the added auxiliary data).

### 3.4. Accuracy Analysis

Accuracy verification aims to compare the optimal prediction map of two different simulation methods with the set of sample pixels used for accuracy verification. Type matching means that the simulation results are correct, and the type inconsistency indicates that the simulation results are wrong. The results are shown in Table 2.

**Table 2.** Accuracy assessment.

Simulation Method	Number of Matching	Proportion
Pre-classified image (GlobeLand30-2015)	5962	75.34%
Simulation results of MCRF	6451	81.52%
Simulation results (with auxiliary data)	6843	86.48%

The overall accuracy of Globeland30 products was high [21], but accuracy was spatially varied in Southeast Asia in part or some land cover types. The proportion of matching the pre-classified image with the sample pixels was 75.34%. The accuracy of the MCRF simulation results was 81.52%, which was 6.18% higher than that of the GlobeLand30-2015 products. The accuracy of matching the co-simulation results with the verification points was 86.48%, which was 11.14% higher than that of the GlobeLand30 products. The simulation results also revealed where the accuracy was relatively low. The area with a low accuracy can be used as the warning area of false classification, which provides a reference for subsequent product improvement.

## 4. Conclusions

In this study, a method is proposed to improve the accuracy of land cover classification products by coupling Geo-Eco zoning and Markov chain geostatistical simulation. This method can be used to evaluate the spatial accuracy variation of land cover classification products and improve the drawbacks of the overall accuracy evaluation of the general confusion matrix method. In this study, the verification points collected from the network are reused. The Geo-Eco zoning attribute data and pre-classified images are set as auxiliary data for co-simulation. The simulation results are more reliable if the simulation area is limited in the Geo-eco zone. The local accuracy of the pre-classified image can be quantified and improved. Therefore, the coupling of Geo-Eco zoning and Markov chain geostatistical simulation can enhance the accuracy of Globeland30 data by more than 10%.

However, some deficiencies should be improved in the future.

- (1) In this study, the image size that can be processed is limited because of the high operation cost of the algorithm. The experimental area only contains one Geo-eco zone. In future studies, the algorithm should be optimized, and a GPU-parallel acceleration method should be used to increase the amount of data that can be processed and make the algorithm more practical.
- (2) Existing data are limited, thus only the attribute data related to elevation are tested, and the impact of other types of auxiliary data are yet to be described. The role of Geo-Eco zoning in geostatistical simulation should be further explored. For example, geoscience knowledge on Geo-Eco zoning can be applied and combined with verification points to generate reasonable transiograms and further reduce the number of sample pixels required by the algorithm.
- (3) The case study only tested in one site with a high density of the sample points. The method needs to be verified with a wider scope and more example sites in the future.

**Author Contributions:** Conceptualization, L.Z.; Data curation, J.L. and Y.L.; Funding acquisition, L.Z.; Investigation, J.L.; Methodology, L.Z. and J.L.; Resources, T.J.; Software, Y.L.; Validation, T.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by The National Key Research and Development Program of China under Grant (NO.2016YFB0501404) and state key laboratory of remote sensing science open fund projects (OFSLRSS201927).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** The authors gratefully acknowledge the support of NGCC (National Geomatics Center of China).

