Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition

Thanh Noi Phan *, Verena Kuch and Lukas W. Lehnert

Department of Geography, Ludwig-Maximilians-University Munich, Luisenstr.37, 80333 Munich, Germany; v.kuch@campus.lmu.de (V.K.); lehnert.lu@lmu.de (L.W.L.)

* Correspondence: tn.phan@iggf.geo.uni-muenchen.de; Tel.: +49-0-89-2180-6748

check for updates

Abstract: Land cover information plays a vital role in many aspects of life, from scientific and economic to political. Accurate information about land cover affects the accuracy of all subsequent applications, therefore accurate and timely land cover information is in high demand. In land cover classification studies over the past decade, higher accuracies were produced when using time series satellite images than when using single date images. Recently, the availability of the Google Earth Engine (GEE), a cloud-based computing platform, has gained the attention of remote sensing based applications where temporal aggregation methods derived from time series images are widely applied (i.e., the use the metrics such as mean or median), instead of time series images. In GEE, many studies simply select as many images as possible to fill gaps without concerning how different year/season images might affect the classification accuracy. This study aims to analyze the effect of different composition methods, as well as different input images, on the classification results. We use Landsat 8 surface reflectance (L8sr) data with eight different combination strategies to produce and evaluate land cover maps for a study area in Mongolia. We implemented the experiment on the GEE platform with a widely applied algorithm, the Random Forest (RF) classifier. Our results show that all the eight datasets produced moderately to highly accurate land cover maps, with overall accuracy over 84.31%. Among the eight datasets, two time series datasets of summer scenes (images from 1 June to 30 September) produced the highest accuracy (89.80% and 89.70%), followed by the median composite of the same input images (88.74%). The difference between these three classifications was not significant based on the McNemar test ($p > 0.05$). However, significant difference ($p < 0.05$) was observed for all other pairs involving one of these three datasets. The results indicate that temporal aggregation (e.g., median) is a promising method, which not only significantly reduces data volume (resulting in an easier and faster analysis) but also produces an equally high accuracy as time series data. The spatial consistency among the classification results was relatively low compared to the general high accuracy, showing that the selection of the dataset used in any classification on GEE is an important and crucial step, because the input images for the composition play an essential role in land cover classification, particularly with snowy, cloudy and expansive areas like Mongolia.

Keywords: land cover classification; Google Earth Engine (GEE); Landsat; image composition; Mongolia; Random Forest (RF)

1. Introduction

The most apparent indicator for surficial changes of the Earth, no matter the type, is land cover [1]. Recent studies have reported that the ongoing land use/cover change (LUCC) is having an increasingly negative impact on various aspects of the Earth's surface, such as terrestrial ecosystems, water balance,

biodiversity and climate [2–5]. Among these, the effects of LUCC on the terrestrial ecosystem receive the most attention from researchers, as ecosystems play a crucial role in the global carbon cycle [6–9]. Covering one-third of the world's land surface [10], grassland is an important component of the terrestrial ecosystem. For instance, the vast steppe ecosystems of Mongolia, which are located between the Siberian taiga and the Central Asian deserts, are considered as one of the valuable ecosystems in the world. However, due to climate change (e.g., the warming trend, the increasing frequency in extreme climate events), land cover change (e.g., increasing grazing density and mining), as well as the changing policies of governments, the grassland and the pastoralists that they support are being negatively affected [11–15]. In the literature, many studies have reported that in temperate grassland areas, grazed areas tend to have greater biodiversity than ungrazed areas [16–19]. This suggests that better grassland management would result in better grassland for both health and rural livelihoods [20]. Therefore, accurate, current and long-term information of land use/cover maps is highly demanding in Mongolia, for not only economic development but also the governmental policies [21–23].

Remotely sensed data has been recognized as one of the most important data sources for land cover mapping and for monitoring land cover change over time with Landsat being the most frequently used data source [24–26]. Other sources of data for LUCC studies are Satellite Pour l'Observation de la Terre (SPOT) [27], Synthetic Aperture Radar (SAR) [28,29], Moderate Resolution Imaging Spectroradiometer (MODIS) [30,31] and Sentinel 2 [32–35]. In many applications dealing with support decision making (e.g., environmental monitoring), high spatiotemporal resolution and long term data are essential [36,37]. To date, Landsat is the only operational system that can provide high spatial resolution (30 m), temporal resolution (16 days – with a single satellite and eight days, if data from both satellites are combined) and continuity over 30 years [38].

When mapping land cover over a large area, there are two main challenges—"big data" needs to be processed and the availability of cloud-free images over a large area. For example, to cover the whole Mongolian Country, at least 125 Landsat tiles are required (Figure 1). It would be very labor intensive if we process data using the traditional methods, from searching, filtering, downloading and mosaicking to preprocessing, such as cloud masking or atmospheric correction. Such "big data" are not only labor intensive but also require significant storage capacity and access to high power computing. Furthermore, due to cloud cover, it is not easy to achieve clear images for a large area in a short time period (e.g., monthly composition in Figure 1). To reduce the time span over which cloud-free mosaics are created, partly cloudy images have to be downloaded and preprocessed, creating more difficulty and much heavier workloads [39]. Furthermore, in areas such as Mongolia, snow cover does not allow to reliably detect the underlying land cover, reducing the availability of suitable data. Therefore, to create a cloud (and snow) free composite image for the entire Mongolia, several hundred Landsat scenes are required (Figure 1).

Google Earth Engine (GEE), a cloud-based computing platform, can solve the most significant problems with respect to land cover mapping of large areas. Users can analyze all available remotely sensed images using a web-based Integrated Development Environment (IDE) code editor without downloading these data to the local machine. In this way, users can easily access, select and process large volumes of data for a large study area [40]. Besides the fast processing, another important aspect that makes GEE more and more popular is the availability of several packages with lots of algorithms simplifying access to remote sensing tools for both expert and non-experts. According to Tamiminia et al. [41], since 2013 the number of publications using GEE steadily increased. Among the available datasets in GEE, data from optical satellites and particularly the approximately 40 years long time series of Landsat, have been the most frequently used. It is reported that GEE has been applied to various areas, ranging from agriculture, forestry, ecology to economics and medicine [41,42]. Among these, forest and vegetation studies were the most frequent application disciplines, followed by land use and land cover studies.
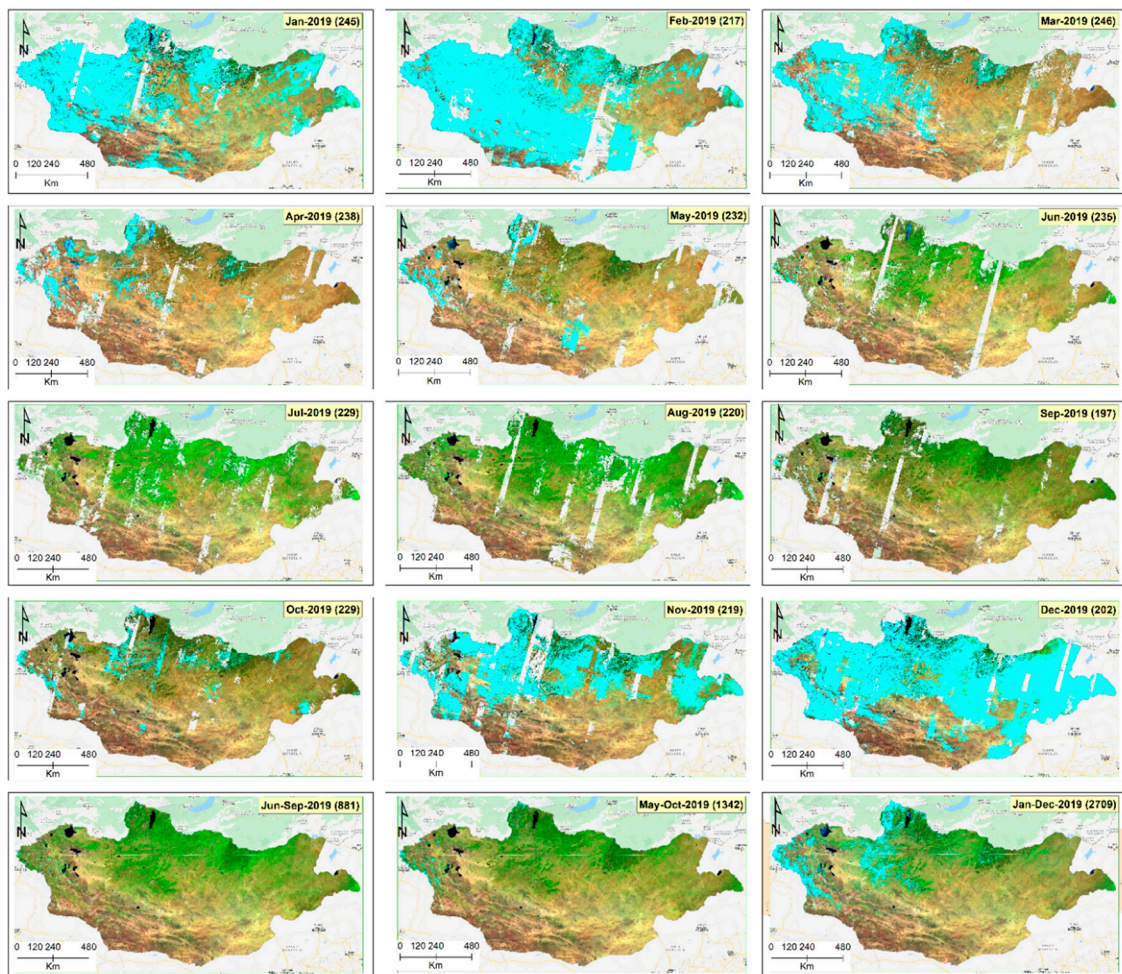
Figure 1. Monthly median natural composite (7-5-3) of Landsat 8 surface reflectance data over Mongolia in 2019. The number of Landsat scenes used for the composition is shown in parentheses. Bright blue represents snow cover.

As mentioned above, Mongolia is covered by 125 Landsat tiles, however, to create a free cloud and snow composite imagery more than 800 scenes are required (e.g., Figure 1, Jun-Sep composition imagery). Therefore, GEE is considered the best option and presents great opportunities for mapping land cover in Mongolia.

After identifying all remotely sensed scenes which could provide data for a specific land use/cover study, the first critical step is the combination of these datasets. In the literature, there are two composition methods that have been widely applied for land cover classification using multi-temporal Landsat images. One is using the temporal aggregation method, that is, use the metrics, such as mean, median and min/max, derived from time series images [43–47]. Another is making a composition of time series data from all the (cloud free) available Landsat images [48–50]. Obviously, these two methods are physically different. However, the effects of these composition methods on the land cover classification accuracy have not been fully exploited to date and there are many studies that simply select as many images as possible until the gaps are filled, without concern for the effect of different year/season that might affect the classification accuracy. For instance, the most popular strategy for selecting input images for an annual cloud free composite is to use images acquired over three years [44–47].

