Scientific Research Publishing

Development of Global Cropland Agreement Level Analysis by Integrating Pixel Similarity of Recent Global Land Cover Datasets

Anjar Dimara Sakti[1,2], Wataru Takeuchi[1], Ketut Wikantika[2]

[1]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan
[2]Center for Remote Sensing, Institut Teknologi Bandung, Bandung, Indonesia
Email: anjar@iis.u-tokyo.ac.jp, wataru@iis.u-tokyo.ac.jp, Ketut@gd.itb.ac.id

Abstract

Global cropland monitoring is important when considering tactical strategies for achieving food sustainability. Different global land cover (GLC) datasets providing cropland information have already been published and they are used in many applications. The different data input methods, classification techniques, class definitions and production years among the different GLC datasets make them all independently useful sources of information. This study attempted to produce a cropland agreement level (CAL) analysis based on the integration of several cropland datasets to more accurately estimate cropland area distribution. Estimating cropland area and how it has changed on a national level was done by converting the level of cropland agreement into percentages with an existing cropland fraction map. A pre-analysis showed that the four GLC datasets used in the 2005 and 2010 groups had similar year input data acquisitions. Therefore, we placed these four datasets (GlobCover, MODIS LC, GLCNMO and ESACCI LC) into 2005 and 2010 year-groups and selected them to process dataset integration through a CRISP approach. The results of this process proposed four agreement levels for this CAL analysis, and the model correlation was converted into percentage values. The cropland estimate results from the CAL analysis were observed along with FAO data statistics and showed the highest accuracy, with a 0.70 and 0.71 regression value for 2005 and 2010 respectively. In the cropland area change analysis, this CAL change analysis had the highest level of accuracy when describing the total size of cropland area change from 2005 and 2010 when compared to other individual original GLC datasets.

Keywords

Global Land Cover, Pixel Comparison, Cropland Agreement Level

1. Introduction

Global trends of the last 25 years show that for the developing world as a whole, the share of undernourished people among the total population has decreased significantly. Some important factors that have affected this reduction have been stability in economic growth and the massive development of irrigation systems, which have caused an escalation in agricultural productivity [1]. Sustainable development goals (SDGs), as part of the development of a global pledge, are targeting to eradicate hunger by 2030 and ensure sustainable food production systems [2]. To achieve this target, the availability of accurate information about global and regional cropland distribution that can be monitored periodically is becoming important for the construction of strategies for achieving food sustainability targets [3]. Satellite-derived cropland dataset can be one alternative source of information since it directly correlates with food resource distribution [4] and water requirements [5].

Over the last two decades, many institutions have published different global land cover (GLC) datasets that provide cropland class information. Since the first one released in 1993, the trend in the improvement of the production of GLC datasets has been in an increase of spatial resolution and accuracy [6]. However, each GLC dataset has a different production year, data input method, classification technique, class definition and accuracy distribution value [7] [8]. These datasets are released as independent datasets, which make them incomparable, especially for a multitemporal analysis [9].

To evaluate the differences between each GLC dataset, researchers have tried to analyze land cover class agreement using a relative pixel comparison approach. The results of this kind of analysis produce a spatial agreement analysis that is effective for determining regions with levels of high confidence as a reference [10] [11]. Previous pixel comparison analyses were carried out by comparing GLC datasets, which does not take into consideration when the data was collected. This condition makes the differences that result from a comparison analysis ambiguous as the differences may be a result of real physical changes to the cropland area during that period of time (not because of the different characteristics or classification system of the dataset). Therefore, the method for selecting GLC datasets for a comparison analysis must be reliable and consider time.

The goal of this research was to integrate the combined cropland classes from current GLC datasets to produce a 1 km cropland agreement level (CAL) analysis and its changes (year 2005 and 2010) through two main processes of dataset harmonization and pixel comparison. This study also focused on analyzing the potential use of a cropland agreement level for cropland area estimate and cropland change phenomena.

The presented model of cropland agreement was, to our knowledge, the first study to compare the four current versions of GLC datasets (GlobCover, MODIS LC, GLCNMO and ESA CCI LC) that were focused on worldwide cropland

classes while taking into consideration data collection time. There was no pre-vious GLC dataset comparison [12] [13] that considered data collection time and used similar two-year term GLC datasets. It was the first study to estimate total cropland area on a national level by converting the levels of agreement into per-centage values using a correlation model between the CAL analysis and an IIASA cropland fraction map.

2. Materials

2.1. Global Land Cover (GLC) Datasets

During this study, we explored all existing GLC datasets in the low spatial reso-lution category (300 - 1000 m) and distributed those GLC datasets according to data collection time. After distributing the GLC datasets on a timeline, we grouped the data based on the proximity of data collection time and got five groups of GLC datasets. **Figure 1** shows the five groups of data in the timeline of GLC datasets. This grouping strategy made those datasets more comparable for pixel comparison and integration analysis. Within each group, the maximum difference in data collection time was two years. The reason for setting two years as the maximum time difference was to minimize any extreme changes that could occur in cropland fields over long periods.

To create this CAL analysis, the GLC datasets were put through two processes, which were a harmonization and pixel comparison. In the harmonization process, we standardized the analysis depth for all GLC datasets within five groups. Those five groups consisted of 14 GLC datasets from the seven GLC da-taset versions listed below:

1) IGBP GLCC v 2.0 (1993 dataset) using the International Geosphere-Biosphere Programme (IGBP) classification system produced by the United States Geolog-ical Survey (USGS) [14];

2) UMd (1993 dataset) using the simplified IGBP classification system devel-oped by the University of Maryland [15];
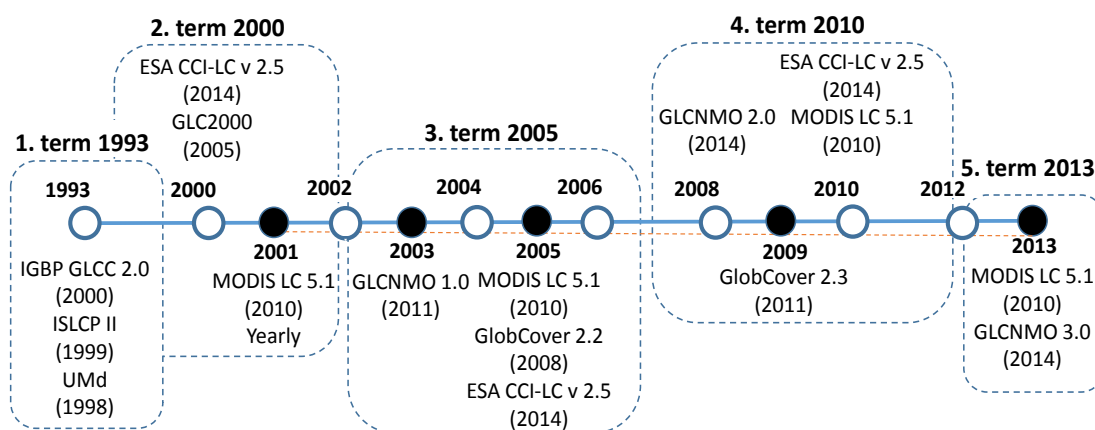

Figure 1. Timeline of the GLC datasets in five groups of data, based on the closeness of time collection, making the pixel comparison process more reliable in regards to time.

3) GLC2000 (2000 dataset) using the Land Cover Classification System (LCCS) of Food and Agricultural Organizations (FAO) generated by the European Commission's Joint Research Center (EC-JRC) [16];

4) GlobCover V 2.2 and V 2.3 (2005 and 2009 datasets) using the FAO LCCS created by the European Space Agency (ESA) [17];

5) MODIS LC (MCD12Q1) collection 5.1 (2001, 2005, 2010 and 2013 datasets) using the IGBP classification system produced by Boston University [18];

6) GLCNMO V.1 and V.2 (2003 and 2008 datasets) using the FAO LCCS created by Chiba University [19] [20];

7) ESA CCI-LC v 2.5 (2000, 2005 and 2010 datasets) using the FAO LCCS generated by the ESA [21], in 2017 ESA have published CCI-LCmaps from 1993-2015 that can be useful for long year monitoring of land cover change.

For the pixel comparison process, we used datasets from the 2005 and 2010 groups from the same GLC dataset versions. These GLC datasets were the GlobCover, MODIS LC,GLCNMO and ESA CCI LC datasets. The same GLC datasets from each group were paired to produce two CAL analyses for further analysis of cropland change. Table 1 shows the characteristics of each of the GLC datasets in the 2005 and 2010 groups that were used in the pixel comparison process.

