Article

Estimating Global Cropland Extent with Multi-year MODIS Data

Kyle Pittman [1],*, **Matthew C. Hansen** [1], **Inbal Becker-Reshef** [2], **Peter V. Potapov** [1], and **Christopher O. Justice** [2]

[1] South Dakota State University Geographic Information Science Center of Excellence, Brookings, SD 57007, USA; E-Mails: Matthew.Hansen@sdstate.edu (M.C.H.); Peter.Potapov@sdstate.edu (P.V.P.)

[2] Department of Geography, University of Maryland, College Park, MD 20742, USA; E-Mails: ireshef@hermes.geog.umd.edu (I.B.-R.); justice@hermes.geog.umd.edu (C.O.J.)

* Author to whom correspondence should be addressed; E-Mail: Kyle.Pittman@sdstate.edu; Tel.: +1-605-688-6591; Fax: +1-605-688-5227.

Abstract: This study examines the suitability of 250 m MODIS (MODerate Resolution Imaging Spectroradiometer) data for mapping global cropland extent. A set of 39 multi-year MODIS metrics incorporating four MODIS land bands, NDVI (Normalized Difference Vegetation Index) and thermal data was employed to depict cropland phenology over the study period. Sub-pixel training datasets were used to generate a set of global classification tree models using a bagging methodology, resulting in a global per-pixel cropland probability layer. This product was subsequently thresholded to create a discrete cropland/non-cropland indicator map using data from the USDA-FAS (Foreign Agricultural Service) Production, Supply and Distribution (PSD) database describing per-country acreage of production field crops. Five global land cover products, four of which attempted to map croplands in the context of multiclass land cover classifications, were subsequently used to perform regional evaluations of the global MODIS cropland extent map. The global probability layer was further examined with reference to four principle global food crops: corn, soybeans, wheat and rice. Overall results indicate that the MODIS layer best depicts regions of intensive broadleaf crop production (corn and soybean), both in correspondence with existing maps and in associated high probability matching thresholds. Probability thresholds for wheat-growing regions were lower, while

areas of rice production had the lowest associated confidence. Regions absent of agricultural intensification, such as Africa, are poorly characterized regardless of crop type. The results reflect the value of MODIS as a generic global cropland indicator for intensive agriculture production regions, but with little sensitivity in areas of low agricultural intensification. Variability in mapping accuracies between areas dominated by different crop types also points to the desirability of a crop-specific approach rather than attempting to map croplands in aggregate.

Keywords: agriculture; cropland mapping; global; MODIS

List of Abbreviations:

MODIS	**MODerate Resolution Imaging Spectroradiometer**
USDA	**United States Department of Agriculture**
FAS	**Foreign Agricultural Service**
CADRE	**Crop Condition Data Retrieval and Evaluation**
LACIE	**Large Area Crop Inventory Experiment**
AgRISTARS	**Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing**
GLAM	**Global Agriculture Monitoring project**
UNFAO	**United Nations Food and Agricultural Organization**
GIEWS	**Food Security Global Information and Early Warning System**
USAID	**United States Agency for International Development**
FEWS	**Famine Early Warning System**
MARS	**Monitoring Agriculture with Remote Sensing**
GMFS	**Global Monitoring of Food Security**
IRSA	**Institute of Remote Sensing Applications**
IWMI	**International Water Management Institute**

1. Introduction

One of the most fundamental uses of earth observation satellite data is the mapping and monitoring of croplands. Much of the global food supply is dependent on the cultivation of a few select crops produced during each local growing season, including corn, wheat, soybeans and rice. Over the past several decades, numerous governmental agencies have utilized satellite remote sensing data sets in order to monitor and quantify these crop types.

Within the United States, some of the early attempts at cropland mapping and monitoring using remotely sensed data included the LACIE (Large Area Crop Inventory Experiment) and AgRISTARS (Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing) programs [1]. Since these initial pathfinding endeavors, numerous national operational agriculture monitoring programs employing remote sensing data have been implemented. Most of these programs focus on mapping cropland extent, crop type, crop condition and production [2]. At the global scale, fewer such

endeavors exist. The United Nations Food and Agricultural Organization (UNFAO) uses remotely sensed data in its Food Security Global Information and Early Warning System (GIEWS), as does the United States Agency for International Development (USAID) Famine Early Warning System (FEWS); The United States Department of Agriculture's Foreign Agricultural Service (FAS) Production Estimates and Crop Assessment Division (PECAD) program employs remotely sensed data; the European Commission operates the Monitoring Agriculture with Remote Sensing (MARS) and Global Monitoring of Food Security (GMFS) programs; China's Institute of Remote Sensing Applications (IRSA) within the Chinese Academy of Sciences has an expanding global capability with its Chinese Cropwatch Program.

