

Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets

Steffen Fritz,¹ Liangzhi You,² Andriy Bun,¹ Linda See,¹ Ian McCallum,¹ Christian Schill,³ Christoph Perger,⁴ Junguo Liu,⁵ Matt Hansen,⁶ and Michael Obersteiner¹

Received 19 November 2010; revised 5 January 2011; accepted 12 January 2011; published 23 February 2011.

[1] This paper presents a methodology for the creation of a cropland map for Africa through the combination of five existing land cover products: GLC-2000, MODIS Land Cover, GlobCover, MODIS Crop Likelihood and AfriCover. A synergy map is created in which the products are ranked by experts, which reflects the likelihood or probability that a given pixel is cropland. The cropland map is then calibrated with national and sub-national crop statistics using a novel approach. Preliminary validation of the map was undertaken and the results are presented. The resulting cropland map has an accuracy of 83%, which is higher than the accuracy of any of the individual maps. The cropland map is freely available at agriculture.geo-wiki.org. **Citation:** Fritz, S., L. You, A. Bun, L. See, I. McCallum, C. Schill, C. Perger, J. Liu, M. Hansen, and M. Obersteiner (2011), Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets, *Geophys. Res. Lett.*, 38, L04404, doi:10.1029/2010GL046213.

1. Introduction

[2] Global cropland cover provides vital baseline information of agricultural production in many spatially explicit models and in applications such as the United Nations' Millennium Ecosystem Assessment (2005), and the Global Environmental Outlook (GEO, 2007). Furthermore, land cover and in particular cropland influences climate - it has been shown that cropland management itself can have an influence on the global climate [Lobell *et al.*, 2006]. Changes in land cover may also lead to biophysical effects on the climate through changes in evapotranspiration, albedo and surface roughness [Brovkin *et al.*, 2006].

[3] Accurate spatial information of cropland is particularly important for crop monitoring of food security and early warning, and satellite derived land cover datasets have been widely used for this purpose. A major challenge in addressing future global food security is to determine the spatial distribution as well as the identification of areas where additional food or non-food crops can be grown. Such information is also essential for global economic land-use models which

evaluate the impact of a certain policy (e.g., impact of a strict EU biofuel target on deforestation, food security etc.) see Havlik *et al.* [2010]. It has been shown both globally [Fritz and See, 2008], and in particular for Africa [Fritz *et al.*, 2010], that large discrepancies between current continental and global land cover maps exist both in terms of overall area and spatial distribution. Ramankutty *et al.* [2008] estimated that the global cropland area lies between 1.22 and 1.71 billion hectares which translates into a 40% difference between the upper and lower estimates. Different medium to coarse resolution global land cover datasets exist that identify cropland, i.e. GLC-2000, MODIS Land Cover, GlobCover, MODIS Crop Likelihood, but each was derived using a different classification algorithm and furthermore it does not have particularly high accuracy for cropland estimation [Friedl *et al.*, 2002; Bicheron *et al.*, 2008].

[4] Recent approaches to the development of cropland maps have focused on using national [Pittman *et al.*, 2010] and sub-national statistics [Ramankutty *et al.*, 2008], but there have been few attempts at integrating the current products, especially for agriculture. An exception is the work by Jung *et al.* [2006], who used a fuzzy agreement scoring approach to explore the synergies between global land cover products for carbon cycle modeling. This paper extends the concept of the synergy map [Jung *et al.*, 2006], by including expert judgment in ranking the land cover products as well as the inclusion of regional and national products derived from higher resolution satellite data. A novel approach to calibration based on national and sub-national information is also introduced and the validation of the resulting cropland map is presented.

2. Data and Methods

2.1. Data Sources

[5] Five land cover products are used in this paper: GLC-2000, MODIS Land Cover, GlobCover, MODIS Crop Likelihood and AfriCover. Table 1 provides the details of these datasets. Although there are other cropland products available [e.g., Ramankutty *et al.*, 2008; Ramankutty and Foley, 1998], they are not included in this analysis because they are either rather old or use the products listed above in their development and are at a coarser resolution. Instead we chose to use the original land cover products in this analysis.