In the recent literature, based on our best knowledge, no study has compared the effect on classification accuracy from different selecting strategies for annual image composites, as well as the

effect of different composition methods (i.e., median metric versus single date image or time series image data).

Consequently, our aim is to provide the reader with a summary of the effects of different composition methods (metrics, time series and the different annual composition strategies) and the input feature selection (spectral and auxiliary variables) on the accuracy of land cover classification. A pilot test site (approximately one Landsat tile) over Ulaanbaatar (Mongolia) with different cloud/snow cover conditions and the seasonally changing land cover was selected. We used Landsat 8 surface reflectance data and the GEE cloud computing platform to investigate the comparisons.

2. Materials and Methods

2.1. Study Area

To show more detail about the different effects of image selection, we use one Landsat footprint (path 131 row 27) over Ulaanbaatar (Mongolia) as a case study (Figure 2). It should be mentioned that, to reduce the effect of the edge, we buffer 3 km inside the footprint to use as the study area.
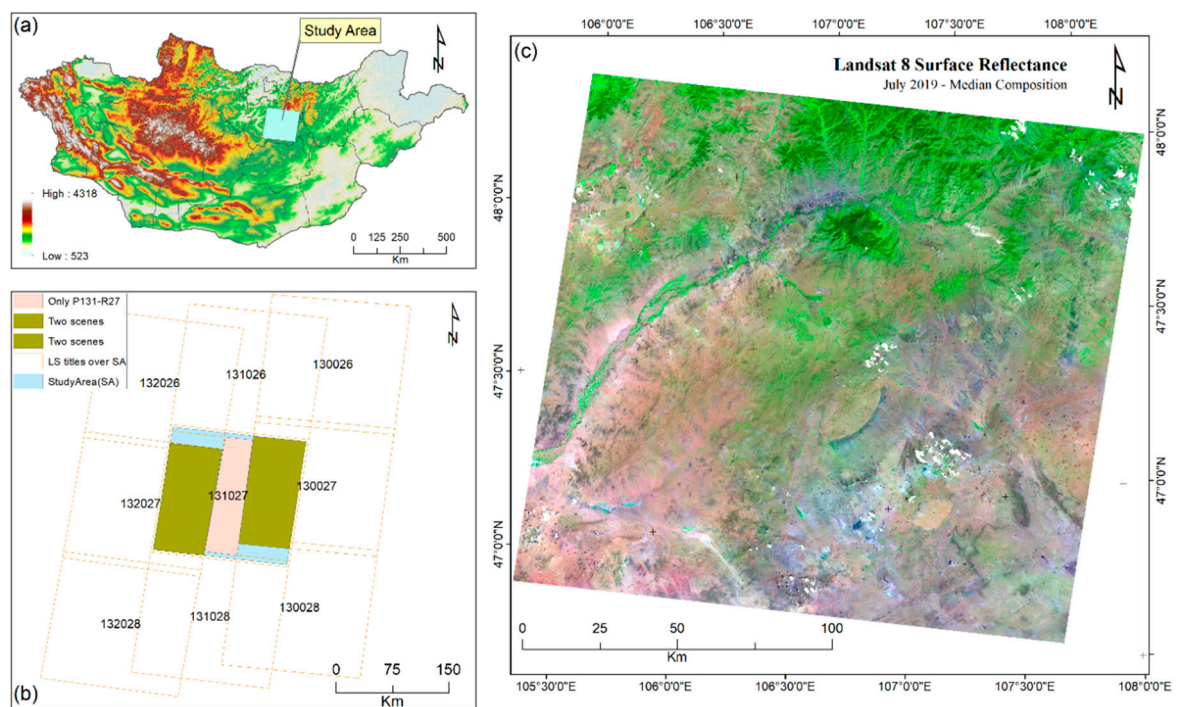
Figure 2. (**a**) The location of the study area in Mongolia with elevation derived from SRTM (30m) data, (**b**) The Landsat scenes over the study area—different colors represents different numbers of Landsat scenes covering the study area, (**c**) Natural composite (7-5-3) of the median image in July 2019 of the study area.

2.2. Data Used

2.2.1. Landsat-8 Surface Reflectance Tier 1 data (L8sr)

To focus on the effect of image selection on land cover classification, we minimized all other effects encompassing for example, those arising from the differences between RS sensors, for example, the difference between Landsat OLI and previous Landsat sensors [51], the effect of preprocessing (e.g., atmospheric correction), the effect of different footprints or different acquisition dates [52]. Consequently, we selected all Landsat 8 atmospherically corrected surface reflectance scenes available on the GEE platform for the year 2019 (except Dataset 4 which also encompassed the images from June to September 2018).

Our study area is covered by either a single scene (path 131, row 27) or can be merged from nine scenes (Figure 2). It is worth mentioning that, in GEE, if the users do not specifically select the path and row of the scenes, it will automatically select all the images that intersect with the study area's boundary. That is why we have a large different number of images in each datasets (Table 1). Obviously, the selection method (automatic vs. specific selection of path and row of scenes) has an effect on the quality and characteristic of the composite images that are going to be used for classification.

Table 1. The composition of the eight datasets used for classification and comparison.

Dataset	Description	No. L8 Images Used for Composition	No. Reflectance Bands	No. Auxiliary Variables
Dataset 1	All the available data from L8sr in 2019 were selected to calculate the median image for classification.	196	7	13
Dataset 2	Only images between 1st of June and 30th of September 2019 were selected to calculate the median image.	61	7	13
Dataset 3	Only images with cloud cover less than 30% were used for median calculation.	130	7	13
Dataset 4	Median image was composited from June to September of two years: 2018 and 2019.	126	7	13
Dataset 5	The best single scene (p131r27) covering the entire study area was selected based on the lowest cloud cover percentage.	1	7	13
Dataset 6	Median image of all p131r27 images between 1 June and 30 September 2019.	7	7	13
Dataset 7	Time series images of Collection 2.	61	28	13
Dataset 8	Time series images of single scene cover study area (p131r27) between 1 June and 30 September 2019.	7	28	13

Ten popular auxiliary datasets, namely the Normalized Difference Vegetation Index (NDVI), Enhanced vegetation index (EVI), Soil Adjusted Vegetation Index (SAVI), Modified Secondary Soil-Adjusted Vegetation Index (MSAVI2), Normalized Difference Water Index (NDWI), modified NDWI (mNDWI), Normalized Difference Water Body Index (NDWBI), Normalized Difference Built-Up Index (NDBI), simple ration (SR) and Entropy were derived and calculated from the L8sr data in order to increase the accuracy of land cover classification [53,54]. The definitions and formulas of these variables are presented in Table A1. In addition, topographic variables, such as elevation, slope and aspect, have been shown to be related with the distribution of the land cover type in many studies [55,56]. Therefore, in this study, the 30 m elevation, slope and aspect derived from the Shuttle Radar Topography Mission (SRTM) data were used as auxiliary variables for the classification. The number of auxiliary variables used within each dataset is shown in Table 1.

2.2.2. Training and Validation Sample Data

Eight land cover types dominate the area of investigation—(1) Agriculture (AG), (2) Burned Area (BA), (3) Bare land (BL), (4) Grassland (GA), (5) Mixed Grassland (GRm), (6) Residence (RE), (7) Forest (FR) and (8) Water. All training and validate samples were collected based on manual visual interpretation of high-resolution images from Google Earth and Planet Lab [57]. This method is widely applied and reported in the literature [58–60]. Furthermore, to minimize the effect of spatial autocorrelation yet still capture the gradient of land cover type within each class (e.g., shallow water versus deep water; low, medium and high grassland cover) we selected training samples as small polygons (i.e., this polygon contains a number of relatively homogenous pixels of a given land cover types) and validation samples as points. If validation samples are selected close to the training polygons, it is more likely to miss the detection of overfitting. Consequently, we randomly selected validation points with the constraint of being at least 100 m away from the closest training polygon to reduce spatial autocorrelation. In the next step, the (initially) selected training polygons and validation

points were downloaded from GEE and uploaded to Planet Lab. These samples were further adjusted and corrected by the co-authors. Finally, we had 513 training polygons and 1039 validation points.

2.3. Methods

We created eight datasets on the GEE platform with respect to their temporal aggregation and time series stack methods (Figure 3). Dataset 1 to Dataset 6 are median imageries, Dataset 7 and 8 are time series data that have the same input component images as Dataset 2 and Dataset 6, respectively. The input images (the temporal aggregation and time series) were selected based on different strategies in order to assess the effect of different selections on classification accuracy (Table 1). These datasets, as well as the auxiliary variables, have the same spatial resolution (30 m) as the original resolution of Landsat 8 images.
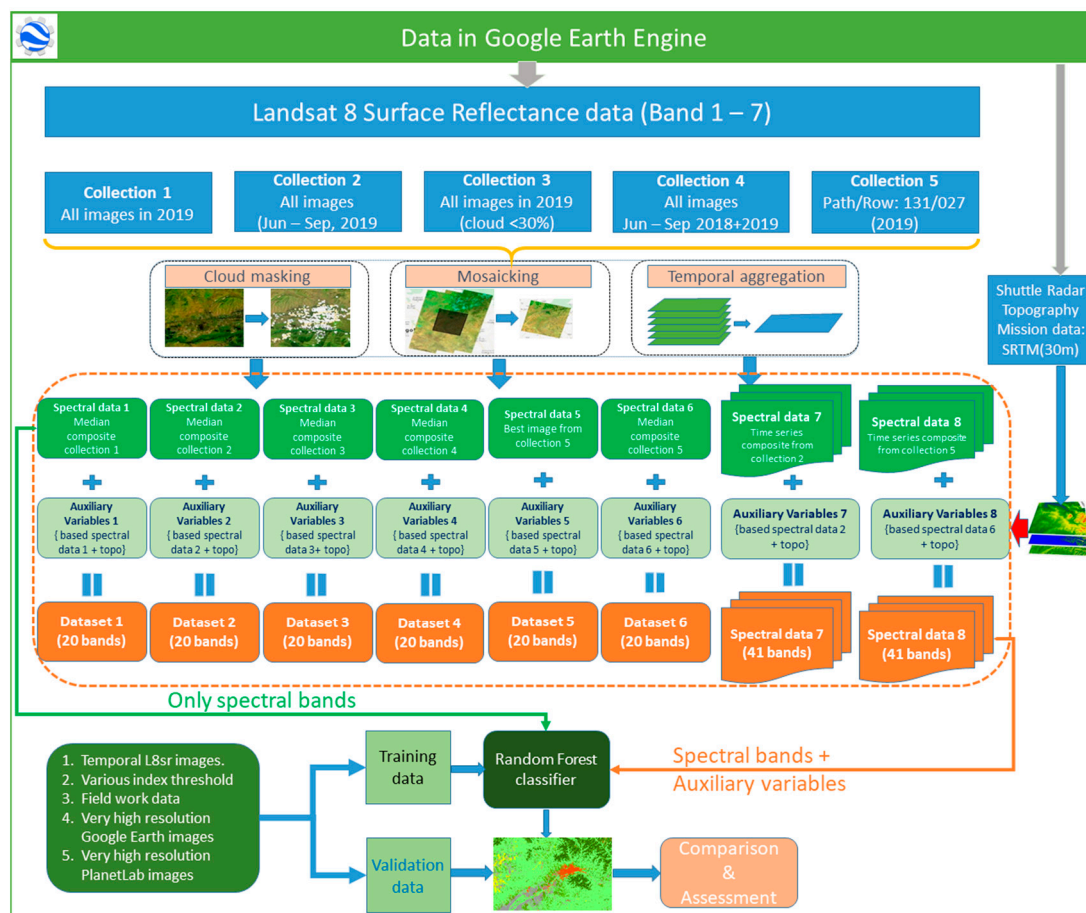

Figure 3. Flowchart of the study.

