

Letter

# Mapping Fragmented Impervious Surface Areas Overlooked by Global Land-Cover Products in the Liping County, Guizhou Province, China

Jing Zhao  and Narumasa Tsutsumida \* 

Graduate School of Global Environmental Studies, Kyoto University, Kyoto 606-8501, Japan; zhao.jing.75a@st.kyoto-u.ac.jp

\* Correspondence: tsutsumida.narumasa.3c@kyoto-u.ac.jp; Tel.: +81-75-753-6368

Received: 11 April 2020; Accepted: 8 May 2020; Published: 11 May 2020



**Abstract:** Imperviousness is an important indicator for monitoring urbanization and environmental changes, and is evaluated widely in urban areas, but not in rural areas. An accurate impervious surface area (ISA) map in rural areas is essential to achieve environmental conservation and sustainable rural development. Global land-cover products such as MODIS MCD12Q1, ESA CCI-LC, and Global Urban Land are common resources for environmental practitioners to collect land-cover information including ISAs. However, global products tend to focus on large ISA agglomerations and may not identify fragmented ISA extents in less populated regions. Land-use planners and practitioners have to map ISAs if it is difficult to obtain such spatially explicit information from local governments. A common and consistent approach for rural ISA mapping is yet to be established. A case study of the Liping County, a typical rural region in southwest China, was undertaken with the objectives of assessing the global land-cover products in the context of rural ISA mapping and proposing a simple and feasible approach for the mapping. This approach was developed using Landsat 8 imagery and by applying a random forests classifier. An appropriate number of training samples were distributed to towns or villages across all townships in the study area for classification. The results demonstrate that the global land-cover products identified major ISA agglomerations, specifically at the county seat; however, other fragmented ISAs over the study area were overlooked. In contrast, the map created using the developed approach inferred ISAs across all townships with an overall accuracy of 91%. A large amount of training samples together with geographic information of towns or villages is the key suggestion to identify and map ISAs in rural areas.

**Keywords:** impervious surface area mapping; global land products; rural areas; Landsat; random forests classifier; Liping County; southwest China

## 1. Introduction

The development of impervious surface areas (ISAs) has profound impacts on both urban and rural regions [1,2]. Impervious surfaces (ISs) usually refer to artificial surfaces such as pavements and roofs, through which a fluid cannot pass [3,4]. Imperviousness was originally defined as an artificial condition that affects water quality and urban runoff [5,6]. More recently, imperviousness was recognized as a key indicator to monitor urban extent [7,8] and understand human–environment interactions [4,9]. The expansion of ISAs has the potential to substantially impact local hydrology and water environment [10], accelerate loss of biodiversity [11], contribute to climate change [12], and threaten food security [13]. Spatially explicit information on ISAs is essential to support watershed management, urban/rural planning, and land use regulation [14,15]. Monitoring ISAs is essential for all practitioners, policymakers, and researchers who are concerned with anthropogenic impacts on land [4].

Many previous studies have attempted to depict the extents of ISAs in urban areas in both local and global scales [7,8,16]. However, only a few studies have focused on ISAs in less populated areas. ISAs are a consequence and critical indicator of human activities exploiting the natural environment. Therefore, an accurate estimation of ISAs is considered essential in urban areas for understanding urbanization processes and in rural areas for understanding human–environmental interactions. Previous studies have reported environmental impacts caused by the expansion of ISAs in rural areas in China [17,18]. Liu et al. [19] reported a simultaneous outmigration-led depopulation and rural residential land expansion in rural areas in China, causing an excessive anthropogenic impact on the environment. Kahn [20] reported on the impacts of suburbanization, that is, suburban area development and expansion caused by the population shift from the city center to suburbs in many developed countries. These studies suggest that such an expansion of ISAs in rural areas is not directly related to the population size, but is triggered by unobserved socio-geographic factors [19,21]. ISAs in rural areas tend to expand at the cost of abundant ecological services such as food, water, and wood provisioning [22,23]. Thus, accurate ISA mapping in such areas is an important starting point to evaluate human–environment interactions towards sustainable rural land management.

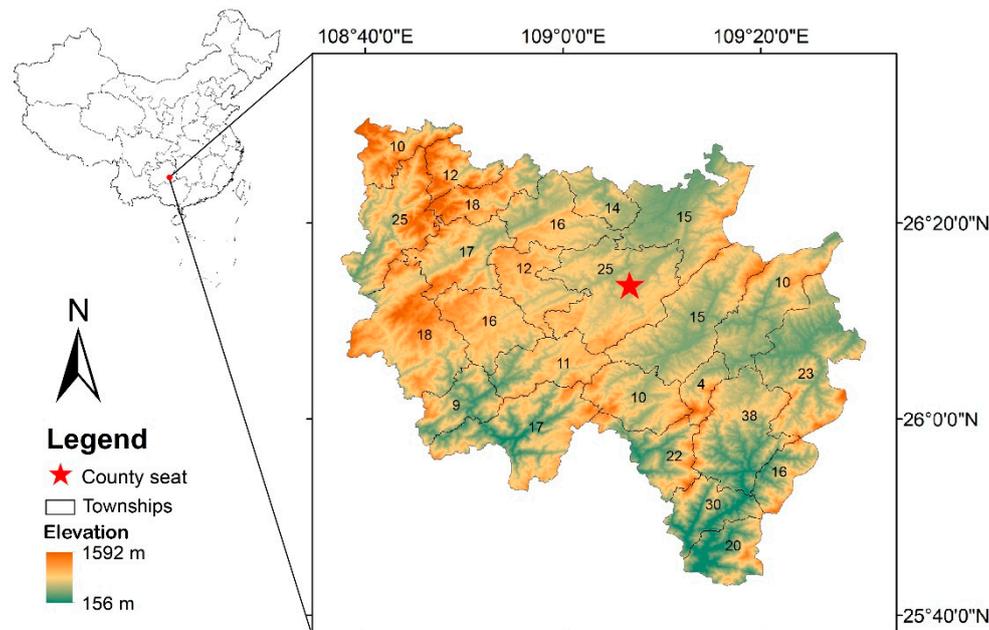
Rural land-cover information can be obtained directly and easily if global land-cover products are reliable. Global land-cover datasets, such as Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type/Dynamics products (MCD12Q1) [24], Global Urban Land (GUL) maps [8], and European Space Agency Climate Change Initiative-Land Cover (ESA CCI-LC) [25], that describe terrestrial land surfaces in different characteristic thematic land-cover categories have been published to help users understand land covers at the global and regional levels. However, these global land-cover datasets are likely to be retrieved from different remote sensing data and processed by different methods, leading to inconsistent interpretations and conclusions on the same object [26]. Hence, much effort has gone into improving the accuracy and consistency of global land-cover products for a better interpretation of global environmental issues [26–28]. However, the issues of these products were addressed only at a global level. It is still uncertain how well these global land-cover products work at a local level, particularly in a rural setting. As ISAs in rural areas tend to be small and fragmented, it is imperative to assess the performance of global land-cover products regarding ISA mapping in rural areas.