[6] Cropland data at the country level are reported by the Food and Agriculture Organization of the United Nations (FAO). Sub-national level data have been assembled from various national statistics publications and agricultural census surveys by IFPRI (International Food Policy Research Institute). IFPRI's sub-national crop database covers all major crops in the world and includes mostly the second-level sub-

¹IIASA, Laxenburg, Austria.

²International Food Policy Research Institute, Washington, DC, USA.

³Felis, University of Freiburg, Freiburg, Germany.

⁴Fachbereich Geoinformatik, Fachhochschule Wiener Neustadt, Wiener Neustadt, Austria.

⁵School of Nature Conservation, Beijing Forestry University, Beijing, China.

⁶Geographic Information Science Center of Excellence, South Dakota State University, Brookings, South Dakota, USA.

Table 1. Comparison of Land Cover Products

Characteristics	GLC-2000	MODIS Land Cover	GlobCover	MODIS Crop Likelihood	AfriCover
Reference year	2000	2005	2005	2000–2008	1982–2000
Producer	JRC	University of Boston	ESA/JRC	University of Maryland	FAO
Resolution	1 km	500 m	300 m	250 m and 1 km	30 meter
Reference	<i>Bartholomé and Belward</i> [2005]	<i>Friedl et al.</i> [2002]	<i>Bicheron et al.</i> [2008]	<i>Pittman et al.</i> [2010]	(http://www.fao.org/waicent/faoinfo/sustdev/Eldirect/EIre0053.htm)

national units (e.g. a district in Kenya) but also some third-level administrative units if available [You *et al.*, 2009]. In some locations, favorable biophysical conditions allow multiple cropping for certain crops. Cropping intensity is defined as the ratio between harvested crop areas (where double or triple cropping areas are counted twice or three times respectively) and the physical areas where the crops are planted. The cropping intensity data by country is compiled from various sources, including national statistics, agricultural census, household surveys, and expert opinions [You *et al.*, 2009]. We calculated physical areas by dividing the harvested area by the corresponding cropping intensity. We then sum all these physical areas by crops and get the total cropland for a certain administrative unit.

[7] The final cropland by sub-national units serves as a lower bound for the validation of cropland extent from satellite data. There are two reasons for adopting this approach. Firstly, the crop list may miss minor crops produced in a particular region, and secondly, that account has not been taken of mixed cropping (i.e. planting two or more crops at the same plot within a year) due to data availability. Pasture and areas of low intensity cropping are not included in the cropland extent map.

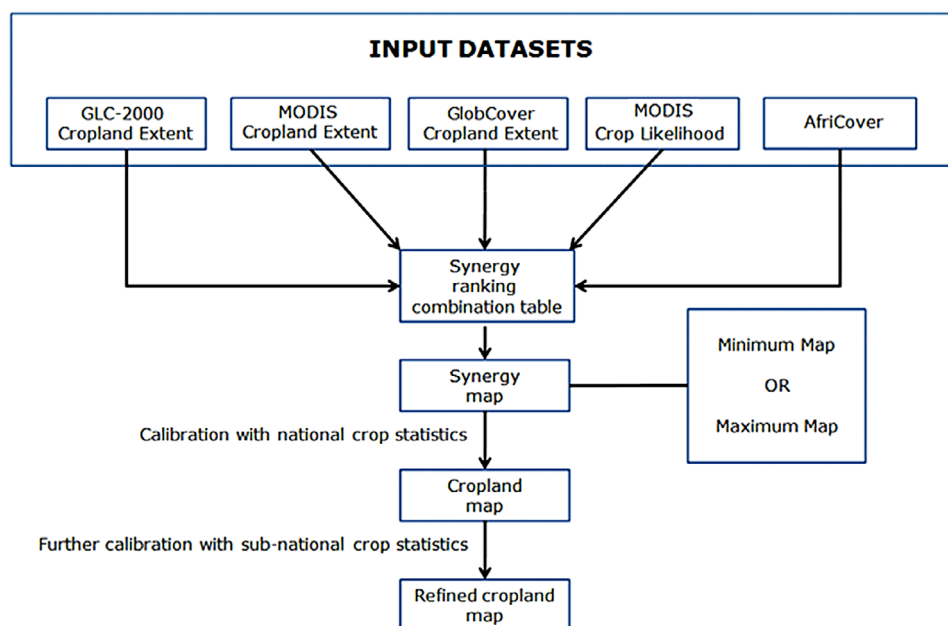
2.2. Methodology

[8] The methodology consists of a series of steps, which are shown in Figure 1 and are described in each of the sections that follow. Before starting, all the land cover products were standardized to a resolution of 1 km.