2.4. Random Forest Classifier

To date, RF is considered one of the most widely used algorithms for land cover classification using remote sensing data [55–57,61–66]. According to Mahdianpari et al. [67] and Xia et al. [68], the reasons for RF receiving considerable interest over the last two decades are—(1) Good handling of the outliers and noisier datasets; (2) Good performance with high dimensional and multi-source datasets; (3) Higher accuracy than other popular classifiers, such as SVM, kNN or MLC in many applications [69,70]; and (4) Increasing the processing speed by selecting important variables [71].

Another factor making RF more popular than other machine learning algorithms (e.g., SVM) is that only two parameters (ntree and mtry) are required to be optimized, facilitating the application of RF [72]. A meta-analysis of 349 GEE peer-reviewed articles over the last 10 years shows that the RF

algorithm is the most frequently used classification algorithm for satellite imagery [41]. Considering all these reasons, we chose RF for the present study.

Based on the recommendations of previous studies [62,73] and pretests from our data, we selected 100 trees (ntree = 100), while mtry was set to the default value (square root of the total number of features).

2.5. Accuracy Assessment, Comparison and Statistical Testing

Foody [74] highly recommends that the kappa coefficient should not be routinely used in the assessment and comparison of the accuracy of thematic maps derived from image classification. Therefore, in the present study, for each classification accuracy assessment, we used the popular measures extracted from confusion matrix reports, such as overall accuracy (OA), producer accuracy (PA) and user accuracy (UA). However, according to Foody [75], a confusion matrix (i.e., OA, PA and UA) only provides information for an "estimate" of classification accuracy, thus only a tentative conclusion can be made. This is especially the case when we compare different classification results with small accuracy differences. Janssen & van der Wel [76] state that it is necessary to compare the accuracy of different classifications in a statistically rigorous manner. For example, if two classification maps, M1 and M2, of a region were created by two different classifiers (or different input datasets), it is required to assess whether the difference in the accuracy of these maps is significant or not. In the literature, there are a number of studies using kappa coefficients to assess the significant difference between classifications via z value.

$$z = \frac{\kappa_1 - \kappa_2}{\sqrt{\delta_{\kappa_1}^2 + \delta_{\kappa_2}^2}},\tag{1}$$

where, κ_1, κ_2, δ_{κ_1} and δ_{κ_2} are the estimated kappa coefficients and the associated estimates of the standard error of kappa for M1 and M2, respectively [77,78].

According to Foody [66], Equation (1) should be replaced by the equation:

$$z = \frac{\rho_{0_1} - \rho_{o_2}}{\sqrt{\delta_{\rho_{01}}^2 + \delta_{\rho_{02}}^2}},\tag{2}$$

where, ρ_0 and δ_{ρ_0} are the proportion correct and the standard error, respectively.

It is worth to mention that Equations (1) and (2) can be applied if the samples used are independent. However, in remote sensing applications, when comparing the difference between maps, the same ground truth (validation/ references) dataset is usually used. In this case, the McNemar test [79] can indicate whether the difference in classification results is significant [75]. Though the non-parametric McNemar test was based on a 2 by 2 confusion matrix, the remote sensing image based classification confusion matrix (which is often more than two classes) can be collapsed to the size of 2 by 2 by focusing on corrected and uncorrected pixels of classified output [80]. In doing so, the McNemar test calculates the z value:

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}},\tag{3}$$

where f_{12} is the number of corrected samples in classified result one while result two is uncorrected and f_{21} is the number of corrected samples in classified result two while result one is uncorrected. z is following the Chi^2 distribution with five degrees of freedom which has been used to estimate the significance of differences in accuracy among classifications in our study (95% confidence level).

2.6. Effects of Differences Among Classifications on the Spatial Estimation of Land Use Classes

Assuming that each classification will differ with respect to the spatial occurrences of the different land use classes, we calculated the mode (majority class) of each pixel in all 8 classifications. The mode classification image was then compared to the classification featuring the highest overall accuracy as a

reference. In addition, starting from the mode classification, we calculated the number of pixels for each class that have been assigned in all classifications only to this class, as well as the number of pixels that have been assigned to this class and one other class, and so forth. This will provide us with a class-wise estimation of the uncertainty arising from the combination strategy used to create cloud- and snow-free images in GEE.

3. Results

3.1. Overall Accuracy of Different Datasets With and Without Auxiliary Variables

We applied the same training sample and validation data points to classify and assess the accuracy of the land cover maps.

As mentioned in the Methods section, aside from the spectral bands of Landsat images, we also used additional variables to test whether they increase the accuracy of the land cover maps.

The results of classification accuracy assessment (Table 2) show that if only spectral feature bands from L8sr were used, a moderate to high agreement with reference data can be achieved (OA ranges from 77.6% to 85.27%). However, when the additional auxiliary variables (10 spectral indices + 3 topographic indices) were included in the model, the OAs increased by approximately 4.1% to 7.7%. This is consistent for all datasets. The highest increases were observed with Dataset 1, Dataset 2, Dataset 5 and Dataset 6 with more than 7.5%.

Table 2. The overall accuracy (%) from the classification results for different datasets.

Data	Only Spectral Bands		Spectral + Auxiliary Variables	
	No. Bands	OA	No. Bands	OA
Dataset 1	7	78.25	20	85.95
Dataset 2	7	81.23	20	88.74
Dataset 3	7	80.46	20	85.08
Dataset 4	7	80.17	20	84.31
Dataset 5	7	77.66	20	85.27
Dataset 6	7	78.15	20	85.66
Dataset 7	28	85.08	41	89.80
Dataset 8	28	85.27	41	89.70

This increase can be explained by the order of variable importance (Figure 4) of the 20 input features band (Dataset 1 to Dataset 6) and 41 features band of Dataset 7 and Dataset 8. In all datasets, elevation always ranked as the most important variable in the classification, followed by the entropy, B7 and B1 (except Dataset 6, where B7 was ranked at 8th position).

Regarding time series data (Dataset 7 and Dataset 8), the 3 most important variables were elevation, entropy and Band 5 of July composite image. Spectral indices were ranked very low in all datasets. Regarding the three topography indices, elevation always featured high importance values in all datasets, whereas aspect and slope contributed only marginally to the models in all datasets.