When published data are not sufficient for the applications of rural land-use planning, land-use planners and practitioners have to infer ISAs at a local level from the remotely sensed imagery produced by a variety of image processing protocols. With the advent of enormous volumes of satellite data, remotely sensed imagery is widely used in acquiring information on ISAs [4]. For example, Liu et al. [8] used Landsat imagery to infer global urban extents. Wang et al. [29] used Landsat 8 imagery to map seasonal land covers in Beijing. Verma et al. [30] employed very high-resolution satellite imagery and a deep learning approach to map urban slums in Mumbai, and Liu et al. [31] mapped rural ISAs using historical aerial photographs and questionnaire surveys. As rural regions are paid less attention when compared with urban areas, data such as aerial photographs are not always available, and thus open access remotely sensed data are becoming essential for mapping. Some studies obtained rural ISA information from land-use surveys conducted by governments or by researchers themselves [31,32]. However, such ISA information is not always spatially explicit, up-to-date, and easily accessible. There is a need to identify ISAs in rural areas with a simple and feasible classification strategy to obtain reliable information.

Hence, the objectives of this study are to assess the performance of global land-cover products in the context of rural ISA mapping and to demonstrate a simple and feasible approach for mapping of rural ISAs, which can be beneficially utilized by a variety of environmental practitioners. The Liping County located in southwest China, which is a typical mountainous rural area in China, was selected as the area for this case study.

## 2. Study Area

The Liping County in the Guizhou Province is a typical mountainous rural county in southwest China in terms of economic status and geography [33]. Liping County consists of 26 townships administering 33 towns and 390 villages [34] (Figure 1). Most of the towns and villages are located in relatively lower flat areas in valleys and are surrounded by farmlands. It covers approximately 4433 km<sup>2</sup> with a population of 393,000 in 2017 [34]. With the process of rapid urbanization in China, many local young people have migrated to eastern coastal urbanized cities, resulting in a drastic depopulation of approximately 18% from 2006 to 2017 [34]. The migrants, however, face a social barrier in becoming citizens of their destined cities owing to the household registration system, which effectively makes them unwilling to give up their rural land quota and motivates them to build new bigger houses in their original hometown [17], causing land-cover conversions from agricultural to ISAs in rural areas. Furthermore, the local government in China might have huge ambitions to develop new constructions, which may lead to excessive development of ISAs [35]. In the absence of a spatially explicit ISA map for the Liping County, concerns on efficient rural land management arise. Therefore, practitioners and policymakers require an accurate ISA map for the Liping County to prepare a sustainable land-use plan that can be utilized to prevent excessive future land developments.



**Figure 1.** Location of the Liping County. Numbers indicate the number of villages and/or towns in each township (390 villages and 33 towns, in total). Elevation data from [36].

## 3. Materials and Methods

### 3.1. Data Collection and Preprocessing

#### 3.1.1. Global Land-Cover Products

To assess the performance of global land-cover products for ISA mapping in the Liping County, three global land-cover datasets were utilized: MODIS/Terra and Aqua Combined Land Cover Type Yearly Global 500 m SIN Grid Version 006 (MODIS MCD12Q1) [24], ESA Climate Change Initiative Land Cover Project (ESA CCI-LC) [25], and high-resolution multi-temporal mapping of Global Urban Land (GUL) [8], with spatial resolutions of 500 m, 300 m, and 30 m, respectively. Considering the mapped years that were available across these products, the year 2015 was selected to ensure consistency in comparison. Imperviousness-related classes are labeled as “urban and built-up” in

MODIS MCD12Q1, “urban areas” in ESA CCI-LC, and “urban” in GUL. The legends of “urban” related classes in the three global products are all associated with impervious and/or artificial surfaces, providing a similarity of definition between them (Table 1). Therefore, these comparable classes were considered as ISAs, and the classification was simplified in these global land-cover products into two classes: “ISs” and “other” (i.e., land covers except for ISs).

**Table 1.** Summary of the imperviousness-related classes in the global land-cover products used in this study. ISA, impervious surface area; MODIS MCD12Q1, Moderate Resolution Imaging Spectroradiometer Land Cover Type/Dynamics products; GUL, Global Urban Land; ESA CCI-LC, European Space Agency Climate Change Initiative-Land Cover.

Global Products	Spatial Resolution	Classification Method	Imperviousness Related Class	Definition of the Class	Reference
MODIS MCD12Q1 version 6	500 m	Supervised classification	Urban and built-up	At least 30% of ISAs including building materials, asphalt, and vehicles.	[24]
ESA CCI-LC	300 m	Spectral unsupervised and supervised classification	Urban areas	Artificial surfaces and associated areas.	[25]
GUL	30 m	Normalized Urban Areas Composite Index (NUACI)	Urban	ISs, that is, artificial cover and structures such as pavement, concrete, brick, stone, and other man-made impenetrable cover types.	[8]

### 3.1.2. Satellite Imagery

Landsat archive from United States Geological Survey (USGS)—a free, open, and long archiving satellite data—is widely used in land cover/use monitoring [37]. As the Landsat archive contains data over the last 30 years, it has a huge potential for use in rural ISA mapping for the past as well. Landsat 8 at 30 m spatial resolution for the year 2015 was selected for a consistent comparison with the other global land-cover products. To demonstrate the availability of the Landsat archive, the Landsat 8 Surface Reflectance Tier 1 was used to produce the ISA map. This dataset is the atmospherically corrected surface reflectance data from the Landsat 8 OLI/TIRS sensors [38]. Using the cloud mask provided by quality control band values found in the dataset, the annual median composite of imagery in 2015 with bands 1–7 (b1: ultra-blue, b2: blue, b3: green, b4: red, b5: near-infrared, b6: shortwave infrared 1, and b7: shortwave infrared 2) at the spatial resolution of 30 m was preprocessed on the Google Earth Engine Platform (GEE) [39].

