

Global Change Biology (2015) 21, 1980–1992, doi: 10.1111/gcb.12838

Mapping global cropland and field size

STEFFEN FRITZ¹, LINDA SEE¹, IAN MCCALLUM¹, LIANGZHI YOU^{2,3}, ANDRIY BUN¹, ELENA MOLTCHANOVA^{1,4}, MARTINA DUERAUER¹, FRANSIZKA ALBRECHT^{5,6}, CHRISTIAN SCHILL⁷, CHRISTOPH PERGER¹, PETR HAVLIK¹, ALINE MOSNIER¹, PHILIP THORNTON⁸, ULRIKE WOOD-SICHTA³, MARIO HERRERO⁹, INBAL BECKER-RESHEF¹⁰, CHRIS JUSTICE¹⁰, MATTHEW HANSEN¹⁰, PENG GONG¹¹, SHETA ABDEL AZIZ¹², ANNA CIPRIANI^{13,14}, RENATO CUMANI¹⁵, GIULIANO CECCHI¹⁶, GIULIA CONCHEDDA¹⁵, STEFANUS FERREIRA¹⁷, ADRIANA GOMEZ¹⁸, MYRIAM HAFFANI¹⁹, FRANCOIS KAYITAKIRE²⁰, JAITEH MALANDING¹⁴, RICK MUELLER²¹, TERENCE NEWBY²², ANDRE NONGUIERMA²³, ADEAGA OLUSEGUN²⁴, SIMONE ORTNER²⁵, D. RAM RAJAK²⁶, JANSLE ROCHA²⁷, DMITRY SCHEPASCHENKO¹, MARIA SCHEPASCHENKO²⁸, ALEXEY TEREKHOV²⁹, ALEX TIANGWA³⁰, CHRISTELLE VANCUTSEM²⁰, ELODIE VINTROU³¹, WU WENBIN³², MARIJN VAN DER VELDE¹, ANTONIA DUNWOODY¹, FLORIAN KRAXNER¹ and MICHAEL OBERSTEINER¹

¹International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria, ²Key Laboratory of Agri-informatics, Ministry of Agriculture/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China, ³International Food Policy Research Institute (IFPRI), 2033 K Street, NW, Washington, DC 20006, USA, ⁴School of Mathematics and Statistics, University of Canterbury, Christchurch 8140, New Zealand, ⁵GeoVille GmbH, Innsbruck, Austria, ⁶Department of Geography and Regional Research, University of Vienna, Vienna, Austria, ⁷Albert-Ludwig University, Freiburg, Germany, ⁸CGIAR Research Programme on Climate Change, Agriculture and Food Security, International Livestock Research Institute, 00100 Nairobi, Kenya, ⁹Commonwealth Scientific and Industrial Research Organisation, 306 Carmody Road, St Lucia, 4068 QLD, Australia, ¹⁰Department of Geography, University of Maryland, 2181 LeFrak Hall, College Park, MD 20742, USA, ¹¹Ministry of Education Laboratory for Earth System Modeling, Center for Earth System Science, Tsinghua University, Beijing 100084, China, ¹²Faculty of Agriculture, Ain Shams University, Cairo, Egypt, ¹³Dipartimento di Scienze Chimiche e Geologiche, Università degli Studi di Modena e Reggio Emilia, Large S. Eufemia 19, 41121 Modena, Italy, ¹⁴Lamont Doherty Earth Observatory of Columbia University, Palisades, NY 10964, USA, ¹⁵Food and Agriculture Organization of the UN, Viale delle Terme di Caracalla, 00100 Rome, Italy, ¹⁶Food and Agriculture Organization of the UN, Sub-regional Office for Eastern Africa, CMC Road, P.O. Box 5536, Addis Ababa, Ethiopia, ¹⁷GeoTerraImage (Pty) Ltd., 477 Witherite Street, Pretoria 0184, South Africa, ¹⁸University of Natural Resources and Life Sciences (BOKU), Vienna, Austria, ¹⁹Centre National de la Cartographie et de la Télédétection, BP 200, Tunis, Tunisia, ²⁰Joint Research Centre of the European Commission (JRC), Via E. Fermi, 2749, Ispra, VA I-21027, USA, ²¹USDA/NASS, 3251 Old Lee Hwy, Rm 305, Fairfax, VA 22030, USA, ²²Agricultural Research Council, Private Bag X79, Pretoria 0001, South Africa, ²³UNECA, ECA - Menlik II Avenue, P.O. Box 3001, Addis Ababa, Ethiopia, ²⁴Department of Geography, University of Lagos, Lagos, Nigeria, ²⁵University of Applied Sciences, Wiener Neustadt, Austria, ²⁶Space Applications Centre (ISRO), Ahmedabad, Gujarat 380015, India, ²⁷Geoprocessing Lab, School of Agricultural Engineering, University of Campinas, Campinas, Brazil, ²⁸Russian Institute of Continuous Education in Forestry, Pushkino, Russia, ²⁹Institute of Space Research, 15 Shevchenko Street, Almaty 50010, Ukraine, ³⁰South Sudan Food Security Technical Secretariat, South Sudan National Bureau of Statistics, P.O. Box 137 May Street, Juba, South Sudan, ³¹CIRAD, Maison de la Télédétection, 500 Rue JF Breton, Montpellier 34090, France, ³²Institute of Agricultural Resources & Regional Planning, Chinese Academy of Agricultural Sciences, No. 12, Zhongguancun South Street, Haidian District, Beijing 100081, China

Abstract

A new 1 km global IIASA-IFPRI cropland percentage map for the baseline year 2005 has been developed which integrates a number of individual cropland maps at global to regional to national scales. The individual map products include existing global land cover maps such as GlobCover 2005 and MODIS v.5, regional maps such as AFRICOVER and national maps from mapping agencies and other organizations. The different products are ranked at the national level using crowdsourced data from Geo-Wiki to create a map that reflects the likelihood of cropland. Calibration with national and subnational crop statistics was then undertaken to distribute the cropland within each country and subnational unit. The new IIASA-IFPRI cropland product has been validated using very high-resolution satellite imagery via Geo-Wiki and has an overall accuracy of 82.4%. It has also been compared with the EarthStat cropland

Correspondence: Dr Steffen Fritz, tel. +43 2236 807 353, fax: +43 2236 807 599, e-mail: fritz@iiasa.ac.at and Dr Liangzhi You, e-mail: l.you@cgiar.org

product and shows a lower root mean square error on an independent data set collected from Geo-Wiki. The first ever global field size map was produced at the same resolution as the IIASA-IFPRI cropland map based on interpolation of field size data collected via a Geo-Wiki crowdsourcing campaign. A validation exercise of the global field size map revealed satisfactory agreement with control data, particularly given the relatively modest size of the field size data set used to create the map. Both are critical inputs to global agricultural monitoring in the frame of GEOGLAM and will serve the global land modelling and integrated assessment community, in particular for improving land use models that require baseline cropland information. These products are freely available for downloading from the <http://cropland.geo-wiki.org> website.

Keywords: agricultural intensity, cropland, data fusion, field size, land cover, synergy map

Received 29 July 2014; revised version received 30 November 2014 and accepted 8 December 2014

Introduction

The population of the Earth recently exceeded the 7 billion mark (Tollefson, 2011) and is projected to reach 9 billion by 2050 (Roberts, 2011). As wealth rises globally, there will be an increased shift towards a meat-based diet and more processed foods (Godfray *et al.*, 2010), which will put added pressure on the food supply system, particularly in terms of rising demand for grains as livestock feed (Tilman *et al.*, 2002). Feeding this increasing population and coping with increased food demand from shifting diets in an environmentally sustainable way poses a major global challenge. Although population growth is decelerating in many industrialized countries, Africa will continue to see the highest rates of fertility and population growth in the future (Tollefson, 2011). This region of the world is also one where food security continues to be a serious and widespread issue (FAO, 2011a) and is likely to increase under an uncertain, changing climate (Brown & Funk, 2008; Wheeler & von Braun, 2013).