2.2. Reference Data

This study used a relative accuracy assessment, which was the use as a reference of another dataset that was considered to use accurate datasets. We explored the reference data in order to convert agreement levels into percentage values. As reference data, a 1 km global IIASA-IFPRI cropland percentage map with a 2005 baseline year [22] was developed by integrating a number of individual global and regional cropland maps. This IIASA-IFPRI cropland percentage map was validated by high-resolution satellite imagery via Geo-Wiki (http://www.geo-wiki.org) [23] with an overall accuracy result of 82.4%. The reason for choosing the IIASA data was because this dataset was a recent one that gave a highly accurate overview of cropland area distribution in percentage values.

3. Methods

To achieve the study's first goal, there were two main methods for producing the CAL analysis, which were the harmonization and pixel comparison processes. The different characteristics and classification systems for each dataset made these two processes important for getting reliable CAL analysis results. During the harmonization process, we standardized all GLC datasets to produce comparable cropland classes for the pixel comparison process. The result of the cropland harmonization process was evaluated by comparing the result to each group. A balanced distribution verified that the data was standardized and comparable. Finally, datasets within the same group were overlaid using the CRISP approach for a pixel comparison processes to produce the CAL analysis

Table 1. Characteristics of the four GLC datasets used in the pixel comparison analysis.

Dataset name	Data provider	Satellite sensor	Spatial resolution	Time of data collection	Projection/ Datum	Input data	Classification technique	Classification scheme	Data for Validation	Overall accuracy
GlobCover 2005 v 2.2	ESA, MEDIAS-France /POSTEL	Envisat (MERIS)	300 m (1/360°)	December 2004-June 2006	Plate-Carrée with a Geographic (Lat/lon)/ WGS_1984	MERIS L1B data	Unsupervised classification	FAO LCC/ 22 Classes	SPOT-VEGETAT ION NDVI, and Virtual/Google Earth	Globally 73.2%
GlobCover 2009 v 2.3	ESA, Univ. Catholique de Louvain			January-December 2009		MERIS FR mosaics (13 Spectral bands)	Classification using local/regional expertise			Globally 67.1%
MODIS LC (v.5.1) (MCD12Q1)	Boston University, NASA	Terra Aqua (MODIS)	500 m (1/240°)	Yearly since 2001-2013	MODIS sinusoidal/ WGS_1984	MODIS L2/L3, MODIS 16-day EVI, MODIS 8-day LST, DEM	Supervised classification using decision tree, neural networks	IGBP/ 17 Classes	High resolution land cover information	Globally 75%
GLCNMO (v.1) 2003	GSI Japan, Univ. Chiba, ISCGM	Terra Aqua (MODIS)	1000 m (1/120°)	January-December 2003	Geographic (Lat-lon)/ WGS_1984	500-m MODIS 16-day, Landsat ETM+, DMSO-OLS	Supervised classification	FAO LCCS/ 20 Classes	Google Earth and regional existing maps	Globally 76.5%
GLCNMO (v.2) 2008			500 m (1/240°)	January-December 2008						Globally 77.9%
ESA CCI-LC (v 2.5) 2005	ESA	MERIS Full, SPOT-VGT (Reduced Resolution)	300 m (1/360°)	2003-2007 epoch	Geographic (Lat-lon)/ WGS_1984	MERIS 10-year LC map as baseline SPOT-VGT time series	Unsupervised classification	FAO LCCS/ 22 classes	2600 Primary Sampling Units (PSUs) and other reference	Globally 74.4%
ESA CCI-LC (v 2.5) 2010				2008-2012 epoch						

As the fourteen GLC datasets from each group had different mapping standards, we applied the following cross-walking methods to get comparable products:

3.1. Re-Projection

To facilitate different projections, these data sets were co-registered and re-projected to a geographic (latitude-longitude) image with map datum WGS 84, which is a pseudocylindrical equal-area map projection [24]. This projection has been used in many GLC datasets. GlobCover, GLCNMO and ESA CCI-LC are GLCs that have used this projection. Its use was also supported by the use of this projection system in similar studies of GLC datasets comparisons [25].

3.2. Rescaling Method

The resampling process had an important role in standardizing the different pixel resolutions of the GLC datasets because during the comparison process the GLC datasets needed to have the same pixel size. Different rescaling processes were a special concern in this study as a mistake in the resampling process could have caused changes in the class area. When considering the goals of this study, which was to resample all GLC datasets to a 1 km resample target, this study attempted to adopt and combine some important parts of the resampling process from another study with an aim of minimizing errors from the resampling process.

For the cropland rescaling process, we used two resampling processes, which were the nearest neighbor and maximum area methods. We used these two techniques because both would not change the value of cells. For datasets that have coarser resolutions, direct resampling using the nearest neighbor method might have caused a non-ignorable disagreement between the original and the rescaled dataset [9]. This is based on the fact that the footprint of the sensor is not at the same location during each revisit [26]. The maximum area method, however, has been proven a more powerful approach for aggregating discrete land cover data [27]. Furthermore, it also tends to give a smoother result than the nearest neighbor method because the new value resulting from this approach is obtained based on the most common values around the pixel. The GLC datasets that already have a 1 km resolution did not need a rescaling process, which minimized the changes from resampling the results data [26].

Taking into consideration the facts from this previous explanation, the steps that were taken in the rescalling process during this study were as follows: 1) Global land cover datasets that have a large resolution, such as the GlobCover and ESA CC-LC (300 m), MODIS LC (500 m), and GLMNO 2008 V.2 (500 m) datasets, were resampled on a grid with a 250 m resolution using the nearest neighbor method. 2) After that, the entire GLC 250 m grid was aggregated using a majority area method to 1 km for all datasets. 3) For GLMNO 2003 v1 since they were already in a 1 km spatial resolution, the datasets were kept in their

original resolution without a rescaling process.

3.3. Legend Harmonization

For this study, legend harmonization played a crucial role because the focus of this study was on analyzing cropland classes among several GLC datasets. Differences in the definitions and characteristics of each class correlating with the cropland in each individual GLC would produce ambiguities among the comparisons of the GLC datasets. To overcome this problem, we evaluated all the classes used by the different GLC datasets and then analyzed those classes to determine how they correlated with a standardized cropland classification system, which was the Land Cover Classification System (LCCS). The rationale for choosing LCCS rather than IGBP was: 1) LCCS was developed after an analysis of existing relevant FAO nomenclature documents [26] that can explain some categories of cropland classes in detail. 2) MODIS LC v 6, one of the GLC products that use IGBP, would soon change to LCCS with their next dataset [18].

One step in the legend harmonization process was to convert the original class numbers to LCCS-labels. Table 2 shows the conversion results for some of the original class numbers to cropland LCCS-labels. The class definition followed the LCCS hierarchy based on the dichotomous phase and the modular-hierarchical phase [28]. As in Table 2, the cropland class of the MODIS LC, LCCS-label is "A11-A3". A11 is the dichotomous phase of "Cultivated and Managed Land" and A3 is the modular-hierarchical phase of "Herbaceous Crop". The fact that the original class descriptions from the GLC datasets had a great impact on choosing the type of dichotomous phase and modular-hierarchical phase, we adopted the descriptions of the original classes from the GLC datasets from a previous study [29] [30]. For this conversion, we grouped all classes that correlate with cropland into a new cropland class, including mosaic vegetation cropland that had a smaller cropland percentage when compared to vegetation. Since all the GLC datasets in the five year groups would be harmonized for comparison proposes, Table 2 shows information about the type of class that was categorized as "cropland" based on the LCCS-label conversion status of seven GLC datasets.

4. Result and Discussion

4.1. Thematic Similarity

Following the M. C. Hansen and B. Reed [31] strategy, we evaluated the harmonization results for the cropland classes by analyzing cropland pixel similarity in each dataset group. A successful result in the harmonization process would produce a balanced proportion among the GLC datasets in the same groups of GLC data. Figure 2 shows the proportion of the cropland classes from the harmonization results and Figure 3 shows the spatial distribution of harmonized cropland area.

Table 2. The conversion results for the original cropland classes converted to LCCS-labels for all GLC datasets.