3.2. The Effect of Different Composition Datasets on Land Cover Classification Accuracy

In the following, we only report results based on spectral bands (1 to 7) and auxiliary variables to focus on the effects of different composition strategies on the classification accuracy. In general, the maps show that all classifications have a consistent pattern of land cover types with some slight differences (Figure 5). For instance, there is more agricultural land at the southwest corner in Datasets 1, 3, 4 and 6. Some additional agricultural land occur at the lower right corner of classification of Dataset 5. Based on datasets 1, 3, 4 and 6 the models also classified more grassland at the southwest corner.
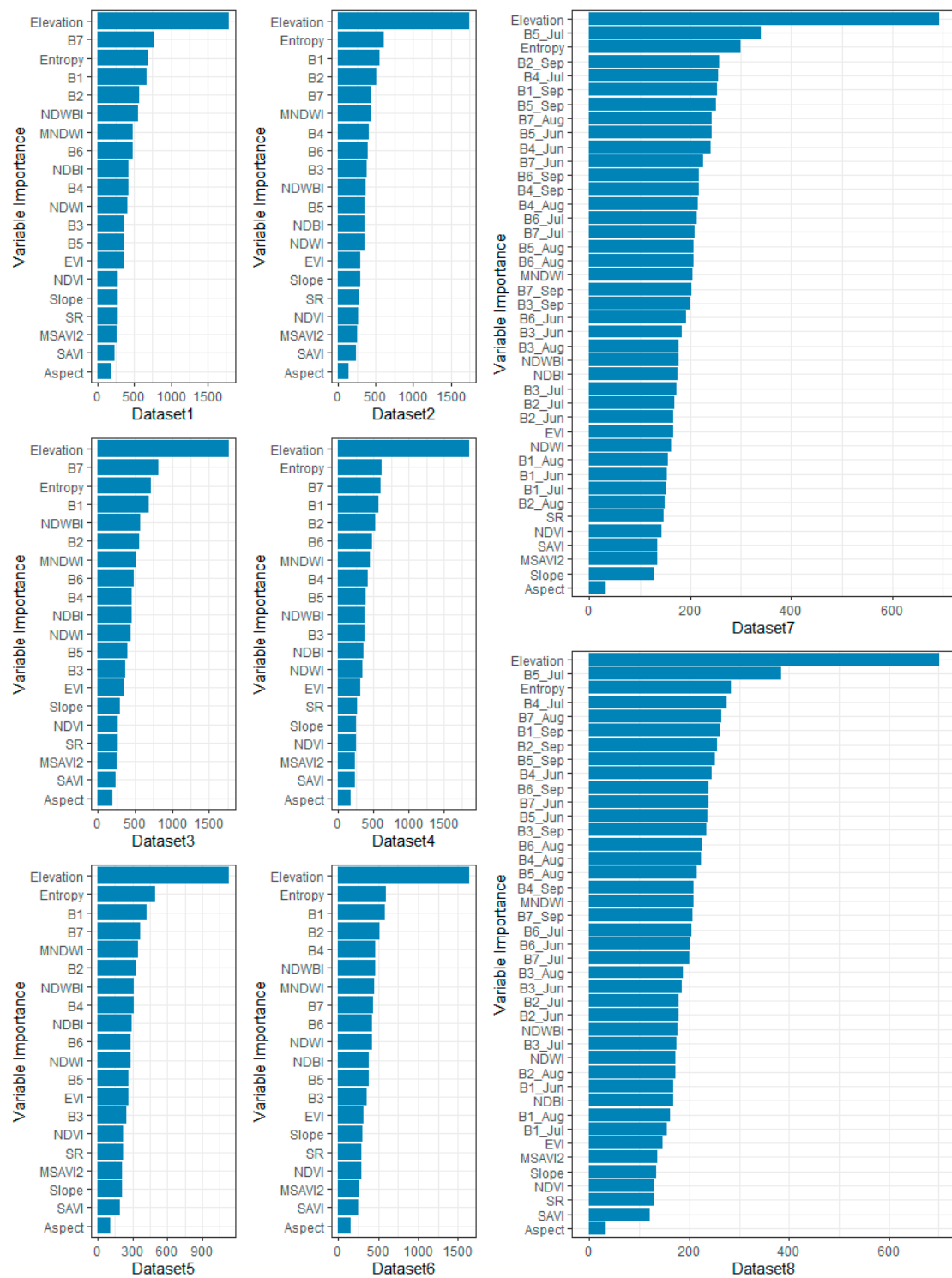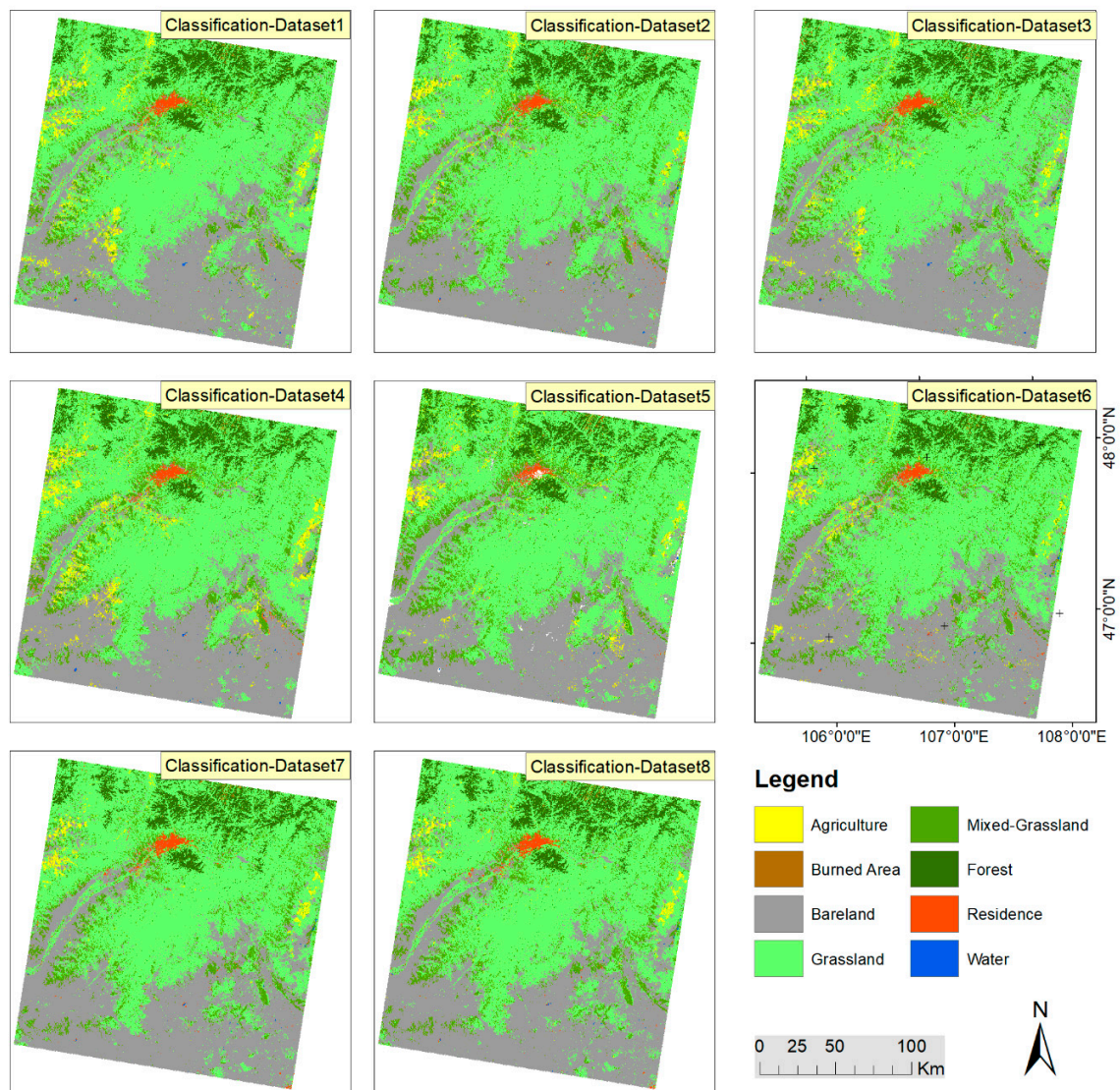
Figure 4. Variable importance in the Random Forest (RF) models trained on the eight datasets.

Figure 5. Land cover classification of the eight datasets.

In general, all datasets produced high accuracies (OA ranges from 84.31% to 89.80%, Table 3). On average, burned area (BA) and residence area were classified with the highest accuracy, followed by grassland, mixed grassland and water. The lowest accuracy was observed with agriculture and forest. These low accuracies are caused by the very low PA of Dataset 4, Dataset 5 and Dataset 6 for agriculture and forest class.

Among all datasets, Dataset 7 and Dataset 8 produced the highest OA (89.80 and 89.70%, respectively), followed by Dataset 2 (88.74), Dataset 1 (85.95%), Dataset 6 (85.66%), Dataset 5 (85.27%), Dataset 3 (85.08%) and the lowest OA was with Dataset 4 (84.31%). However, these results do not show whether these differences are significant. Therefore, the McNemar test (Table 4) was applied to indicate the significant difference between classification results.

Regarding the difference between Dataset 1 and Dataset 2, which were median composited from all the images in 2019 (196 images) and the images from June to September in 2019 (61 images) over the study area, accuracies of Dataset 2 (88.74%) were higher than those of Dataset 1 (85.95%) and the difference was significant ($p < 0.05$, Table 4).

Table 3. User's (UA), producer's (PA) and overall (OA) accuracies of land cover classes from the classification results of the eightifferent datasets.

		AG	BA	BL	GR	GRm	RE	FR	WA	OA
Dataset 1	PA	70.54	85.71	86.59	94.00	82.35	97.60	75.00	80.00	85.95
	UA	88.76	100.00	78.68	76.05	92.31	98.39	89.66	95.65	
Dataset 2	PA	75.89	94.64	89.94	94.80	86.27	97.60	72.12	87.27	88.74
	UA	92.39	100.00	80.50	80.07	90.72	99.19	97.40	100.00	
Dataset 3	PA	69.64	83.04	86.03	93.60	83.33	98.40	69.23	81.82	85.08
	UA	90.70	100.00	78.97	72.22	92.39	99.19	90.00	100.00	
Dataset 4	PA	53.57	95.54	81.01	94.80	85.29	98.40	68.27	83.64	84.31
	UA	78.95	99.07	73.98	76.70	87.00	97.62	91.03	100.00	
Dataset 5	PA	66.67	93.75	87.71	94.40	85.29	97.60	62.34	52.38	85.27
	UA	89.16	100.00	73.71	80.00	93.55	98.39	77.42	95.65	
Dataset 6	PA	53.57	96.43	92.74	93.20	81.37	99.20	71.15	76.36	85.66
	UA	88.24	100.00	75.45	78.19	88.30	99.20	88.10	100.00	
Dataset 7	PA	80.36	93.75	94.97	94.80	85.29	99.20	78.85	69.09	89.8
	UA	97.83	100.00	82.13	84.34	90.63	100.00	85.42	100.00	
Dataset 8	PA	78.57	91.96	94.41	95.60	87.25	99.20	80.77	65.45	89.7
	UA	98.88	100.00	83.25	84.45	89.90	99.20	83.17	100.00	

Table 4. Matrix of McNemar test showing the statistical significance of differences between all classification pairs. McNemar's test values (z^2) are on the left side of the diagonal, *p*-values on the right side. Bold value indicates significantly different classification ($p < 0.05$).

	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6	Dataset 7	Dataset 8
Dataset 1		**<0.05**	0.208	0.161	0.480	0.805	**<0.05**	**<0.05**
Dataset 2	**6.05**		**<0.05**	**<0.05**	**<0.05**	**<0.05**	0.313	0.353
Dataset 3	1.59	**9.89**		0.537	0.886	0.620	**<0.05**	**<0.05**
Dataset 4	1.97	**16.28**	0.38		0.713	0.227	**<0.05**	**<0.05**
Dataset 5	0.50	**9.0**	0.02	0.14		**<0.05**	**<0.05**	**<0.05**
Dataset 6	0.06	**9.85**	0.25	1.46	7.01		**<0.05**	**<0.05**
Dataset 7	**15.09**	1.02	**19.84**	**22.10**	**38.21**	7.75		0.858
Dataset 8	**13.46**	0.86	**18.00**	**21.78**	**37.07**	**15.75**	0.03	

Comparing the results of Dataset 2 and Dataset 4, which were composited from all images between June and September of one year (2019, 61 images) and two years (2018 + 2019, 126 images), respectively, Dataset 2 always produces higher accuracy of all land cover types (except PA of burned area and residence class) both regarding PA and UA (Table 3) and this difference (Dataset 2 and Dataset 4) is significant ($p < 0.05$). Therefore, if images within one year (or season) can fill all the gaps (e.g., due to cloud), then the Dataset 2 should be applied instead of Dataset 4.

Comparing between the median composition images and time series images that use the same input images, Dataset 2 versus Dataset 7, Dataset 6 versus Dataset 8, as shown in Table 3, the results of Dataset 2 and Dataset 7 were slightly different (OA = 88.74% and 89.80%, respectively). The largest difference was observed in Water, Agriculture and Bare land classes. Water has higher accuracy in Dataset 2, agriculture and bare land have higher accuracies in Dataset 7. However, this difference is not significant ($p = 0.31$). In contrast, the difference between Dataset 6 and Dataset 8 is significant ($p < 0.05$). This indicates that the automatic selection (by study area, i.e., Dataset 1) and the manual select scenes (by path and row, that is, Dataset 6) together with composition methods (median vs. time series composition) have an effect on the classification results. Therefore, in applications both the trade-off between median and time series composition, as well as the image included in the collection should be taken into account.

