

Technical Note

Usability Study to Assess the IGBP Land Cover Classification for Singapore

Nanki Sidhu ^{1,*} , Edzer Pebesma ¹  and Yi-Chen Wang ² 

¹ Institute for Geoinformatics, Heisenbergstrasse 2, 48149 Muenster, Germany; edzer.pebesma@uni-muenster.de

² Department of Geography, National University of Singapore, 1 Arts Link, Singapore 117570, Singapore; geowyc@nus.edu.sg

* Correspondence: n_sidh01@uni-muenster.de; Tel.: +49-17631532286

Received: 7 September 2017; Accepted: 11 October 2017; Published: 22 October 2017

Abstract: Our research focuses on assessing the usability of the International Geosphere Biosphere Programme (IGBP) classification scheme provided in the MODIS MCD12Q1-1 dataset for assessing the land cover of the city-state, Singapore. We conducted a user study with responses from 33 users by providing them with Google Earth images from different parts of Singapore, asking survey-takers to classify these images according to their understanding by the IGBP definitions provided. We also conducted interviews with experts from major governmental agencies working with satellite imagery, which highlighted the need for a detailed land classification for Singapore. In addition to the qualitative analysis of the IGBP land classification scheme, we carried out a validation of the MCD12Q1-1 remote sensing product against SPOT-5 imagery for our study area. The user study revealed that survey-takers were able to correctly classify urban areas, as well as densely forested areas. Misclassifications between *Cropland* and *Mixed Forest* classes were highest and were attributed by users to the broad terminology of the IGBP of the two land cover class definitions. For the accuracy assessment, we obtained validation points using weighted and unweighted stratified sampling. The overall classification accuracy for all 17 IGBP land classes is 62%. Upon selecting only the four most occurring IGBP land classes in Singapore, the classification accuracy improved to 71%. Validation of the MCD12Q1-1 against ground truth for Singapore revealed less-common land classes that may be of importance in a global context but are sources of error when the same product is applied at a smaller scale. Combining the user study with the accuracy assessment gives a comprehensive overview of the challenges associated with using global-level land cover data to derive localized land cover information specifically for smaller land masses like Singapore.

Keywords: MODIS; global classification; land cover classification; user study; validation

1. Introduction

Earthly phenomena are widely captured through satellite imagery and global scale data are available in large scale big data repositories such as the Google Earth engine or the Amazon Cloud. More specifically, with an increased focus on land use and land cover monitoring of different areas in the world, it becomes essential for researchers and policy-makers, to access and understand the vast amount of data we have. One of the major challenges for users is to have an in-depth knowledge of the data they are working with. This is especially true for users who may not be experts in the domain of remote sensing or working with classified land cover data. For example, non-expert users may depend on remotely-sensed data for use in policy-support and decision-making. Thus, it is also crucial to know the “fitness for use” of the datasets to be used, for instance, for urban planning, monitoring deforestation, observing climate cycles, or supporting the conception of policies for governmental agencies.

One of the most effective ways to test how suitable a particular dataset is for a certain purpose is to involve users in the process of this evaluation [1]. It is essential for users to be able to understand the suitability of the data they are seeking. When it comes to using remote sensing data, various factors, such as the spatial, temporal, and spectral resolutions are important to note, especially if we need to understand and process the finished results accurately. Thus, through a user study we can assess the “expertise” of the users and their level of knowledge, as well [1,2]. Engaging users not only provides us with an understanding of the users’ behaviour, but also helps to evaluate the quality of the datasets that are available. Thus, we obtain a comprehensive overview of understanding data-user interaction. Providing accurate statistics of a thematic map assists users in selecting a dataset for an appropriate application. This is especially important for cases where thematic maps are used for policy-based planning and decision-making.

In our case, we want to explore the quality of the land classification scheme provided by the Moderate Resolution Imaging Spectrometer (MODIS) MCD12Q1-1. This product classifies the globe into different land cover types and we want to evaluate how well it captures land use and land cover for areas the size of our study area, Singapore and, by extension, smaller areas, in general. It is known to work well for vast forested areas such as the Brazilian Amazon [3,4], however, it is less clear whether global classification schemes, such as the International Geosphere Biosphere Programme (IGBP) [5] or the Land Cover Classification Scheme (LCCS) [6], can be effectively used for monitoring the urban dynamics of a country like Singapore. We were motivated by the challenge of the representation of smaller land masses in existing global classified datasets. The main question that arises is how well or how poorly these areas are represented. Global land cover datasets are useful for smaller countries as they provide an important overview of their landscape, especially in relation to the rest of the local region, continent, and the globe. In the case of Singapore, we were faced with a lack of open data, whereby users did, to a certain extent, depend on existing open datasets, such as MODIS and Landsat to assess the transforming physical environment of the island and urban planning.

It is important to treat the IGBP land cover definitions and the spatial information of the MCD12Q1-1 as two components that, together, constitute the classified dataset. The user study explicitly focuses on the ambiguity of the IGBP land class definitions for different land classes contained in the MCD12Q1-1. In addition to the user study, which focuses on the qualitative evaluation of the IGBP land class definitions, we also carried out a validation of the MCD12Q1-1 dataset using ground truth imagery. The accuracy assessment validates the spatial accuracy of any remote sensing product containing thematic information. The user study and the accuracy assessment were carried out independent to one another, yet they complement each other by providing a comprehensive overview of the MCD12Q1-1, in terms of the land class information it contains and the locational accuracy of the pixels.

The goal of this research is to engage users in an exercise with the aim of classifying Singapore into the existing classes of the MCD12Q1 IGBP land cover classification [7]. This is done with the goal of making users understand the existing definitions of certain land cover classification systems, as well as giving us an idea of how clear, concise, and accurately Singapore is classified by the IGBP land classification scheme. The survey also provided an opportunity to evaluate the level of detail offered in the MCD12Q1-1 product, in terms of the definitions provided and how they specifically fit within the landscape of Singapore, which is one of the most urbanised countries in the world. The validation of the MCD12Q1-1 provides a quantification of the degree of agreement between the remotely-sensed data and ground reality. The statistics from the outcome of the validation can also be utilised by users to gauge the suitability of the MCD12Q1-1 to map the land cover of Singapore.