2.2.1. Creation of a Synergy Map

[9] The basic idea behind the synergy approach is to give each pixel a score based upon the agreement between the different land cover products used in the analysis [Jung *et al.*, 2006]. The synergy approach is extended in this paper in two ways. The first modification allows the producer of the final map to rank the land cover products before they are combined (either based on expert judgement, timeliness or accuracy assessment) rather than giving all products the same weight. The more agreement between the maps, the higher the likelihood or probability that cropland exists at that pixel. Figure 2a illustrates how this ranking works using three different land cover products. Land cover product #1 is given the highest ranking by experts followed by land cover products #2 and #3. The black pixels are areas of cropland in each land cover product. A pixel is assigned a value of 1 if all three maps agree that cropland exists. A value of 2 is assigned to areas of cropland where the land cover products #1 and #2 agree but land cover product #3 does not. With three land cover products, there are 7 possible combinations as shown in Table 2, which provides a simple illustration. However, this methodology can be applied to any number of land cover products where five maps have been used in the development of the final cropland map outlined in this paper.

[10] The second modification to the synergy methodology of Jung *et al.* [2006] is referred to here as priority ranking, which allows for the incorporation of regional products at places where confidence is the highest. Thus maps with very high confidence, i.e. those based on high resolution data in

**Figure 1.** Overview of the methodology for the creation of a cropland map.

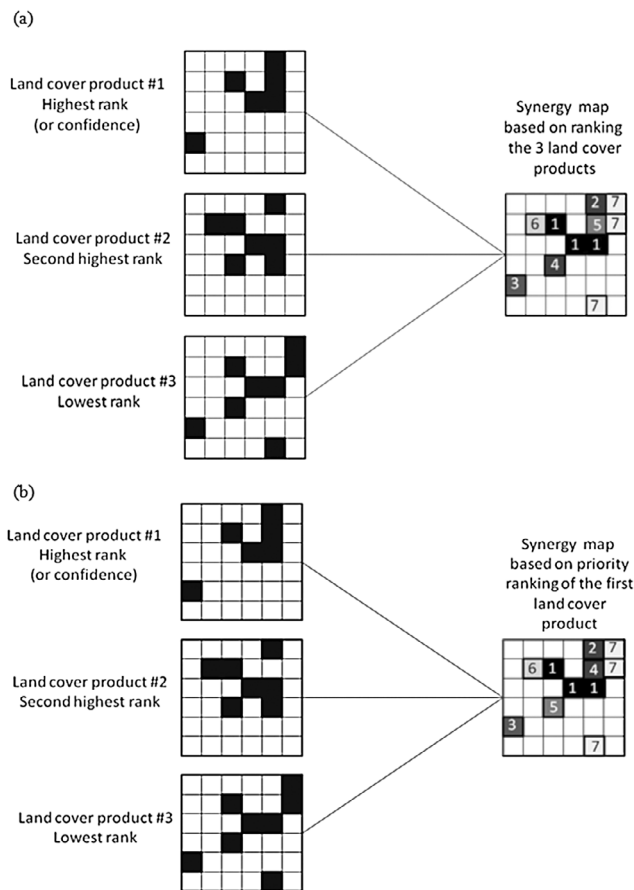


Figure 2. Creating a synergy map based on (a) ranking the land cover products and (b) giving priority to the top rank which means that the layer is always used first even if there is no agreement with another map at a specific pixel. The black pixels are areas of cropland in each product.