An interesting result is the comparison between Dataset 5 (single date imagery), Dataset 6 (median composite from the 7 images) and Dataset 8 (time series image of the 7 images) of the single scene (path 131, row 27). Although, there is an apparently small difference between OA of Dataset 5 and Dataset 6 (85.27% vs. 85.66%), they are significantly different. Looking at all the eight datasets, as shown in Table 4, there are significant differences ($p < 0.05$) of classification results between time series data

(Dataset 7 and Dataset 8) with all other compositions, except Dataset 2. Furthermore, the classification result of Dataset 2 is also significantly different from other composition's classification results (except the results of Dataset 7 and Dataset 8). It is suggested that the input images for the compositions play a crucial role in producing significantly higher accuracies of land cover maps. In this case, Dataset 2, Dataset 7 and Dataset 8 were composited from images between June and September within one year. It is worth to remind that Dataset 4, also composited from all images between June and September 2019, however, it includes images between June and September in 2018. That is why, it produced significant lower ($p < 0.05$) accuracy in comparison to those of Dataset 2, Dataset 7 and Dataset 8.

3.3. Variation of Land Cover Types Derived from Different Datasets

From all 8 single classifications, we calculated the mode classification symbolizing for each pixel the class with the highest agreement within the single classifications (Figure 6a). Only small differences between the mode classification and the classification on Dataset 7 featuring the highest accuracy occurred (Figure 6b showing differences in the spatial subset around Ulaanbaatar). If differences between mode classification and all other classifications are assessed, large discrepancies are obvious (Figure 6c for the same spatial subset and Figure 6d for a graphical analysis based on the entire area). Especially at the borders of Ulaanbaatar, the classifications results differed heavily. All classifications agreed widely for the spatial estimation of forest occurrence. Overall, 76% of the pixels assigned to the forest class in the mode image were also assigned to forest in all other classifications. Another 22% of the forest pixels in the mode classification were classified into one another class in at least one classification. For all other classes the coincidence among the classifications was much lower. The lowest coincidence among the classifications was observed for Agriculture, where only 12% of the mode classification pixels were assigned to Agriculture in all other classes.
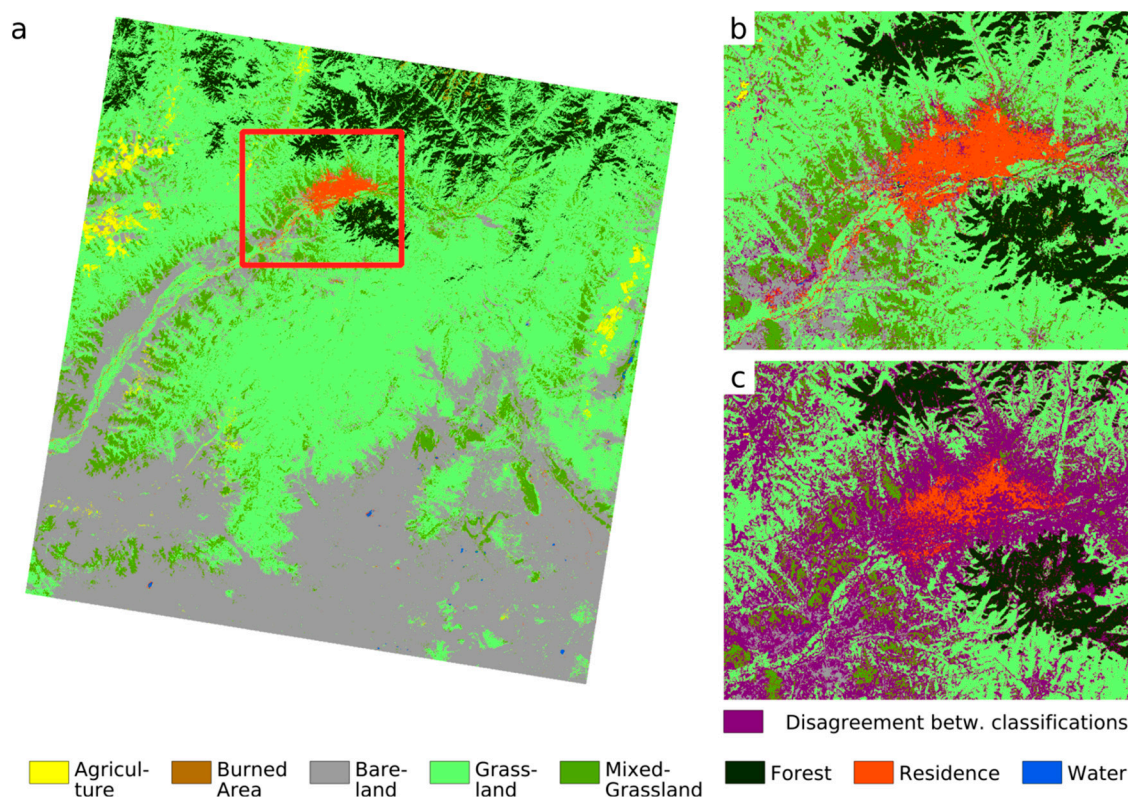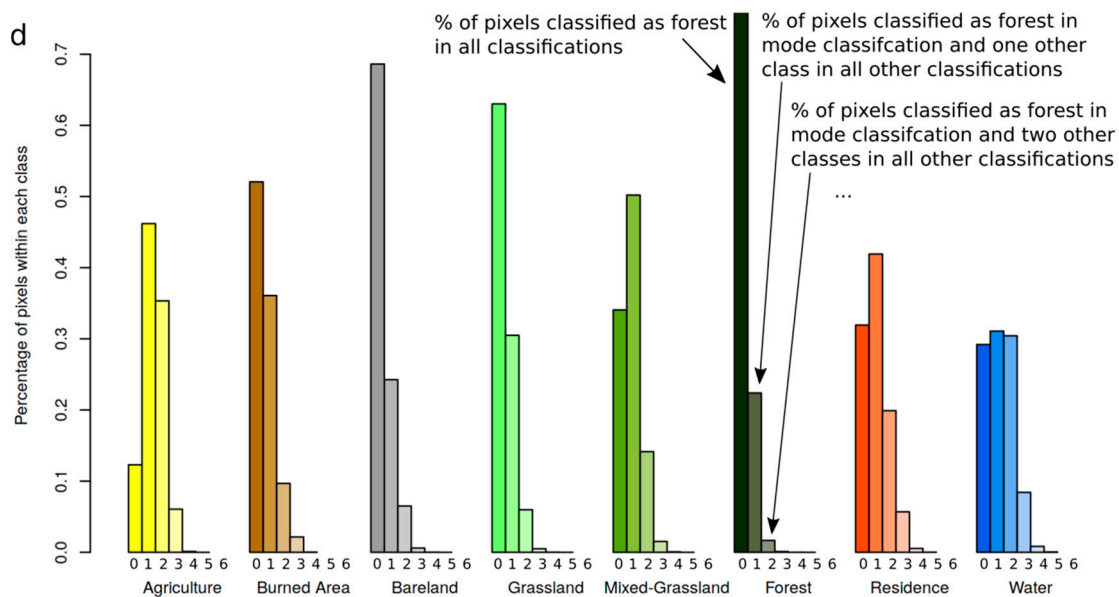

Figure 6. *Cont.*

Figure 6. (**a**) Mode of classifications giving the class with the highest frequency per pixel among all 8 classifications. (**b**) Agreement between mode of classifications and the classification with the highest OA (classification based on Dataset 7) for the area around Ulaanbaatar indicated by the red box in (**a**). Pink areas mark disagreement. (**c**) Same as (**b**) but showing disagreement between all single classifications. In this respect, all pixels are colored in pink that are classified into at least two different classes considering all classifications. (**d**) Frequency of class numbers observed among all classifications given as class wise percentage of pixels in mode classification. The bars marked with zero show the percentage of pixels with agreement in all classifications. For instance, 76% of forest pixels (dark green bar) in mode classification are classified as forest in all other classifications; 22% of forest pixels in mode classification are put into another class than forest in at least one other classification; 2% of forested pixels in mode classification are put into 2 other classes. Please note that numbers in (**d**) are calculated for the entire area of the classifications.

4. Discussion

In this study, we test different composition methods (metrics, time series and the different annual/seasonal composition strategies) to generate spectral input data for land cover classifications and investigate how this procedure affects the classification accuracy. Our results show that accuracies were generally high, but, although all classifications were trained with and validated against the same reference dataset, the strategies used to select images led to significantly different results.

Prior to the availability of GEE, many different composition methods were proposed to solve the problem of cloud cover. For example, Senf et al. [81] used fusion techniques to simulate data for the missing areas. Inglada et al. [82] used linear interpolation to fill the invalid pixels by using the previous and/or following cloud-free-date pixels to map land cover over France. However, since GEE has become available, the most popular method for filling gaps in cloudy images is to use median metrics (temporal aggregation method). Carrasco et al. [39] report that the advantage of this method is the significant reduction of data volume, resulting in an easier and faster analysis. For our study, the volume of the median data was approximately 4.3 GB, meanwhile the volume of the time series data (Dataset 7 and Dataset 8) were 20.8 GB and 13.5 GB, respectively. Obviously, such large data volumes are not easy to handle and analyze on a personal computer (PC) with remote sensing software. However, on the GEE platform in particular, this task can be completed easily; hundreds of images can be rapidly processed [83]. Using the median composition method, the input images (typically annual composition) are created in a pixel-wise manner by taking the median value (i.e., DN, TOA or reflectance) from all cloud-free pixels of the image collection. It is worth mentioning that, besides the median method, in GEE there are also other methods available such as mean, minimum, maximum,

standard deviation and percentile. However, in the literature, the most popular method is the median reducer. For our case, we also tested the mean composition (the result is not shown), but this method produced lower accuracy compared to the median composition method. Therefore, the median composition was selected in this study. Most importantly, if the input images used for composition are optimally selected, the median composite image (with low data volume) can produce results which are as good as those based on time series composites (Dataset 2 vs. Dataset 7, Table 4), whereas a significantly different classification result could be achieved if the median image is not optimally created (e.g., Dataset 6 and Dataset 8).

The strategies for selecting images in a collection vary between studies as a consequence of cloud cover and types of land cover. Richards & Belcher [47] used images over three years to create composites (i.e., images from 1999–2001 and 2014–2016 to map land cover in 2000 and 2015, respectively). Although, the land cover has only two classes (Vegetated Cover and Unvegetated Cover) and the accuracy was assessed based on 293 validation points, the overall error rate was (with 9.2% and 7.5%) quite high for the classification in 2000 and 2015, respectively. This is consistent with our results, in that all images available in a year produces lower accuracy than seasonal image composition (i.e., Dataset 1 vs. Dataset 3 produced lower accuracy compared to Dataset 2); in addition, composition image from multi-years' data produce land cover maps with lower accuracy than those from a single year (i.e., Dataset 2 vs. Dataset 4). This is consistent with the recommendation of Frantz et al. [84] that the tradeoff between different years (i.e., land cover might change) and large different days in the same years (phenological consistency might be lost) should be taken into consideration. In this case, using data from the same year is preferable. Nevertheless, the most popular method to select images is based on the seasonal image composition. For example, to map land cover for Central Asia, Hu & Hu [46] selected only images between April and July in the span of three years (target year ±1 year) for mapping land cover in 2001 and 2017. Meanwhile, Nyland et al. [44], selected images between July 1st to September 30th (to minimize the effect of snow cover) for land cover mapping in the transitional zone between physiographic provinces in the West Siberian Plains and on the Central Siberian Plateau.