A number of recommendations for feeding the world's growing population have been offered (Stuart, 2009; Godfray *et al.*, 2010; Foley *et al.*, 2011), which include the following: closing the yield gap (i.e. the difference between the potential yield and that achieved by farmers, Lobell *et al.*, 2009), which could increase production by as much as 58% (Foley *et al.*, 2011); reducing food wastage, which is estimated at around one-third of food produced for human consumption (FAO, 2011b); and making changes to diets, shifting production from feed to food. Achieving these goals both practically and sustainably requires providing access to inputs and technologies that will improve yields and conserve both water quality and quantity, and broader economic and policy changes (Bekunda *et al.*, 2010; Foley *et al.*, 2011). However, these goals entail the need to accurately quantify the amount of land currently under cultivation and to have a clear knowledge of the spatial distribution of this and other land cover types globally. Referred to as cropland, this is defined as arable land and permanent crops based on

the definition from FAO. FAO further defines arable land as land under temporary agricultural crops (multiple-cropped areas are counted only once), temporary meadows for mowing or pasture, land under market and kitchen gardens and land temporarily fallow (<5 years). The abandoned land resulting from shifting cultivation is not included in this category (see <http://faostat.fao.org/site/375/default.aspx>). Permanent crops include land that is cultivated with long-term crops which do not have to be replanted for several years (such as cocoa and coffee); land under trees and shrubs producing flowers, such as roses and jasmine; and nurseries (except those for forest trees, which should be classified under 'forest'). Permanent meadows and pastures are excluded from land under permanent crops (see <http://faostat.fao.org/site/375/default.aspx>).

The spatial extent of cropland is available from generic global land cover products such as the GLC-2000 (Fritz *et al.*, 2003), MODIS (Friedl *et al.*, 2010) or GlobCover (Bicheron *et al.*, 2008), but these products do not agree in terms of the overall amount of cropland or in terms of the spatial distribution (Fritz *et al.*, 2011). Specific cropland products are available that use national and subnational crop statistics in their development, for example the Agricultural Lands in the Year 2000 EarthStat cropland product (Ramankutty *et al.*, 2008) and a MODIS-derived cropland probability layer (Pittman *et al.*, 2010). However, the uncertainty in these products also remains high. In fact, Ramankutty *et al.* (2008) have attempted to quantify this uncertainty at the global scale, estimating that global cropland extent varies between 1.22 and 1.71 billion hectares, or by more than 40%.

One method of improving the spatial representation of cropland extent is to adopt a hybrid or data fusion approach whereby existing land cover products and cropland extent maps are integrated into a single cropland percentage layer. Modifying the synergy approach of Jung *et al.* (2006), Fritz *et al.* (2011) developed a hybrid percentage cropland map for Africa, which combined four existing land cover products (GLC-2000,

MODIS v.5, GlobCover 2005 and AFRICOVER) and the MODIS cropland likelihood layer (Pittman *et al.*, 2010). Validation of the map revealed an improved accuracy over individual products in characterizing cropland extent. To further develop of this product, a number of additional national and regional cropland extent maps were obtained through a data sharing exercise initiated via a cropland workshop (See *et al.*, 2012). The methodology developed in Fritz *et al.* (2011) has been further enhanced and applied globally to produce the 1 km global hybrid percentage cropland product (hereafter referred to as the IIASA-IFPRI (International Institute for Applied Systems Analysis-International Food Policy Research Institute) cropland product. This study will outline the enhancements to the methodology employed here, in particular the use of crowdsourced data collected via Geo-Wiki (Fritz *et al.*, 2009, 2012), to determine the ranking of individual products based on how well they match the crowdsourced data on a national level when integrated. The IIASA-IFPRI product produced for the baseline year 2005 is compared with the EarthStat cropland product of Ramankutty *et al.* (2008), which was produced for the year 2000, and shows improvements in terms of overall accuracy but also in terms of capturing the spatial distribution of cropland.

The study also presents the first global field size map at a 1 km² resolution. Information about field size can be used as a proxy for human development as research has shown that there is a positive relationship between field size and farm size (Levin, 2006) and farm size and income (Berry, 1972). This suggests that change in field size might be used as one indicator for human development and can be used to complement household surveys, which are resource intensive and prone to bias. Field size is also useful for agricultural monitoring, in particular to determine which sensor resolution is required to monitor agriculture in different areas. For example, in areas with large field sizes, MODIS resolution imagery can be used to monitor agriculture while small and very small field sizes would require Landsat and very high-resolution imagery. Moreover, it can be used as a proxy for agricultural or mechanization (Kuemmerle *et al.*, 2013) as well as a proxy for species richness in cropland (Geiger *et al.*, 2010). Various approaches to field size mapping have been utilized in the past, for example interpolation of LUCAS (Land Use/Cover Area Frame Statistical Survey, <http://www.lucas-europa.info>) samples to create a European field size map for 2009 (Kuemmerle *et al.*, 2013) and automated field extraction from Landsat imagery for the United States (Yan & Roy, 2014). However, the field size map presented here is the first global-wide product.

Materials and methods

Data

Cropland input data sets. The IIASA-IFPRI cropland map integrates numerous data sets, many of which were contributed as part of an ongoing data sharing exercise in which experts in the fields of remote sensing, land cover and agriculture were brought together at a cropland workshop (See *et al.*, 2012). These individual products are outlined in Table 1.

The national, regional and global products are of varying resolutions from 30 m to 1 km. The data sets at a resolution of <1 km were aggregated to a common geographical grid to produce percentage cropland by pixel at a 1 km resolution. For the GLC-2000, which was already at a 1 km resolution and indicates only the presence/absence of cropland, the legend definition was used to assign a cropland percentage. For

Table 1 Cropland products used in the IIASA-IFPRI cropland map

| Scale | Products | Source |
|----------|--|---|
| Global | Cropland masks from: <ul style="list-style-type: none"> • GLC-2000 • MODIS (for 2005) • GlobCover 2005 • GEOCOVER (only for parts of South America, the middle East, parts of Ethiopia) • Cropland probability layer | Fritz <i>et al.</i> (2003), Friedl <i>et al.</i> (2010), Bicheron <i>et al.</i> (2008), Contributed at a workshop (See <i>et al.</i> , 2012) and Pittman <i>et al.</i> (2010) |
| Regional | <ul style="list-style-type: none"> • Cropland mask from CORINE land cover • Cropland mask for Africa • AFRICOVER (Burundi, DRC, Eritrea, Kenya, Rwanda, Senegal, Somalia, Tanzania, Uganda) | Vancutsem <i>et al.</i> (2012); Contributed at a workshop (See <i>et al.</i> , 2012) but freely available from FAO |
| National | <ul style="list-style-type: none"> • Australia • Brazil (summer crops) • Burkina Faso • China • Egypt • Gambia • India (northern part) • Lesotho • Mali • South Africa • Sudan/South Sudan • Swaziland • Tunisia • USA | Contributed at a workshop (See <i>et al.</i> , 2012). Maps for Australia and USA are also freely available. |

example, in the GLC-2000, cropland is defined as 60–100% in the class ‘cultivated and managed’. Therefore, the minimum, average and maximum values were assigned to each pixel, which resulted in a minimum, average and maximum map for these products.

Cropland statistics for map calibration. The IIASA-IFPRI cropland product has been calibrated using national statistics provided by FAO (2005) and subnational cropland data collected by You *et al.* (2014). Cropland statistics for 4 years (2003–2006) were averaged to provide a more stable estimate for the year 2005. In areas where double or triple cropping is practiced, a crop-specific cropping intensity factor, that is the ratio between harvested crop areas and the physical areas where the crops are planted, was first applied to adjust the physical area. These intensity factors were collected from local experts (e.g. agronomists) and agricultural census or farming system literature. We did not explicitly account for fallow land or failed harvests. Although we recognize that fallow or crop losses are important aspects of the cropping system, such data are not available, particularly at a global scale. These areas were then summed to produce the total cropland for each administrative unit.

Although there is recognition that agricultural statistics in some countries are poor and that there has been in decline in quality over time (World Bank *et al.*, 2011; FAO *et al.*, 2012; Carfagna *et al.*, 2013), FAOSTAT contains the most comprehensive cropland database available and is used by many researchers and international institutes for global and regional studies. To ensure comparability across countries/regions and over time, we have used FAO country data to calibrate the hybrid product.