combination with information from the ground such as a TM-based visual interpretation, are always ranked higher than the other land cover products, even if they have no or little agreement with the priority ranked product. This concept is illustrated in Figure 2b where the effect of the priority ranking can be seen. There are two pixels, labeled with values of 4 and 5, where 4 in this situation means that land cover products #2 and #3 agree on cropland but land cover product #1 does not while a value of 5 indicates that only land cover product #1 indicates cropland at that pixel. This is also illustrated in Table 2 in the right hand column. The order of these rankings is switched from Figures 2a to 2b because land cover product #1 has a higher priority in the ranking and therefore this pixel gets a value of 4 instead of 5 when applying the priority ranking method. This algorithm would most likely be used if there is a regional product available that has been derived from higher resolution images together with ground data so confidence in this product would be higher than the rest.

[11] For the cropland map created for Africa, five different land cover maps were used in order of rank or confidence: AfriCover, MODIS Crop Likelihood, MODIS Land Cover, GlobCover and the GLC-2000 land cover products. When five different products are combined, there are 30 different possible combinations or rankings from 1 to 30, see Table 3.

AfriCover is given priority ranking over the other land cover products. The remaining land cover maps were then ranked based on how accurate they are perceived to be in cropland estimation. In this case we assumed that the more recent product was indeed perceived as more accurate (1st most recent MODIS Crop Likelihood, 2nd most recent MODIS Land Cover, 3rd most recent GlobCover and least recent GLC-2000). A synergy map was then created after implementing the ranking.

2.2.2. Creating a Cropland Map Through Calibration With National Crop Statistics

[12] To create a map of cropland, the synergy map must be calibrated using crop statistics. Where only national crop statistics are available, the process is as follows for one country at a time:

[13] 1. Cells with the highest ranking are selected and the sum of the cropland area is calculated taking the minimum value of cropland as specified in the legend definitions of each land cover product. Where the synergy map shows agreement of 2 or more land cover products, the average cropland area would be used. For example, in Africa the GLC-2000 uses 50% for the minimum cropland while MODIS uses 60%. If both maps agreed that a given pixel contained cropland, a minimum value of 55% would be used.

[14] 2. The cells with the next highest ranking are added to the previously selected cells and the total area of the selection is calculated.

[15] 3. Step 2 is repeated until a total area of the selected cells (summed up on a country level) match the national statistics as closely as possible.

[16] This process creates a cropland map based on the minimum value of cropland specified in the definitions of cropland in different land cover products. It is also possible to create these maps using the maximum definition and the average. An illustration of the algorithm is provided in Figure 3 for a simple example. The synergy map pictured has pixels with ranks between 1 to 4. The total areas of cropland belonging to these classes are 59, 75, 36 and 77, respectively. The national statistics for that country indicate an area of 180 so the classes are summed until they approach this value, which would be classes 1 to 3, giving a total value of 170. Hence the new statistics on cropland for that country would be 170, and the resulting map is a combination of the three highest ranked areas, or areas with the highest likelihood of containing cropland.

2.2.3. Calibrating the Synergy Map With Sub-National Crop Statistics

[17] The detailed sub-national crop database compiled by IFPRI [You *et al.*, 2009] contains statistics for nearly all crops, but in some sub-national units there may only be a few or no statistics available. However, it is still possible to make use

Table 2. Synergy Map Ranking Combinations When Combining Three Land Cover Products

Rank Value	Meaning for Ranking	Meaning for Priority Ranking
1	All land cover products agree	All land cover products agree
2	LC #1 and LC #2 agree	LC #1 and LC #2 agree
3	LC #1 and LC #3 agree	LC #1 and LC #3 agree
4	LC #2 and LC #3 agree	Only LC #1 indicates cropland
5	Only LC #1 indicates cropland	LC #2 and LC #3 agree
6	Only LC #2 indicates cropland	Only LC #2 indicates cropland
7	Only LC #3 indicates cropland	Only LC #3 indicates cropland

Table 3. Synergy Map Ranking Combinations When Combining Five Land Cover Products^a