As far as we know, no study has investigated the difference between selecting images based on the seasonal image and selecting as many images as possible to fill all the missing pixels due to cloud and snow cover. Our results show that (Tables 3 and 4), using the same composition method (i.e., median) for different input images (image collection) might produce significantly different land cover classification accuracy. For example, of Dataset 1, Dataset 2 and Dataset 4, which have different input images, Dataset 2 produced the highest accuracy (88.74%) and was significantly more different than Dataset 1 and Dataset 4. The classification accuracy of Dataset 1 is higher than that of Dataset 4. These different results can be explained because most land cover types in our study area have strong seasonality, such as grassland, mixed grassland and agriculture. Even the water might seasonally change, that is, it could be snow/ice in winter and bare land in summer. Furthermore, due to the changing condition of clouds, the median composition of input images from different years (i.e., Dataset 4) could affect the phenology information, for example, the missing pixels in June 2018 could be filled by pixels from July 2019. Such effects would be large for study areas where land cover types have a strong seasonality, like in our study site. That is why the classification result of Dataset 2 was significantly more accurate than that of Dataset 4. Another image input selection strategy is shown in Dataset 1 and Dataset 3, where images were selected based on the percentage of cloud cover threshold. In our case, we selected all scenes over the study area in 2019 with cloud cover less than 30%. This selection highly depends on cloud cover conditions and is therefore not recommended, particularly when the study aims to produce seasonal land cover maps. For example, in Dataset 1 and Dataset 3 of the present study (Table 1), there were 196 and 130 images, in which 61 and 37 images were acquired between June and September 2019, respectively. This means that there were 135 useless images in Dataset 1 and 93 useless images Dataset 3. That is why the results of Dataset 1 and Dataset 3, although they used a very high number of images, produced significantly lower accuracies compared to Dataset 2.

The selection of images is also crucial regarding the usage of images from multiple paths and rows. To our knowledge, this has not been discussed in the literature yet. As in our study area, the final composite can be either covered by data from a single row/path combination or by merging data from multiple rows and paths (Figure 2). In GEE, if images are selected based on a study area, the engine will automatically consider data from 9 Landsat footprints (Dataset 1, Dataset 2, Dataset 3, Dataset 4 and Dataset 7). For our other datasets, we specifically selected the input images so that the scenes come from the same Landsat footprint (path 131–row 27). The results (Table 4) show that if these image collections (automatic selection vs manual selection) were used for time series composition, the classification results were similar such as Dataset 7 and Dataset 8. However, if these image collections were used for median composition, the classification results were significant different (Dataset 2 vs. Dataset 6). Therefore, in GEE application, where the median composition method is widely applied, this different selection issue should be taken into account when processing input images for land cover classifications.

Another point that should be mentioned is the improvement of classification accuracy after incorporating auxiliary variables as additional predictors (Table 2). In this study, we used 13 auxiliary variables (10 spectral and 3 topography indices). The 10 spectral indices were calculated based on the correspondent spectral data input (Figure 3). For example, "auxiliary variable 1" was calculated from "spectral data 1." Particularly, in Dataset 7 and Dataset 8, we tested the "time series indices" (i.e., monthly auxiliary variable indices), however, the performance of these indices were not as good as the indices calculated from the median images. Therefore, the auxiliary variables of Dataset 7 and Dataset 8 were used as those of Dataset 2 and Dataset 6. The result (Tables 3 and 4) indicates that selecting input images as well as the auxiliary variables for the classification can significantly improve the classified results. Popular spectral indices (e.g., NDVI, EVI, SAVI) have been reported to increase the accuracy of land cover classification using remotely sensed images. However, in this study, these indices always had a low rank compared to other feature bands (Figure 4). The reason is that vegetation is the dominant cover type in the study area and it changes (green to brown) quickly. Due to the climatic gradient across our study area, these changes are not consistent in time. Furthermore, the agricultural activities across the area regarding planting and harvesting times change according to the crops and the climate. This is consistent with the studies of Abdi [85] and Zha et al. [86].

NDBI and mNDWI are two other popular indices, which also ranked very low. For NDBI, this can be explained by the residence cover area in Mongolia, which is not easy to be separated from bare land. In the case of mNDWI, the index could not help to separate water and shadowed surfaces (Figure A1). This is in line with the study of Feyisa et al. [87].

In this study, Elevation and Entropy are always the two most important variables regardless of the dataset (Figure 4). As shown in Figures 5 and A1, there is a consistent distribution of land cover types (Figure 5) and elevation (Figure A1), it is suggested that elevation can help in separating grassland and other vegetation or grassland and bare land, as well as bare land versus resident area. Entropy can help to distinguish between grassland, agricultural land and mixed-grassland.

In addition to the accuracy reported in the confusion matrix, looking at the classification results (land cover maps), there is more grassland and agricultural land (Figure 5, west-south corner) in classification results of Dataset 1, Dataset 3 and Dataset 4. Obviously, these differences could lead to high uncertainties for any subsequent application (e.g., biomass estimation) using the map as input. This is again confirming that image selection plays a crucial role in land cover classification accuracy, particularly, with a specific target land cover class.

If the consistency of the classifications over space is directly assessed, we find surprisingly high differences among the classifications. Although the OAs of all single classifications were above 84%, for many classes the majority of the pixels were assigned to another class in at least one classification. This causes high uncertainties arising solely from the method how the input data is preprocessed and how many and which predictors are used. This clearly shows that the question of how to generate a

cloud- and snow free input dataset for LUCC in GEE is an important and crucial step, whose effects on the classification results is generally underestimated.

5. Conclusions

In a vast country like Mongolia, both cloud and snow cover are present and the land cover changes over the course of the season. It is not possible to use a single date or even monthly composite images to map land cover. Mosaicking, stacking and filtering (selecting) the images are always required in order to create a satisfactory land cover map from the remotely sensed images. These steps are more easily accessible since the availability of GEE, as it would otherwise be very costly and labor intensive. Our pilot experiment study shows that different strategies to select the input images produce different classification results (with OA ranging from 77.66% to 89.80%). The results show that in order to achieve a highly accurate and realistic land cover map, it is not enough simply to select all the available satellite images (i.e., Dataset 1), filter the images based on the cloud cover threshold (i.e., Dataset 3), or fill in the cloud cover pixels with pixels from the same season in another year (i.e., Dataset 4). The optimal solution to select images for the collection (the input image for the median or time series composition) is based on cloud/snow cover and the land cover types. Regarding the composition methods (i.e., median vs. time series), our results show that if the input images (image collection) are optimally selected (i.e., collection 2 of Dataset 2), the median composite image (low data volume) can produce equally high accuracy as time series composite data does (Dataset 2 vs Dataset 7, Table 4). Therefore, we highly recommend that when implementing land cover classification on the GEE platform, the median composition method should be given priority over the time series composition, because it is not only filling the gaps due to cloud and snow cover but also reducing data volume (i.e., easier and faster analysis) and producing as high accuracy as time series (multi-temporal) image data.

To date, both Landsat (TM, ETM+ and OLI) and Sentinel-2 (A and B) are available in GEE, meaning that all the above mentioned issues (i.e., cloud cover, snow cover, lack of close time acquisition images of Landsat 8 solely) can be solved by integrating Landsat 7, Landsat 8 and Sentinel-2. However, it should be noted that in order to integrate these different sensor images, the differences between the band wavelengths of these sensors need to be taken into account. A handful of studies in the literature combine Landsat7, Landsat 8 and Sentinel-2 for land cover classification while considering the mentioned differences. Therefore, future studies should investigate the optimal method to integrate these different sensor images in order to produce the best land cover map for large areas like Mongolia.

Appendix A

Table A1. Selected auxiliary variables.

| Index	Formula	References
NDVI	(NIR − RED)/(NIR + RED)	Rouse et al. [88]
EVI	2.5 * ((NIR − RED)/(NIR + 6 * RED − 7.5 * BLUE + 1))	Liu & Huete [89]
SAVI	(NIR − RED)/(NIR + RED + 0.5) * (1.5)	Huete [90]
MSAVI2	$(2 * NIR + 1 − SQRT((2 * NIR + 1)^2 − 8 * (NIR − RED)))/2$	Qi et al. [91]
NDWI	(NIR − SWIR)/(NIR + SWIR)	Gao [92]
mNDWI	(GREEN − SWIR)/(GREEN + SWIR)	Xu [93]
NDWBI	(GREEN − NIR)/(GREEN + NIR)	McFeeters [94]
NDBI	(SWIR − NIR)/(SWIR + NIR)	Zha et al. [86]
SR	NIR/RED	Birth & McVey [95]
Entropy	entropy of the NIR band were selected from the 4 × 4 local window	Jia et al. [96]
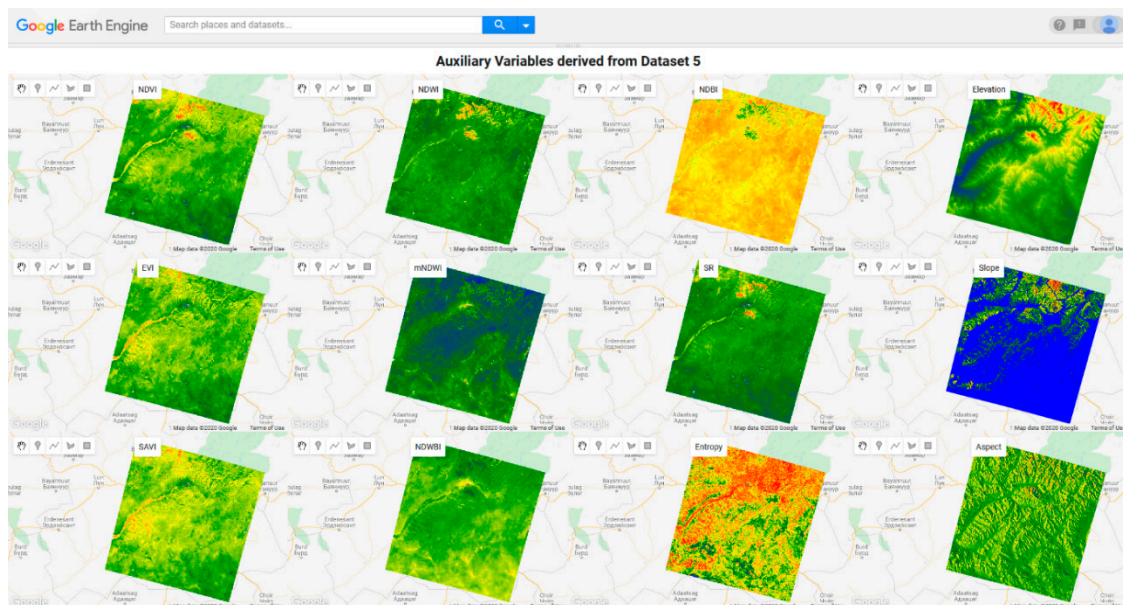
Figure A1. An example of auxiliary variables derived from Dataset 5. Color from blue to red presents the value from low to high of auxiliary variables.