There are several complicating factors involved in characterizing croplands at the global scale. First, the spatial extent of croplands is highly variable between and within nations. Depending on the historical, political, social and technological context of agricultural development and natural factors such as landscape pattern, cropland characteristics such as field size can be highly variable, even for the same crop type. Second, intensification in the form of management practices such as the use of fertilizer varies greatly, especially between developed and developing nations. Third, each crop type has a specific growth phenology and structure, with significant seasonal variation between and even within individual crop types. Fourth, cropland is a land use and can be confused with natural vegetation cover types, such as cereal grains versus tall-grass prairie.

To overcome these limitations, high-temporal earth observation coverage at fine spatial scales is desired. However, such a capability at the global scale does not exist. Higher spatial resolution data such as Landsat do not have the temporal coverage to enable cropland monitoring at the global scale. Landsat-like data with higher repeat imaging rates, such as Indian Remote Sensing's AWiFS instrument, have been used at national scales [3] but are not acquired globally. Near daily observational coverage is available only with coarse spatial resolution data such as NASA's MODIS (MODerate Resolution Imaging Spectroradiometer) sensor. While MODIS will not be a viable option for cropland mapping over many regions due to its inability to resolve smaller field sizes, it can provide a means for indicating cropland presence over large areas [2].

The MODIS sensor, aboard the Aqua and Terra satellites provides a valuable compromise between high temporal frequency and high spatial resolution. MODIS contains seven bands designed specifically to map the land surface with 250 m and 500 m spatial resolutions. The data are free and provide daily acquisitions above 30 degrees latitude [4]. The data also undergo a standard processing procedure to surface reflectance, enabling their use for global-scale applications [5].

There has been a variety of attempts to map global croplands either independently or as part of multi-class landcover classifications. Ramankutty *et al*. mapped both early 1990's [6] and year 2000 [7] global croplands at a ~10 km resolution using satellite data alongside national and local agricultural inventory data. The IGBP DISCover project [8] used 1 km Advanced Very High Resolution Radiometer data to produce a global land cover product that included a cropland class, as did Hansen *et al*. [9] with the University of Maryland (UMD) land cover map. Friedl *et al*. [10] advanced these efforts in creating the standard MODIS land cover product (MOD12Q1) which included a cropland class. Bartholomé *et al*. [11] used SPOT VEGETATION data to produce the Global Landcover 2000 (GLC2000) product, consisting of 21 regional land cover classes, including a cultivated area/cropland class. The Globcover initiative then used 300 m ENVISAT MERIS data to

produce a global multi-class land cover map for 2005–2006 [12]. A discrete 10 km characterization of irrigated [13] and rainfed cropland [14] extent was generated by the International Water Management Institute (IWMI) using a variety of processing algorithms and data set inputs. Global cropland extent can be derived by combining the 9 rainfed cropland and 8 irrigated cropland classes of the respective products.

The purpose of this study is to examine the ability of 250 m MODIS data for mapping global cropland extent and to evaluate the results for major crop production countries and regions. The analysis advances the topic of global cropland characterization in a number of ways. First, the product is made at a 250 m spatial resolution, the finest-scale global cropland product derived using synoptic inputs. Second, nine years of MODIS inputs are used to create a data set insensitive to short-term interannual variability in attempting to depict core cropland production areas. Third, the method employs high-resolution data sets covering the majority of the globe's crop production areas for model calibration. Fourth, the initial output is a per pixel probability layer that is adjusted to match regional cropland area statistics. Fifth, unlike the majority of previous global land cover maps where cropland was mapped in the context of multiple land cover classes, this effort is aimed solely at mapping cropland extent. Results reveal where MODIS is an appropriate data source for cropland indicator mapping at the global scale.

2. Methods

2.1. Training Data

The training data for this project were based on high resolution satellite-derived classification products. The primary source was the Geocover Land Cover product, derived from the Landsat global Geocover data set [15]. These data were augmented by the UN Food and Agriculture Organization AfriCover [16], USDA National Agricultural Statistics Service Cropland Data Layer [3], United States National Land Cover Database [17], Agriculture and Agri-Food Canada [18], South Africa State of the Environment [19], and European CORINE Landcover [20] data sets. The products that existed as discrete landcover classifications were simply reclassified to a binary cropland/no cropland classification scheme. The input sources that contained continuous classification of crop cover were thresholded so that all MODIS pixels with greater than 50% cropland coverage were classified as cropland while all other pixels were classified as no cropland. All products were then reprojected and resampled to the 250 m MODIS global grid. While the training data represented 80% of the earth's land surface, excluding Antarctica, some gaps remained. To fill these gaps, training data created using photointerpretion of Landsat imagery were added for Russia, parts of Central Africa and Southeast Asia.

2.2. MODIS Data

The input dataset consisted of multiyear MODIS data covering four of the seven MODIS land bands, one thermal band, and NDVI (Normalized Difference Vegetation Index). The MODIS land band data were sourced from the Collection 5 MOD09 standard product [21] covering 2000–2008. The MODIS inputs consisted of 250 m band 1 (620–670 nm, red), band 2 (841–876 nm, near infrared),

500 m band 3 (459–479 nm, blue), band 7 (2,105–2,155 nm, midinfrared), and 1 km band 31 (10.780–11.280 μm, thermal) data. The NDVI was calculated using bands 1 and 2. The MODIS land bands were designed for land cover monitoring applications such as this one and the addition of a thermal band enabled the binning of the land bands based on local seasonality.