Class	GLC-2000 Pixel	MODIS Land Cover Pixel	GlobCover Pixel	MODIS Crop Likelihood Pixel	AfriCover Pixel
1	1	1	1	1	1
2	0	1	1	1	1
3	1	1	0	1	1
4	1	0	1	1	1
5	1	1	1	0	1
6	0	1	0	1	1
7	0	0	1	1	1
8	0	1	1	0	1
9	1	1	0	0	1
10	1	0	1	0	1
11	0	0	0	1	1
12	0	0	1	0	1
13	0	0	0	0	1
14	1	1	1	1	0
15	0	1	1	1	0
16	1	1	0	1	0
17	1	0	1	1	0
18	1	1	1	0	0
19	1	1	0	1	0
20	1	0	1	1	0
21	0	1	0	1	0
22	0	0	1	1	0
23	1	0	0	1	0
24	0	1	1	0	0
25	1	1	0	0	0
26	1	0	1	0	0
27	0	0	0	1	0
28	0	1	0	0	0
29	0	0	1	0	0
30	1	0	0	0	0

^aWhere 1 indicates the presence of cropland and 0 indicates no cropland present.

of these data by applying some simple modifications to the cropland map created from the national calibration. The process is illustrated in Figure S1 of the auxiliary material for a country with 3 sub-national units represented by dark grey, light grey and white.¹ Numbers in each pixel show the area of the cropland. Sub-national cropland areas are calculated from the nationally calibrated map and compared to the sub-national statistics using these simple rules:

[18] 1. When no data are available at sub-national level, the cropland areas are taken from the national level map, which is the white zone in Figure S1.

[19] 2. When the sub-national level statistical data results in higher cropland areas than the one derived from the national analysis (pixel areas summed up on a sub-national level from previous analysis – see section 2.2.2), then the sub-national cropland values are used for this zone, shown as light grey in Figure S1.

[20] 3. When the national level analysis yields a higher area of cropland than the sub-national level, the algorithm outlined in section 2.2.2 is rerun for such zones. However, the sub-national statistics are used in place of the pixels summed from the nationally calibrated map, with the requirement that the national crop totals must be preserved.

2.2.4. Validation of the Cropland Map

[21] The cropland map was validated against a sample of 2553 points systematically distributed over Africa at each lat/long intersection. These vector points were placed on top of Google Earth as KML files within agriculture.geo-wiki.org and the validation was done as the percentage of crop-

land visible within a 1 km pixel. The outline of the pixel appears directly on top of Google Earth when the validation tool embedded into agriculture.geo-wiki.org is used. Using Google Earth as a tool for validation of land cover has been undertaken for the validation of other land cover products [Pekkarinen *et al.*, 2009; Bicheron *et al.*, 2008] and is particularly appealing in remote areas of Africa where no other datasets for validation are currently available. The majority of the very high resolution images were from the year 2003 to 2007. Even though not ideal we still consider this validation dataset sufficient to validate our cropland map. In the future Google Earth/Google Earth Engine will make an increasing amount of very high resolution time series available which will allow to improve the validation datasets. The validation was undertaken by experts trained in the recognition of cropland areas. Three different confidence levels were defined: 1 = very confident (corresponding to very high resolution Google Earth images (60 cm to 1 m); 2 = medium confidence (high resolution – 2 to 10 m) or Landsat in non-complex landscapes; and 3 = least confident (30 m Landsat data). The experts were a group of young scientists who have been specifically trained in the validation of land cover during the summer of 2010. Validation was undertaken at each latitude/longitude intersection across Africa, and distributed between the scientists. In some cases, multiple validations were undertaken at the same point.

3. Results

[22] The validated cropland map is provided in Figure S2. A confusion matrix was produced differentiating between cropland and non-cropland. The overall accuracy for all 2553 pixels was 82.7%, the omission errors for cropland

¹Auxiliary materials are available in the HTML. doi:10.1029/2010GL046213.