References

1. Herold, M. Assessment of the Status of the Development of the Standards for the Terrestrial Essential Climate Variables. In *Land. Land Cover*; FAO: Rome, Italy, 2009.
2. Koschke, L.; Fürst, C.; Frank, S.; Makeschin, F. A multi-criteria approach for an integrated land-cover-based assessment of ecosystem services provision to support landscape planning. *Ecol. Indic.* **2012**, *21*, 54–66. [CrossRef]
3. Sterling, S.M.; Ducharne, A.; Polcher, J. The impact of global land-cover change on the terrestrial water cycle. *Nat. Clim. Chang.* **2012**, *3*, 385–390. [CrossRef]
4. Salazar, A.; Baldi, G.; Hirota, M.; Syktus, J.; McAlpine, C. Land use and land cover change impacts on the regional climate of non-Amazonian South America: A review. *Glob. Planet. Chang.* **2015**, *128*, 103–119. [CrossRef]
5. Niquisse, S.; Cabral, P.; Rodrigues, Â.; Augusto, G. Ecosystem services and biodiversity trends in Mozambique as a consequence of land cover change. *Int. J. Biodivers. Sci. Ecosyst. Serv. Manag.* **2017**, *13*, 297–311. [CrossRef]

6. Beer, C.; Reichstein, M.; Tomelleri, E.; Ciais, P.; Jung, M.; Carvalhais, N.; Rodenbeck, C.; Arain, M.A.; Baldocchi, D.; Bonan, G.B.; et al. Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science* **2010**, *329*, 834–838. [CrossRef]

7. Pan, Y.; Birdsey, R.A.; Fang, J.; Houghton, R.; Kauppi, P.E.; Kurz, W.A.; Phillips, O.L.; Shvidenko, A.; Lewis, S.L.; Canadell, J.G.; et al. A large and persistent carbon sink in the world's forests. *Science* **2011**, *333*, 988–993. [CrossRef]

8. Reichstein, M.; Bahn, M.; Ciais, P.; Frank, D.; Mahecha, M.D.; Seneviratne, S.I.; Zscheischler, J.; Beer, C.; Buchmann, N.; Frank, D.C.; et al. Climate extremes and the carbon cycle. *Nature* **2013**, *500*, 287–295. [CrossRef]

9. Ahlström, A.; Xia, J.; Arneth, A.; Luo, Y.; Smith, B. Importance of vegetation dynamics for future terrestrial carbon cycling. *Environ. Res. Lett.* **2015**, *10*. [CrossRef]

10. Bengtsson, J.; Bullock, J.M.; Egoh, B.; Everson, T.; O'Connor, T.; O'Farrell, P.J.; Smith, H.G.; Lindborg, R. Grasslands—more important for ecosystem services than you might think. *Ecosphere* **2019**, *10*, e02582. [CrossRef]

11. Fernández-Giménez, M.E.; Batkhishig, B.; Batbuyan, B. Cross-boundary and cross-level dynamics increase vulnerability to severe winter disasters (dzud) in Mongolia. *Glob. Environ. Chang.* **2012**, *22*, 836–851. [CrossRef]

12. Reid, R.S.; Fernández-Giménez, M.E.; Galvin, K.A. Dynamics and Resilience of Rangelands and Pastoral Peoples around the Globe. *Annu. Rev. Environ. Resour.* **2014**, *39*, 217–242. [CrossRef]

13. Khishigbayar, J.; Fernández-Giménez, M.E.; Angerer, J.P.; Reid, R.S.; Chantsallkham, J.; Baasandorj, Y.; Zumberelmaa, D. Mongolian rangelands at a tipping point? Biomass and cover are stable but composition shifts and richness declines after 20 years of grazing and increasing temperatures. *J. Arid Environ.* **2015**, *115*, 100–112. [CrossRef]

14. Fernández-Giménez, M.E.; Venable, N.H.; Angerer, J.; Fassnacht, S.R.; Reid, R.S.; Khishigbayar, J. Exploring linked ecological and cultural tipping points in Mongolia. *Anthropocene* **2017**, *17*, 46–69. [CrossRef]

15. Dashpurev, B.; Bendix, J.; Lehnert, L. Monitoring Oil Exploitation Infrastructure and Dirt Roads with Object-Based Image Analysis and Random Forest in the Eastern Mongolian Steppe. *Remote Sens.* **2020**, *12*, 144. [CrossRef]

16. McNaughton, S.J. Grazing as an optimization process: Grass-ungulate relationships in the Serengeti. *Am. Nat.* **1979**, *113*, 691–703. [CrossRef]

17. Tilman, D.; Wedin, D.; Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **1996**, *379*, 718–720. [CrossRef]

18. Tilman, D.; Reich, P.B.; Knops, J.; Wedin, D.; Mielke, T.; Lehman, C. Diversity and productivity in a long-term grassland experiment. *Science* **2001**, *294*, 843–845. [CrossRef] [PubMed]

19. Jacobs, S.M.; Naiman, R.J. Large African herbivores decrease herbaceous plant biomass while increasing plant species richness in a semi-arid savanna toposequence. *J. Arid Environ.* **2008**, *72*, 891–903. [CrossRef]

20. Leisher, C.; Hess, S.; Boucher, T.M.; Beukering, P.; Sanjayan, M. Measuring the impacts of community-based grasslands management in Mongolia's Gobi. *PLoS ONE* **2012**, *7*, e30991. [CrossRef]

21. Skole, D.S.; Justice, C.O.; Janetos, A.; Townshend, J.R.G. A land cover change monitoring program: A strategy for international effort. In *Mitigation and Adaptation Strategies for Global Change*; Kluwer: Amsterdam, The Netherlands, 1997; pp. 1–19.

22. Lautenbacher, C.C. The Global Earth Observation System of Systems: Science Serving Society. *Space Policy* **2006**, *22*, 8–11. [CrossRef]

23. Bontemps, S.; Herold, M.; Kooistra, L.; van Groenestijn, A.; Hartley, A.; Arino, O.; Moreau, I.; Defourny, P. Revisiting land cover observation to address the needs of the climate modeling community. *Biogeosciences* **2012**, *9*, 2145–2157. [CrossRef]

24. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [CrossRef]

25. Li, C.; Gong, P.; Wang, J.; Zhu, Z.; Biging, G.S.; Yuan, C.; Hu, T.; Zhang, H.; Wang, Q.; Li, X. The first all-season sample set for mapping global land cover with Landsat-8 data. *Sci. Bull.* **2017**, *7*, 508–515. [CrossRef]

26. Wulder, M.A.; White, J.C.; Loveland, T.R.; Woodcock, C.E.; Belward, A.S.; Cohen, W.B.; Fosnight, E.A.; Shaw, J.; Masek, J.G.; Roy, D.P. The global Landsat archive: Status, consolidation and direction. *Remote Sens. Environ.* **2016**, *185*, 271–283. [CrossRef]

27. Disperati, L.; Virdis, S.G.P. Assessment of land-use and land-cover changes from 1965 to 2014 in Tam Giang-Cau Hai Lagoon, central Vietnam. *Appl. Geogr.* **2015**, *58*, 48–64. [CrossRef]

28. Reiche, J.; Verbesselt, J.; Hoekman, D.; Herold, M. Fusing Landsat and SAR time series to detect deforestation in the tropics. *Remote Sens. Environ.* **2015**, *156*, 276–293. [CrossRef]

29. Zhu, Z.; Woodcock, C.E.; Rogan, J.; Kellndorfer, J. Assessment of spectral, polarimetric, temporal and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. *Remote Sens. Environ.* **2012**, *117*, 72–82. [CrossRef]

30. Wan, B.; Guo, Q.; Fang, F.; Su, Y.; Wang, R. Mapping US Urban Extents from MODIS Data Using One-Class Classification Method. *Remote Sens.* **2015**, *7*, 10143–10163. [CrossRef]

31. Xin, Q.; Olofsson, P.; Zhu, Z.; Tan, B.; Woodcock, C.E. Toward near real-time monitoring of forest disturbance by fusion of MODIS and Landsat data. *Remote Sens. Environ.* **2013**, *135*, 234–247. [CrossRef]

32. Thanh Noi, P.; Kappas, M. Comparison of Random Forest, k-Nearest Neighbor and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* **2017**, *18*, 18. [CrossRef]

33. Lambert, M.-J.; Traoré, P.C.S.; Blaes, X.; Baret, P.; Defourny, P. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sens. Environ.* **2018**, *216*, 647–657. [CrossRef]

34. Rapinel, S.; Mony, C.; Lecoq, L.; Clément, B.; Thomas, A.; Hubert-Moy, L. Evaluation of Sentinel-2 time-series for mapping floodplain grassland plant communities. *Remote Sens. Environ.* **2019**, *223*, 115–129. [CrossRef]

35. Furberg, D.; Ban, Y.; Nascetti, A. Monitoring of Urbanization and Analysis of Environmental Impact in Stockholm with Sentinel-2A and SPOT-5 Multispectral Data. *Remote Sens.* **2019**, *11*, 2408. [CrossRef]

36. Kuenzer, C.; Ottinger, M.; Wegmann, M.; Guo, H.; Wang, C.; Zhang, J.; Dech, S.; Wikelski, M. Earth observation satellite sensors for biodiversity monitoring: Potentials and bottlenecks. *Inter. J. Remote Sens.* **2014**, *35*, 6599–6647. [CrossRef]

37. Mack, B.; Leinenkugel, P.; Kuenzer, C.; Dech, S. A semi-automated approach for the generation of a new land use and land cover product for Germany based on Landsat time-series and Lucas in-situ data. *Remote Sens. Lett.* **2017**, *8*, 244–253. [CrossRef]

38. Wulder, M.A.; White, J.C.; Goward, S.N.; Masek, J.G.; Irons, J.R.; Herold, M.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sens. Environ.* **2008**, *112*, 955–969. [CrossRef]