Training data collected via Geo-Wiki. A data set suitable for training purposes was collected via the Geo-Wiki application

(Fritz *et al.*, 2009, 2012) through a crowdsourcing competition in which land cover information was gathered for the validation of a map of land availability for biofuels (Perger *et al.*, 2012; Fritz *et al.*, 2013). Geo-Wiki uses Google Earth imagery, which varies in resolution from 15 m to 10 cm. In this competition, around 53 000 sample pixels with 36 000 unique locations at a 1 km resolution were collected globally, where users were asked to choose the dominant land cover type from among ten simple land cover classes including one class for cropland and one for a mosaic of cropland and natural vegetation (see Table S1). From this total, 13 963 unique locations featured cropland and were used to rank the input data sets outlined previously. This ranking was then input into the creation of a layer that indicates where cropland is most likely to occur. Users who took part in the competition were also asked to indicate the field size for those pixels that were identified as having cropland or a cropland/natural vegetation mosaic land cover type. Field size was specified as very small, small, medium and large using high-resolution imagery, where examples were provided to guide the user. This data set has been used to create the global map of field size. Figure 1 shows the location of the training data used in the calibration of the global cropland map and in the development of the field size map. Appendix S1 shows how field size varies by region.

Validation data collected via Geo-Wiki. Independent data were required for the validation of the IIASA-IFPRI cropland product. A validation data set was developed using two crowdsourcing campaigns, where the sample locations coincided with those used by Tsinghua University in developing their 30 m global land cover product (Zhao *et al.*, 2014). We first stratified this sample for cropland by eliminating locations where cropland is not found, choosing only those places where the average temperature is >3 °C, and the average annual precipitation is above 200 mm based on the

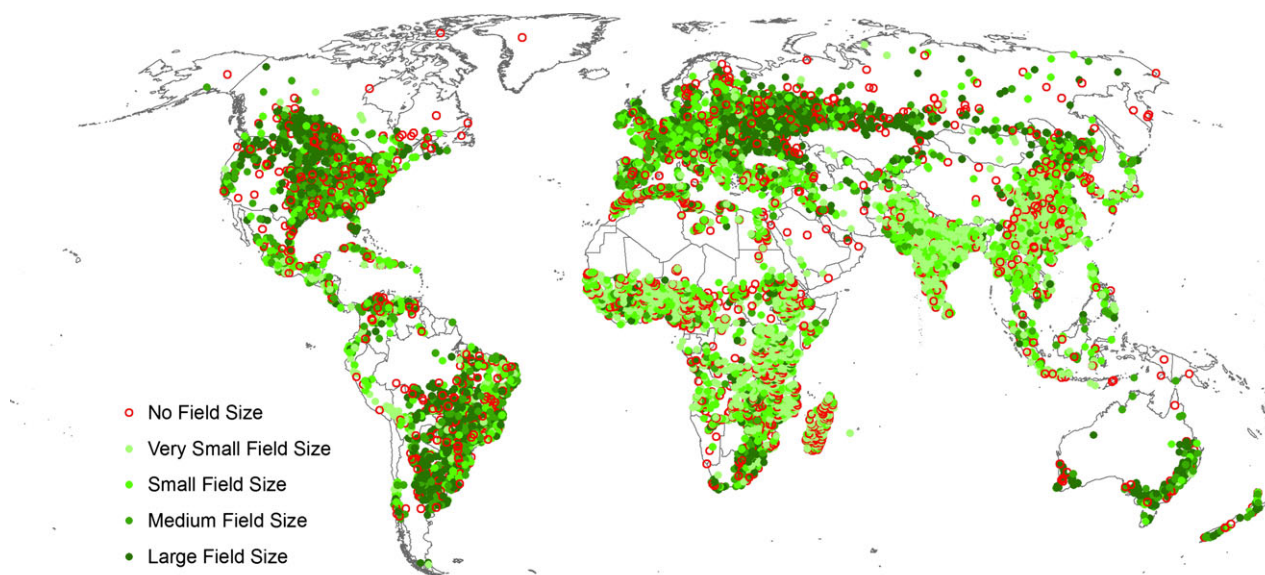


Fig. 1 The location of the data collected via Geo-Wiki that was used in the development of the IIASA-IFPRI cropland map and the field size map.

WorldClim database (Hijmans *et al.*, 2005). Using this stratification, we ensured sampling in places where cropland can potentially occur and did not sample in deserts or very cold areas. To account for irrigated areas, we added samples from areas in which cropland occurred based on either the IIASA-IFPRI map or the EarthStat cropland map where precipitation was below 200 mm. Two Geo-Wiki crowdsourcing campaigns were run independently to validate these locations, where users were asked to identify up to three different land cover types visible in a 1 km² pixel and the percentage of each type. These two independent crowdsourced data sets were then compared as part of a quality control exercise. Where agreement at the same location was >65% (see Appendix S2 for more details on how agreement was calculated), the sample was included in the final validation data set, which consisted of 10 626 cropland percentage pixels (cropland 3674 and noncropland 6952). Figure 2 shows the global distribution of the validation data while Figure S1 contains the full distribution.

We then aggregated the IIASA-IFPRI cropland product to a resolution of 10 km² to match that of the EarthStat cropland product (Ramankutty *et al.*, 2008), assigning the mean value of the 1 km² pixels to the resultant 10 km² pixel. We ensured an exact geographical match between the aggregated hybrid cropland product and the EarthStat cropland map. A cropland disagreement map was then produced by subtracting one map from the other. A total of 1033 10 km² cropland percentage pixels were then created using Geo-Wiki, based on a stratified sample in the disagreement areas identified above. Each pixel was classified by two experts from a total pool of ten experts. These experts are research staff at IIASA or reliable validators from the Geo-Wiki network with a range of backgrounds including remote sensing, geospatial sciences, forestry, geology and biology. Within each 10 km² pixel, each

expert identified cropland percentage within nine subpixels. The total cropland percentage for the pixel was the mean of all nine cells. In those cases where the absolute value of the difference between the mean value from expert 1 and the mean value from expert 2 differed by more than 50%, a third expert revisited those pixels. The third expert's average value, along with the closest value from the two original experts, was then used to form the new average pixel value. This occurred 56 times where difficulties in interpretation may have occurred due to coarser resolution imagery, the presence of tree crops rather than forest cover and abandoned lands or pasture which may have been confused with cropland.

Methodology

Development of the IIASA-IFPRI cropland map. The IIASA-IFPRI cropland map was created by combining individual land cover products. Using a convergence of evidence approach, the products were integrated to produce a synergy map. A synergy map is a single layer that characterizes the agreement between input layers and was originally proposed by Jung *et al.* (2006) in the context of different land cover data sets and carbon cycle modelling. Where all the input layers agree, for example on the presence of cropland, the agreement between products is high and the likelihood of cropland being present is likewise high. The synergy map therefore captures the agreement between the layers and provides a spatial representation of cropland likelihood. Product integration is not by equal weighting but rather by ranking, which is driven by comparison with crowdsourced data. Finally, the synergy map is transformed into a percentage cropland map by calibration with national and subnational statistics. These steps are described in more detail below.

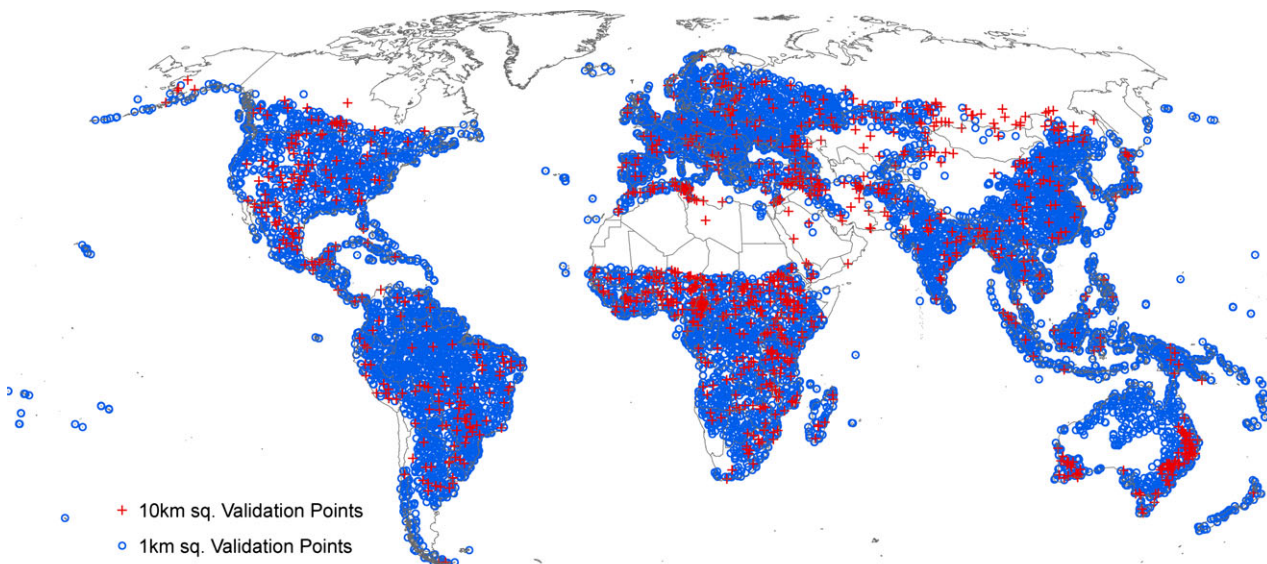


Fig. 2 The location of the validation data for assessing the performance of the IIASA-IFPRI cropland map.