Global Datasets (Number of class)	LCC Level	LCC Label
A. GLCC v.2/ IGBP DIS (IGBP)		
Croplands (12)	A11-A3	Herbaceous Crop(s)
Cropland/Natural Vegetation Mosaic (14)	A11-A3	Herbaceous Crop(s)
	A12	Natural and Semi-natural Primarily Terrestrial Vegetation
B. UMd (IGBP)		
Croplands (11)	A11-A3.B5	Continuous Field(s) Of Herbaceous Crop(s)
C. GLC2000 (LCCS)		
Cultivated and managed areas (16)	A11	Cultivated and Managed Terrestrial Area(s)
	A23	Cultivated Aquatic or Regularly Flooded Area(s)
Mosaic: Cropland/Tree Cover/Other Natural Vegetation (17)	A11	Cultivated and Managed Terrestrial Area(s)
	A12-A3.A20.B2	Closed to Open Trees
	A12	Natural And Semi-natural Primarily Terrestrial Vegetation
Mosaic: Cropland/Shrub and/or Herbaceous cover (18)	A11	Cultivated and Managed Terrestrial Area(s)
	A24-A4.A20	Closed to Open Shrubland (Thicket)
	A12-A2.A20	Herbaceous Closed to Open Vegetation
D. MODIS LC (IGBP)		
Croplands (12)	A11-A3	Herbaceous Crop(s)
Cropland/ Natural Vegetation Mosaic (14)	A11-A3	Herbaceous Crop(s)
	A12	Natural and Semi-natural Primarily Terrestrial Vegetation
E. GlobCover (LCCS)		
Post-flooding or irrigated croplands (or aquatic) (11)	A11-A1XXXXD3	Irrigated Tree Crop
	A11-A2XXXXD3	Irrigated Shrub Crop
	A11-A3XXXXD3	Irrigated Herbaceous Crop
	A11-A3XXXXD2	Post Flooding Herbaceous Crop
	A11-A2XXXXD1	Rainfed Shrub Crop
Rainfed croplands (14)	A11-A1XXXXD1	Rainfed Tree Crop
	A11-A3XXXXD1	Rainfed Herbaceous Crop
Mosaic cropland/vegetation (grassland/shrubland/forest) (20)	A11	Cultivated and Managed Terrestrial Area(s)
	A12	Natural and Semi-natural Primarily Terrestrial Vegetation
Mosaic vegetation (grassland/shrubland/ forest)/cropland (30)	A12	Natural and Semi-natural Primarily Terrestrial Vegetation
	A11	Cultivated and Managed Terrestrial Area(s)
F. GLCNMO (LCCS)		
Cropland (11)	A11-A3	Herbaceous Crop(s)
Paddy field (12)	A11-A4.A5	Graminoid Crops//Non-Graminoid Crops
Cropland /Other Vegetation Mosaic (13)	A11	Cultivated and Managed Terrestrial Area(s)
	A12	Natural And Semi-Natural Primarily Terrestrial Vegetation
	A23	Cultivated Aquatic or Regularly Flooded Area(s)

Continued

| G. ESA CCI-LC (LCCS)		
Cropland, rainfed (10)	A11-A2XXXXD1	Rainfed Shrub Crop
	A11-A1XXXXD1	Rainfed Tree Crop
	A11-A3XXXXD1	Rainfed Herbaceous Crop
Cropland, irrigated or post-flooding (20)	A11-A1XXXXD3	Irrigated Tree Crop
	A11-A2XXXXD3	Irrigated Shrub Crop
	A11-A3XXXXD3	Irrigated Herbaceous Crop
	A11-A3XXXXD2	Post Flooding Herbaceous Crop
Mosaic cropland (>50%)/natural vegetation (tree, shrub, herbaceous cover) (<50%) (30)	A11	Cultivated and Managed Terrestrial Area(s)
	A12	Natural and Semi-natural Primarily Terrestrial Vegetation
Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%)/cropland (<50%) (40)	A12	Natural and Semi-natural Primarily Terrestrial Vegetation
	A11	Cultivated and Managed Terrestrial Area(s)
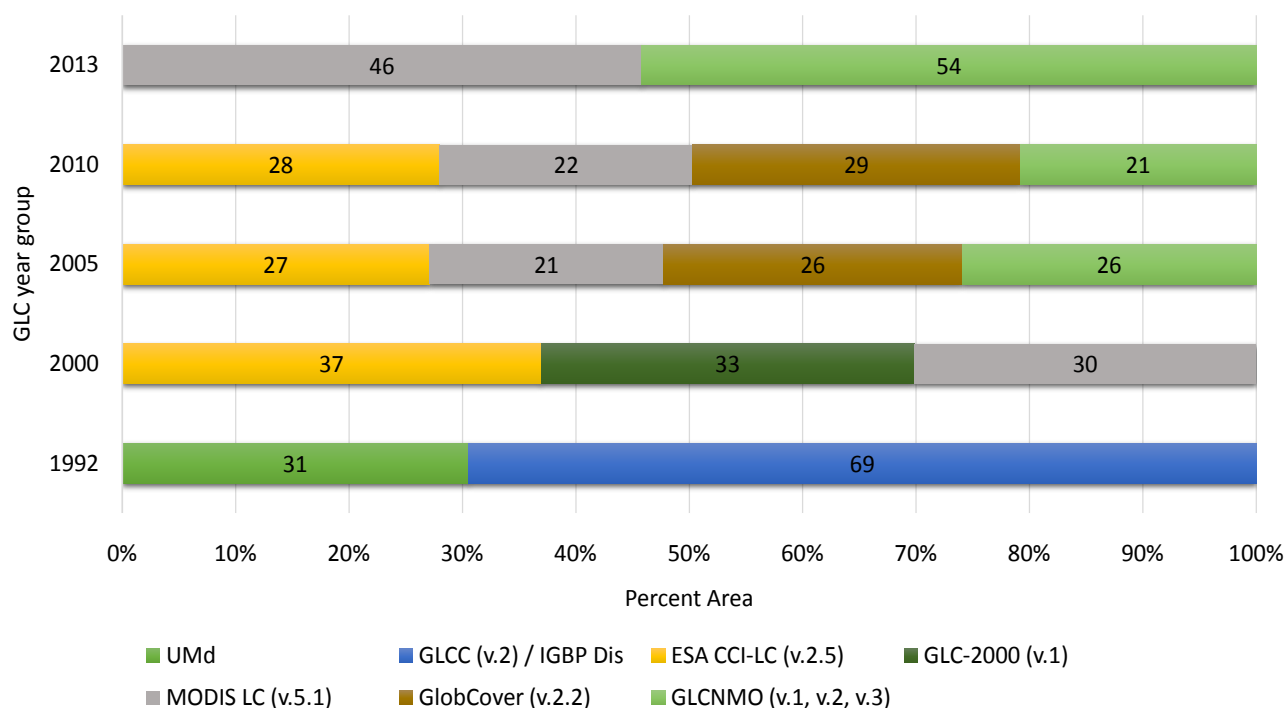
Figure 2. Proportion value of cropland analyses as the result of the harmonization of the five groups.

In general, four out of five groups showed a balanced proportion of cropland classes. The subtraction value of the highest from the lowest percentage in each group were 7%, 6%, 6%, 8%, 8% for 2000, 2005, 2010 and 2013 group respectively. However, in the 1992 group, there was a large difference in cropland classes between UMd and GLCC of 38%. The absence of mosaic agricultural classes in the UMd datasets was the cause of this difference [15]. This shows that the differences in the goals and focus of an analysis caused conspicuous differences between cropland and vegetation by classifying more area as a vegetation
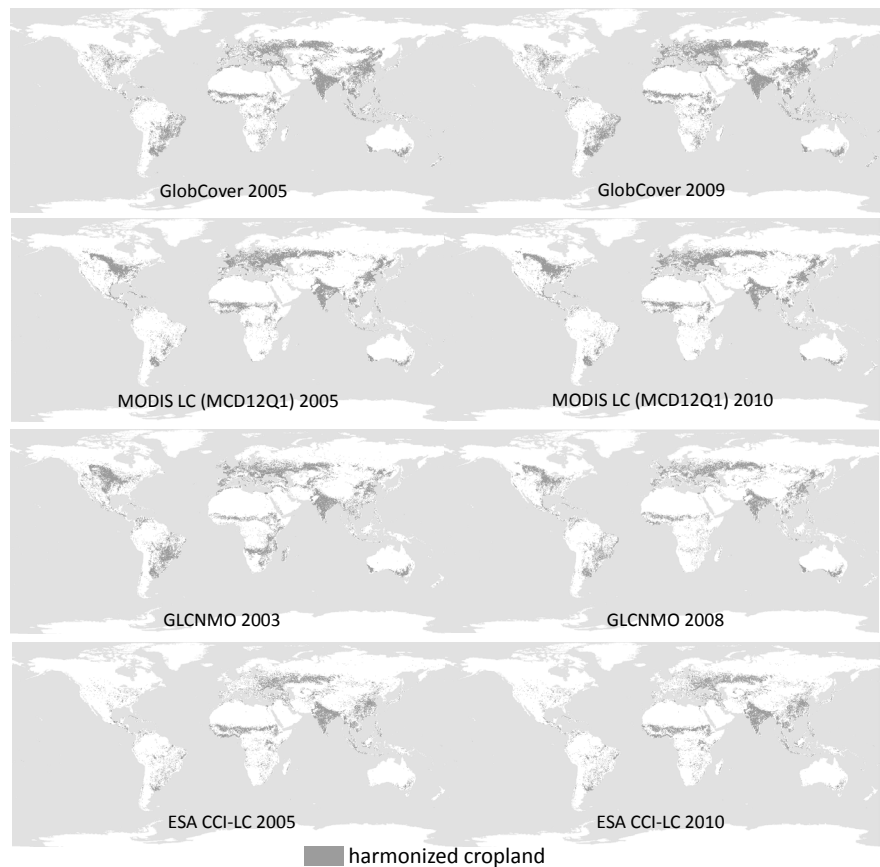
Figure 3. Spatial distribution of harmonized cropland area (cropland and mosaic cropland) from four selected GLC datasets in the two year-groups of 2005 and 2010. Only groups of GLC datasets that are located within these two year-groups are used for the cropland class comparison process.

class. This explanation shows that the harmonization result was acceptable especially for the 2005 and 2010 groups. So, both were eligible to be analyzed in the next step.