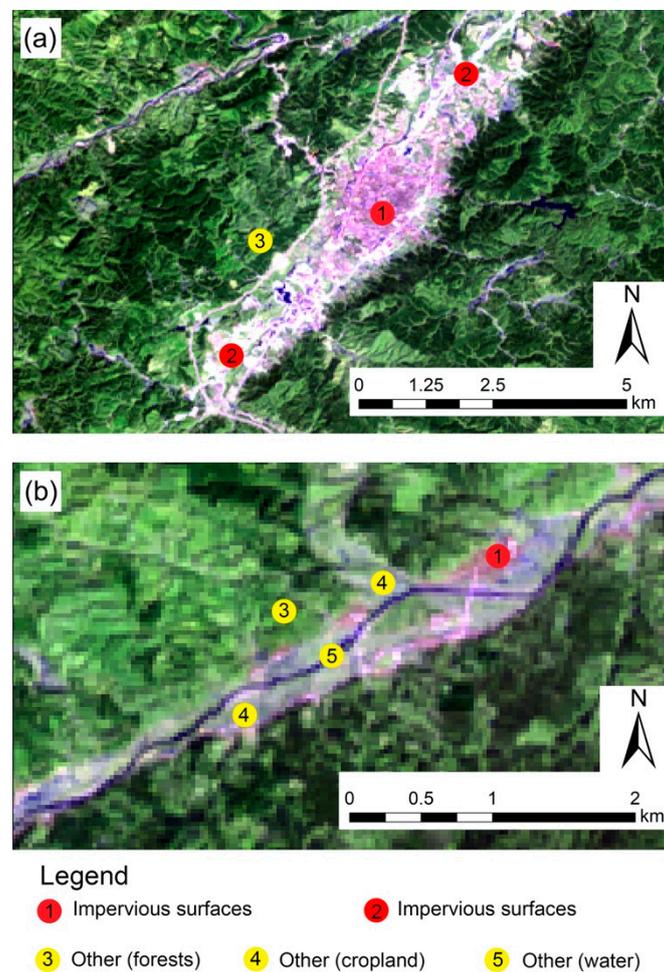
### 3.2. Supervised Classification

The random forests (RF) classifier was used to classify land covers from Landsat 8 surface reflectance imagery into two classes: “ISs” and “other”. “IS(s)” is defined as an artificial land covered with impenetrable materials such as asphalt, concrete, tile, brick, and stone, corresponding with the definitions used in the global land-cover products (Table 1) [8,24,25]. “Other” refers to a land covered with surfaces other than impervious surfaces such as vegetation, soils, and water.

RF, a non-parametric and robust supervised classification method, is widely applied in land-cover classification [40,41]. RF is developed from the decision tree classifier by randomly selecting a subset of variables through replacement [42]. A decision tree divides a dataset into smaller subdivisions based on the defined threshold at each branch [43]. However, a decision tree is not always robust as different training data may result in different trees. RF has been developed to address this issue by randomly selecting a subset of training data and growing many trees, like that of a “forest”.

### 3.2.1. Training Data Collection

A training sample was collected for the RF classifier with the use of a false-color composite image with the combination of RGB = 7-6-4 bands to highlight the ISAs visually. For example, in Figure 2, the colors represent different land types: purple, white, green, and cyan indicate roof materials, pavements, forests, and cropland, respectively. Thus, the pixels were randomly selected and manually labeled with colors. The accuracy of classification results improves with the increase in the size of the training sample [44]. The training sample collection fulfills sufficiency criteria as a considerable amount of data was collected using a polygon-based sampling method with sufficient spatial coverage within the study area. The polygon-based training method performs better than the pixel-based method if the classes are spectrally heterogeneous [45]. Thus, a total of 624 sample polygons, including 217 for “ISs” and 407 for “other” (Figure 3), were created. It was assumed that the ISAs were distributed in all townships. Sample collection was designed such that it included all townships, that is, each township has at least one polygon. The existence of villages and towns was supplementally confirmed using OpenStreetMap.



**Figure 2.** A false-color composite for the creation of training samples (a) at the county seat, and (b) at a village.

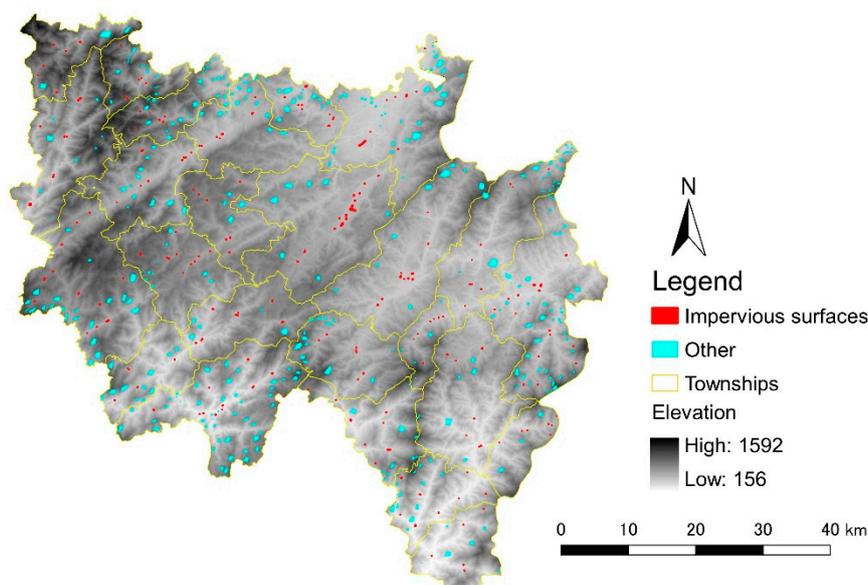


Figure 3. Spatial distribution of training sample polygons.