Ranking of individual cropland products and creation of a synergy map. In the development of the African IIASA-IFPRI cropland product, the individual cropland products were ranked using expert knowledge on a global basis (Fritz *et al.*, 2011), so a single ranking matrix was applied to transform the individual cropland layers into a synergy map. Here, in contrast, the analysis operates at a national level, so a ranking table of products is created for each country using crowdsourced data from Geo-Wiki, which are expressed as a binary cropland/noncropland value per pixel. For each country, the percentage agreement between the crowdsourced data and each individual product available for that country was calculated. This agreement was used to rank the products to develop the synergy map. In this way, the synergy map was tailored to each country. The ranking procedure is explained in more detail in Appendix S3. At the same time that the synergy map was created, minimum, average and maximum cropland percentage maps were calculated based on the individual input maps. For example, suppose we have a country where only global maps were available, namely GLC-2000 (with a minimum cropland definition of 60%, an average of 80% and a maximum of 100%), GlobCover (minimum 70%, average of 85% and maximum of 100%) and MODIS (minimum of 80%, average of 90% and maximum of 100%) and all agreed that cropland is present in the synergy map, then the minimum, average and maximum map would be calculated as the average of the minimum, average and maximum cropland values of the three products, respectively, that is minimum of 70%, an average of 85% and a maximum of 100%. The average map is used in the calibration with FAO statistics, as described in the next step while the minimum map is used in the visual correction. Some experimentation was undertaken with the maximum map, but in the end, it was sufficient to calibrate the cropland layer using only the average and minimum maps.

Calibration of the global cropland extent map. Consistency with FAO statistics is a requirement as many models, including IIASA's GLOBIOM economic land use model (Havlík *et al.*, 2011; Havlik *et al.*, 2014), use data sets from various international bodies. This includes, for example, data on land resources from FAO such as the amount of area under forest, grassland, cropland, and UN projections which relate to land use change such as meat consumption and technological innovation, all of which are consistent with national statistics. Therefore, the synergy map and the average cropland percentage map were used together with cropland statistics from FAO and IFPRI for the year 2005 to produce the IIASA-IFPRI cropland map. Starting at the national level, the crop statistics were applied on a country-by-country basis to allocate the cropland areas starting with those areas that have the highest probability of cropland as indicated in the synergy map. The area was calculated based on the average cropland percentage map. This allocation process continues until the total area in the cropland map matches the national statistics as closely as

possible. A further calibration was then applied using sub-national statistics to further refine the distribution of cropland areas across each country. The full details can be found in Fritz *et al.* (2011).

Visual correction for mismatch with FAO statistics. Once the calibration was complete, a visual inspection was undertaken on a country-by-country basis. The cropland map was overlaid onto Google Earth imagery and for those countries where the FAO statistics were too low and cropland was not sufficiently captured (i.e. omission errors), the calibration procedure described above was repeated except that the minimum percentage cropland map was employed instead of the average. This served to allocate cropland to a larger area in a given country. This was generally the exception rather than the rule, but it helped to improve the accuracy of the final product. In a few cases, the minimum cropland map resulted in too much allocation of cropland areas and therefore an average value between the minimum and the average was used. The maximum cropland map always resulted in too little cropland extent and was therefore not used.

Development of a global map of field size. As mentioned previously, field size from 1 (very small) to 4 (large) was collected during a Geo-Wiki campaign using high-resolution satellite imagery. Information for approximately 53 000 samples of 1 km² pixels were gathered at around 36 000 unique locations. Of these unique locations, 13 963 samples contained cropland and therefore the size of the field. This sample data set of ordinal values was then interpolated using inverse weighted distance, treating the ordinal values as a continuum, to create a global map of field size at a 1 km² resolution, which was then clipped using the IIASA-IFPRI cropland map.

Validation of the global cropland product and comparison with the EarthStat cropland product. The IIASA-IFPRI percentage cropland product was first reclassified into a binary map of cropland/no cropland, where a cropland percentage greater than zero was assigned to the cropland category. The validation data set described previously was used. Overall accuracy and the allocation and quantity disagreement were calculated (Foody, 2010; Pontius & Millones, 2011).

A comparison with the Earthstat cropland map was then undertaken, which involved aggregating the IIASA-IFPRI product to a 10 km² resolution to match that of the EarthStat cropland product (Ramankutty *et al.*, 2008). The aggregation assigned the average value of the 1 km² pixels to the resultant 10 km² pixel and produced an exact geographical match between the aggregated IIASA-IFPRI cropland product and the EarthStat cropland product. The root mean square error (RMSE) between the data gathered via Geo-Wiki (expressed as percentage cropland) and the values in the IIASA-IFPRI cropland and the EarthStat cropland products were calculated.

Validation of the field size map. The field size map was validated using two different approaches. As the field size map was created through simple interpolation of data collected as part of a crowdsourcing campaign, the first validation approach was to examine the trustworthiness of the crowd. Five hundred sample locations were selected randomly from the original crowdsourced data set, and these were then independently validated by three experts using Geo-Wiki to produce a data set of field size controls. The experts were provided with the images randomly on Geo-Wiki and the process continued until all 500 samples had been validated at least twice by different experts. Images with a coarse resolution (i.e. Landsat) were then removed from the sample as well as those with no evidence of cropland. The Spearman's rank correlation coefficient was then calculated between the answers from the crowd and the experts.

The second validation approach involved sampling five hundred new locations randomly from the interpolated field size map, which were not at the original locations of the crowdsourced data used in the interpolation. The values obtained from the interpolated field size map were then compared with a second field size control data set created by the three experts where each sample was validated at least twice by different experts. Once again coarse resolution samples were removed as well as those with no evidence of cropland, and the Spearman's rank correlation coefficient was calculated.

Results

Figure 3 presents the IIASA-IFPRI cropland map while the global field size map is shown in Fig. 4. Both products are currently downloadable from the 'Cropland' branch of Geo-Wiki (<http://www.geo-wiki.org>).

Validation of the IIASA-IFPRI cropland product

The overall accuracy of the IIASA-IFPRI cropland map is 82.4%. The total disagreement (17.6%) can be further broken down into two components, that is the allocation disagreement (which reflects the presence of commission and omission errors in the same category or a spatial allocation error) of 9.4% and the quantity disagreement (which reflects a disagreement in the proportions of the categories between the map and the reference data) of 8.2%. Thus, the overall disagreement shows errors of both kinds split more or less evenly. The quantity disagreement may partly be the result of the calibration process as the allocation of cropland is constrained by the national statistics. The allocation disagreement, on the other hand, may partly be a result of combining individual maps that are actually all incorrect at the same location but which represent the best information available for that area. This error might be minimized in the future by obtaining more accurate national maps that will rank higher than global products when creating the synergy map.