4.2. Cropland Level Agreement (CAL) Analysis

After assigning the harmonized dataset target and proving the balance proportion value, we compared and observed for pixel similarity within the four GLC datasets in the two year-groups by overlaying the datasets using the CRISP approach, which is based upon cross-walking between classes [9]. Since we only focused on one cropland class (not analyzing multiple classes), utilizing a CRISP approach that matches using one-to-one mapping was sufficient [12] [32]. Another technique is the Fuzzy approach, which can allow an overlap between legend definitions to be taken into consideration. It requires expert knowledge to quantify uncertainty in the classification and transition zones of boundaries [33] [34]. In the previous study explained by McCallum *et al.* [12], the result of this pixel comparison technique could indicate an agreement level among these datasets.

The comparison result of these cropland classes produced four agreement levels, which in this study were also referred to as the Cropland Agreement Level (CAL) analysis. The four levels that were obtained from the CAL analysis in this study were as follows (Figure 4):

Level 1: No agreement for pixels with a unique aggregated class in each data set.

Level 2: Low agreement for pixels where only two of the four data sets were in agreement.

Level 3: Medium agreement for pixels where three of the four data sets were in agreement.

Level 4: Full agreement where all the four data sets within a pixel were in agreement.

Since we focused on analyzing cropland class similarity within four GLC datasets in two groups of data from 2005 and 2010, we produced two CAL analyses from those years.

To study this CAL analysis more deeply, we divided the study area into seven test sites according to size. Those study areas were South America, North America, Europe, Africa, Australia, Russia and Asia. Table 3 shows the percentage calculation result for four CAL analysis levels for seven test sites. Four GLC datasets were used and the underlined value shows the highest percentage value whereas the bolded value shows the lowest percentage for the site area.

Comparison results for the seven sites in the two years of analysis showed that Europe was indicated as the area with highest full agreement level whereas
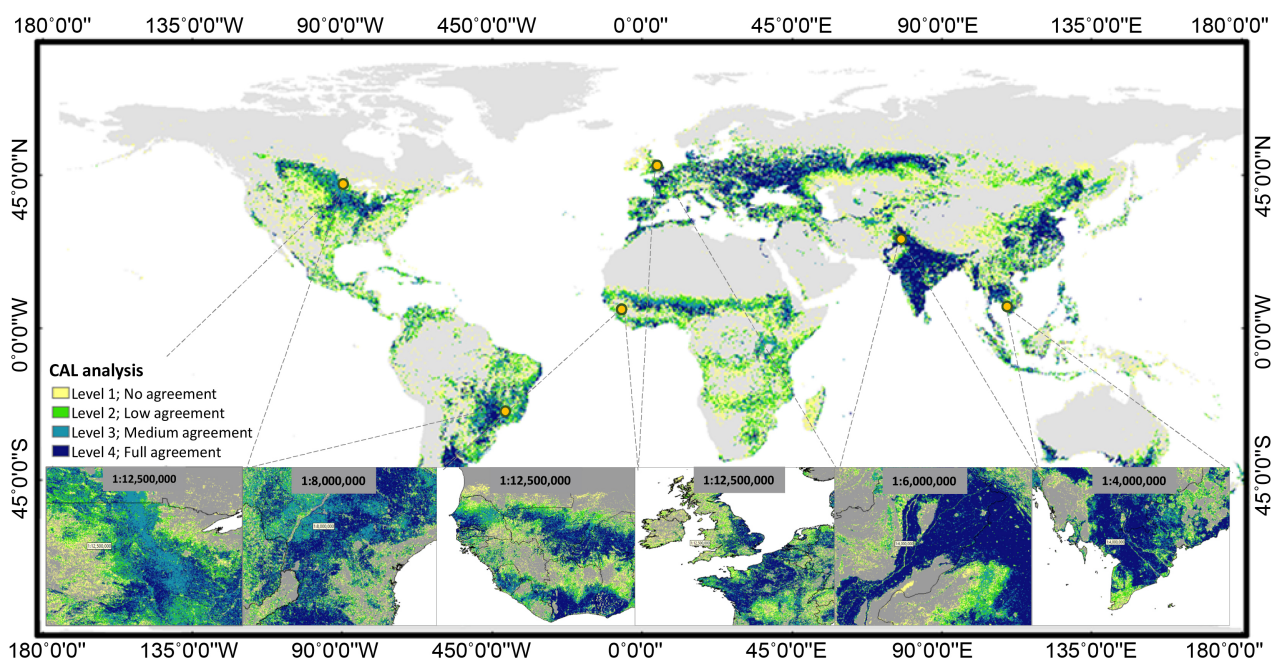

Figure 4. Cropland Agreement Level (CAL) analysis 2005, based on pixel comparison analysis between four selected recent GLC datasets.

Table 3. Agreement percentage of the CAL analysis across the seven test sites in 2005 and 2010.

Test Sites	No Agreement		2 of 4 Agree		3 of 4 Agree		Full Agreement	
	2005	2010	2005	2010	2005	2010	2005	2010
South America	0.32	0.27	0.24	0.23	<u>0.25</u>	<u>0.24</u>	0.19	0.26
Nort America	0.44	0.32	<u>0.26</u>	<u>0.28</u>	0.20	0.26	0.10	0.15
Europe	**0.21**	**0.18**	0.19	**0.16**	0.24	0.19	<u>0.36</u>	<u>0.47</u>
Africa	0.40	0.39	0.29	0.26	0.22	0.19	0.08	0.16
Australia	<u>0.60</u>	<u>0.49</u>	**0.17**	0.21	**0.17**	0.21	**0.06**	**0.09**
Russia	0.32	0.26	0.20	0.19	0.22	**0.17**	0.27	0.38
Asia	0.33	0.27	0.21	0.20	0.19	0.19	0.27	0.34

Australia, Russia and Africa had the three highest no agreement level areas. A short time period for cultivation and a small cropland area combined with a large area of vegetation were the causes of a mix in cropland classes. This triggered different classification results in each GLC dataset. Besides this, there were also random changes in levels two and three of the 2005 and 2010 CAL analyses.

4.3. Correlation Factor between CAL and Existing Cropland Fraction

To obtain a correlation model between the CAL analysis and cropland percentage, first we analyze the trend correlation between the CAL analysis with the original cropland classes from the four selected GLC datasets (Table 4). The dominant distribution of "cropland" classes (class numbers 10 and 20 for ESA CCI-LC, numbers 11 and 14 for GlobCover, numbers 11 and 12 for GLCNMO and number 12 for MODIS LC) is in level 4, compared with the "mosaic cropland" that is dominant in level 1 and level 2, indicating that the four levels in the CAL analysis correlate with the amount of actual cropland area in one pixel.

To investigate the meaning of each CAL analysis level in percentage values, we analyze the pixel correlation between the CAL analysis and the IIASA-IFPRI cropland percentage map by using a 2D Scatter Plot (Figure 3). Results show that level 4 of the CAL analysis mainly corresponds with 80% of cropland fraction and with the same approach, the following correlations were also obtained: level 3 with 40%, level 2 with 20% and level 1 with 10% (Table 5). Figure 5 shows a comparison between the CAL model and IIASA-IFPRI map using a 2D Scatter Plot. To estimate cropland area from the CAL analysis, the pixel area from each agreement level is multiplied by the correlation percentage.