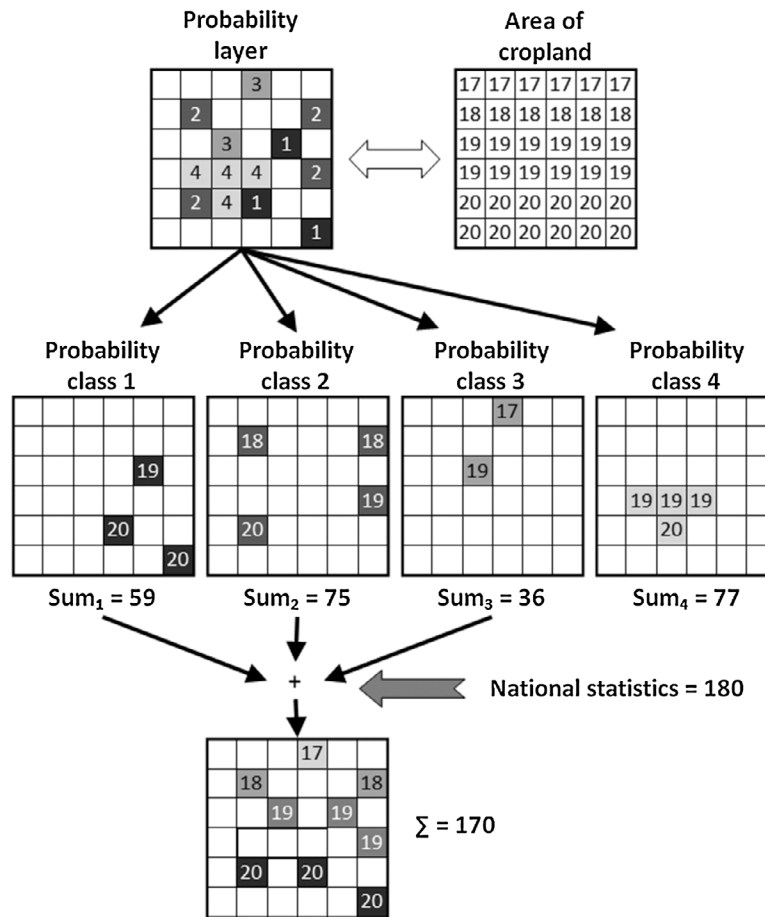


Figure 3. Creating a cropland map through calibration of the synergy map with national crop statistics.

and no cropland were 61.8% and 8.0% respectively, and the commission errors were 49.7% for cropland and 12.4% for no cropland. When confidence level 3 was excluded, the number of validated pixels reduced to 2125, but the results show a robust pattern (i.e. overall accuracy of 83.1%, omission errors of 62.5% (cropland) and 7.6% (no cropland) and commission errors of 49.8% (cropland) and 12.1% (no cropland). Finally, when just those pixels are included with very high confidence, the overall accuracy increases to 85.7% although the sample size decreases quite substantially to 1889 pixels.

[23] High accuracies are achieved in eastern Africa (probably due to the inclusion of AfriCover), whereas the accuracy is lower in Western Africa (Burkina Faso, Ivory Coast, and Ghana). A case study in South Africa has been chosen which shows the 4 different cropland input maps, the probability layer, and the resulting calibrated synergy map (Figure S3).

4. Summary and Discussion

[24] We synergistically combined existing cropland data sets to improve estimates of croplands in Sub-Saharan Africa, reducing local discrepancies with national and sub-national crop statistics. The resultant cropland map ranges from 83–86% accuracy. The methodology is flexible in that it can easily incorporate new products as they become available, in particular with respect to new Landsat based products that will be developed over Africa (e.g., AfriCover Senegal). An accuracy of 83% represents both an improvement over the

original input products and an independent dataset (the M3-Cropland layer produced by *Ramankutty et al.* [2008]). If the same error matrix is applied to the M3-Cropland layer the resulting accuracy is 69% (although the pixel size is larger and hence not directly comparable). However, the accuracy assessment also shows that there is still room for improvement. The method described here obviously depends highly upon the accuracy of the input layers as well as the reliability of the sub-national statistics. As more high quality data becomes available in the near future, this product is expected to improve further.