### 3.2.2. Model Setup

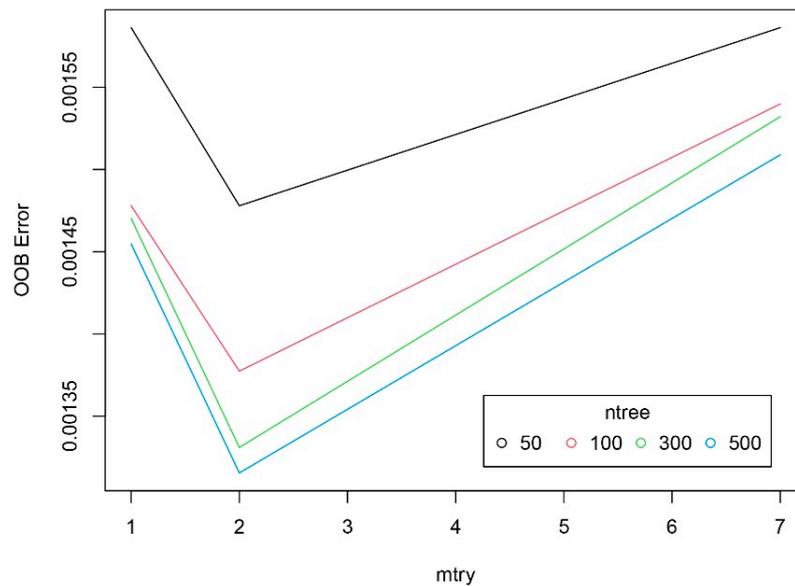
The RF classifier was implemented using *randomForest* function in the *randomForest* package of R software [46]. The variables for the RF classifier are shown in Table 2. The predictor variables were seven bands from Landsat 8 Surface Reflectance Tier 1, which were processed on GEE, and the response variable was a binary variable defined by the land covers: “ISs” (coded as 1) and “other” (coded as 0). RF has a good performance in coping with complex interactions and highly correlated predictor variables by randomly selecting a subset of predictor variables [42]. As a result, the trees are less similar and less correlated, and the final prediction can be more accurate. It is noted that the *multi.collinear* function in *rfUtilities* package confirmed no multi-collinearity among the predictor variables [47,48].

Table 2. Variables used for the random forests (RF) classifier.

Types of Variables	Variables	Description
Response	Land cover types	1: ISs, an artificial land covered by impenetrable materials such as asphalt, concrete, tile, brick, and stone; 0: Others, land covered with surfaces except for impervious surfaces.
Predictor	b1	Band 1 (ultra blue) surface reflectance; wavelength: 0.435–0.451 $\mu\text{m}$ .
Predictor	b2	Band 2 (blue) surface reflectance; wavelength: 0.452–0.512 $\mu\text{m}$ .
Predictor	b3	Band 3 (green) surface reflectance; wavelength: 0.533–0.590 $\mu\text{m}$ .
Predictor	b4	Band 4 (red) surface reflectance; wavelength: 0.636–0.673 $\mu\text{m}$ .
Predictor	b5	Band 5 (near-infrared) surface reflectance; wavelength: 0.851–0.879 $\mu\text{m}$ .
Predictor	b6	Band 6 (shortwave infrared 1) surface reflectance; wavelength: 1.566–1.651 $\mu\text{m}$ .
Predictor	b7	Band 7 (shortwave infrared 2) surface reflectance; wavelength: 2.107–2.294 $\mu\text{m}$ .

Two essential parameters *ntree* and *mtry* need to be set for the RF classifier [41]; *ntree* indicates the number of trees to grow and *mtry* denotes the number of predictor variables randomly sampled at each split of tree node. Figure 4 illustrates the result of some preliminary tests based on out-of-bag (OOB) errors with different parameters. The OOB error is an effective internal estimate to monitor the error of RF, which helps tune parameters [42]. This test indicated that *mtry* = 2 with all the *ntree*

numbers resulted in the lowest OOB error, and larger *ntree* tended to reduce the OOB error. From this exploratory test, we set the parameters of *ntree* as 500 and *mtry* as 2 for the RF classifier.



**Figure 4.** Out-of-bag (OOB) error of the random forests classifier with different parameter values of *mtry* and *ntree*.

### 3.2.3. Accuracy Assessment

The ISA classification map was assessed by a confusion matrix and the overall, precision, and recall accuracy measures [49], which were calculated as follows:

$$Precision = \frac{tp}{tp + fp} \quad (1)$$

$$Recall = \frac{tp}{tp + fn} \quad (2)$$

$$Overall\ accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (3)$$

where *tp*, *fp*, *tn*, and *fn* indicate the number of cases of true positive, false positive, true negative, and false negative, respectively. Here, the class of ISs is positive, while the class of “other” is negative in this study.

Stratified random sampling was employed for collecting validation sample points over the study area [50]. The targeted sample size is calculated as follows:

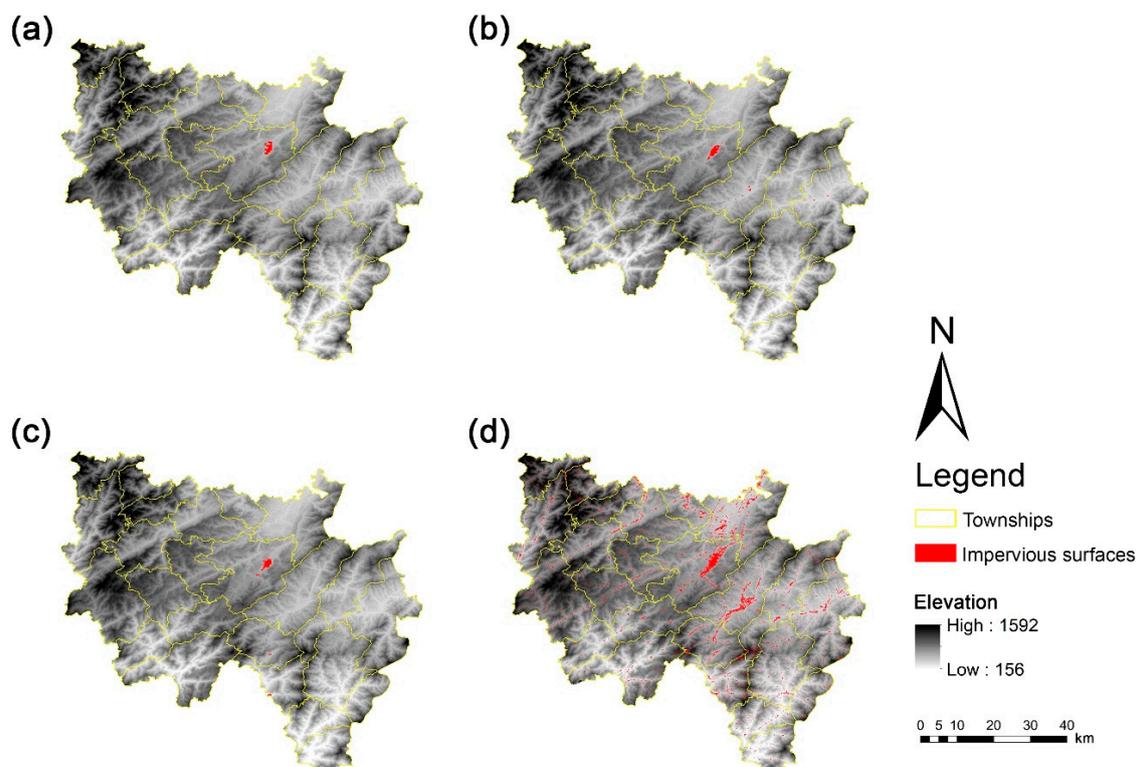
$$n = \left( \frac{\sum w_i s_i}{S(\hat{\delta})} \right)^2 \quad (4)$$

where  $S(\hat{\delta})$  is the standard error of the estimated overall accuracy that is required,  $w_i$  is the mapped proportion of the area of the class  $i$ , and  $s_i$  is the standard deviation of stratum  $i$ . A target standard error for the overall accuracy was set at 0.02, and it was conjectured that the recall accuracy will be 0.85 for “ISs” and 0.9 for “other”. On this basis, the total validation sample size was determined as 227 for this study. Proportional allocation of the validation sample size is a popular sample allocation strategy, but produces inaccurate recall accuracy for rare classes [50]. Hence, to investigate “ISs” more intensively, sample points were allocated equally, as 113 for “ISs” and 114 for “other”.

The validation sample was visually recorded and collected from very fine spatial resolution imagery in Google Earth to assure the independence of the accuracy assessment from the training scheme. It should be noted that very fine spatial resolution images of some parts of the Liping County were not available in Google Earth for 2015. To deal with this issue, images from 2013 and 2017 were used complementarily to conjecture the 2015 land covers in these areas by confirming a stable land cover over the period. In this case, it is a reasonable approach to obtain sample data in regions that are not covered by the available very fine spatial resolution images, as IS tends to be developed quickly and is not often converted into other land-cover types [51].

#### 4. Results

ISA maps for the Liping County produced using the global land-cover products (MODIS MCD12Q1, ESA CCI-LC, and GUL) and the RF classifier are shown in Figure 5. The global products depicted ISAs at the county seat, the largest ISA in the Liping County, but underestimated them in the other regions. ESA CCI-LC represented a few distributions of small ISAs in the eastern regions; however, the map was far from accurate. As illustrated in Figure 1, there are 390 villages and 33 townships located in the Liping County, and each township has at least four villages according to the statistical yearbook published by the Liping County [34]. As the three “urban” classes (Table 1) are defined as all major artificial ISs, the ISAs are underestimated by the global land-cover products in almost all townships. Moreover, the sizes of ISA agglomerations at the county seat are different among the three global products. For example, MODIS MCD12Q1 identified the smallest total ISA (375 ha) among the three global products, while ESA CCI-LC showed the largest ISA (517 ha). GUL inferred a total of 467 ha of ISAs. As the estimation errors were clearly significant, accuracy assessments were not conducted for the global product maps. In contrast, RF inferred a wide range of sizes of ISAs in the semi-urban area around the county seat and rural areas across all the townships.



**Figure 5.** ISA maps in the Liping County for 2015 found by (a) Moderate Resolution Imaging Spectroradiometer Land Cover Type/Dynamics products (MODIS MCD12Q1), (b) European Space

Agency Climate Change Initiative-Land Cover (ESA CCI-LC), (c) Global Urban Land (GUL), and (d) estimated by random forests classifier.

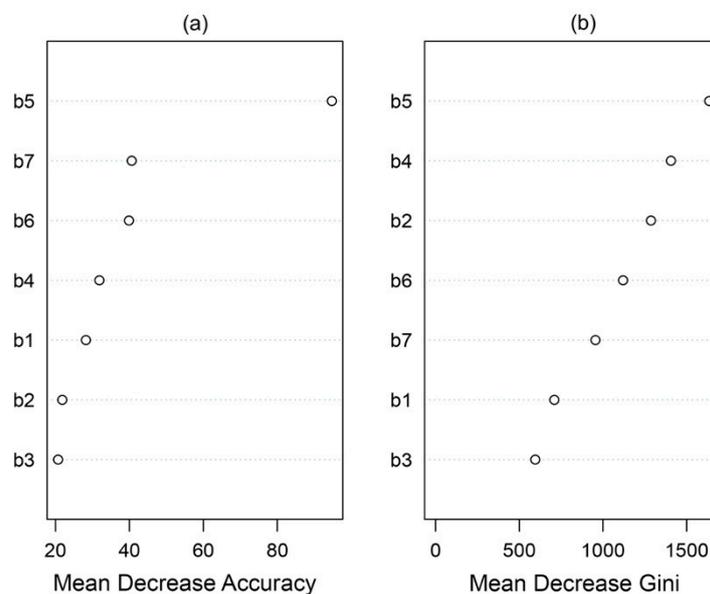
Table 3 summarizes the results from the four maps obtained using the RF classifier and global land-cover products. The RF method estimated that the ISAs represented 1.3% of the land cover in the Liping County, an order of magnitude higher than that of the global products, all of which identified  $\leq 0.12\%$ .