4.4. Comparison of Cropland Area Estimates from CAL with the FAO Data

To evaluate cropland estimation results on the national level, cropland areas derived from the CAL analysis are compared to the cropland area estimates from

Table 4. Trend correlation between the CAL analysis with the original cropland classes from the four selected GLC datasets.

GLC cropland classes	CAL analysis Level 1	Level 2	Level 3	Level 4
GlobCover				
11 Post-flooding or irrigated croplands (or aquatic)	0.2	0.5	1.4	5.7
14 Rainfed croplands	2.5	4.3	8.4	14.8
20 Mosaic cropland/vegetation	3.0	4.9	7.8	10.9
30 Mosaic vegetation/cropland	6.5	10.2	11.1	7.7
MODIS LC				
12 Croplands	2.1	7.4	18.3	34.3
14 Cropland/Natural Vegetation Mosaic	6.5	10.6	11.7	9.0
GLCNMO				
11 Cropland	12.5	13.9	19.2	28.6
12 Paddy field	0.3	0.3	0.7	2.2
13 Cropland/Other Vegetation Mosaic	6.4	5.2	4.8	5.8
ESA CCI-LC				
10 Cropland, rainfed	3.7	11.4	22.3	27.3
20 Cropland, irrigated or post-flooding	0.4	0.9	1.9	5.5
30 Mosaic cropland (>50%)/natural vegetation (<50%)	1.9	4.0	4.6	3.6
40 Mosaic natural vegetation (>50%)/cropland (<50%)	3.1	4.6	3.2	1.5

Table 5. Correlation factor between CAL analysis and IIASA cropland fraction.

CAL analysis	IIASA cropland fraction									
	10%	**20%**	30%	**40%**	50%	60%	70%	**80%**	90%	100%
Level 1	**41.1**	25.5	16.4	11.1	7.9	4.7	2.5	2.3	4.9	16.4
Level 2	34.4	**35.1**	29.1	24	18.6	12.8	7.5	5.8	9.2	26.6
Level 3	19.6	28.1	34.6	**37**	35.1	31.6	25.6	19	21.7	27.5
Level 4	5	11.3	19.8	27.9	38.3	50.9	64.4	**72.9**	64.2	29.5
Total	100	100	100	100	100	100	100	**100**	100	100

FAO-stat 2005 as a statistical data reference. Based on the definition from FAO statistics, cropland is defined as "arable land and permanent crops" [35]. Overall correlation also is observed for 2005 and 2010 respectively and shows some proximity between the CAL analysis and the FAO with a 0.70 and 0.71 regression value (Figure 6).

We also analyze the accuracy of the four selected GLC datasets to FAO statistics. We divide countries into two groups based on the subtraction value of the cropland area. The groups are 1) small, which have a cropland area from 5000 to 140,000 km², and 2) medium to large, which have cropland area from 180,000 to
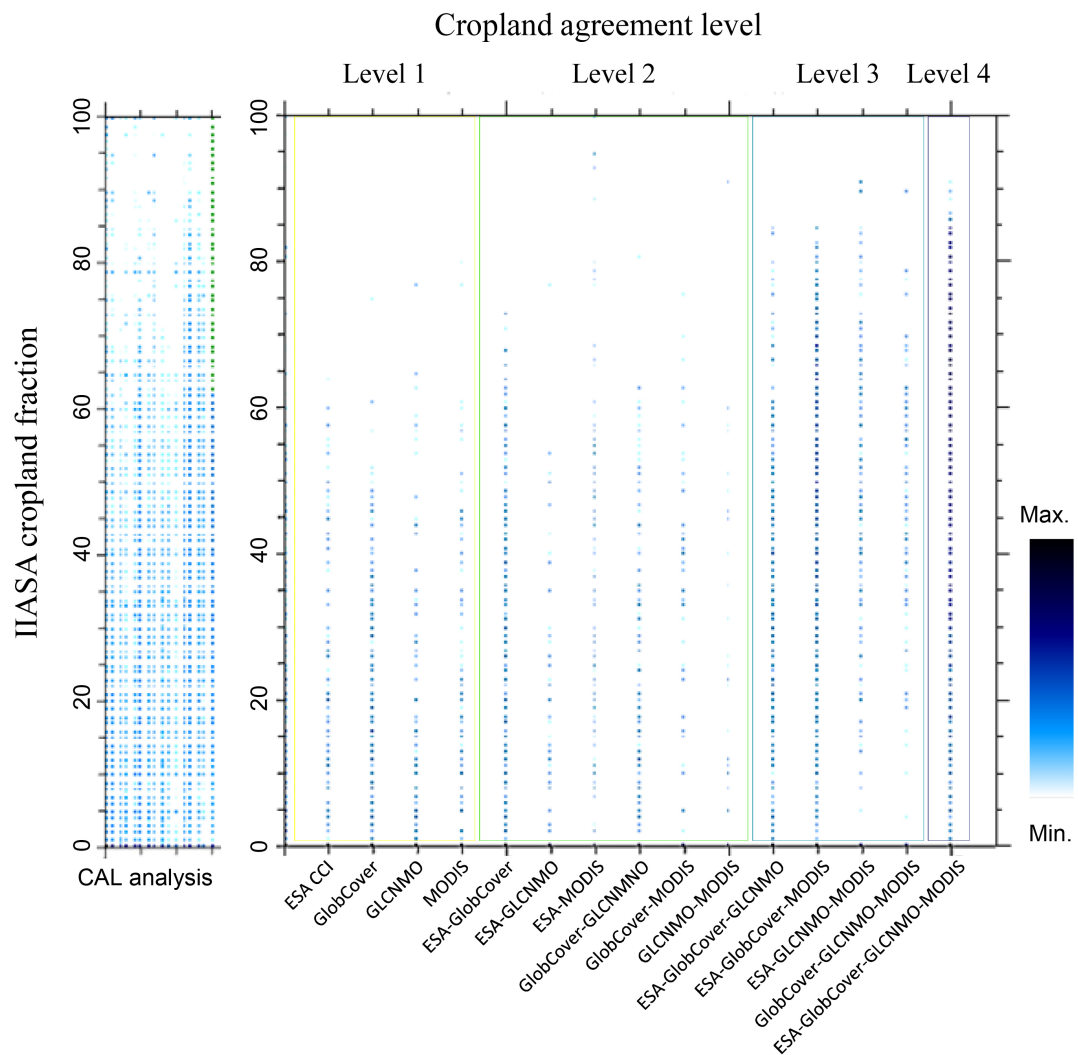
Figure 5. Comparison between the CAL analysis and IIASA cropland fraction using scatter plot.

1,700,000 km² (**Figure 7**). In this framework, the regression value and relative error to the FAO (%) have been observed as 0.574; 42.2 for small cropland area countries and 0.858; 29.8 for medium to large cropland area countries between FAO-stat and the CAL analysis (**Table 6**). Those values promote the CAL analysis as the most accurate dataset compared to FAO-stat within all datasets in estimating cropland area. Good correlation of the results to FAO statistics is still not enough for quality assessment of the product. To provide accurate validation it is worth to compare the product with some more precise country level cropland map which have higher resolution. Globland30 [36] and Unified Cropland layer [37] as one of the alternative for the analysis in next study.

4.5. Cropland Agreement Level (CAL) Change Analysis

The CAL change analysis is used to study the potentiality of cropland area changes monitored between 2005 and 2010. **Figure 8** shows global cropland agreement level change. The area that experience the increase and decrease in
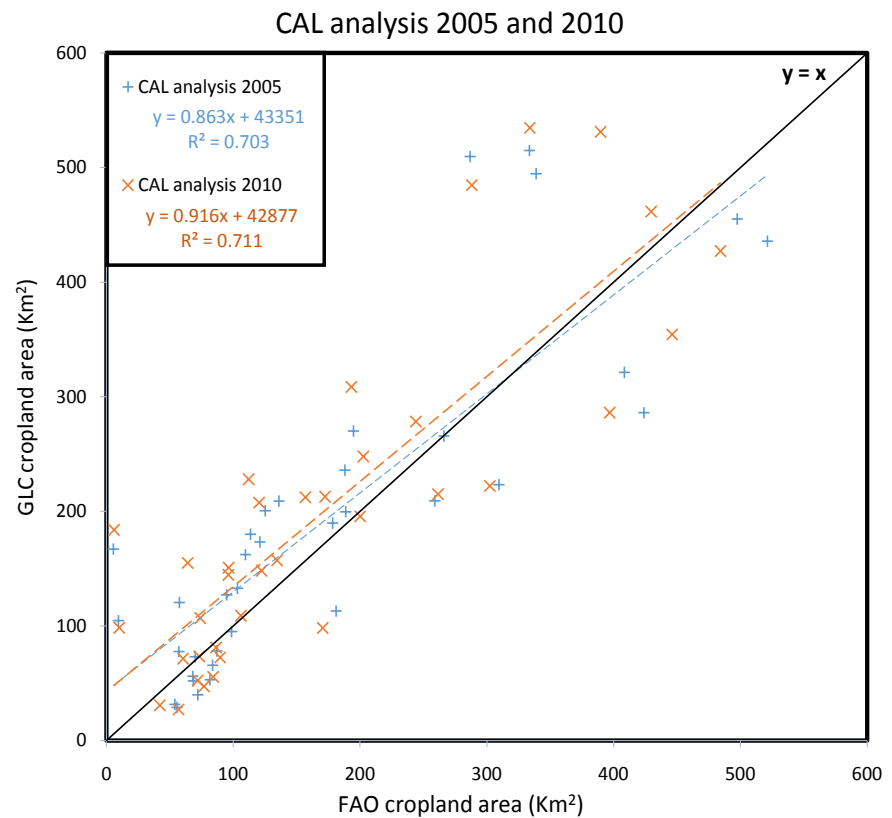
Figure 6. Regression model of cropland area estimates derived from two years of the CAL analysis with FAO cropland area statistics.