To aid in avoidance of cloud cover, the MODIS daily acquisitions were converted to 16-day composites, resulting in 23 composites per band for each of the 9 years of data [22]. Compositing is an approach used to improve data quality and reduce data volumes of time-series datasets by preferentially selecting cloud-free observations for each pixel within a given time interval [23]. Multi-year metrics were created to produce a single dataset of a representative phenology for a given pixel and band for a particular 16-day composite period across the 9 year period. This was done by ranking, by pixel, all years for each composite period and band and using the ranked data to create a set of 39 metrics, similar to those of Hansen *et al*. [24]. Table 1 lists the multi-year metrics used as the independent variables for the cropland mapping analysis.

Table 1. MODIS metrics used for mapping global cropland extent.

Mean of the 3 least reflective channel 1 (red) composites
Mean of channel 1 (red) in 3 warmest composites
Mean of channel 1 (red) in 3 greenest composites
Mean of the 3 **least reflective** channel 2 (NIR) composites
Mean of channel 2 (NIR) in 3 warmest composites
Mean of channel 2 (NIR) in 3 greenest composites
Mean of the 3 warmest channel 31 (thermal) composites
Mean of channel 31 (thermal) in 3 greenest composites
Mean of the 3 **least reflective** channel 7 (SWIR) composites
Mean of channel 7 (SWIR) in 3 warmest composites
Mean of channel 7 (SWIR) in 3 greenest composites
Mean of the 3 greenest (NDVI) composites
Mean of NDVI in 3 warmest composites
Mean of the 6 **least reflective** channel 1 (red) composites
Mean of channel 1 (red) in 6 warmest composites
Mean of channel 1 (red) in 6 greenest composites
Mean of the 6 **least reflective** channel 2 (NIR) composites
Mean of channel 2 (NIR) in 6 warmest composites
Mean of channel 2 (NIR) in 6 greenest composites
Mean of the 6 warmest channel 31 (thermal) composites
Mean of channel 31 (thermal) in 6 greenest composites
Mean of the 6 **least reflective** channel 7 (SWIR) composites
Mean of channel 7 (SWIR) in 6 warmest composites
Mean of channel 7 (SWIR) in 6 greenest composites
Mean of the 6 greenest (NDVI) composites
Mean of NDVI in 6 warmest composites
Mean of the 12 **least reflective** channel 1 (red) composites
Mean of channel 1 (red) in 12 warmest composites
Mean of channel 1 (red) in 12 greenest composites

Table 1. *Cont.*

Mean of the 12 **least reflective** channel 2 (NIR) composites
Mean of channel 2 (NIR) in 12 warmest composites
Mean of channel 2 (NIR) in 12 greenest composites
Mean of the 12 warmest channel 31 (thermal) composites
Mean of channel 31 (thermal) in 12 greenest composites
Mean of the 12 **least reflective** channel 7 (SWIR) composites
Mean of channel 7 (SWIR) in 12 warmest composites
Mean of channel 7 (SWIR) in 12 greenest composites
Mean of the 12 greenest (NDVI) composites
Mean of NDVI in 12 warmest composites

2.3. Classification Tree Algorithm

The S-Plus statistical package was used to implement the classification tree analysis [25]. Classification trees are a supervised approach using binary splits to create successively purer subsets. The classification tree uses a deviance reduction measure to split the data into more homogeneous subsets. The reduction in deviance (*D*), is calculated:

$$D = D_s - D_t - D_u$$

where *s* is the parent node and *t* and *u* are the splits from *s*. Each possible split across all input metrics is tested until *D* is maximized. The data are then divided and the process repeated for the two newly created subsets. To calculate deviance, a log likelihood is used as follows:

$$D_i = -2 \Sigma n_{ik} \, logp_{ik}$$

where *n* is the number of pixels of class *k* in node *i* and *p* is the probability distribution of class *k* in node *i*.

Bagging, a process whereby multiple runs of decision trees are generated using sampling with replacement, allows for more reliable results than that of a single tree [26]. Repeatedly sampling the data and growing multiple tree models limits the impact of overfitting within any given single model. A median or mean result from multiple model runs may be used as the final result. For this study, the 39 multiyear MODIS metrics were used as the independent variables, while the yes/no global Landsat-derived training dataset was the dependent variable. Twenty-five samples were taken from the training data, with replacement, each representing 20% of the available training, or approximately 16% of the earth's land surface. Because of processing limitations, perfect trees were not used in the bagging approach. A threshold was implemented requiring a minimum deviance explained of 0.00125% of the root deviance in order for new splits to be created. The resulting trees had an average of approximately 800 terminal nodes. For each terminal node, the probability of cropland was retained and used to label the output pixel. The 25 output probabilities were then ranked and the median result of the 25 used as the final result, ranging from 1 to 99% probability of cropland class membership. The per pixel probability map was used as both a confidence measure as well as the basis for applying different thresholds to match reported national-scale cropland area estimates.