**Table 3.** Total IS areas and their proportion of total area in the Liping County estimated from the global land-cover products and the random forests classifier for 2015.

	RF	MODIS MCD12Q1	ESA CCI-LC	GUL
ISAs (ha)	5797	375	517	467
The proportion of ISAs in the total area (%)	1.3	0.08	0.12	0.11

Figure 6 presents the variable importance of the RF classifier. Mean decrease accuracy indicates how much the accuracy decreases if a variable is excluded, while mean decrease Gini suggests the decrease of node impurity when a variable is selected. Both of the variable importance measures indicated that band 5, near-infrared (0.851–0.879  $\mu\text{m}$ ), contributed the most to the final accuracy, which is reasonable as vegetation, water, and impervious areas are likely to be distinct in this wavelength range.

Table 4 summarizes the OOB assessment of the RF classifier. The OOB-based overall accuracy achieved 0.99, indicating the goodness of fit of the model. The confusion matrix assessed by independent validation samples indicated that the RF map achieved an overall accuracy of 0.91, a precision of 0.84, and a recall of 0.97, indicating relatively high accuracy (Table 5). These assessments indicated a similar tendency of the accuracy, suggesting robustness of the model.



**Figure 6.** The variable importance of the random forests classifier by (a) Mean Decrease Accuracy and (b) Mean Decrease Gini.

**Table 4.** Out-of-bag (OOB) confusion matrix for the random forests classifier.

		Observations		
		ISs	Other	Total
Predictions	ISs	3860	117	3977
	Other	56	125,198	125,254
	Total	3916	125,315	129,231
OOB overall accuracy: 0.99; precision: 0.97; recall: 0.99				

**Table 5.** Confusion matrix for the random forests classification map assessed by validation samples.

		Ground Reference		
		ISs	Other	Total
Classification Map	ISs	95	18	113
	Other	3	111	114
	Total	98	129	227
Overall accuracy: 0.91; precision: 0.84; recall: 0.97				

## 5. Discussion

The results show considerable differences between global land-cover products and an RF classification map in terms of the ISA estimation in the Liping County in China. All global land-cover products underestimated the extent of the ISAs and only inferred a few ISA agglomerations including the largest ISA at the county administrative center. In contrast, the RF estimated the distribution of ISAs in the lower and flat areas of valleys in a wide range of regions including all the townships. The classification accuracy also indicated an acceptable result. Even though a relatively precise ISA map was produced, some classification errors still exist. The aim of this study was not to develop the best classification model in terms of accuracy, but to demonstrate a simple and feasible classification method that can be utilized by experts as well as non-experts who have an interest in producing ISA maps of rural areas. The RF classifier was found to provide acceptable performance in mapping ISAs in the Liping County. It is evident that such a simple approach performs better than the existing global land-cover products in the context of mapping ISAs in less populated areas.

### 5.1. Reasons for Global Products Overlooking Rural ISAs

Unlike the large extents of ISs in urban areas, the ISAs in rural areas are likely to be small and fragmented. Global land-cover products tend to have coarse resolution and mainly focus on the large urban ISA extents. These characteristics might be the foremost reasons for the lack of recognition of ISAs in rural areas by the global land-cover products. As summarized in Table 1, the MODIS MCD12Q1 product employed supervised classification [24], while ESA CCI-LC uses spectral unsupervised and supervised algorithms that input variables associated with spectral properties and then integrates the two classification results based on objective rules for land-cover classification [25]. Both products employ per-pixel classification with a coarse spatial resolution (over 300–500 m), and hence, some rural ISAs may be smaller than such a minimum mapping unit, resulting in a failure of mapping. Vintrou et al. [52] reported a similar problem when using coarse-resolution global land-cover products to map the small extents of croplands in Mari. Therefore, spatial resolution is important in determining the mapping performance. In this sense, GUL has a resolution of 30 m [8] and the accuracy pertaining to spatial resolution is less. However, the protocols used for GUL may not perform well in mapping rural ISAs. This product mainly focuses on large urban agglomerations with the use of The Defense Meteorological Program (DMSP) operational line-scan system (OLS) (DMSP-OLS) nighttime lights, whose spatial resolution is 1 km to mask urban coverage, leading to the disregard of small impervious agglomerations in rural areas. Moreover, as global land-cover products collect

training or validation data at a global level, they are more likely to neglect local scale entities. In contrast to the global products, in this study, one particular area was focused on; thus, reference data that were distributed more widely throughout the area of interest could be collected rather than focusing only on large urban ISA extents.

### 5.2. Recommendations for Rural ISA Mapping

The simple training data collection strategy with a RF classifier utilized in this study was able to improve rural ISA mapping. Compared with the global land-cover products, the mapping results were greatly improved using the medium resolution (30 m) Landsat imagery. This strategy is proposed to achieve rural ISA mapping based on comprehensively collected training data, particularly by collecting samples at every administrative unit by referring to the actual location of villages or towns and creating a large training sample to achieve the required classification accuracy. The suitability of algorithms can vary in terms of the study area and satellite imagery, and the selection of algorithms is not restricted. The RF method could be replaced by other supervised classification algorithms such as maximum likelihood (ML), support vector machines (SVMs), or any other classification method. Although the accuracy of the RF classification was the best throughout our experiments (Table A1), all of the other classification methods were able to estimate the fragmented ISAs in the Liping County. Non-parametric algorithms such as RF and SVM tend to outperform the parametric algorithm like ML in land-cover classification [41]. As SVM is more flexible and complicated, requiring a careful parameter selection, the RF classifier was used to take advantage of its robustness and accuracy for this case study. This strategy can be implemented in other rural regions that have small and fragmented ISAs. In future studies, rural ISA mapping in other areas using this proposed strategy will be explored.