CAL change analysis especially around southern Africa actually categorized as mixed cropland with pasture area. It can be happened since some GLC mentioned that area as mosaic cropland class (**Figure 3**).

Almost all of the CAL analysis and the four GLC datasets are not able to produce similar cropland area change data comparable to the cropland change statistical data from the FAO. However, the CAL analysis has the highest proximity of cropland change compared to the other GLC datasets.

To give a visual description of cropland area changes, we chose Bali Island, Indonesia as a test site. According to FAO statistics, Indonesia is a country that has had massive cropland area expansion (**Figure 9**). Bali was chosen because within five years (from 2005-2010) it had large changes in cropland area, both in regards to an expansion of cropland area and shrinkage in cropland area caused by changes in land use. Massive development in the tourism sector caused changes in land use from cropland to housing, especially in Denpasar [38].

Moreover, the development of three new reservoirs in Bali and East Java built within the 2005-2010 period expanded the cropland area around the reservoirs [39]. **Figure 9** shows that in ESA CCI LC and GlobCover, almost all of the cropland area in 2005 and 2010 had no change, whereas in the MODIS data, there is a decline in cropland change. In contrast, GLCNMO shows an expansion in cropland area. Cropland change results from the CAL analysis can accommodate
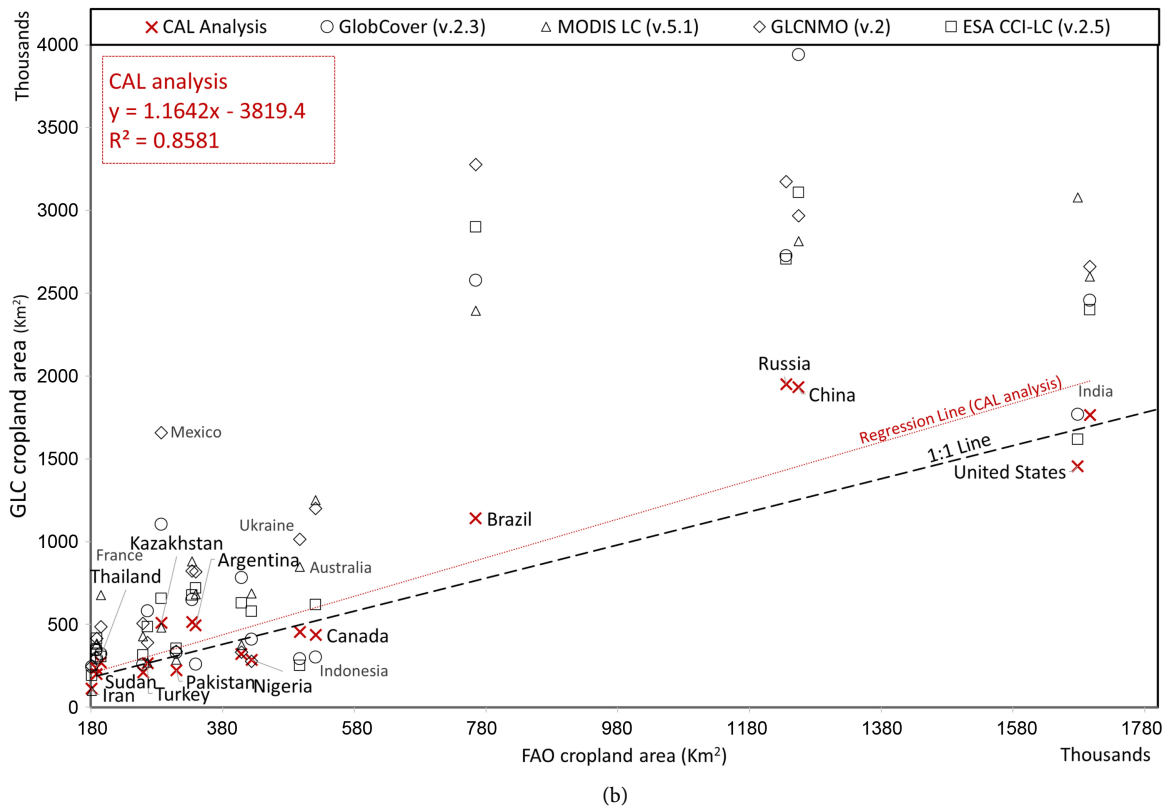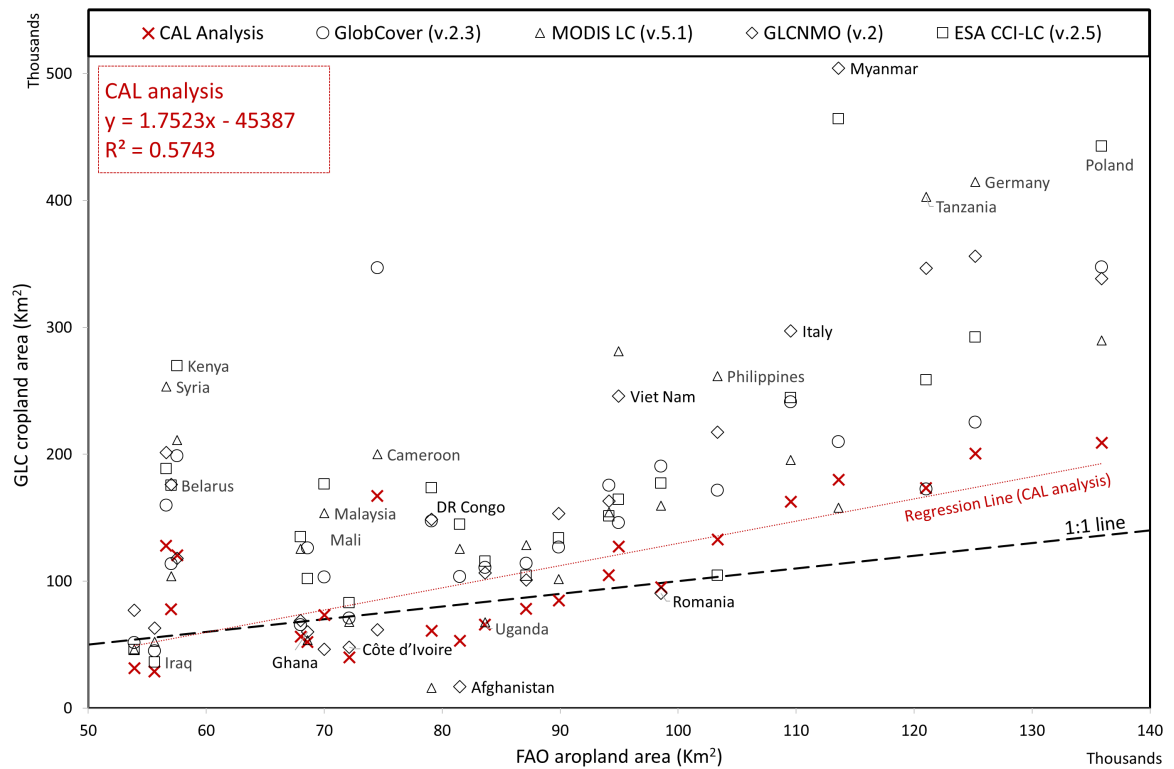
Figure 7. Comparison on the national level between the CAL analysis and the four original GLC datasets with cropland area estimates from the FAO in (a) small cropland area countries and (b) medium to large cropland area countries.
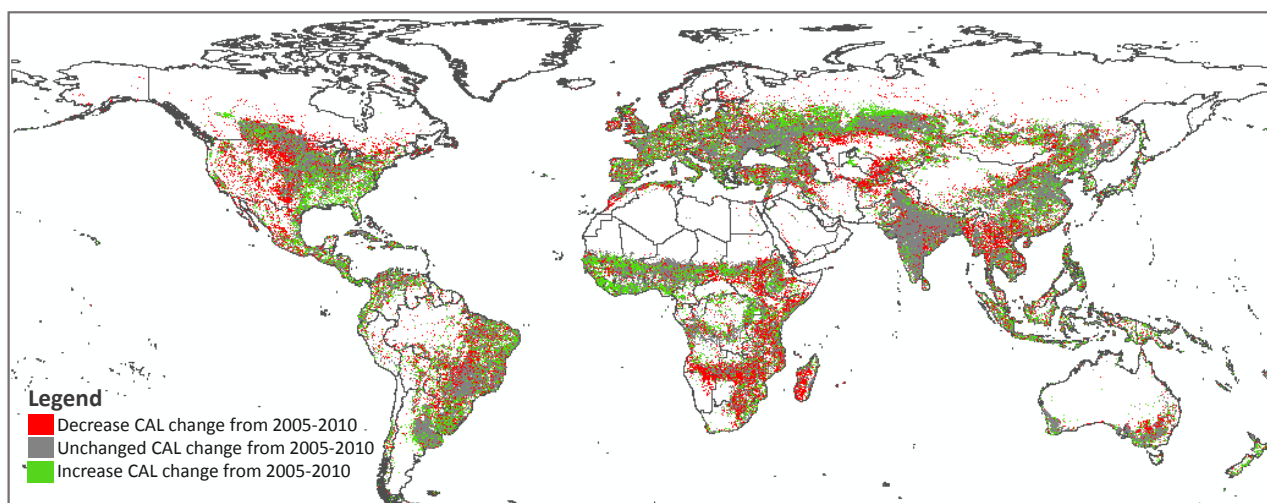
Figure 8. Global Cropland Agreement Level (CAL) change analysis from 2005 to 2010.
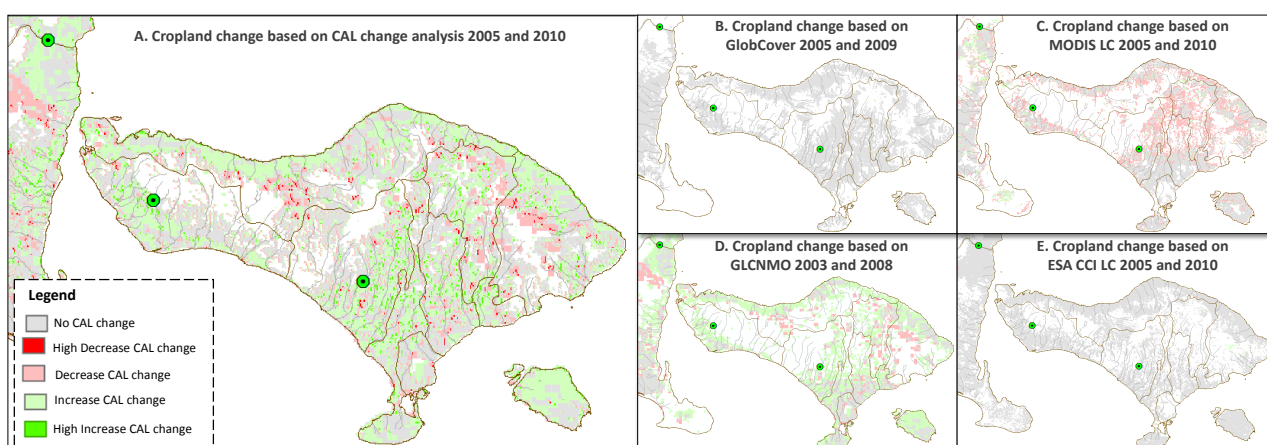