2. Materials and Methods

2.1. Datasets and Study Area

The dataset we are using and evaluating is a MODIS-MCD12Q1 global land cover product. The product contains five different land classification systems, however, our research focuses on the IGBP system (MCD12Q1-1) [5,7,8], which divides the land cover of the Earth into 17 land classes (Table 1 and Figure S1).

Table 1. MCD12Q1 International Geosphere Biosphere Programme (IGBP) land cover classification classes.

Class Code	Land Cover Class
0	Water
1	Evergreen Needleleaf Forests
2	Evergreen Broadleaf Forests
3	Deciduous Needleleaf Forests
4	Deciduous Broadleaf Forests
5	Mixed Forests
6	Closed Shrublands
7	Open Shrublands
8	Woody Savannas
9	Savannas
10	Grasslands
11	Permanent Wetlands
12	Croplands
13	Urban and Built-up
14	Cropland/Vegetation mosaic
15	Snow and Ice
16	Barren or Sparsely Vegetated
254	Unclassified
255	Fill Value

The spatial resolution of the dataset is 500 m. The product is derived from imagery from both MODIS Aqua and Terra satellites. There is a new dataset compiled at the end of each calendar year. Our work uses the classification generated on 31 December 2008. Singapore was covered by clouds for long periods and the year 2008 is a time where we could acquire relatively clear images through most of the year.

The study area for our research is Singapore (Figure 1). The main motivation behind selecting Singapore as a study area is to find out how well an area the size of Singapore can be represented by a global remote sensing dataset. This also highlights the question of how we derive finer details from existing land cover definitions to improve upon the limitations that most land cover classification systems face. As seen from Figure 1, Singapore has a highly urbanized landscape, which is interspersed with protected areas consisting of primary forests and wetlands (CCR and Bukit Timah), as well as urban green spaces. Land cover analysis is, moreover, inherently uncertain in Singapore due to its constantly changing physical boundaries. Land reclamation is indeed a common practice in the city-state [9,10].

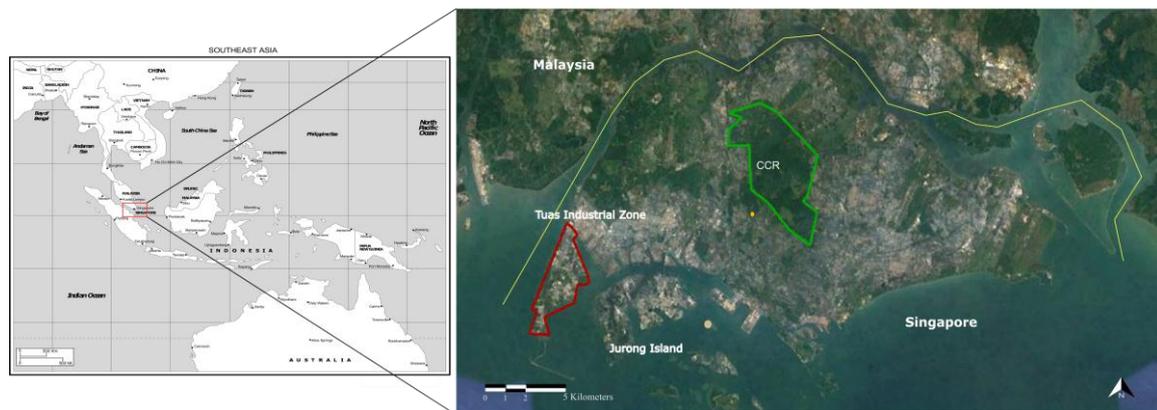


Figure 1. The study area, Singapore (Located at: 1.3521°N , 103.8198°E). The Tuas industrial site is highlighted in red and the Central Catchment Reserve (CCR) and the Bukit Timah Natural Reserve are in green. Malaysian and Singaporean waters are delineated by the yellow line.

2.2. Usability Studies

In order to ethically carry out the user study and interviews in Singapore we were required to obtain clearance from the National University of Singapore Institutional Review Board (NUS-IRB). The approval date for the same is 11 April 2016, with approval code NUS-2935.

To collect the responses for our user survey, we used the open source platform Limesurvey [11]. Upon the creation of each survey, a unique user link is made available to the users which, in our case, can be found in [11]. Through the user link they could gain access to the survey. The link was circulated amongst potential survey-takers at the National University of Singapore, Singapore Land Authority, Urban Redevelopment Authority and Centre for Remote Imaging Sensing and Processing, Singapore, via email. The link was kept live for a period of two months, during which time users could openly access the survey link. The questionnaire was kept anonymous as the name of the users was not relevant to the work, however, we did ask for the organizations they are affiliated with as that provides an idea of their extent of experience with land cover information. We received 33 complete responses out of 89 attempts. The survey questions consisted of Google Earth(2016) (Google Inc., Mountain View, CA, USA) images of subsections of Singapore with each image containing different landscapes. Each image has a spatial extent of 300–500 m. Some images consisted of a single land cover while others contained a mix of different cover types. Each image further varied in terms of brightness zoom level and angle in order to ensure randomisation and the receipt of meaningful responses. The questions were in the form of multiple choice, with eight answer options. The answer options provided were the land cover labels from the MODIS IGBP land cover classification (Table 1), more specifically, the land classes that occur in Singapore according to the MCD12Q1-1 dataset. By providing maximum information to a user, we reduce the scope for biased responses.