2.4. Thresholds

The FAS Production, Supply and Distribution (PSD) database [27] was used to find a suitable probability threshold for creating a yes/no cropland map out of the continuous result. PSD data come from a variety of sources including official country statistics, international organizations and agricultural attaches at United States embassies around the world. For this project, the median harvested area from 2000 to 2008 for production field crops (barley, corn, cotton, oats, rice, rye, sorghum, soybeans and wheat) was used to determine cropland area per country. One limitation is the lack of formal documentation in the global FAS PSD database for areas that are double-cropped and double-counted towards area estimates. This will, by definition, lead to an overestimation of spatial extent for countries that have significant double-cropping in their annual crop area estimates. Per-country cropland areas from the FAOSTAT database [28], which does explicitly account for double-cropped areas, were compared to the FAS PSD areas in order to assess the potential impact of double-cropped areas on the final product. Table 2 shows the eight countries in which FAS PSD production cropland area exceeded FAOSTAT arable land area, indicating the presence of a double-cropping system. While modeling of single and double cropping systems with MODIS is possible [29], such an analysis was beyond the scope of this work. Users of the new 250 m cropland data layer may apply a custom threshold using FAOSTAT data or any other ancillary reference that may facilitate the use of the probability layer in mapping croplands within a more local context. Before thresholds were calculated, a 250 m MODIS-derived water mask was applied [30] to the product.

Table 2. Countries in which FAS Production Cropland area exceeds FAO arable land area, indicating the presence of double cropping.

Country	FAO Area	FAS Area	% Diff	FAS Threshold	FAO Threshold
Argentina	25,456,125	26,711,333	4.7%	42	44
Bangladesh	7,996,000	12,009,556	33.4%	1	10
Vietnam	6,444,438	8,815,444	26.9%	12	15
Philippines	4,942,625	6,745,667	26.7%	10	13
Egypt	2,937,875	3,294,222	10.8%	16	22
Nepal	2,345,625	3,335,556	29.7%	14	21
Turkmenistan	1,790,000	1,979,778	9.6%	51	54
Tajikistan	765,750	884,111	13.4%	68	71

Three sets of thresholds were created for evaluation purposes. First, a single global threshold was calculated to match global cropland area as reported by FAS PSD data. Next, countries were grouped into regions: Latin America (less Brazil and Argentina), Africa, Europe, Central Asia and South/East Asia (less China and India). For each of these regions, a separate matching threshold was calculated. Eight countries representing the largest single-nation agricultural producers were not aggregated into regional groupings: Argentina, Australia, Brazil, Canada, China, India, Russia and the United States. Finally, for all other individual countries (with the exception of the European Union, where only aggregate statistics were available in the PSD database), the probability threshold most closely matching the PSD production crop harvested area was calculated. These per country thresholds were used to create the final cropland extent product.

2.5 Evaluation

Matching thresholds were evaluated at national, regional and global scales, as well for nations and regions containing a dominant crop type. A high probability threshold indicated that the global model had captured a relatively unambiguous global cropland signal while a low threshold indicated lower confidence in the included crop cover.

The application of the final per country matching thresholds resulted in a discrete croplands indicator map with all pixels of a probability greater than or equal to the matching threshold being classified as croplands and all pixels with a probability less than the threshold classified as not croplands. Though the use of thresholds guarantees the cropland areas will correspond precisely to the comparison databases, it does not necessarily locate those cropland areas in the correct geographic locations within the country or region being mapped. In order to determine the spatial accuracy of the product, it was evaluated against five of the global land cover studies mentioned previously (Ramankutty 2000, IGBP DISCover, UMD, MOD12Q1, GLC2000). Consideration was given to using only the 2000-era products (Ramankutty, MOD12Q1, GLC2000) for evaluation due to the potential impact of global landcover change since the derivation of the 1992–1993 products (IGBP DISCover, UMD). However, the use of all 5 products significantly increases the dynamic range of the validation reference map for many regions. The Globcover product derived from MERIS data was not included as, unlike the AVHRR, MODIS and VEGETATION-derived products, it did not include equal data input density across the globe. For example, data coverage of North America was less than Europe [12]. The IWMI products were also not included due to the inclusion of several mixed cropland/natural vegetation classes at the coarse 10km spatial resolution. It should be noted that with the exception of the Ramankutty 2000 product, these heritage datasets all attempted to map croplands globally in the context of multiclass land cover classifications. Because they did not target croplands specifically, it is not a perfect comparison and the evaluation is thus somewhat limited.

The five selected data sets were reprojected to the 250 m Sinusoidal grid matching the MODIS result using a nearest neighbor resampling and then reclassified into a yes/no cropland product. All mixed classes representing less than 50% cropland were classified as not cropland, and the Ramankutty product was thresholded with all pixels greater than 50% crop coverage labeled croplands. A composite map was then made with values from 0 to 5, where 5 indicated universal agreement on cropland labeling, a 4 meant 4 of the 5 agreed, down to 0 where none of the comparison products called a pixel cropland. These various agreement masks were used to evaluate the MODIS 250 m cropland mask.