Comparison of the aggregated IIASA-IFPRI products with the EarthStat cropland map

Figure 5 presents a disagreement map between the aggregated IIASA-IFPRI cropland product and the EarthStat cropland product, in particular those areas where cropland is present in one product but not the other and vice versa. The grey areas show regions where both products indicate cropland of some percentage between 1% and 100%, and thus there is

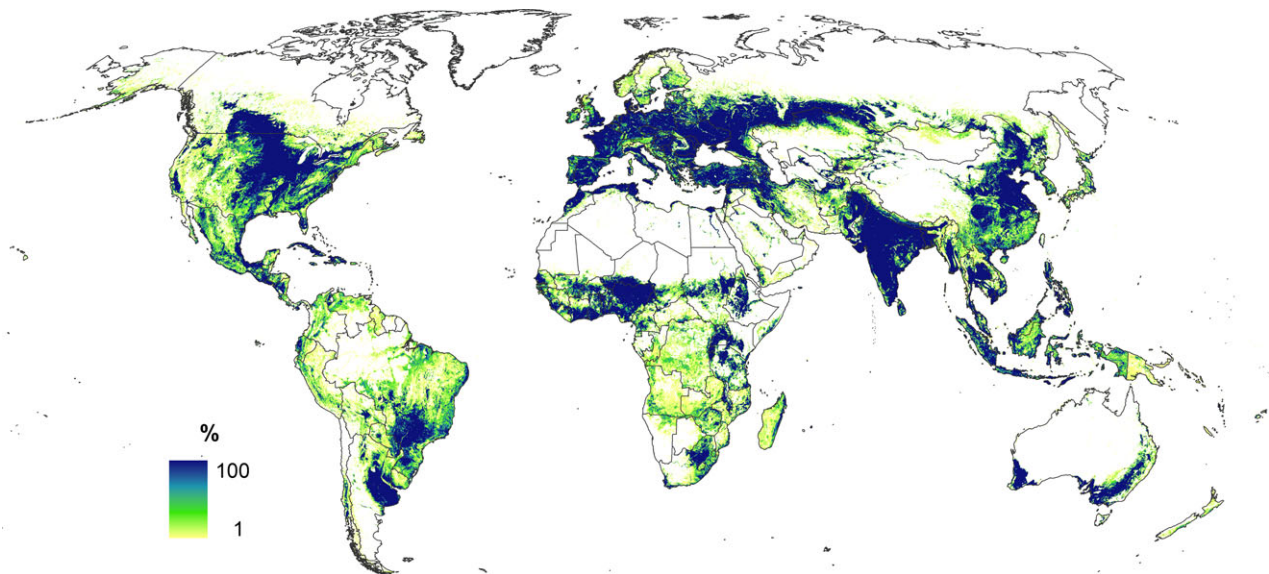


Fig. 3 The IIASA-IFPRI cropland percentage map calibrated with national and subnational cropland statistics from FAO and IFPRI.

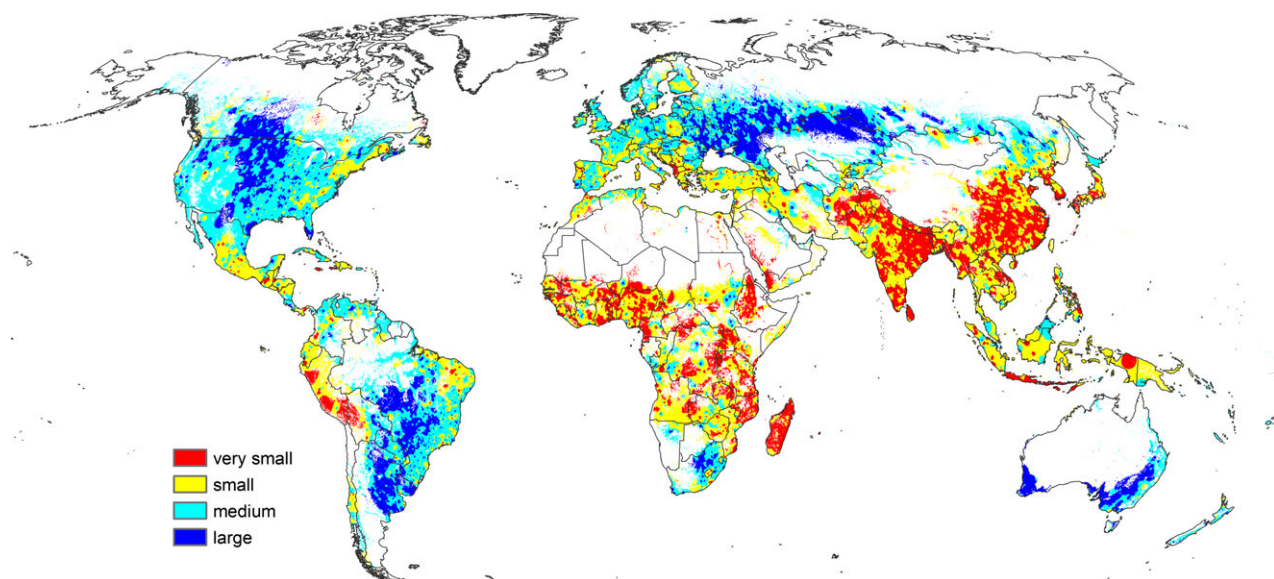


Fig. 4 A global map of field size.

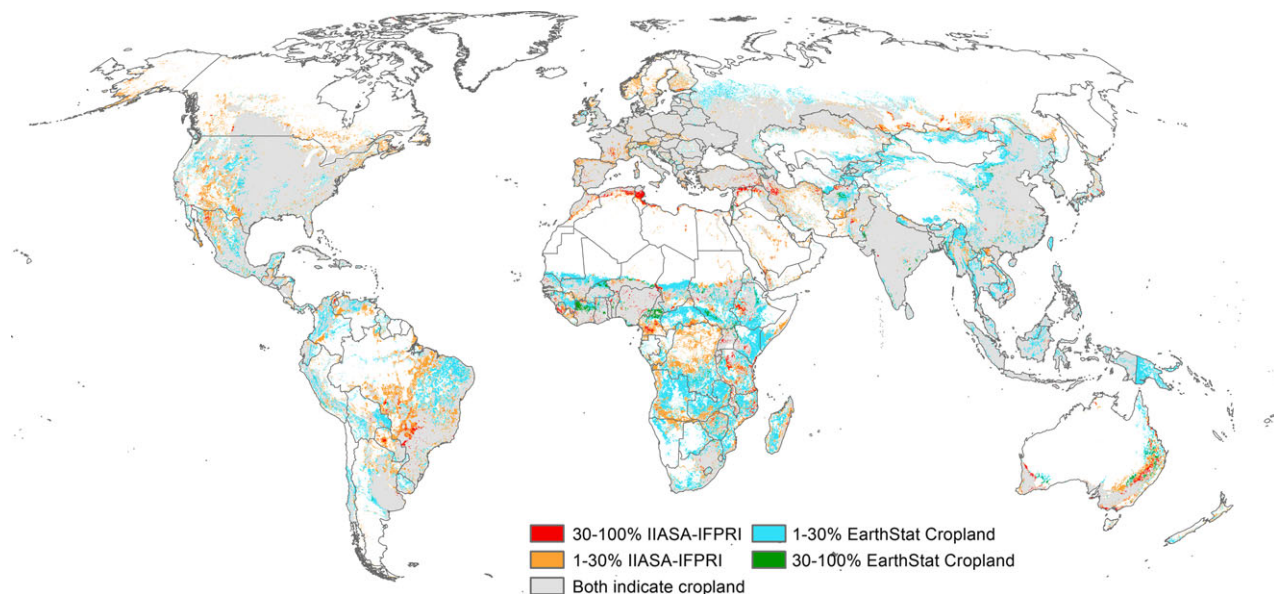


Fig. 5 A disagreement map between the EarthStat cropland and aggregated IIASA-IFPRI cropland products. Grey shading shows areas where both products indicate the presence of cropland to some percentage value. Red and orange shading indicates areas where the IIASA-IFPRI product indicates the presence of cropland which is absent in the EarthStat cropland product while shades of green indicate the opposite situation.

general agreement on the presence of cropland in most parts of North and Central America, Europe, India, China, Southeast Asia and parts of Australia. However, there are considerable areas of disagreement in parts of Central, Southern and Northern Africa, Brazil, Papua New Guinea and parts of other countries throughout the world. Figure 6 shows the African disagreement map in more detail. Superimposed on this map is a random sample drawn from both the EarthStat cropland

and IIASA-IFPRI cropland maps from those areas where they disagree to indicate which product is correct at that location. The distribution shows that the IIASA-IFPRI cropland product is correct twice as often as the EarthStat cropland product in those areas.