At the end of the survey we also included an open ended question about how participants might use a global land cover dataset similar to MCD12Q1-1, which provides us with insight into what the survey-takers and users of remote-sensing data living and working with Singapore-specific data think about a global-level classification of the country.

Interviews

In order to support this survey, we also conducted personal interviews with four experts from organizations who work with remotely-sensed data for Singapore. They were affiliated to the National University of Singapore (NUS), Urban Redevelopment Authority (URA), National Parks (NP), and the Centre for Remote Imaging Sensing and Processing, Singapore. The interviews were informally

structured, but highly focused. The questions were slightly varied to adapt to the background and experience of each interviewee. The questions asked are as follows:

- What is your affiliation and the field of work you are in?
- In your research or professional life, have you ever worked with remote sensing data?
- What kind of remote sensing datasets have you worked with most frequently?
- Have you ever worked with global land cover classification schemes such as MCD12Q1-1 or GlobCover (2009)?
- Looking at this MCD12Q1-1 image, what is your opinion on what Singapore looks like according to this dataset?
- In your opinion, how recent and up-to-date do these classifications seem?
- How often do you think this data needs to be upgraded?
- How would you use such an image?
- What are the specific challenges of using remote sensing data for Singapore?

The interviewees did not participate in the user survey.

2.3. Labelling and Classification Scheme

The URA is one of the main governmental agencies that develop the land use plans for the island. The URA provides an updated land use plan for the country every 10–15 years. These plans are openly accessible at the URA website [12]. Some of the classes as listed in these master plans were residential, commercial, business parks and transportation sites and the interpretation of these different classes can be found in [13].

In addition to the user study validation, we propose a hierarchical labelling structure that divides Singapore into land classes that are derived from the URA plan and may be integrated into the MCD12Q1-1.

Owing to the highly urbanized landscape of Singapore, we commenced with two broad categories, namely “Urban” and “Non-Urban”. We further divide “Urban” into residential, Central Business District (CBD), and defence areas (e.g., military zones). The “Non-Urban” category contains national parks (primary forests and wetlands) and water bodies (Figure 2).

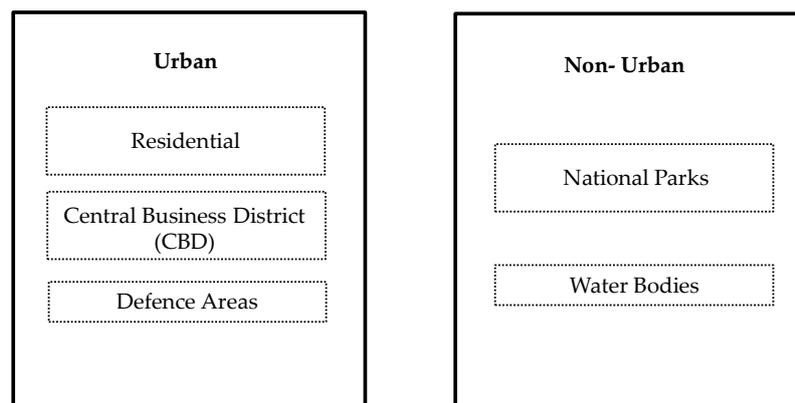


Figure 2. A simple design for the creation of a future land cover classification scheme for Singapore.

2.4. Accuracy Assessment of the MCD12Q1-1 for Singapore Using Ground Truth Imagery

An accuracy assessment may be defined as the process of testing the degree of agreement between a classified map and ground truth. With the availability of numerous global land cover classifications, such as the MCD12Q1-1 [5], GlobCover (2009) [14], and GLC2000 [15], and validation efforts have been undertaken to evaluate their thematic accuracy at global and regional levels. These studies address

existing challenges in land use and land cover mapping, such as inconsistencies in land class definitions, comparison of legend information from different classification schemes, and the errors arising due to the different methods employed for data collection and compilation of remotely-sensed datasets.

Ground truth data used within accuracy assessments may be in the form of field data or high-resolution satellite imagery.

For our validation process, we obtained Satellite Pour l'Observation de la Terre (SPOT-5) imagery for Singapore (Figure 3). The spatial resolution of the image is 10 m in the four spectral bands, namely, Band 1 (green: 0.50–0.59 μm), Band 2 (red: 0.61–0.68 μm), Band 3 (near infrared: 0.78–0.89 μm), and Band 4 (short-wave infrared: 1.58–1.75 μm).