39. Carrasco, L.; O'Neil, A.; Morton, R.; Rowland, C. Evaluating Combinations of Temporally Aggregated Sentinel-1, Sentinel-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine. *Remote Sens.* **2019**, *11*, 288. [CrossRef]

40. 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]

41. Tamiminia, H.; Salehi, B.; Mahdianpari, M.; Quackenbush, L.; Adeli, S.; Brisco, B. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 152–170. [CrossRef]

42. Kumar, L.; Mutanga, O. Remote Sensing of Above-Ground Biomass. *Remote Sens.* **2017**, *9*, 935. [CrossRef]

43. Beckschäfer, P. Obtaining rubber plantation age information from very dense Landsat TM & ETM + time series data and pixel-based image compositing. *Remote Sens. Environ.* **2017**, *196*, 89–100. [CrossRef]

44. E Nyland, K.; EGunn, G.; IShiklomanov, N.; NEngstrom, R.; AStreletskiy, D. Land Cover Change in the Lower Yenisei River Using Dense Stacking of Landsat Imagery in Google Earth Engine. *Remote Sens.* **2018**, *10*, 1226. [CrossRef]

45. Xie, S.; Liu, L.; Zhang, X.; Yang, J.; Chen, X.; Gao, Y. Automatic Land-Cover Mapping using Landsat Time-Series Data based on Google Earth Engine. *Remote Sens.* **2019**, *11*, 3023. [CrossRef]

46. Hu, Y.; Hu, Y. Land Cover Changes and Their Driving Mechanisms in Central Asia from 2001 to 2017 Supported by Google Earth Engine. *Remote Sens.* **2019**, *11*, 554. [CrossRef]

47. Richards, D.R.; Belcher, R.N. Global Changes in Urban Vegetation Cover. *Remote Sens.* **2019**, *12*, 23. [CrossRef]

48. Griffiths, P.; van der Linden, S.; Kuemmerle, T.; Hostert, P. A Pixel-Based Landsat Compositing Algorithm for Large Area Land Cover Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2088–2101. [CrossRef]

49. Zhu, Z.; Woodcock, C.E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **2014**, *144*, 152–171. [CrossRef]

50. Hermosilla, T.; Wulder, M.A.; White, J.C.; Coops, N.C.; Hobart, G.W. Disturbance-Informed Annual Land Cover Classification Maps of Canada's Forested Ecosystems for a 29-Year Landsat Time Series. *Can. J. Remote Sens.* **2018**, *44*, 67–87. [CrossRef]

51. Roy, D.P.; Kovalskyy, V.; Zhang, H.K.; Vermote, E.F.; Yan, L.; Kumar, S.S.; Egorov, A. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sens. Environ.* **2016**, *185*, 57–70. [CrossRef]

52. Griffiths, P.; van der Linden, S.; Kuemmerle, T.; Hostert, P. Erratum: A pixel-based landsat compositing algorithm for large area land cover. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2088–2101.

53. Denize, J.; Hubert-Moy, L.; Betbeder, J.; Corgne, S.; Baudry, J.; Pottier, E. Evaluation of Using Sentinel-1 and -2 Time-Series to Identify Winter Land Use in Agricultural Landscapes. *Remote Sens.* **2018**, *11*, 37. [CrossRef]

54. Kupidura, P. The Comparison of Different Methods of Texture Analysis for Their Efficacy for Land Use Classification in Satellite Imagery. *Remote Sens.* **2019**, *11*, 1233. [CrossRef]

55. Li, X.; Chen, W.; Cheng, X.; Wang, L. A Comparison of Machine Learning Algorithms for Mapping of Complex Surface-Mined and Agricultural Landscapes Using ZiYuan-3 Stereo Satellite Imagery. *Remote Sens.* **2016**, *8*, 514. [CrossRef]

56. Jin, Y.; Liu, X.; Chen, Y.; Liang, X. Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: A case study of central Shandong. *Inter. J. Remote Sens.* **2018**, *39*, 8703–8723. [CrossRef]

57. Planet Satellite Imagery Products. 2018. Available online: https://www.planet.com (accessed on 15 June 2020).

58. Hansen, M.C.; Roy, D.P.; Lindquist, E.; Adusei, B.; Justice, C.O.; Altstatt, A. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sens. Environ.* **2008**, *112*, 2495–2513. [CrossRef]

59. Bwangoy, J.B.; Hansen, M.C.; Roy, D.P.; Grandi, G.D.; Justice, C.O. Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices. *Remote Sens. Environ.* **2010**, *114*, 73–86. [CrossRef]

60. De Sousa, C.; Fatoyinbo, L.; Neigh, C.; Boucka, F.; Angoue, V.; Larsen, T. Cloud-computing and machine learning in support of country-level land cover and ecosystem extent mapping in Liberia and Gabon. *PLoS ONE* **2020**, *15*, e0227438. [CrossRef]

61. Millard, K.; Richardson, M. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sens.* **2015**, *7*, 8489–8515. [CrossRef]

62. Cánovas-García, F.; Alonso-Sarría, F.; Gomariz-Castillo, F.; Oñate-Valdivieso, F. Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery. *Comput. Geosci.* **2017**, *103*, 1–11. [CrossRef]

63. Maxwell, A.E.; Strager, M.P.; Warner, T.A.; Ramezan, C.A.; Morgan, A.N.; Pauley, C.E. Large-Area, High Spatial Resolution Land Cover Mapping Using Random Forests, GEOBIA and NAIP Orthophotography: Findings and Recommendations. *Remote Sens.* **2019**, *11*, 1409. [CrossRef]

64. Kelley, L.C.; Pitcher, L.; Bacon, C. Using Google Earth Engine to Map Complex Shade-Grown Coffee Landscapes in Northern Nicaragua. *Remote Sens.* **2018**, *10*, 952. [CrossRef]

65. Teluguntla, P.; Thenkabail, P.S.; Oliphant, A.; Xiong, J.; Gumma, M.K.; Congalton, R.G.; Yadav, K.; Huete, A. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.* **2018**, *144*, 325–340. [CrossRef]

66. Amani, M.; Mahdavi, S.; Afshar, M.; Brisco, B.; Huang, W.; Mohammad Javad Mirzadeh, S.; White, L.; Banks, S.; Montgomery, J.; Hopkinson, C. Canadian Wetland Inventory using Google Earth Engine: The First Map and Preliminary Results. *Remote Sens.* **2019**, *11*, 842. [CrossRef]

67. Mahdianpari, M.; Salehi, B.; Mohammadimanesh, F.; Motagh, M. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band and TerraSAR-X imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 13–31. [CrossRef]

68. Xia, J.; Falco, N.; Benediktsson, J.A.; Du, P.; Chanussot, J. Hyperspectral Image Classification With Rotation Random Forest Via KPCA. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1601–1609. [CrossRef]

69. Rodriguez-Galiano, V.F.; Chica-Rivas, M. Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models. *Inter. J. Digital Earth* **2012**, *7*, 492–509. [CrossRef]

70. Abdel-Rahman, E.M.; Mutanga, O.; Adam, E.; Ismail, R. Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 48–59. [CrossRef]

71. Van Beijma, S.; Comber, A.; Lamb, A. Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. *Remote Sens. Environ.* **2014**, *149*, 118–129. [CrossRef]

72. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Inter. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]

73. Ghimire, B.; Rogan, J.; Galiano, V.R.; Panday, P.; Neeti, N. An Evaluation of Bagging, Boosting and Random Forests for Land-Cover Classification in Cape Cod, Massachusetts, USA. *GISci. Remote Sens.* **2013**, *49*, 623–643. [CrossRef]

74. Foody, G.M. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. *Remote Sens. Environ.* **2020**, *239*. [CrossRef]

75. Foody, G.M. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 627–633. [CrossRef]

76. Janssen, L.L.F.; van der Wel, F.J.M. Accuracy assessment of satellite derived land-cover data: A review. *Photogramm. Eng. Remote Sens.* **1994**, *60*, 419–426.

77. Congalton, R.G.; Oderwald, R.G.; Mead, R.A. Assessing Landsat classification accuracy using discrete multivariate-analysis statistical techniques. *Photogramm. Eng. Remote Sens.* **1983**, *49*, 1671–1678.

78. Smits, P.C.; Dellepaine, S.G.; Schowengerdt, R.A. Quality assessment of image classification algorithms for land cover mapping: A review and a proposal for a cost based approach. *Int. J. Remote Sens.* **1999**, *20*, 1461–1486. [CrossRef]

79. Agresti, A. *An Introduction to Categorical Data Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2007; pp. 1–372.

80. Momeni, R.; Aplin, P.; Boyd, D. Mapping Complex Urban Land Cover from Spaceborne Imagery: The Influence of Spatial Resolution, Spectral Band Set and Classification Approach. *Remote Sens.* **2016**, *8*, 88. [CrossRef]

81. Senf, C.; Leitão, P.J.; Pflugmacher, D.; van der Linden, S.; Hostert, P. Mapping land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. *Remote Sens. Environ.* **2015**, *156*, 527–536. [CrossRef]

82. Inglada, J.; Vincent, A.; Arias, M.; Tardy, B.; Morin, D.; Rodes, I. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sens.* **2017**, *9*, 95. [CrossRef]

83. Wulder, M.A.; Coops, N.C.; Roy, D.P.; White, J.C.; Hermosilla, T. Land cover 2.0. *Inter. J. Remote Sens.* **2018**, *39*, 4254–4284. [CrossRef]

84. Frantz, D.; Röder, A.; Stellmes, M.; Hill, J. Phenology-adaptive pixel-based compositing using optical earth observation imagery. *Remote Sens. Environ.* **2017**, *190*, 331–347. [CrossRef]

85. Abdi, A.M. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GISci. Remote Sens.* **2020**, *57*, 1–20. [CrossRef]

86. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [CrossRef]

87. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated water extraction index: A new technique for surface water mapping using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35. [CrossRef]

88. Rouse, J.; Hass, R.; Schell, J.; Deering, D. Monitoring vegetation systems in the great plains with ERTS. In *Third ERTS Symposium*; NASASP-351 I: Greenbelt, MD, USA, 1973; pp. 309–317.

89. Liu, H.Q.; Huete, A. A feedback based modification of the NDV I to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 457–465. [CrossRef]

90. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]

91. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]

92. Gao, B.-C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water fromspace. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]

93. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [CrossRef]

94. McFeeters, S.K. The use of the normalized difference water index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]

95. Birth, G.; McVey, G. Measuring the Color of Growing Turf with a Reflectance Spectrophotometer. *Agron. J.* **1968**, *60*, 640–643. [CrossRef]

96. Jia, K.; Wei, X.; Gu, X.; Yao, Y.; Xie, X.; Li, B. Land cover classification using Landsat 8 operational land imager data in Beijing, China. *Geocarto Int.* **2014**, *29*, 941–951. [CrossRef]