## 6. Conclusions

This study highlights the importance of mapping ISAs in rural areas (or less populated areas) that have rarely been focused on. It was found that most of the rural ISAs are not depicted in the global land-cover products because they overlook small and fragmented ISAs. This study demonstrates a simple and feasible training sample strategy with a RF classifier to obtain accurate rural ISA maps. When compared with the global land-cover products, this simple strategy significantly improved the ISA map obtained for the Liping County, a typical rural area in southwest China. Owing to its simplicity, this approach can be applied to any other rural regions. The spatial extent and distribution of the fragmented ISAs in rural areas is an important indicator to monitor the anthropogenic impact on natural resources. This approach provides a valuable means of mapping the fragmented ISAs in rural areas for efficient rural land development and planning.

**Author Contributions:** J.Z.: data curation, writing—original draft preparation, methodology, software, formal analysis, visualization, investigation. N.T.: conceptualization, methodology, validation, supervision, writing—reviewing and editing, project administration, funding acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research were funded by the Kyoto University Foundation and the National Geographic Foundation.

**Acknowledgments:** We thank three anonymous reviewers for providing helpful comments to improve an earlier version of this manuscript. We thank the ESA CCI Land Cover project, MCD12Q1 project, Global Urban Land project, and Liping Bureau of Statistics for providing data.

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

## Appendix A

**Table A1.** A comparison of different classifiers.

	RF	ML	SVM
ISAs (ha)	5797	10084	6464
Overall accuracy	0.91	0.83	0.89
Precision	0.84	0.67	0.81
Recall	0.97	0.97	0.95

Note: the training data was the same for all the classifiers. Three parameters were set for support vector machine (SVM) by the *tune* function in the R *e1071* package [53]: *cost* as 4, *gamma* as 0.6, and *kernel* as radial. ML, maximum likelihood; RF, random forests.

## References

- Lambin, E.F.; Turner, B.; Geist, H.J.; Agbola, S.B.; Angelsen, A.; Bruce, J.W.; Coomes, O.T.; Dirzo, R.; Fischer, G.; Folke, C.; et al. The causes of land-use and land-cover change: Moving beyond the myths. *Glob. Environ. Chang.* **2001**, *11*, 261–269. [[CrossRef](#)]
- Seto, K.C.; Sánchez-Rodríguez, R.; Fragkias, M. The New Geography of Contemporary Urbanization and the Environment. *Annu. Rev. Environ. Resour.* **2010**, *35*, 167–194. [[CrossRef](#)]
- Schueler, T.R. The importance of imperviousness. *Watershed Prot. Tech.* **1994**, *1*, 100–111.
- Weng, Q. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sens. Environ.* **2012**, *117*, 34–49. [[CrossRef](#)]
- Arnold, C.L.; Gibbons, C.J. Impervious Surface Coverage: The Emergence of a Key Environmental Indicator. *J. Am. Plan. Assoc.* **1996**, *62*, 243–258. [[CrossRef](#)]
- Weng, Q. Modeling urban growth effects on surface runoff with the integration of remote sensing and GIS. *Environ. Manag.* **2001**, *28*, 737–748. [[CrossRef](#)]
- Seto, K.C.; Fragkias, M.; Güneralp, B.; Reilly, M.K. A Meta-Analysis of Global Urban Land Expansion. *PLoS ONE* **2011**, *6*, e23777. [[CrossRef](#)]
- Liu, X.; Hu, G.; Chen, Y.; Li, X.; Xu, X.; Li, S.; Pei, F.; Wang, S. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. *Remote Sens. Environ.* **2018**, *209*, 227–239. [[CrossRef](#)]
- Tsutsumida, N.; Saizen, I.; Matsuoka, M.; Ishii, R. Addressing urban expansion using feature-oriented spatial data in a peripheral area of Ulaanbaatar, Mongolia. *Habitat Int.* **2015**, *47*, 196–204. [[CrossRef](#)]
- Lee, J.G.; Heaney, J.P. Estimation of Urban Imperviousness and its Impacts on Storm Water Systems. *J. Water Resour. Plan. Manag.* **2003**, *129*, 419–426. [[CrossRef](#)]
- Seto, K.C.; Güneralp, B.; Hutyrá, L.R. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 16083–16088. [[CrossRef](#)] [[PubMed](#)]
- Kalnay, E.; Cai, M. Impact of urbanization and land-use change on climate. *Nature* **2003**, *423*, 528–531. [[CrossRef](#)] [[PubMed](#)]
- Foley, J.A.; DeFries, R.; Asner, G.P.; Barford, C.; Bonan, G.; Carpenter, S.R.; Chapin, F.S.; Coe, M.T.; Daily, G.C.; Gibbs, H.K.; et al. Global Consequences of Land Use. *Science* **2005**, *309*, 570–574. [[CrossRef](#)] [[PubMed](#)]
- Dougherty, M.; Dymond, R.L.; Goetz, S.J.; Jantz, C.A.; Goulet, N. Evaluation of Impervious Surface Estimates in a Rapidly Urbanizing Watershed. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 1275–1284. [[CrossRef](#)]
- Ma, Q.; He, C.; Wu, J.; Liu, Z.; Zhang, Q.; Sun, Z. Quantifying spatiotemporal patterns of urban impervious surfaces in China: An improved assessment using nighttime light data. *Landsc. Urban Plan.* **2014**, *130*, 36–49. [[CrossRef](#)]
- Wang, B.; Chen, Z.; Zhu, A.-X.; Hao, Y.; Xu, C. Multi-Level Classification Based on Trajectory Features of Time Series for Monitoring Impervious Surface Expansions. *Remote Sens.* **2019**, *11*, 640. [[CrossRef](#)]
- Bai, X.; Shi, P.; Liu, Y. Realizing China's urban dream. *Nature* **2014**, *509*, 158–160. [[CrossRef](#)]
- Li, Y.; Liu, Y.; Long, H.; Cui, W. Community-based rural residential land consolidation and allocation can help to revitalize hollowed villages in traditional agricultural areas of China: Evidence from Dancheng County, Henan Province. *Land Use Policy* **2014**, *39*, 188–198. [[CrossRef](#)]