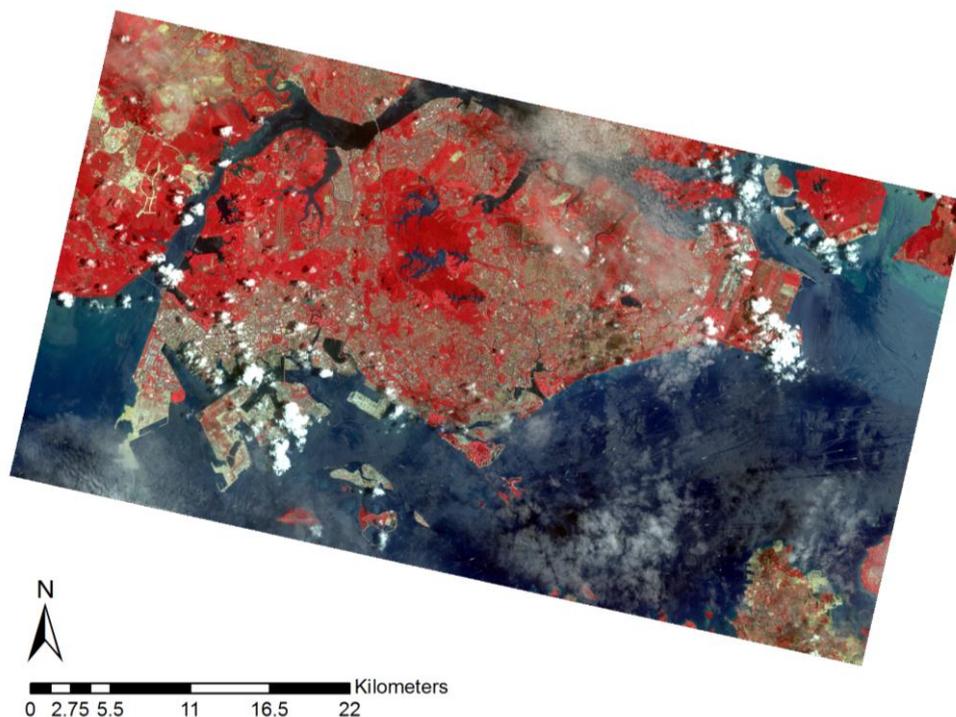


Figure 3. SPOT-5 image showing Singapore and its neighbouring islands, displayed in R = Band 4, G = Band 3, B = Band 2. The image has spatial resolution of 10 m (Lat = 1.524°N, Long = 103.606°E) and the date of acquisition is 14 March 2009.

The initial step to an accuracy assessment is to decide the sampling strategy for the validation points. Some common sampling methods used are: simple random sampling, stratified random sampling, systematic random sampling and cluster sampling. All sampling techniques contain some advantages and disadvantages. For example, simple random sampling increases the chance of missing out on smaller land classes, while systematic random sampling could introduce a sampling bias into the process as samples are selected with prior knowledge of their land class [16]. Thus, for our validation, stratified random sampling was the best fit. This method selects the number of samples proportionately to the size of the classified land cover class. For example, the *Water* and *Urban/Built-Up* land classes are the most abundant for Singapore, thus, the algorithm assigns the largest and second largest validation sample groups to these classes, respectively. For the unweighted stratified sample or equalized stratified sample, each class in the analysis was assigned 100 points.

Validation is carried out initially for all 17 IGBP land cover classes of the MCD12Q1-1 product, using the weighted stratified sampling. This is followed by an accuracy assessment of only the most abundant land cover classes in our study area, namely, *Urban/Built-Up*, *Evergreen Broadleaf Forests*, *Permanent Wetlands*, and *Water*, using both weighted and unweighted stratified sampling. We did not

repeat an equalized stratified sampling for all 17 IGBP classes. A majority of the land classes only contain 1–5 pixels in Singapore, thus, we could not obtain a suitable number of samples for the process.

Through the formulation of an error matrix, we evaluate the following statistics for the accuracy of the IGBP classification of Singapore:

- Overall Accuracy: (No. of correctly classified pixels)/(Total pixels in the assessment). (1)
- Producer's Accuracy: The probability of a ground truth pixel being correctly classified:

$$P.A. (Class i) = \frac{k_{i,i}}{k_i} \quad (2)$$

where $k_{i,i}$ is the diagonal element of the matrix for class i , and k_i is the sum of pixels for class i in the reference image.

- User's Accuracy: The probability of a pixel in our classified map belonging to the correct class:

$$U.A. (Class i) = \frac{k_{i,i}}{k_i} \quad (3)$$

where $k_{i,i}$ is the diagonal element of the matrix for class i , and k_i is the sum of pixels for class i in the classified image.

- Kappa index measure of prediction between a "ground truth" map and any alternative map. This outputs the probability of the actual agreement calculated being better than the agreement expected due to random chance. For example, a Kappa index of 0.45 indicates that overall agreement (from Equation (1)) is 45% better than the agreement due to chance.

3. Results and Discussion

3.1. User Study and Interviews

This section deals with the inferences we were able to draw from the 33 responses from our user study. In addition to the images that were clearly urban or built up land, we were able to highlight some confusion when it came to forested land as well as barren and cropland land classes. This survey is a user-based validation of the MODIS MCD12Q1-1 IGBP classification scheme. Out of the 30 images that users were asked to classify, we look at seven in detail which provided us with some interesting observations.

In images containing a single land cover type, users were able to discern the cover type well. However, in some images (Figures 4iii and 5ii,iv) there is confusion at some points between barren and urbanised areas. According to the definitions provided in the MCD12Q1-1, which the users had continuous access to, it highlights the need for more detailed definitions, as well as a finer spatial resolution classification for an area such as Singapore.

According to the IGBP classification for Singapore most barren areas were also classified as urban or built up land. This is mainly due to the low spatial resolution of the dataset, as well as the logic of the definitions created. Industrial areas within Singapore are large swaths of land which are used as construction sites and holding sites for cement and sand. This is especially true for the Tuas Industrial Zone, which is a hotspot for land reclamation activities within Singapore [9,10]. This inherent understanding can be seen in the results of Figure 5iv where 70% of users classify a stretch of land that falls within an industrial zone as barren. This point once again highlights the coarse level of global definitions for a country-scale analysis.