A sample of 1033 10 km² pixels was then randomly selected from the areas of highest disagreement and classified independently by at least two experts using Geo-Wiki. Before using this sample for comparing the

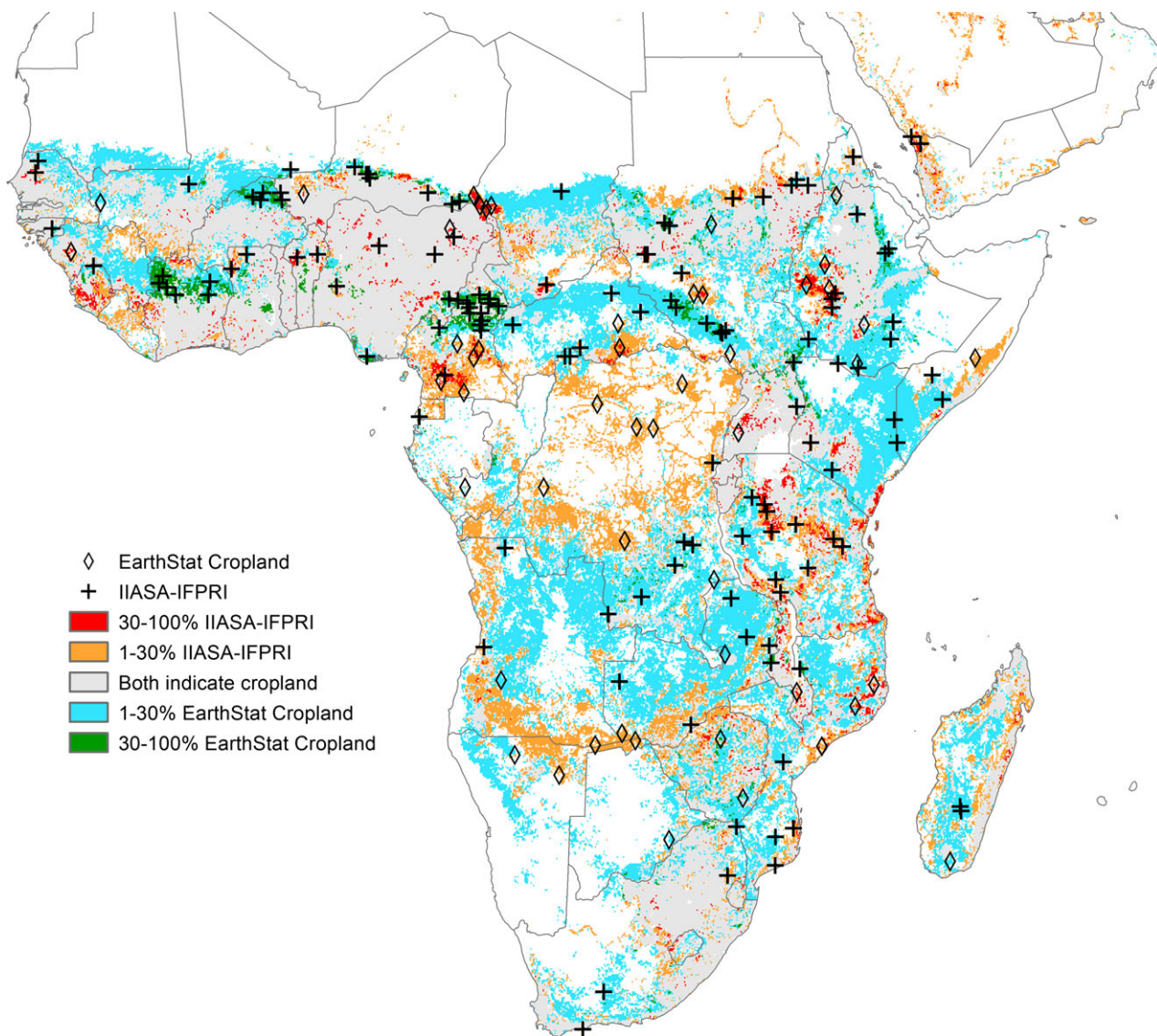


Fig. 6 A disagreement map between the EarthStat cropland and aggregated IIASA-IFPRI cropland products for Africa. Grey shading shows areas where both products indicate the presence of cropland to some percentage value. Red and orange shading indicates areas where the IIASA-IFPRI cropland product indicates the presence of cropland which is absent in the EarthStat cropland product while shades of green indicate the opposite situation. The symbols indicate which product is correct at a particular location of disagreement.

two products, the values from the two experts were compared at each location. Figure 7(a) shows that the two expert estimations are highly correlated ($\rho = 0.93$) and that the RMSE between the estimates from the two independent experts was around 14% (Table 2). The experts clearly found it easier to identify areas with no cropland (RMSE = 7%) while identifying areas of cropland was harder (RMSE = 16%) but still within an acceptable level of agreement. These results, therefore, provide high confidence in the data set derived by the experts using Geo-Wiki.

These expert values were then averaged at each location and used to independently validate the aggregated

IIASA-IFPRI cropland product and the EarthStat cropland product. The results are shown in Table 2 and scatter plots are provided in Figures 7(b,c) including 95% confidence intervals. Due to non-normality of the correlated variables, the confidence intervals for the correlation coefficients were obtained by bootstrapping 1000 samples from the original data set. The results show that the EarthStat cropland map has the highest overall RMSE at 41% and a low Pearson's correlation coefficient ($\rho = 0.14$). The IIASA-IFPRI product does significantly better when compared to the expert-derived data with an overall RMSE of 23% and a Pearson's correlation coefficient of 0.7. Additionally, RMSE

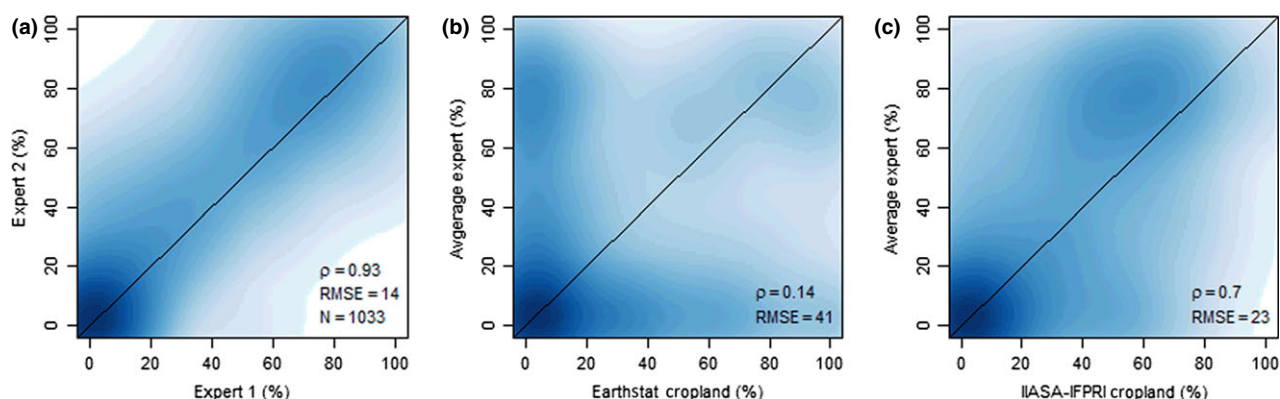


Fig. 7 Scatter plots of (a) the two experts, (b) the EarthStat cropland product and (c) the IIASA-IFPRI cropland map plotted against the average expert cropland percentage pixels. The 1 : 1 line is shown. The 95% confidence intervals are (a) 0.92–0.94, (b) 0.08–0.22 and (c) 0.66–0.73.

Table 2 Consistency assessment in disagreeing areas using per cent RMSE

| | Overall (<i>n</i> = 1033) (%) | No cropland in IIASA-IFPRI cropland map (<i>n</i> = 459) (%) | No cropland in EarthStat cropland (<i>n</i> = 582) (%) |
|-------------------------------|--------------------------------------|--|--|
| IIASA-IFPRI cropland | 23 | 11 | 25 |
| EarthStat cropland | 41 | 40 | 46 |
| Between the two experts | 14 | 7 | 16 |

values were compared over pixels where noncropland pixels were ignored in either the IIASA-IFPRI or EarthStat cropland maps. Although there may be some differences between the products as EarthStat cropland is consistent with cropland statistics for the year 2000 while the IIASA-IFPRI product was calibrated using statistics for the year 2005, we would argue that the main differences between the maps are not due to change in cropland extent over this period. See Appendix S4, which demonstrates that the size of disagreeing areas between the two products is significantly larger than any change in cropland extent between 2000 and 2005 for the majority of countries.