19. Liu, Y.; Liu, Y.; Chen, Y.; Long, H. The process and driving forces of rural hollowing in China under rapid urbanization. *J. Geogr. Sci.* **2010**, *20*, 876–888. [[CrossRef](#)]
20. Kahn, M.E. The environmental impact of suburbanization. *J. Policy Anal. Manag.* **2000**, *19*, 569–586. [[CrossRef](#)]
21. Mieszkowski, P.; Mills, E.S. The Causes of Metropolitan Suburbanization. *J. Econ. Perspect.* **1993**, *7*, 135–147. [[CrossRef](#)]
22. Lambin, E.F.; Meyfroidt, P. Land use transitions: Socio-ecological feedback versus socio-economic change. *Land Use Policy* **2010**, *27*, 108–118. [[CrossRef](#)]
23. Van Vliet, J.; Eitelberg, D.A.; Verburg, P.H. A global analysis of land take in cropland areas and production displacement from urbanization. *Glob. Environ. Chang.* **2017**, *43*, 107–115. [[CrossRef](#)]
24. Sulla-Menashe, D.; Friedl, M.A. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC. 2018. Available online: <https://doi.org/10.5067/MODIS/MCD12Q1.006> (accessed on 23 September 2019).
25. ESA Climate Change Initiative. Land Cover CCI Climate Research Data Package. Available online: <https://maps.elie.ucl.ac.be/CCI/viewer/> (accessed on 20 July 2019).
26. Giri, C.; Pengra, B.W.; Long, J.; Loveland, T. Next generation of global land cover characterization, mapping, and monitoring. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *25*, 30–37. [[CrossRef](#)]
27. Herold, M.; Mayaux, P.; Woodcock, C.; Baccini, A.; Schmillius, C. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sens. Environ.* **2008**, *112*, 2538–2556. [[CrossRef](#)]
28. Congalton, R.; Gu, J.; Yadav, K.; Thenkabail, P.S.; Ozdogan, M. Global Land Cover Mapping: A Review and Uncertainty Analysis. *Remote Sens.* **2014**, *6*, 12070–12093. [[CrossRef](#)]
29. Wang, J.; Li, C.; Hu, L.; Zhao, Y.; Huang, H.; Gong, P. Seasonal Land Cover Dynamics in Beijing Derived from Landsat 8 Data Using a Spatio-Temporal Contextual Approach. *Remote Sens.* **2015**, *7*, 865–881. [[CrossRef](#)]
30. Verma, D.; Jana, A.; Ramamritham, K. Transfer learning approach to map urban slums using high and medium resolution satellite imagery. *Habitat Int.* **2019**, *88*, 101981. [[CrossRef](#)]
31. Liu, Y.; Yang, R.; Long, H.; Gao, J.; Wang, J. Implications of land-use change in rural China: A case study of Yucheng, Shandong province. *Land Use Policy* **2014**, *40*, 111–118. [[CrossRef](#)]
32. Liu, Y.; Luo, T.; Liu, Z.; Kong, X.; Li, J.; Tan, R. A comparative analysis of urban and rural construction land use change and driving forces: Implications for urban–rural coordination development in Wuhan, Central China. *Habitat Int.* **2015**, *47*, 113–125. [[CrossRef](#)]
33. National Bureau of Statistics of China (NBSC). Annual Gross Regional Product by Province. Beijing. Available online: <http://data.stats.gov.cn/english/2018> (accessed on 2 August 2019).
34. Liping Bureau of Statistics. *2017 Nian Liping Tong Ji Nian Jian (Liping Statistical Yearbook 2017)*; Liping Bureau of Statistics: Liping, China, 2018.
35. Zheng, Q.; Deng, J.; Jiang, R.; Wang, K.; Xue, X.; Lin, Y.; Huang, Z.; Shen, Z.; Li, J.; Shahtahmasebi, A. Monitoring and assessing “ghost cities” in Northeast China from the view of nighttime light remote sensing data. *Habitat Int.* **2017**, *70*, 34–42. [[CrossRef](#)]
36. Jarvis, A.; Reuter, H.I.; Nelson, A.; Guevara, E. Hole-Filled SRTM for the Globe Version 4. CGIAR-CSI SRTM 90m Database 2008. Available online: <http://srtm.csi.cgiar.org> (accessed on 15 July 2019).
37. Wulder, M.A.; Loveland, T.R.; Roy, D.P.; Crawford, C.J.; Masek, J.G.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Belward, A.S.; Cohen, W.B.; et al. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* **2019**, *225*, 127–147. [[CrossRef](#)]
38. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. [[CrossRef](#)] [[PubMed](#)]
39. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
40. Rodriguez-Galiano, V.; Ghimire, B.; Rogan, J.; Olmo, M.C.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104. [[CrossRef](#)]
41. Belgiu, M.; Dragut, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
42. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]

43. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984; ISBN 9780412048418.
44. Pal, M.; Mather, P. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* **2003**, *86*, 554–565. [[CrossRef](#)]
45. Chen, D.M.; Stow, D. The effect of training strategies on supervised classification at different spatial resolutions. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 1155–1161.
46. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.
47. Evans, J.S.; Murphy, M.A. Rfutilities. R Package Version 1.0-0. Available online: <https://CRAN.R-project.org/package=rfUtilities> (accessed on 4 May 2020).
48. Tsutsumida, N.; Comber, A.; Barrett, K.; Saizen, I.; Rustiadi, E. Sub-Pixel Classification of MODIS EVI for Annual Mappings of Impervious Surface Areas. *Remote Sens.* **2016**, *8*, 143. [[CrossRef](#)]
49. Powers, D.M.W. Evaluation: From precision, recall and F-factor to ROC, informedness, markedness & correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63. [[CrossRef](#)]
50. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; A Wulder, M. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [[CrossRef](#)]
51. Tsutsumida, N.; Rodríguez-Veiga, P.; Harris, P.; Balzter, H.; Comber, A. Investigating spatial error structures in continuous raster data. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *74*, 259–268. [[CrossRef](#)]
52. Vintrou, E.; Desbrosse, A.; Bégué, A.; Traoré, S.; Baron, C.; Seen, D.L. Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *14*, 83–93. [[CrossRef](#)]
53. Dimitriadou, E.; Hornik, K.; Leisch, F.; Meyer, D.; Weingessel, A. Misc Functions of the Department of Statistics (e1071), TU Wien. Available online: <https://CRAN.R-project.org/package=e1071> (accessed on 3 March 2020).



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).