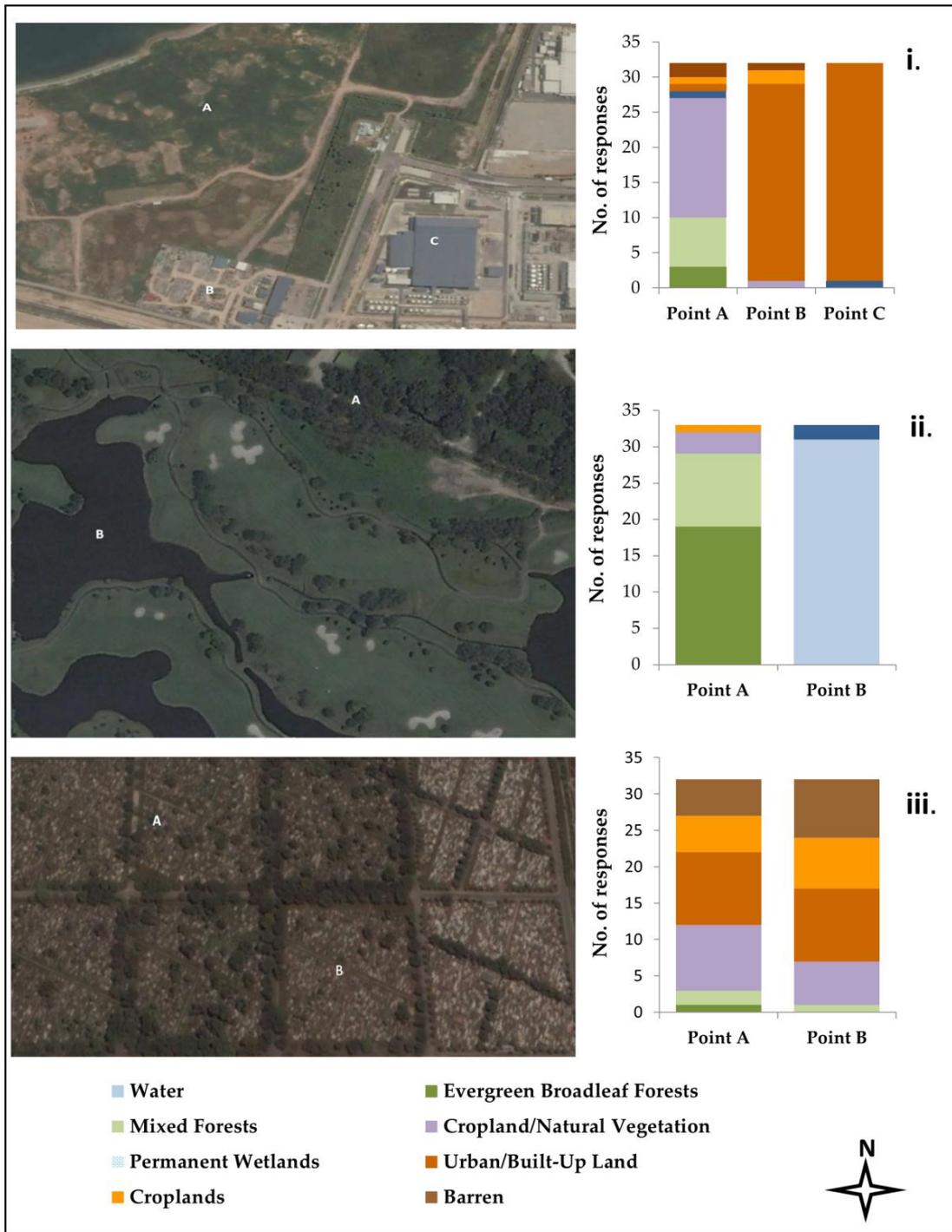


Figure 4. (i–iii) A selection of three out of the 30 question images from the user study. The images are from randomly-selected sites within Singapore. Users were asked to classify the labelled points within each image according to their understanding of the definitions of the IGBP, of which they were provided access to throughout the survey-taking. The 33 responses we received for each image are depicted in the corresponding stacked barcharts.