3. Results and Discussion

The 25 bagged classification trees in the analysis explained 47.9% of the root node deviance in the training dataset. Of the 39 input metrics, just 4 accounted for 65.8% of the explained deviance. The top metric, explaining 13.5% of the root deviance was the average NDVI within the 12 warmest composite periods. A second metric, representing the average red reflectance within the 12 composite periods having the lowest red reflectance, explained a further 9.2% of the root deviance. All metrics were examined for explanatory power and divided into two generic types. The first, referred to as primary

metrics, includes those where data from the band in question were directly used. Secondary metrics included the band that was used for ranking purposes, for example mean red reflectance of the 12 greenest composite periods. In this case, NDVI is a secondary metric used to locate the red reflectance values within the growing season. Table 3 shows the relative contribution of the MODIS inputs, particularly NDVI and red as primary discriminators of cropland extent. The two bands contributing least to the model were band 7, a 500 m shortwave infrared band, and band 31, a 1km brightness temperature band. The coarser resolution of these inputs, as compared to the 250 m red, near-infrared and NDVI inputs, could have limited their value in mapping croplands. When examining secondary ranking metrics, which were greenness as measured by NDVI, and warmness, as measured by band 31, the value of these inputs is evident. Both allow for examining the red, near-infrared and shortwave infrared bands in the context of local growing seasons. For this portion of Table 3, double counting occurs for both the image and ranking inputs, and total relative contribution sums to greater than 100.

Table 3. Percent explained deviance by input MODIS band and NDVI.

MODIS Band	*As primary metric*	*As primary or secondary metric*
Band 1 (Red)	26.41	26.41
Band 2 (NIR)	16.89	16.89
Band 7 (SWIR)	9.92	9.92
NDVI	31.54	53.40
Band 31 (Thermal)	15.24	54.82

Each of the 25 models was applied globally to the 250 m resolution MODIS inputs, yielding 25 results per pixel. These 25 results were then ranked, per pixel, and the median value selected as the final probability. Global total FAS PSD cropland was matched with a threshold of 43%. Given that the training sample was randomly drawn from a data set that covered the vast majority of the study area, it would be expected that the threshold be near 50%. Per-country matching thresholds were then derived to create the final binary crop indicator map. Table 4 shows the matching area threshold for each country/region, as well as the calculated and FAS database cropland areas.

When comparing the global versus national threshold matching approaches, the global probability threshold significantly underestimates crop area in India (global threshold area: 109 million hectares; FAS area: 138 million hectares), while overestimating in Canada (global: 37 million hectares; FAS: 23 million hectares), the United States (global: 117 million hectares; FAS: 98 million hectares) and the European Union (global: 146 million hectares; FAS: 69 million hectares). Figure 1 compares the global and national matching threshold area totals per region/region. A spatial comparison of the global versus national matching area estimates is illustrated in Figure 2. In this figure, the global matching probability threshold of 43% is shown in red and orange, where orange represents areas the global threshold overestimates (areas not counted in the national matching threshold). Regions where the global threshold overestimates national cropland extent include Europe and Australia. Green represents areas of national underestimation when using the global threshold. Areas where the global threshold underestimates cropland extent include West Africa and Southeast Asia. Blue regions are areas of indicated crop probability less than the global threshold of 43% that are likewise not included in the per country thresholds.

Table 4. MODIS cropland probability layer converted to cropland mask by employing a per country area-matching threshold using USDA FAS PSD country data. Calculated areas shown in this table for regions are derived from summing the per country totals. Matching thresholds for regions were calculated independently to ease intercomparison.

Region *Country/Regions*	*Matching Threshold*	*Calculated Area* *(hectares)*	*FAS PSD Area* *(hectares)*
India	35	139,841,931	138,331,222
China	41	113,216,074	114,264,444
United States	49	100,291,610	97,792,333
Russia	43	49,301,727	48,396,333
Brazil	37	41,099,217	41,453,222
Argentina	42	26,882,240	26,711,333
Canada	64	22,501,593	22,627,556
Australia	75	20,308,184	20,363,000
Africa	30	112,756,008	110,901,444
Europe	63	92,203,407	92,959,111
Central Asia	47	78,435,859	77,582,777
South / East Asia	20	76,919,435	77,462,666
Latin America	38	24,002,450	25,023,444