[25] We expect that the improved cropland extent map will also be incorporated into models in order to decrease the current uncertainty which is associated with cropland extent, and that this will lead to an overall improvement in predictions applied in the field of global vegetation modeling, global land use modeling, climate change modeling and earth systems modeling in general.

[26] **Acknowledgments.** We would like to thank Markus Tum, Luke Burns and Adriana Gomez for carrying out the validation. This research was supported by the European Community's Framework Programme (FP7) via EuroGEOSS (226487).

References

Bartholomé, E., and A. S. Belward (2005), GLC2000: A new approach to global land cover mapping from earth observation data, *Int. J. Remote Sens.*, 26, 1959–1977, doi:10.1080/01431160412331291297.

- Bicheron, P., et al. (2008), GLOBCOVER, Products, description, and validation report, Medias-France, Paris. (Available at http://ionial.esrin.esa.int/docs/GLOBCOVER_Products_Description_Validation_Report_I2.1.pdf)
- Brovkin, V., et al. (2006), Biogeophysical effects of historical land cover changes simulated by six Earth system models of intermediate complexity, *Clim. Dyn.*, 26, 587–600, doi:10.1007/s00382-005-0092-6.
- Friedl, M. A., et al. (2002), Global land cover mapping from MODIS: algorithms and early results, *Remote Sens. Environ.*, 83, 287–302, doi:10.1016/S0034-4257(02)00078-0.
- Fritz, S., and L. See (2008), Identifying and quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications, *Global Change Biol.*, 14, 1057–1075, doi:10.1111/j.1365-2486.2007.01519.x.
- Fritz, S., L. M. See, and F. Rembold (2010), Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa, *Int. J. Remote Sens.*, 25, 1527–1532.
- Havlik, P., et al. (2010), Global land-use implications of first and second generation biofuel targets, *Energy Policy*, doi:10.1016/j.enpol.2010.03.030, in press.
- Jung, M., K. Henkel, M. Herold, and G. Churkina (2006), Exploiting synergies of global land cover products for carbon cycle modeling, *Remote Sens. Environ.*, 101, 534–553, doi:10.1016/j.rse.2006.01.020.
- Lobell, D. B., G. Bala, and P. B. Duffy (2006), Biogeophysical impacts of cropland management changes on climate, *Geophys. Res. Lett.*, 33, L06708, doi:10.1029/2005GL025492.
- Pekkarinen, A., L. Reithmaier, and P. Strobl (2009), Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data, *ISPRS J. Photogramm. Remote Sens.*, 64, 171–183, doi:10.1016/j.isprsjprs.2008.09.004.
- Pittman, K., M. C. Hansen, I. Becker-Reshef, P. V. Potapov, and C. O. Justice (2010), Estimating global cropland extent with multi-year MODIS Data, *Remote Sens.*, 2, 1844–1863, doi:10.3390/rs2071844.
- Ramankutty, N., and J. A. Foley (1998), Characterizing patterns of global land use: An analysis of global croplands data, *Global Biogeochem. Cycles*, 12, 667–685, doi:10.1029/98GB02512.
- Ramankutty, N., A. T. Evan, C. Monfreda, and J. A. Foley (2008), Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000, *Global Biogeochem. Cycles*, 22, GB1003, doi:10.1029/2007GB002952.
- You, L., S. Wood, and U. Wood-Sichra (2009), Generating plausible crop distribution and performance maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach, *Agric. Syst.*, 99(2–3), 126–140, doi:10.1016/j.agry.2008.11.003.
- A. Bun, S. Fritz, I. McCallum, M. Obersteiner, and L. See, IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria. (fritz@iiasa.ac.at)
- M. Hansen, Geographic Information Science Center of Excellence, South Dakota State University, Brookings, SD 57007, USA.
- J. Liu, School of Nature Conservation, Beijing Forestry University, Beijing, 100083, China.
- C. Perger, Fachbereich Geoinformatik, Fachhochschule Wiener Neustadt, Wiener Neustadt A-2700, Austria.
- C. Schill, Felis, University of Freiburg, D-79106 Freiburg, Germany.
- L. You, International Food Policy Research Institute, 2033 K St., NW, Washington, DC 20006, USA.