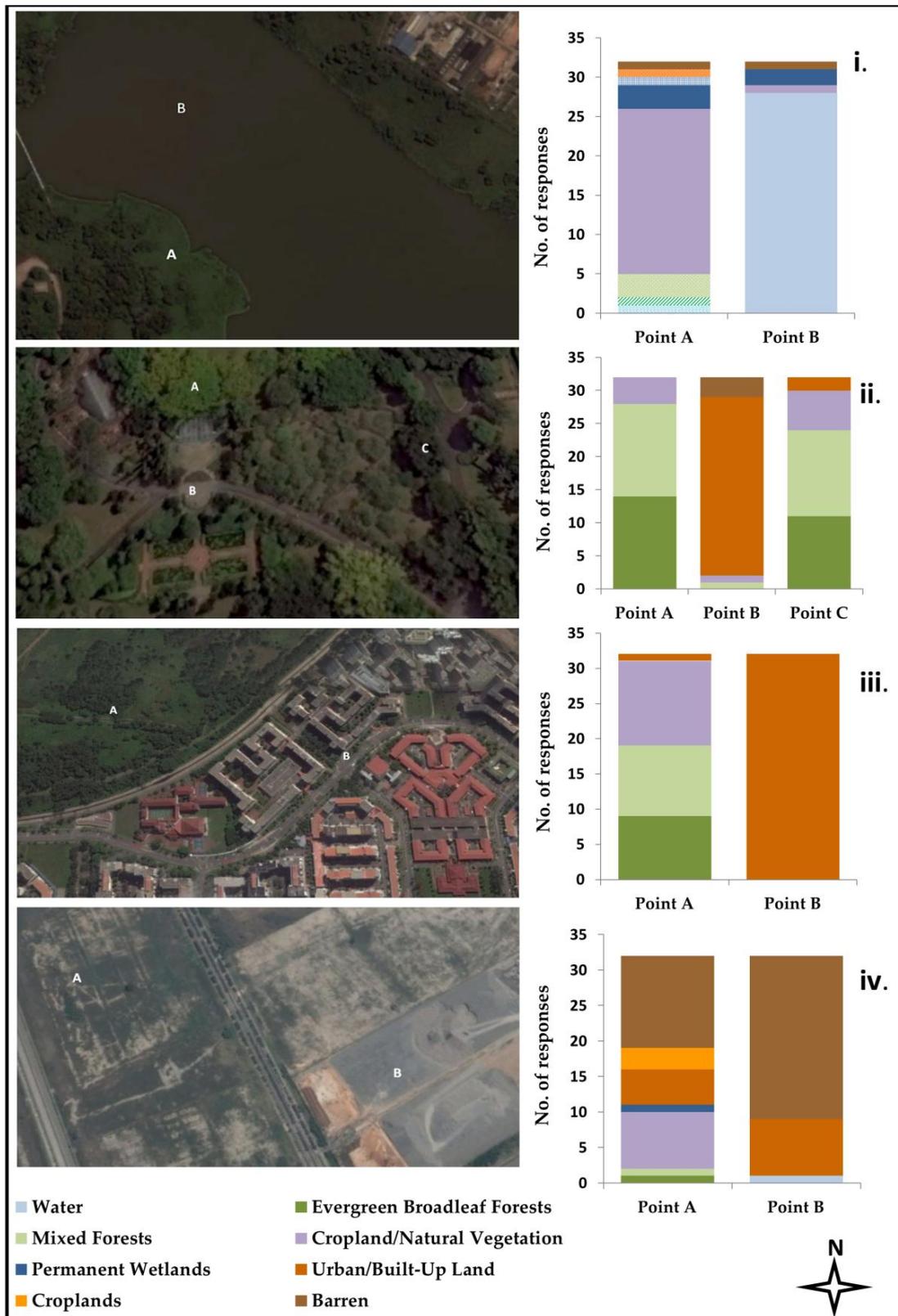


Figure 5. (i–iv) A selection of four out of the 30 question images from the user study. The images are from randomly-selected sites within Singapore. The 33 responses received for each image are depicted in the corresponding stacked bar charts.

Singapore consists of designated protected areas of primary forest, mainly evergreen broadleaf forests [17,18]. As can be seen in Figure 1, these forested areas are mostly untouched due to extremely strict government policies governing the land use and land cover of the country [17,18], and most users correctly classified these areas as *Evergreen Broadleaf Forests* and/or *Mixed Forests*. In Figure 4ii, 70% users assigned the forest classes to Point A, while for Figure 5ii, 85% (Point A) and 73% (Point C) respondents assigned the forest classes. This was especially true in the CCR area and the Bukit Timah National Reserve [19], which are the areas depicted in Figure 1. In some cases, however, we see some amount of confusion between forests and cropland/natural vegetation land cover. In Figure 4i, Point A was classified as *Cropland/Natural Vegetation* by 40% of users and *Evergreen Broadleaf Forest* or *Mixed Forests* by 30%. Similarly in Figure 4iii, 36% users classify Point A as *Cropland/Natural Vegetation* and 57% classify as forests. Most of the users (93%) who took the survey categorized themselves as “experts” or having worked/working with land use and land cover products. Despite the users’ backgrounds and access to the IGBP land definitions throughout the survey, misclassifications revealed the definitions proposed by the IGBP classification scheme were not specific enough for the users’ understanding. The *Cropland/Natural Vegetation* [5] category includes lands with no more than 60% of forest/grasslands/shrublands and this type of terminology leaves space for a subjective analysis of this land class definition, increasing the chances for a misclassification. This sheds light on the need for a more detailed classification for islands the size of Singapore, which, at present, does not exist except as prescribed by the URA, and is primarily based on vector data.

The survey also includes an open-ended question to the respondents regarding how they would apply a global land cover product to study Singapore. We received 32 responses to the question group and the respondents conveyed the advantages of using an open global product especially for analysis of the land cover dynamics at a regional level (Southeast Asia scale). The major disadvantage that most users agree on is the coarse resolution of the dataset. In addition to the results of the user study, the interviews conducted in Singapore were a more focused viewpoint of four candidates in key positions working in different governmental agencies and the National University of Singapore. The major aspect of agreement amongst these four independent interviewees was that global classifications to the likes of the IGBP or LCCS provide a good overview of the different possible land covers present in Singapore and are useful to study the island at a country level, especially to monitor the ever-changing boundary due to the widespread practices of land and water reclamation. They further expressed the need to develop and study a more detailed land cover classification for Singapore, which was also captured through the responses of the survey.

Some of the limitations that were highlighted through the interviews were the need for regular updates to the MODIS datasets as a lot of the smaller islands that were reclaimed by the country were not present in the imagery or were not correctly classified. Furthermore, two out of four interviewees conveyed the importance of remote sensing data being in correspondence with the vector data that is more frequently used and suited for the need to manage a smaller landmass like mainland Singapore and its neighbouring islands.

3.2. Accuracy Assessment

For the validation of the MCD12Q1-1 using SPOT-5 imagery, we performed three accuracy assessment exercises using weighted and unweighted stratified sampling strategies. The selection of different sampling strategies helped us identify different sources of error, which may either be inherent in the MCD12Q1-1 dataset or arise out of the sampling process. We analyse and discuss the descriptive statistics derived through our accuracy assessments in the following sections.

3.2.1. Weighted Stratified Sampling for All 17 IGBP Land Classes

For our initial validation process, we obtained samples using a weighted stratification scheme for all 17-land classes in our study area. The error matrix using the weighted stratified sampling scheme is detailed below (Table 2).

Smaller land cover classes namely, *Open Shrublands*, *Croplands*, *Croplands/Natural Vegetation Mosaic* occur with only one or two pixels in the classified image. The occurrence of these pixels is a source of random error within the MCD12Q1-1 dataset. Additionally, classes such as *Savannas*, *Woody Savannas*, and *Grasslands* do not exist in our study area, yet a small number of pixels appear in the MCD12Q1-1. Hence, the producer's accuracies for the smaller classes could not be evaluated as the classes were not found in the reference image. The overall accuracy of the validation is ~62%. Due to the presence of numerous redundant land classes, the descriptive statistics evaluated in this case do not represent a true validation of the MCD12Q1-1. The process is affected by the difference in support mostly due to the coarse spatial resolution of MCD12Q1-1. Moreover, the MCD12Q1-1 does not include information about the neighbouring areas of Singapore, for example, the Jurong Islands or the Tuas Industrial Zone. As a result, *Water* is frequently misclassified as the *Urban/Built-Up* land class. This problem is specific to the peripheral areas of the island. However, for a country like Singapore, with a continuously changing land boundary, it is important for land cover products to be able to detect dredging, which is the dominant land conversion activity specific to the island.