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

## References

- Chen, J.; Chen, J.; Liao, A. *Remote Sensing Mapping of Global Land Cover*; Sci. Press: Beijing, China, 2016. (In Chinese)
- Townshend, J.; Justice, C.; Li, W.; Gurney, C.; McManus, J. Global land cover classification by remote sensing: Present capabilities and future possibilities. *Remote Sens. Environ.* **1991**, *35*, 243–255. [\[CrossRef\]](#)
- Defries, R.S.; Townshend, J.R.G. Global land cover characterization from satellite data: From research to operational implementation? *Glob. Ecol. Biogeogr.* **1999**, *8*, 367–379. [\[CrossRef\]](#)
- Verburg, P.H.; Neumann, K.; Nol, L. Challenges in using land use and land cover data for global change studies. *Glob. Chang. Biol.* **2011**, *17*, 974–989. [\[CrossRef\]](#)
- Loveland, T.R.; Reed, B.C.; Brown, J.F.; Ohlen, D.O.; Zhu, Z.; Yang, L.; Merchant, J.W. Development of a global land cover characteristics database and IGBP discover from 1 km AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1303–1330. [\[CrossRef\]](#)
- Hansen, M.C.; Defries, R.S.; Townshend, J.R.G.; Sohlberg, R.A. Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sens.* **2000**, *21*, 1331–1364. [\[CrossRef\]](#)
- Bartholomé, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977. [\[CrossRef\]](#)
- Tateishi, R.; Uriyangqai, B.; Al-Bilbisi, H.; Ghar, M.A.; Tsend-Ayush, J.; Kobayashi, T.; Kasimu, A.; Hoan, N.T.; Shalaby, A.; Alsaadeh, B. Production of global land cover data—GLCNMO. *Int. J. Digit. Earth* **2011**, *4*, 22–49. [\[CrossRef\]](#)
- Arino, O.; Perez, J.R.; Kalogirou, V.; Defourny, P.; Achard, F. *GlobCover 2009*; Esa Living Planet Symposium: Bergen, Norway; 2010.
- Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [\[CrossRef\]](#)
- Chen, J.; Cao, X.; Peng, S.; Ren, H. Analysis and Applications of GlobeLand30: A Review. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 230. [\[CrossRef\]](#)
- Gong, P. Mapping essential urban land use categories in china (EULUC-China): Preliminary results for 2018. *Sci. Bull.* **2020**, *65*, 182–187. [\[CrossRef\]](#)
- Giri, C.P. *Remote Sensing of Land Use and Land Cover: Principles and Applications*; CRC Press: Boca Raton, FL, USA, 2012; pp. 254–255.
- Tsendbazar, N.E.; De Bruin, S.; Fritz, S.; Herold, M. Spatial Accuracy Assessment and Integration of Global Land Cover Datasets. *Remote Sens.* **2015**, *7*, 15804–15821. [\[CrossRef\]](#)
- Liu, T.; Chen, X.; Dong, Q.; Cao, X.; Chen, J. Research on the Application of Deep Learning in GlobeLand30-2010 Product Classification Accuracy Optimization. *Remote Sens. Technol. Appl.* **2019**, *34*, 3–11. (In Chinese)
- Makoto, N.; Takashi, M. *A Structural Analysis of Complex Aerial Photographs*; Springer: New York, NY, USA, 1980.
- Civco, D. Knowledge-Based Land Use and Land Cover Mapping. In Proceedings of the Annual Convention of American Society for Photogrammetry and Remote Sensing, Baltimore, MD, USA, 16 March 1989; Volume 3, pp. 276–291.
- Dobson, M.C.; Pierce, L.E. Knowledge-based land-cover classification using ERS-1/JERS-1SAR composites. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 83–99. [\[CrossRef\]](#)
- Wentz, E.A.; Nelson, D.; Rahman, A.; Stefanov, W.L.; Roy, S.S. Expert system classification of urban land use/cover for Delhi, India. *Int. J. Remote Sens.* **2008**, *29*, 4405–4427. [\[CrossRef\]](#)
- Phiri, D.; Morgenroth, J. Remote sensing Review Developments in Landsat Land Cover Classification Methods: A Review. *Remote Sens.* **2017**, *9*, 967.
- Chen, J.; Chen, J.; Gong, P.; Liao, A.; He, C. High resolution remote sensing mapping of global land cover. *Geomat. World* **2011**, *9*, 12–14. (In Chinese)
- Olson, D.M.; Dinerstein, E.; Wikramanayake, E.D.; Burgess, N.D.; Powell, G.W.N.; Underwood, E.C.; D’amico, J.A.; Itoua, I.; Strand, H.E.; Morrison, J.C. Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience* **2001**, *51*, 933–938. [\[CrossRef\]](#)
- Zhu, L.; Sun, Y.; Shi, R.; La, Y.; Peng, S. Exploiting Cosegmentation and Geo-Eco Zoning for Land Cover Product Updating. *Photogramm. Eng. Remote Sens.* **2019**, *85*, 597–611. [\[CrossRef\]](#)