Figure 9. Cropland change of Bali Island, Indonesia (2005 and 2010): (A) Cropland change derived from CAL change analysis, (B) GlobCover, (C) MODIS LC, (D) GLCNMO and (E) ESA CCI LC.

Table 6. Regression value and relative error of the CAL product and the four GLC datasets to FAO statistical data.

GLC datasets	Small cropland area countries		Medium to large cropland area countries	
	R^2	Relative error to FAO (%)	R^2	Relative error to FAO (%)
CAL analysis	0.57	42.2	0.86	29.8
GlobCover (v.2.3)	0.34	85.6	0.65	79.3
MODIS LC (v.5.1)	0.37	104.2	0.79	99.2
GLCNMO (v.2)	0.55	45.1	0.79	75.2
ESA CCI-LC (v.2.5)	0.32	144.5	0.66	77.6

the results from the four GLC datasets, and it also accommodates an area with an extreme level of changes.

5. Conclusion

This study shows that integrating recent GLC datasets can be considered for es-

timating cropland area with the highest accuracy among original datasets. The CRISP approach is also used to analyze per pixel comparisons between cropland map datasets and produce a Crop Agreement Level (CAL) analysis by integrating four GLC datasets in the two year-groups (2005 and 2010). To calculate cropland area from the CAL analysis, an IIASA-IFPRI cropland percentage map is used. The correlation model obtained from the CAL analysis and IIASA comparison successfully estimated the percentage value of four agreement levels. When the correlation between the CAL analysis and cropland percentage is studied, the result shows a good correlation where level 1 correlates with 10%, level 2 with 20%, level 3 with 40% and level 4 with 80% cropland area. The regression value for the CAL analysis is 0.70 - 0.71, this value was the highest compared to other datasets, which are ESA CCI-LC (0.47 - 0.49), GlobCover (0.41 - 0.43), GLCNMO (0.43 - 0.59) and MODIS LC (0.63 - 0.65). Cropland area estimates for each country in 2005 and 2010 show that the CAL analysis is more accommodating for cropland change calculation based on FAO cropland change statistical data.

Acknowledgements

References

[1] FAO, IFAD, and WFP (2015) The State of Food Insecurity in the World 2015. Meeting the 2015 International Hunger Target: Taking Stock of Uneven Progress. FAO, Rome.

[2] FAO (2016) The State of Food and Agriculture. Climate Change, Agriculture and Food Security. FAO, Rome.

[3] See, L., Fritz, S., You, L., Ramankutty, N., Herrero, M., Justice, C., Becker-Reshef, I., Thornton, P., Erb, K., Gong, P., Tang, H., van der Velde, M., Ericksen, P., McCallum, I., Kraxner, F. and Obersteiner, M. (2015) Improved Global Cropland Data as an Essential Ingredient for Food Security. *Global Food Security*, **4**, 37-45. https://doi.org/10.1016/j.gfs.2014.10.004

[4] Leff, B., Ramankutty, N. and Foley, J.A. (2004) Geographic Distribution of Major Crops across the World. *Global Biogeochemical Cycles*, **18**, GB1009.

[5] Doll, P. and Siebert, S. (2002) Global Modeling of Irrigation Water Requirements. *Water Resources Research*, **38**. https://doi.org/10.1029/2001WR000355

[6] Grekousis, G., Mountrakis, G. and Kavouras, M. (2015) An Overview of 21 Global and 43 Regional Land-Cover Mapping Products. *International Journal of Remote Sensing*, **36**, 5309-5335. https://doi.org/10.1080/01431161.2015.1093195

[7] Fritz, S. and See, L. (2008) Identifying and Quantifying Uncertainty and Spatial Disagreement in the Comparison of Global Land Cover for Different Applications.

Global Change Biology, **14**, 1057-1075.
https://doi.org/10.1111/j.1365-2486.2007.01519.x

[8] Ran, Y., Li, X. and Lu, L. (2010) Evaluation of Four Remote Sensing Based Land Cover Products over China. *International Journal of Remote Sensing*, **31**, 391-401. https://doi.org/10.1080/01431160902893451

[9] Bai, Y., Feng, M., Jiang, H., Wang, J., Zhu, Y. and Liu, Y. (2014) Assessing Consistency of Five Global Land Cover Data Sets in China. *Remote Sensing*, **6**, 8739-8759. https://doi.org/10.3390/rs6098739

[10] Giri, C., Zhu, Z. and Reed, B. (2005) A Comparative Analysis of the Global Land Cover 2000 and MODIS Land Cover Data Sets. *Remote Sensing of Environment*, **94**, 123-132. https://doi.org/10.1016/j.rse.2004.09.005

[11] Tsendbazar, N., De Bruin, S., Fritz, S. and Herold, M. (2015) Spatial Accuracy Assessment and Integration of Global Land Cover Datasets. *Remote Sensing*, **7**, 15804-15821.