Figure 1. Comparison of country/regional estimates of cropland extent derived from a single global matching threshold versus per country matching thresholds.
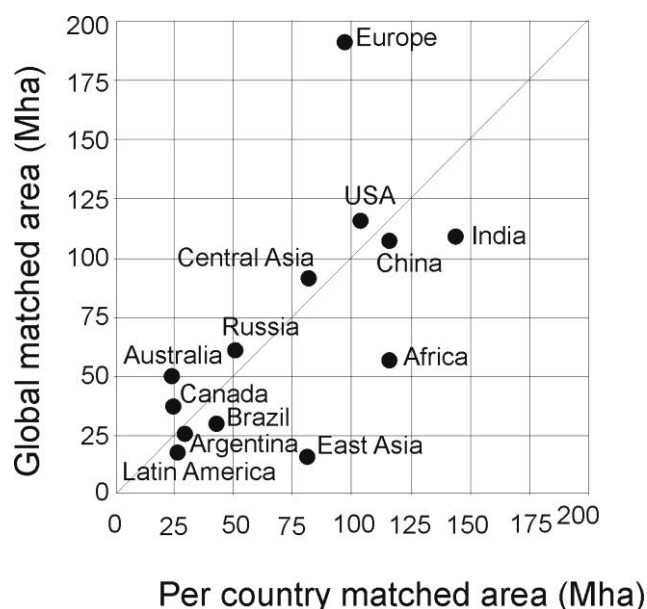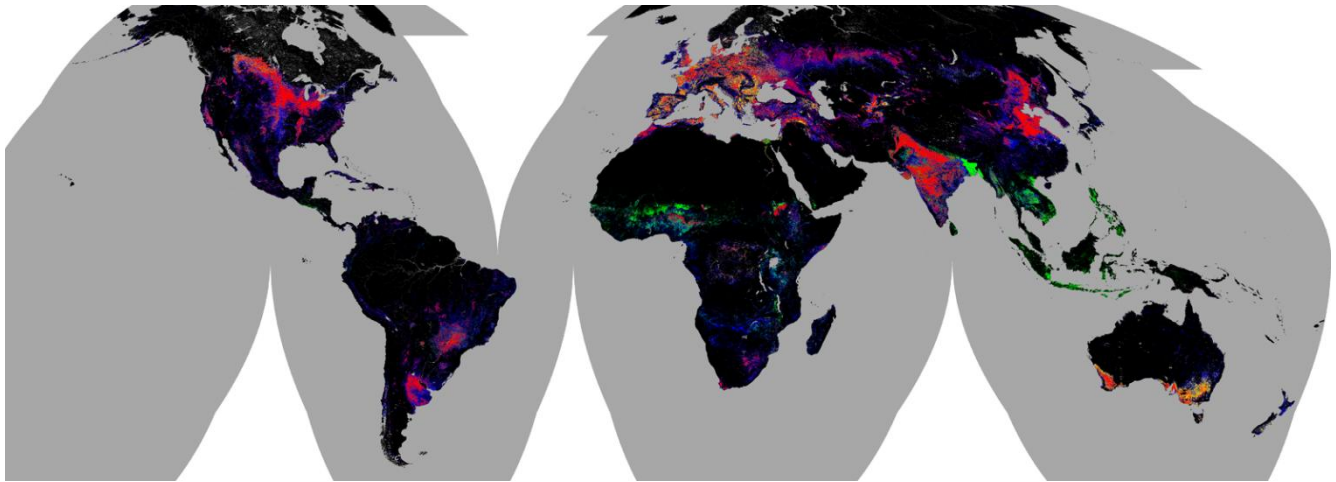
Figure 2. Comparison of cropland extent derived from a single global matching threshold versus per country matching thresholds where red equals global/national cropland (>=43%), orange equals global/not national cropland (>=43%), green equals not global/national cropland (<43%), and blue equals not global/not national cropland (>0% and <43%). Black equals no indicated cropland.

The per country matching product was used as the final map and is shown in Figure 3. In order to better assess the spatial fidelity of the product, it was compared with 5 satellite-derived global land cover products. Global agreement between these heritage data sets is illustrated in Figure 4. Universal agreement between the global cropland extent maps is found in a number of regions, including the U.S. Corn Belt, the Gangetic Plain, the North China Plain and other smaller, but intensively farmed areas such as the Nile Delta. The degree of agreement between the heritage data sets was used as a basis for evaluating the spatial fidelity of the new MODIS 250 m cropland map. For each agreement class in Figure 4, regional masks were created and compared to the final MODIS cropland layer. For each agreement mask (5 out of 5, 4 out of 5, 3 out of 5, 2 out of 5 and 1 out of 5), user's and producer's accuracies were calculated per region per agreement mask. The point at which the cropland user's and producer's accuracies match indicates equal errors of commission and omission, a basis for comparatively assessing the various regions. Figure 5 illustrates the results of this intercomparison. For each region, the probability thresholds from Table 3 matching the MODIS and FAS PSD cropland area estimates are also shown. Lower probability threshold values indicate lower confidence in the cropland labels for a given region. Additionally, the percentage where user's and producer's accuracies match is plotted in relation to the degree of agreement between the 5 heritage cropland map products. In this plot, only the user's and producer's accuracies per ancillary map comparison for the United States are shown. The modeled intersection of the user's and producer's accuracies for the U.S. is found at 63% between the 3 and 4 map-agreement masks.