3.2.2. Weighted Stratified Sampling for Most Abundant IGBP Land Classes in Singapore

In order to pre-correct our validation process, we repeated an accuracy assessment for the MCD12Q1-1, ignoring smaller redundant land classes. The dominant IGBP land cover classes of Singapore are *Urban/Built-Up Land*, *Evergreen Broadleaf Forests*, *Permanent Wetlands*, and *Water*. We narrowed down these relevant IGBP classes on the basis of information available in the master land use and land cover plans devised by the Singapore Land Authority and Urban Redevelopment Authority, which are the local authorities in charge of land planning for the island [12,13,18]. We repeated a weighted stratified random sampling for our validation points, but only for these four land cover classes, as opposed to all 17 IGBP classes in Section 3.2.1.

In this case, our overall classification accuracy improved to ~71%. An interesting observation is the producer's and user's accuracies for the *Permanent Wetlands* class are the poorest. Furthermore, the error matrix (Table 3) also shows that *Evergreen Broadleaf Forests* in the classified image were marked as *Permanent Wetlands* most frequently. This highlights the challenge of using the MCD12Q1-1 dataset to distinguish between mixed landscapes. In Singapore, the wetlands and swamp forests co-exist with the largest expanse of evergreen broadleaf forests in the Central Catchment Reserve (Figure 1). The misclassification was not captured through the user study, which highlights the importance of both qualitative and quantitative evaluations to establish a comprehensive overview of the accuracy of existing land cover classification schemes.

Table 3. Error matrix of four most-occurring IGBP land classes of the MCD12Q1-1 dataset for our study area. Reference data is derived from SPOT-5 imagery for Singapore. Validation points are selected using weighted stratified random sampling scheme.

		REFERENCE DATA				Row Total	Number Classified Correct	User's Accuracy
		Water	Evergreen Broadleaf Forests	Permanent Wetlands	Urban/Built-Up Land			
CLASSIFIED DATA	Water	148	6	4	47	205	148	72.19%
	Evergreen Broadleaf Forests	4	11	8	4	27	11	40.74%
	Permanent Wetlands	2	2	2	1	7	4	28.57%
	Urban Built Up Land	4	8	6	70	88	70	79.55%
	Column Total	158	27	20	122	327		
Number Classified Correct		148	11	2	70		233	
Producer's Accuracy		93.60%	40.74%	10%	57.38%			
Overall Accuracy		71.25%						
Overall Kappa Statistics		0.5017						

3.2.3. Unweighted Stratified Sampling for Most Abundant IGBP Land Classes in Singapore

Following the improvement of the accuracy assessment by considering only the most abundant land cover classes, we repeated the computation using an equalized sampling scheme. The reason for this was to account for random errors that are inherent in the MCD12Q1-1 dataset, as well as account for oversampling within land classes from the weighted sampling method. Using this sampling strategy (Table 4), we obtain ~62% overall agreement between the IGBP classification and ground truth. We notice that our descriptive statistics have improved in comparison to the previous two validation efforts. Despite the lower overall accuracy than in the previous validation run, we further observe that increasing the number of samples within the *Permanent Wetlands* and *Evergreen Broadleaf Forests* leads to an improvement in their producer's and user's accuracies. These improvements may be attributed to the pre-filtering process of disregarding insignificant land classes and also the equal weight distribution to the remaining land classes. An attribute of an unweighted stratified sampling scheme is a reduction in overall accuracy; however, we are able to study individual land classes with higher precision.

Table 4. Error matrix of the four most-occurring IGBP land classes of the MCD12Q1-1 dataset for our study area. Reference data is derived from SPOT-5 imagery for Singapore. Validation points are selected using unweighted stratified random sampling.

		REFERENCE DATA				Row Total	Number Classified Correct	User's Accuracy
		Water	Evergreen Broadleaf Forests	Permanent Wetlands	Urban/Built-Up Land			
CLASSIFIED DATA	Water	79	8	0	13	100	79	79.00%
	Evergreen Broadleaf Forests	16	49	16	19	100	49	49.00%
	Permanent Wetlands	7	16	46	31	100	46	46.00%
	Urban Built Up Land	1	12	12	75	100	75	75.00%
	Column Total	103	85	74	138	400		
Number Classified Correct		79	49	46	75		249	
Producer's Accuracy		76.69%	57.65%	62.16%	54.35%			
Overall Accuracy		62.25%						
Overall Kappa Statistics		0.4967						

4. Conclusions

This research helps users understand a global dataset within the context of a local area which, in our case, is Singapore. We find that in the case of a global classification, smaller islands or countries often fall short of being classified completely and accurately.

In an urban setting, such as Singapore, there has not been land cover conversion in the form of deforestation over many years [17]. However, dredging activities for land reclamation led to a constantly-changing country boundary, intensifying from the year 2004, with the development of the Tuas Industrial Zone in the southwest part of the island [9]. The survey takers and interviewees shed light on the need for an openly-accessible and globally-integrated land classification system for the island state that contains up-to-date land cover information for the island. This was important to note as the Tuas Industrial Zone and Jurong Islands could not be observed in the MCD12Q1-1 (Figure S1) for the year 2008.