Validation of the field size map

The first type of validation was to assess the input data used to create the field size map, that is to assess the trustworthiness of the crowd. The resulting Spearman's rank correlation coefficient of 0.80 indicates that agreement between the crowd and experts is acceptable.

The second form of validation was to compare a randomly chosen set of validation points from the interpolated field size map compared with expert control points, where two experts independently agreed on the validations (although three experts were involved in creating the set). In total, 181 validation points with field sizes ranging from very small (1) to large (4) were used (after removing validation points with coarse resolution imagery and noncropland areas). The results indicate a Spearman's rank correlation coefficient of 0.78. Given that the global map has been developed using only 13 963 points, this agreement is satisfactory.

Discussion and conclusions

This study provided two new products for use by the global modelling and assessment community: a global percentage cropland map at a resolution of 1 km² (the IIASA-IFPRI cropland product) that is consistent with national cropland statistics; and a global field size map at the same resolution. It is envisioned that this IIASA-IFPRI cropland product could become the 'reference' map for 2005 for integrated assessment modelling purposes. The cropland map is also being used by the Group on Earth Observations (GEO) Global Agricultural Monitoring Initiative (GEOGLAM) for their monthly reporting and crop monitoring activities (<http://www.geoglam-crop-monitor.org/>). IFPRI has applied their spatial allocation model approach to disaggregate the global cropland map into spatial distributions of crop type; these are available for viewing in the 'cropland' branch of Geo-Wiki. The field size map will be one input to the improved stratification of agro-environmental areas in the framework of the EU-funded SIGMA (Stimulating Innovation for Global Monitoring of Agriculture and its Impact on the Environment in

support of GEOGLAM) project, as well as a resource for agricultural monitoring more generally.

Both maps have been developed using bottom up approaches, that is integration of existing maps shared by the community, and the development and validation of products driven by crowdsourcing through the availability of very high-resolution satellite imagery. These approaches have both strengths and weaknesses. Data collection via tools such as Geo-Wiki is a low-cost alternative to the more traditional producer-driven collection of validation data using visual interpretation of very high-resolution data sets such as that found on Google Earth (Biradar *et al.*, 2009; Bontemps *et al.*, 2011) or data directly from image providers (Olofsson *et al.*, 2012) and can result in vastly larger quantities of data than have been used to produce and validate these types of maps in the past. The sharing of data promoted through organizations such as GEO and national open data movements are accelerating the exchange of data. These positive trends will fundamentally change the way in which these products are produced in the future.

However, integration of existing maps also has problems in terms of global consistency. Some maps will be less accurate than others, and they will have been produced over different time periods, so the resulting hybrid map really represents the best estimate of cropland given these constraints. As no single product has yet been shown to be consistently accurate in representing percentage cropland, this approach is still better than relying on a single automated classification algorithm using data from one sensor. The collective knowledge that is captured through a hybrid approach will simply be better than any individual product that is currently available. Another important source of uncertainty in the mapping process is that the input maps as well as the resulting IIASA-IFPRI cropland product are static in nature but cropland areas are dynamic, especially in marginal lands where there may be strong variations in rainfall from year to year. However, we would argue that rainfall will have mostly an impact on yield and less impact on area unless there are several years of low rainfall. The IIASA-IFPRI map does not, however, capture active vs. passive fields, that is long periods of fallow, but this is not included in the definition of cropland used here.

Crowdsourcing via Geo-Wiki represents a new way of gathering much larger amounts of data that can be used in the development of new products and in validation. However, there are valid concerns about data quality that have been raised in the literature (Flanagin & Metzger, 2008). Where crowdsourced data have been used in this study, quality control measures have been put in place, for example comparison with data derived by experts to ensure minimum levels of quality and a

minimum level of agreement between multiple observations at a single location.

The field size map should also be considered as a rough first attempt to create such a global product. Inevitably, the size of the sample and the use of interpolation will affect the results when viewed at smaller scales. Moreover, we acknowledge that the data collection method works best when the field sizes are homogeneous within the 1 km² area over which data are collected and does not currently account for variations of field size within the 1 km² areas. However, for modelling and monitoring purposes at a global scale, it represents a reasonable proxy for agricultural mechanization and human development at this stage.

Both global products will continue to be improved in the future. As more and more countries develop the capacity to produce their own national land cover and cropland maps, these can be easily integrated within the hybrid product. However, we must ensure that these national efforts are undertaken with a global perspective, for example the use of standard definitions for cropland as set out by organizations like FAO with their Land Cover Classification System (LCCS) and their Land Cover Meta Language (LCML) or the more recent EAGLE (Eionet Action Group on Land monitoring in Europe) initiative (<http://sia.eionet.europa.eu/EAGLE>), in order to be able to harmonize these products. The IIASA-IFPRI cropland map will also be improved through further contributions of national maps shared through GEOGLAM and the SIGMA project as well as from new mapping efforts such as that recently undertaken by the National Geomatics Centre in China. We are currently producing a new version for 2010. For the moment, the plan is to keep the 1 km² resolution as this is appropriate for integrated assessment modelling at the global scale. We are already using some input cropland layers that were produced using Landsat-type resolution (i.e. some national maps), and we will ingest any new products at this resolution for the 2010 product. The field size map will also be improved through a planned crowdsourcing campaign in the future, in which field size will be sampled more densely.

The differences between the aggregated IIASA-IFPRI product and the EarthStat cropland product clearly show that there are locations where each product is better at capturing cropland. As more data are collected via Geo-Wiki, creating a hybrid product from these two different aggregated maps is a clear possibility. Ultimately, the goal is to produce a cropland product with the highest possible accuracy that meets the needs of the modelling and assessment community. By integrating existing products, we have presented one feasible solution towards attaining this objective.

Acknowledgements

This research was supported by National Basic Research Program of China (973 Program: No.2010CB951502), the CGIAR Research Programme on Climate Change, Agriculture and Food Security (CCAFS) under the Gaming contract (Collecting data via gaming to produce improved land cover products) and the EU FP7 funded GEOCARBON (No. 283080), ERC CrowdLand (No. 612755) and SIGMA (No. 603719) projects. We would also like to acknowledge Ola Ahlqvist, Lucy Bastin, Diego de Abelleira, Pierre Defourny, Carlos Di Bella, Jan de Leeuw, Karlheinz Erb, Jinlong Fan, Giles Foody, Adam Gerrand, Sven Gilliams, Bruno Gérard, Alex Held, Martin Herold, Noriko Hosonuma, Ian Jarvis, Martin Kappas, Patrick Kasangaki, LeeAnn King, Olga Krankina, Tobias Kummerle, Rick Leiterer, Olivier Leo, Tom Loveland, Hawa Mohammed, Joanne Nightingale, An Notenbaert, Pontus Olofsson, Alexander Prischchepov, Li Qiangzi, Bernardo Rudorff, Megan Salmon, Igor Savin, Joe Sexton, Yosio Shimabakuro, Steve Stehman, Pierre Christophe Sibiry Traore, Thorsten Wendt, Alyssa Whitcraft, Stanley Wood, Curtis Woodcock, Bingfang Wu, Xiangming Xiao and Pei Zhiyuan for their participation in the 'Characterizing and Validating Global Agricultural Land Cover Workshop', which initiated this process to create the hybrid land cover map. Other participants who contributed substantially in terms of either cropland or validation data have been acknowledged as co-authors. Support was provided by CCAFS and the Joint Research Centre of the European Union.