[12] McCallum, I., Obersteiner, M., Nilsson, S. and Shvidenko, A. (2006) A Spatial Comparison of Four Satellite Derived 1 km Global Land Cover Datasets. *International Journal of Applied Earth Observation and Geoinformation*, **8**, 246-255. https://doi.org/10.1016/j.jag.2005.12.002

[13] Nakaegawa, T. (2011) Uncertainty in Land Cover Datasets for Global Land-Surface Models Derived from 1-km Global Land Cover Datasets. *Hydrological Processes*, **25**, 2703-2714. https://doi.org/10.1002/hyp.8011

[14] Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z. and Yang, L. (2000) Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1 km AVHRR Data. *International Journal of Remote Sensing*, **21**, 1303-1330. https://doi.org/10.1080/014311600210191

[15] Hansen, M.C., Defries, R.S., Townshend, J.R.G. and Sohlberg, R. (2000) Global Land Cover Classification at 1 km Spatial Resolution Using a Classification Tree Approach. *International Journal of Remote Sensing*, **21**, 1331-1364. https://doi.org/10.1080/014311600210209

[16] Bartholomé, E. and Belward, A.S. (2005) GLC2000: A New Approach to Global Land Cover Mapping from Earth Observation Data. *International Journal of Remote Sensing*, **26**, 1959-1977. https://doi.org/10.1080/01431160412331291297

[17] Bontemps, S., Defourny, P., Van Bogaert, E., Arino, O., Kalogirou, V. and Perez, J.R. (2009) GLOBCOVER 2009 Products Description and Report. Toulouse, France.

[18] Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A. and Huang, X. (2010) MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets. *Remote Sensing of Environment*, **114**, 168-182. https://doi.org/10.1016/j.rse.2009.08.016

[19] Tateishi, R., Uriyangqai, B., Al-Bilbisi, H., Ghar, M.A., Tsend-ayush, J., Kobayashi, T., Kasimu, A., Hoan, N.T., Shalaby, A., Alsaaideh, B., Enkhzaya, T., Gegentana and Sato, H.P. (2011) Production of Global Land Cover Data—GLCNMO. *International Journal of Digital Earth*, **4**, 24-49. https://doi.org/10.1080/17538941003777521

[20] Tateishi, R., Hoan, N.T., Kobayashi, T., Alsaaideh, B., Tana, G. and Phong, D.X. (2014) Production of Global Land Cover Data—GLCNMO 2008. *Journal of Geography and Geology*, **6**, 99-122. https://doi.org/10.5539/jgg.v6n3p99

[21] CCI-LC (2014) Land Cover CCI: Product User Guide Version 2. Louvain-la-Neuve, Belgium.

[22] Fritz, S., See, L., Mccallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Al-

brecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, Hansen, M., Gong, P., Abdel, A.S., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S.C., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D.R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., van der Velde, M., Dunwoody, A., Kraxner, F. and Obersteiner, M. (2015) Mapping Global Cropland and Field Size. *Global Change Biology*, **21**, 1980-1992. https://doi.org/10.1111/gcb.12838

[23] Fritz, S., Mccallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F. and Obersteiner, M. (2009) Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sensing*, **1**, 345-354. https://doi.org/10.3390/rs1030345

[24] See, L., Schepaschenko, D., Lesiv, M., McCallum, I., Fritz, S., Comber, A., Perger, C., Schill, C., Zhao, Y., Maus, V., Siraj, M.A., Albrecht, F., Cipriani, A., Vakolyuk, M., Garcia, A., Rabia, A.H., Singha, K., Marcarini, A.A., Kattenborn, T., Hazarika, R., Schepaschenko, M., van der Velde, M., Kraxner, F. and Obersteiner, M. (2015) Building a Hybrid Land Cover Map with Crowdsourcing and Geographically Weighted Regression. *ISPRS Journal of Photogrammetry and Remote Sensing*, **103**, 48-56. https://doi.org/10.1016/j.isprsjprs.2014.06.016

[25] Jung, M., Henkel, K., Herold, M. and Churkina, G. (2006) Exploiting Synergies of Global Land Cover Products for Carbon Cycle Modeling. *Remote Sensing of Environment*, **101**, 534-553. https://doi.org/10.1016/j.rse.2006.01.020

[26] Tchuenté, A.T.K., Roujean, J.L. and de Jong, S.M. (2011) Comparison and Relative Quality Assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP Land Cover Data Sets at the African Continental Scale. *International Journal of Applied Earth Observation and Geoinformation*, **13**, 207-219. https://doi.org/10.1016/j.jag.2010.11.005

[27] Zhang, N. and Tateishi, R. (2013) Integrated Use of Existing Global Land Cover Datasets for Producing a New Global Land Cover Dataset with a Higher Accuracy: A Case Study in Eurasia. *Advances in Remote Sensing*, **2**, 365-372. https://doi.org/10.4236/ars.2013.24039

[28] Di Gregorio, A. (2016) Land Cover Classification System: Classification Concepts. Software Version 3.

[29] Herold, M., Woodcock, C.E., Di Gregorio, A., Mayaux, P., Belward, A.S., Latham, J. and Schmullius, C.C. (2006) A Joint Initiative for Harmonization and Validation of Land Cover Datasets. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1719-1727. https://doi.org/10.1109/TGRS.2006.871219

[30] Pflugmacher, D., Krankina, O.N., Cohen, W.B., Friedl, M.A., Sulla-Menashe, D., Kennedy, R.E., Nelson, P., Loboda, T.V., Kuemmerle, T., Dyukarev, E., Elsakov, V. and Kharuk, V.I. (2011) Comparison and Assessment of Coarse Resolution Land Cover Maps for Northern Eurasia. *Remote Sensing of Environment*, **115**, 3539-3553. https://doi.org/10.1016/j.rse.2011.08.016

[31] Hansen, M.C. and Reed, B. (2000) A Comparison of the IGBP DISCover and University of Maryland 1 km Global Land Cover Products. *International Journal of Remote Sensing*, **21**, 1365-1373. https://doi.org/10.1080/014311600210218

[32] Herold, M., Mayaux, P., Woodcoc, K.C.E., Baccini, A. and Schmullius, C. (2008) Some Challenges in Global Land Cover Mapping: An Assessment of Agreement and Accuracy in Existing 1 km Datasets. *Remote Sensing of Environment*, **112**, 2538-2556. https://doi.org/10.1016/j.rse.2007.11.013

[33] See, L.M. and Fritz, S. (2006) A Method to Compare and Improve Land Cover Da-

tasets: Application to the GLC-2000 and MODIS Land Cover Products. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1740-1746. https://doi.org/10.1109/TGRS.2006.874750

[34] Fritz, S. and See, L. (2005) Comparison of Land Cover Maps Using Fuzzy Agreement. *International Journal of Geographical Information Science*, **19**, 787-807. https://doi.org/10.1080/13658810500072020

[35] Vancutsem, C., Marinho, E., Kayitakire, F., See, L. and Fritz, S. (2013) Harmonizing and Combining Existing Land Cover/Land Use Datasets for Cropland Area Monitoring at the African Continental Scale. *Remote Sensing*, **5**, 19-41. https://doi.org/10.3390/rs5010019

[36] Chen, J., Ban, Y. and Li, S. (2014) China: Open Access to Earth Land-Cover Map. *Nature*, **514**, 434. https://doi.org/10.1038/514434c

[37] Waldner, F., Fritz, S., Di Gregorio, A., Plotnikov, D., Bartalev, S., Kussul, N., Gong, P., Thenkabail, P., Hazeu, G., Klein, I., Löw, F., Miettinen, J., Dadhwal, V.K., Lamarche, C., Bontemps, S. and Defourny, P. (2016) A Unified Cropland Layer at 250 m for Global Agriculture Monitoring. *Data*, **1**, 3. https://doi.org/10.3390/data1010003

[38] As-syakur, A.R. (2011) Perubahan penggunaan lahan di provinsi Bali. [Land Use Changes in Bali Province, Indonesia.] *Ecotrophic*, **6**, 1-7.

[39] Suputra, D.P.A., Ambarwati, I.G.A. and Tenaya, I.M.N. (2012) Faktor-faktor yang mempengaruhi alih fungsi lahan studi kasus di Subak Daksina , Desa Tibubeneng, Kecamatan Kuta Utara, Kabupaten Badung. [Factors That Affect Land Use Changes. A Study Case in Subak Daksina, Tibubeneng Vilage, Kuta Utara Subdistrict, Badung District, Bali Province.] *E-Journal Agribisnis dan Agrowisata*, **1**, 61-68.