The user survey we undertook focuses specifically on the perspective of the users of land cover maps and the information they derive from classification data. It highlights that broad land class definitions can be a source of confusion for users and increase the probability of misclassifications. Furthermore, the access to the detailed terminology of each IGBP land cover class does not prevent misclassification of land cover types, especially in images with mixed landscapes. The users and

interviews emphasize on the need for improving the level of detail of existing IGBP land cover definitions in the context of Singapore. This is due to the highly focused nature of land cover change and activities in Singapore. The detailed user responses can be found in the supplementary material. The validation of the MCD12Q1-1 dataset against ground truth imagery supports the user study by quantifying the errors in the dataset. The statistics obtained through the validation runs give an overall accuracy of ~62% (Tables 2 and 4) and 71% (Table 3). The low accuracy statistics highlight the challenge of scaling of global level spatial information in MCD12Q1-1, despite efforts to pre-filter for errors by focusing on only the most frequently occurring land classes in our study area. As highlighted in [20], *Urban/Built-Up* and *Permanent Wetlands* classes may be under-represented in a global context. However, these classes are of the highest relevance to users of the MCD12Q1-1 map assessing Singapore and its neighbouring areas, since they are the most frequently occurring land cover classes in Southeast Asia. Hence, it is not suitable to use MCD12Q1-1 to quantify the extent of the *Urban/Built-Up* and *Permanent Wetlands* land classes for our study area.

Some of the challenges that we faced during our research can also be perceived in a broader context. One of the main challenges was the lack of access to open data, specifically for our study area. Thus, it was imperative to include users and officials who work with different types of land cover data to obtain more information. A limitation of the accuracy assessment is a lack of information regarding the training phase of the MCD12Q1 dataset. Thus, errors that may have affected the development phase of the dataset could have affected the validation. At present, research in Singapore relies heavily on vector data (e.g., building footprints, roads, parks) from the SLA to monitor urban expansion. It is challenging to work with vector data in terms of quantifying land change within spatial polygons. There is an opportunity to create and integrate more Singaporean remote sensing data with existing open global level land cover information (e.g., MODIS, GlobCover LCCS), which is used by the scientific community.

Supplementary Materials: The user study questions are openly accessible at <https://classifysingapore.limequery.com>. Supplementary material is attached: **Figure S1**. The MODIS MCD12Q1-IGBP classification for our study area, Singapore.

Acknowledgments: This research is part of a Ph.D. project that is funded by the Deutsche Akademischer Austausch Dienst (DAAD). Funds to publish this in open access were included in the funding. Funding support from the National University of Singapore (grant WBS no. R109-000-190-646) for acquiring the SPOT-5 imagery is appreciated.

Author Contributions: For this article, Nanki Sidhu analysed the data and performed the experiments; Nanki Sidhu and Edzer Pebesma wrote the paper; and Yi-Chen Wang contributed to the design of the survey, organizing interviews, article structure, obtaining the ethical approvals from the NUS-IRB, and the acquisition of the SPOT-5 imagery. All authors contributed to the editing of the article.

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

References

1. Senaratne, H.; Gerharz, L.; Pebesma, E.; Schwering, A. Usability of spatio-temporal uncertainty visualisation methods. In *Bridging the Geographic Information Sciences*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 3–23.
2. Pebesma, E.J.; de Jong, K.; Briggs, D. Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: An air quality example. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 515–527. [CrossRef]
3. Maus, V.; Câmara, G.; Cartaxo, R.; Sanchez, A.; Ramos, F.M.; de Queiroz, G.R. A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3729–3739. [CrossRef]
4. Morton, D.C.; DeFries, R.S.; Shimabukuro, Y.E.; Anderson, L.O.; Del Bon Espírito-Santo, F.; Hansen, M.; Carroll, M. Rapid assessment of annual deforestation in the brazilian amazon using modis data. *Earth Interact.* **2005**, *9*. [CrossRef]

5. International Geosphere Biosphere Programme Land-Cover Classification System. Available online: www.eomf.ou.edu/static/IGBP.pdf (accessed on 1 April 2016).
6. Globcover Land Cover Classification Scheme (LCCS). Available online: http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf (accessed on 1 April 2016).
7. Ganguly, S.; Friedl, M.A.; Tan, B.; Zhang, X.; Verma, M. Land surface phenology from modis: Characterization of the collection 5 global land cover dynamics product. *Remote Sens. Environ.* **2010**, *114*, 1805–1816. [CrossRef]
8. Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. Modis collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [CrossRef]
9. Koh, G.Q. Singapore Finds It Hard to Expand without Sand. PlanetArk, 2005. Available online: <http://planetark.com/dailynewsstory.cfm?newsid=30328> (accessed on 10 January 2015).
10. Koh, G.Q. S'pore Tremors Raise Fear of Building on Reclaimed Land 2007. Available online: <http://www.wildsingapore.com/news/20070304/070309-5.htm> (accessed on 10 January 2015).
11. Limesurvey Open-Source Online Survey Maker Survey. Available online: <https://classifysingapore.limequery.com> (accessed on 1 December 2015).
12. Urban Redevelopment Authority Master Plans. Available online: <https://www.ura.gov.sg/uol/masterplan.aspx?p1=View-Master-Plan> (accessed on 1 October 2016).
13. Urban Redevelopment Authority Master Plans Legend. Available online: https://www.ura.gov.sg/maps/ura/planning/masterplan/MPWS_2014_for_map_legend.htm#a13 (accessed on 1 October 2016).
14. Arino, O. GlobCover 2009. Available online: http://epic.awi.de/31046/1/Arino_et_al_GlobCover2009-a.pdf (accessed on 20 October 2017).
15. Bartholeme, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977.
16. Köhl, M.; Magnussen, S.S.; Marchetti, M. *Sampling Methods, Remote Sensing and GIS Multiresource Forest Inventory*; Springer Science & Business Media: Berlin, Germany, 2006.
17. National Environmental Agency, Singapore. Available online: <http://www.nea.gov.sg/> (accessed on 1 June 2015).
18. National Parks Board, Singapore. Available online: <https://www.nparks.gov.sg/> (accessed on 14 May 2015).
19. Lum, S.K.Y.; Lee, S.K.; LaFrankie, J.V. *Bukit Timah Forest Dynamics Plot, Singapore*; University of Chicago Press: Chicago, IL, USA, 2004; pp. 464–473.
20. Herold, M.; Mayaux, P.; Woodcock, C.E.; Baccini, A.; Schmullius, 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]



© 2017 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/>).