References

- Bekunda M, Sanginga N, Woormer PL (2010) Chapter four – restoring soil fertility in sub-Saharan Africa. In: *Advances in Agronomy*, Vol. Volume 108 (ed. Sparks DL), pp. 183–236. Academic Press, San Diego CA.
- Berry RA (1972) Farm size distribution, income distribution, and the efficiency of agricultural production: Colombia. *The American Economic Review*, **62**, 403–408.
- Bicheron P, Defourny P, Brockmann C *et al.* (2008) *Globcover: Products Description and Validation Report*. Medias France, Toulouse France.
- Biradar CM, Thenkabail PS, Noojipady P *et al.* (2009) A global map of rainfed cropland areas (GMRCAs) at the end of last millennium using remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, **11**, 114–129.
- Bontemps S, Defourny P, van Bogaert E, Arino O, Kalogirou V, Perez JR (2011) GLOBCOVER 2009: products description and validation report.
- Brown ME, Funk CC (2008) Food security under climate change. *Science*, **319**, 580–581.
- Carfagna E, Pratesi M, Carfagna A (2013) Methodological developments for improving the reliability and cost-effectiveness of agricultural statistics in developing countries. In: *Proceedings of the 59th World Statistics Congress of the International Statistical Institute*. International Statistical Institute, The Hague, the Netherlands.
- FAO (2005) FAOSTAT agricultural data.
- FAO (2011a) *The State of Food Insecurity in the World: How Does International Price Volatility Affect Domestic Economies and Food Security?* Food and Agriculture Organization of the United Nations, Rome.
- FAO (2011b) *Global Food Losses and Food Waste – Extent, Causes and Prevention*. FAO UN, Rome, Italy.
- FAO, World Bank, United Nations Statistical Commission (2012) *Action Plan of the Global Strategy to Improve Agricultural and Rural Statistics*. FAO UN, Rome, Italy.
- Flanagin A, Metzger M (2008) The credibility of volunteered geographic information. *GeoJournal*, **72**, 137–148.
- Foley JA, Ramankutty N, Brauman KA *et al.* (2011) Solutions for a cultivated planet. *Nature*, **478**, 337–342.
- Foody GM (2010) Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sensing of Environment*, **114**, 2271–2285.
- Friedl MA, Sulla-Menasha D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X (2010) MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, **114**, 168–182.
- Fritz S, Bartholomé E, Belward A, Hartley A, Stibig H-J, Eva H, Mayaux P (2003) *Harmonisation, Mosaicing and Production of the Global Land Cover 2000 Database (Beta Version)*. Office for Official Publications of the European Communities, Luxembourg.
- Fritz S, McCallum I, Schill C *et al.* (2009) Geo-Wiki.Org: the use of crowdsourcing to improve global land cover. *Remote Sensing*, **1**, 345–354.
- Fritz S, You L, Bun A *et al.* (2011) Cropland for sub-Saharan Africa: a synergistic approach using five land cover data sets. *Geophysical Research Letters*, **38**, L04404.
- Fritz S, McCallum I, Schill C *et al.* (2012) Geo-Wiki: an online platform for improving global land cover. *Environmental Modelling & Software*, **31**, 110–123.
- Fritz S, See L, van der Velde M *et al.* (2013) Downgrading recent estimates of land available for biofuel production. *Environmental Science & Technology*, **47**, 1688–1694.
- Geiger F, Bengtsson J, Berendse F *et al.* (2010) Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. *Basic and Applied Ecology*, **11**, 97–105.
- Godfray HCJ, Beddington JR, Crute IR *et al.* (2010) Food security: the challenge of feeding 9 billion people. *Science (New York, N.Y.)*, **327**, 812–818.
- Havlik P, Valin H, Herrero M *et al.* (2014) Climate change mitigation through livestock system transitions. *Proceedings of the National Academy of Sciences*, **111**, 3709–3714.
- Havlik P, Schneider UA, Schmid E *et al.* (2011) Global land-use implications of first and second generation biofuel targets. *Energy Policy*, **39**, 5690–5702.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Jung M, Henkel K, Herold M, Churkina G (2006) Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment*, **101**, 534–553.
- Kuemmerle T, Erb K, Meyfroidt P *et al.* (2013) Challenges and opportunities in mapping land use intensity globally. *Current Opinion in Environmental Sustainability*, **5**, 484–493.
- Levin G (2006) Farm size and landscape composition in relation to landscape changes in Denmark. *Danish Journal of Geography*, **106**, 45–59.
- Lobell DB, Cassman KG, Field CB (2009) Crop yield gaps: their importance, magnitudes, and causes. *Annual Review of Environment and Resources*, **34**, 179–204.
- Olofsson P, Stehman SV, Woodcock CE *et al.* (2012) A global land-cover validation data set, part I: fundamental design principles. *International Journal of Remote Sensing*, **33**, 5768–5788.
- Perger C, Fritz S, See L, Schill C, dervan Velde M, McCallum I, Obersteiner M (2012) A campaign to collect volunteered geographic information on land cover and human impact. In: *GI Forum 2012: Geovisualisation, Society and Learning* (eds Jekel T, Car A, Strobl J, Griesebner G), pp. 83–91. Herbert Wichmann Verlag, Berlin/Offenbach.
- Pittman K, Hansen MC, Becker-Reshef I, Potapov PV, Justice CO (2010) Estimating global cropland extent with multi-year MODIS data. *Remote Sensing*, **2**, 1844–1863.
- Pontius RG, Millones M (2011) Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, **32**, 4407–4429.
- Ramankutty N, Evan AT, Monfreda C, Foley JA (2008) Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, **22**.
- Roberts L (2011) 9 Billion? *Science*, **333**, 540–543.
- See L, Fritz S, Thornton P *et al.* (2012) Building a Consolidated Community Global Cropland Map | Earthzine. Available at: <http://earthzine.org/2012/01/24/building-a-consolidated-community-global-cropland-map/> (accessed 01 February 2013).
- Stuart T (2009) *Waste: Uncovering the Global Food Scandal*. Penguin, London, UK.
- Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S (2002) Agricultural sustainability and intensive production practices. *Nature*, **418**, 671–677.
- Tollefson J (2011) Seven billion and counting. *Nature*, **478**, 300.
- Vancutsem C, Marinho E, Kayitakire F, See L, Fritz S (2012) Harmonizing and combining existing land cover/land use datasets for cropland area monitoring at the African continental scale. *Remote Sensing*, **5**, 19–41.
- Wheeler T, von Braun J (2013) Climate change impacts on global food security. *Science*, **341**, 508–513.
- World Bank, FAO, United Nations Statistical Commission (2011) *Global Strategy to Improve Agricultural and Rural Statistics*. World Bank, Washington, DC.
- Yan L, Roy DP (2014) Automated crop field extraction from multi-temporal Web Enabled Landsat Data. *Remote Sensing of Environment*, **144**, 42–64.
- You L, Wood-Sichra U, Fritz S, Guo Z, See L, Koo J (2014) Spatial Production Allocation Model (SPAM) 2005 Beta Version. International Food Policy Research Institute, Washington, DC, USA and the International Institute for Applied Systems Analysis, Laxenburg, Austria. Available at: <http://mapspam.info> (accessed July 2014).
- Zhao Y, Gong P, Yu L (2014) Towards a common validation sample set for global land cover mapping. *International Journal of Remote Sensing*, **35**, 4795–4814.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. The spatial distribution of validation points across South America and Africa. Green indicates validation points where agreement between multiple answers was >65% while red indicates agreement <65%. Only those validation points where agreement was greater than 65% were used in the paper to ensure a minimum quality in the validation data. Globally, 36% of the total validation points had an agreement higher than 65%. In South America 43% were used while in Africa 33% were used. The lower number used in Africa is due to the fact that there are smaller fields and more heterogeneous landscapes so the interpretation of satellite imagery is more difficult.

Table S1. The 10 land cover classes in the Geo-Wiki simple legend. These were chosen to be consistent with the generalized land cover classes proposed by Herold *et al.* (2008), which allows for comparison of different land cover products.

Appendix S1. Quantifying spatial variation in field size by region.

Table S2. I^2 statistics by region.

Figure S2. Frequency distribution of field size by region and the respective estimate of I^2 . Regions are listed in Table S2.

Appendix S2. Calculation of agreement.

Table S3. Crowdsourced answers from two users for the same location.

Appendix S3. The ranking procedure and production of the synergy map.

Table S4. The ranking of products by country. Only those countries with more than 400 crowdsourced data points are included.

Table S5. Synergy table for Russia where 4 input layers are available. Grey shading highlights those combinations with a tied ranking as the accuracy values for MODIS and GlobCover are the same in this example.

Appendix S4. Comparison of change in cropland area from 2000 to 2005 compared with the size of the disagreement between the IIASA-IFPRI and EarthStat cropland products.

Table S6. Countries ranked in descending order by cropland area in 2005 showing the difference in areas between 2005 and 2000 based on FAOSTAT, the area of disagreement between the IIASA-IFPRI and EarthStat cropland products and the latter two quantities in relation to one another, expressed as a percentage.