Figure 3. Global MODIS 250 m cropland product based on per country matching thresholds. Final cropland extent is shown in red, with probability values outside of final mask shown in green.
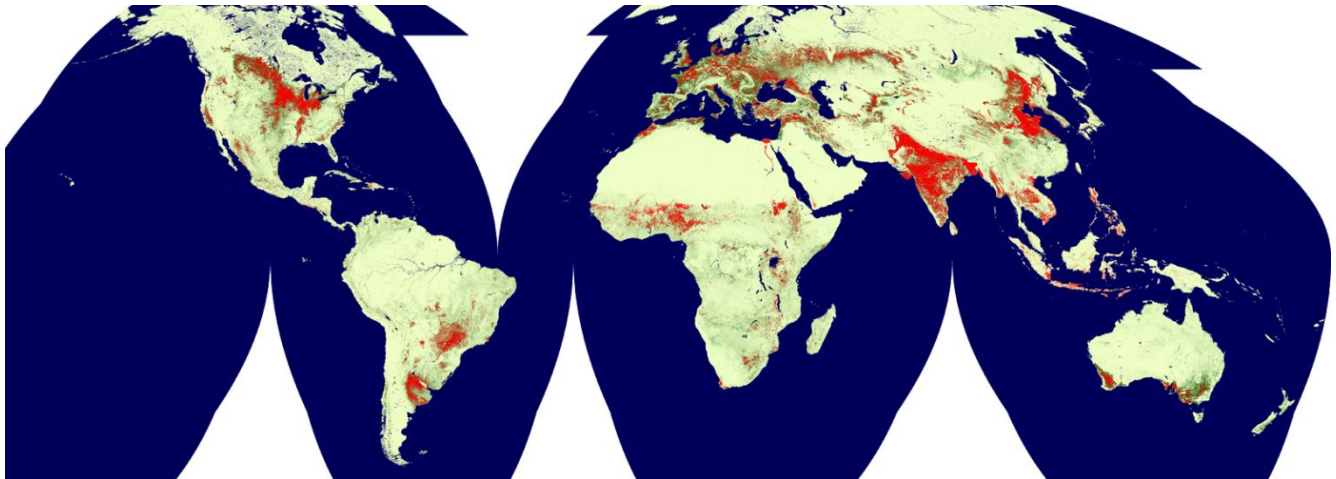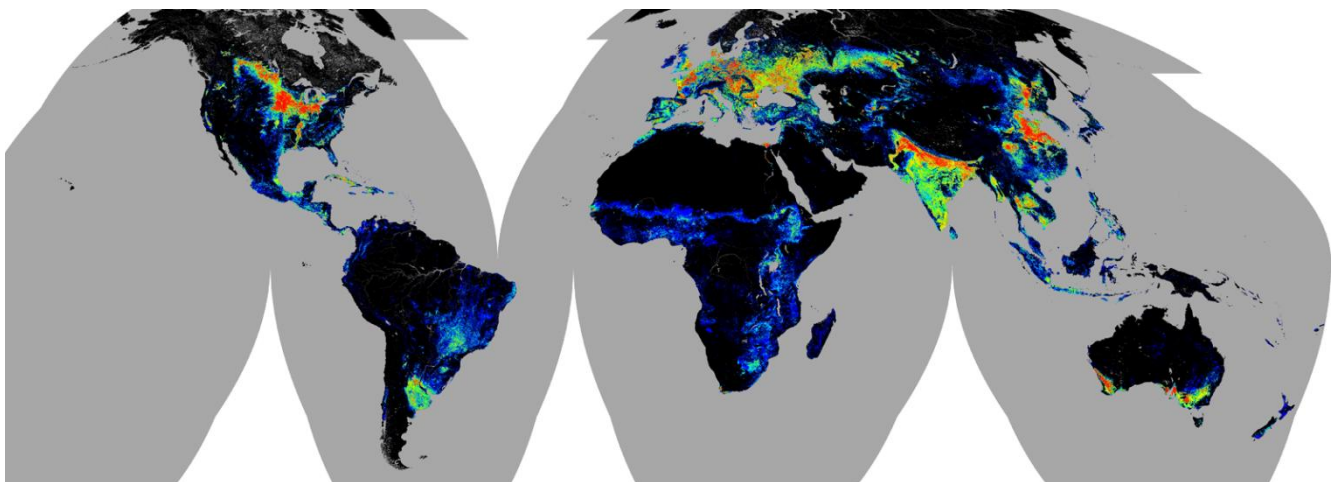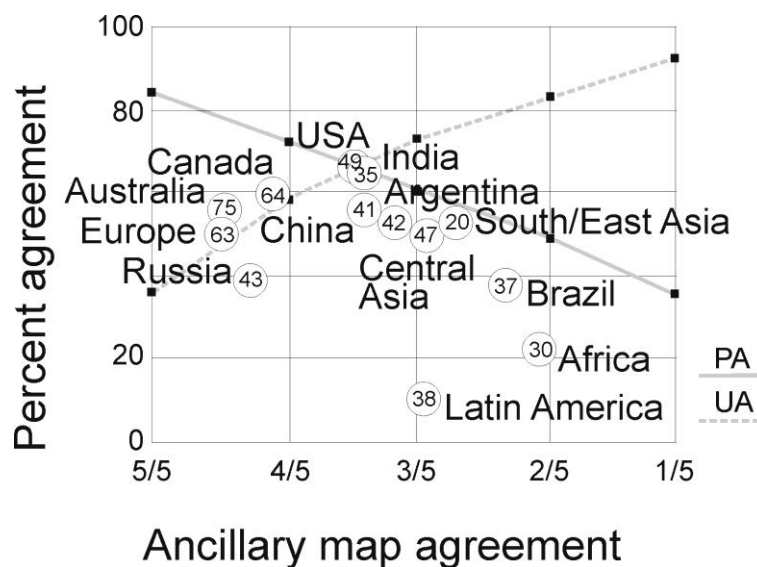