24. Van der Meer, F. Remote-sensing image analysis and Geostatistics. *Int. J. Remote Sens.* **2012**, *33*, 5644–5676. [[CrossRef](#)]
25. De Bruin, S. Predicting the areal extent of land-cover types using classified imagery and geostatistics. *Remote Sens. Environ.* **2000**, *74*, 387–396. [[CrossRef](#)]
26. Carvalho, J.; Soares, A.; Bio, A. Improving satellite images classification using remote and ground data integration by means of stochastic simulation. *Int. J. Remote Sens.* **2006**, *27*, 3375–3386. [[CrossRef](#)]
27. Tang, Y.; Atkinson, P.M.; Wardrop, N.A.; Zhang, J. Multiple-point geostatistical simulation for post-processing a remotely sensed land cover classification. *Spat. Stat.* **2013**, *5*, 69–84. [[CrossRef](#)]
28. Li, W. Transiogram: A spatial relationship measure for categorical data. *Int. J. Geogr. Inf. Sci.* **2006**, *20*, 693–699.
29. Li, W.; Zhang, C. A Markov chain geostatistical framework for land-cover classification with uncertainty assessment based on expert-interpreted pixels from remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2983–2992. [[CrossRef](#)]
30. Li, W.D.; Zhang, C.; Willig, M.R.; Dey, D.K.; Wang, G.; You, L. Bayesian Markov Chain Random Field Cosimulation for Improving Land Cover Classification Accuracy. *Math. Geosci.* **2015**, *47*, 123–148. [[CrossRef](#)]
31. Li, W. Transiograms for Characterizing Spatial Variability of Soil Classes. *Soil Sci. Soc. Am. J.* **2007**, *71*, 881–893. [[CrossRef](#)]
32. Xing, H.; Chen, J.; Zhou, X. A Geoweb-Based Tagging System for Borderlands Data Acquisition. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 1530–1548. [[CrossRef](#)]
33. Mayaux, P.; Eva, H.; Gallego, J.; Strahler, A.H.; Herold, M.; Agrawal, S.; Naumov, S.; De Miranda, E.E.; Di Bella, C.M.; Ordoyne, C. Validation of the global land cover 2000 map. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1728–1739. [[CrossRef](#)]
34. Bontemps, S.; Defourny, P.; Bogaert, E.V.; Santoro, M.; Kirches, G.; Wevers, J.; Boettcher, M.; Brockmann, C.; Lamarche, C. GlobCover—Products Description and Validation Report. Available online: <https://epic.awi.de/> (accessed on 22 November 2020).
35. Friedl, M.A.; McIver, D.K.; Hodges, J.C.F.; Zhang, X.Y.; Muchoney, D.; Strahler, A.H.; Woodcock, C.E.; Gopal, S.; Schneider, A.; Cooper, A. Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens. Environ.* **2002**, *83*, 287–302. [[CrossRef](#)]
36. Friedl, M.A.; Muchoney, D.; McIver, D.; Gao, F.; Hodges, J.C.F.; Strahler, A.H. Characterization of North American land cover from NOAA-AVHRR data using the EOS MODIS land cover classification algorithm. *Geophys. Res. Lett.* **2000**, *27*, 977–980. [[CrossRef](#)]
37. Olofsson, P.; Stehman, S.V.; Woodcock, C.E.; Sulla-Menashe, D.; Sibley, A.M.; Newell, J.D.; Friedl, M.A.; Herold, M. A global land-cover validation data set, part I: Fundamental design principles. *Int. J. Remote Sens.* **2012**, *33*, 5768–5788. [[CrossRef](#)]
38. Zhao, Y.Y.; Gong, P.; Yu, L.; Wang, X. Towards a common validation sample set for global land-cover mapping. *Int. J. Remote Sens.* **2014**, *35*, 4795–4814. [[CrossRef](#)]