Figure 4. Agreement amongst five heritage cropland data sets, where red equals 5 out of 5 agreement, orange equals 4 out 5, green 3 out of 5, cyan 2 out of 5, and blue 1 out of 5. Each zone of agreement was converted to a mask for intercomparison with the MODIS 250 m cropland layer.

For regions that have (1) relatively higher FAS PSD matching thresholds, (2) higher matching user's and producer's accuracies for (3) agreement masks that represent a majority of the ancillary data sets, confidence in the product is correspondingly higher. From Figure 5, countries/regions such as Canada, Australia and Europe exhibit relatively unambiguous cropland signatures that are readily captured by the MODIS 250m product. On the other hand, Latin America (less Brazil and Argentina), Africa and Brazil have low matching thresholds and the lowest user's/producer's accuracies derived where less than 3 of the ancillary maps agree. The lack of agreement among the heritage maps indicates a lower overall ability to map these areas with coarser resolution data. The 250 m MODIS map has a relatively low matching threshold, also an indication of mapping limitations. South/East Asia has a low FAS PSD matching threshold, but one that has some spatial fidelity as shown by its relatively higher user's/producer's accuracy value. In this case, adjusting the probability layer per

country appears to be beneficial as the lower threshold captures cropland distributions that are to a degree congruent with heritage map layers. A globally applied threshold would not quantify the majority of South/East Asian cropland. Russia is an anomalous case, where the FAS PSD area is considerably less than that mapped by the heritage cropland maps. This in effect caps the producer's accuracy for Russia. One possible explanation for the lower FAS PSD area estimate for the 2000's compared to the 90's, which is the period of data inputs for three of the five heritage comparison products, is abandonment following the break-up of the Soviet Union [31]. It is estimated that 30–40% of Russian cropland area was lost from 1992 to 2005 [32].

> **Figure 5.** Modeled matching user's and producer's accuracies (percent agreement axis) for regions of common agreement amongst five heritage cropland data layers (ancillary map agreement axis). Only the data points for the USA are shown to illustrate the variation of commission and omission errors. Each data point has the probability threshold value independently derived using the FAS PSD database to create the MODIS 250 m cropland layer.

While the heritage layers cannot be taken as truth, this intercomparison points out some product strengths. First, agro-industrial regions, and intensive agricultural zones in general, are more clearly quantified. Coarse resolution cropland maps concur on the locations of the major production areas of the globe. Agreement is less certain in the following regions which also have lower FAS PSD matching thresholds: Russia, Brazil, Africa and Latin America (less Brazil and Argentina). While this does not necessarily mean that the new map is inaccurate in these regions, there is no clear verdict as to its spatial fidelity.

Subsets of the heritage map agreement zones and the MODIS 250m cropland layer are shown in Figure 6 for the U.S. Corn Belt, Argentina/Uruguay/Brazil border, North China Plain, and the India/Pakistan/China border. Two positive attributes of the new product are evident. First, the 250 m data afford improved depiction of landscape heterogeneity. Settlement patterns are evident in all examples, and smaller agricultural centers, such as Mendoza in western Argentina (Figure 6d) are clearly delineated. Second, the value of the probability layer is evident. Red delineates the area that

matched the FAS PSD cropland area estimates. Probability of cropland class membership below the matching threshold is shown in green. For example, an adaptable use of the threshold could be used to delineate more of the rice growing regions along the lower Yangtze River in China (Figure 6f).

> **Figure 6.** For (a) U.S. Corn Belt, (c) Argentina/Uruguay/Brazil, (e), North China Plain, and (g) Pakistan/India/China, agreement amongst five heritage cropland data sets, where red equals 5 out of 5 agreement, orange equals 4 out 5, green 3 out of 5, cyan 2 out of 5, and blue 1 out of 5. For the same respective regions in (b), (d), (f), (h), MODIS 250 m global cropland product based on per country matching thresholds. Final cropland extent is shown in red, with probability values outside of final mask shown in green. Each subset is 1,843 km by 1,843 km in size.
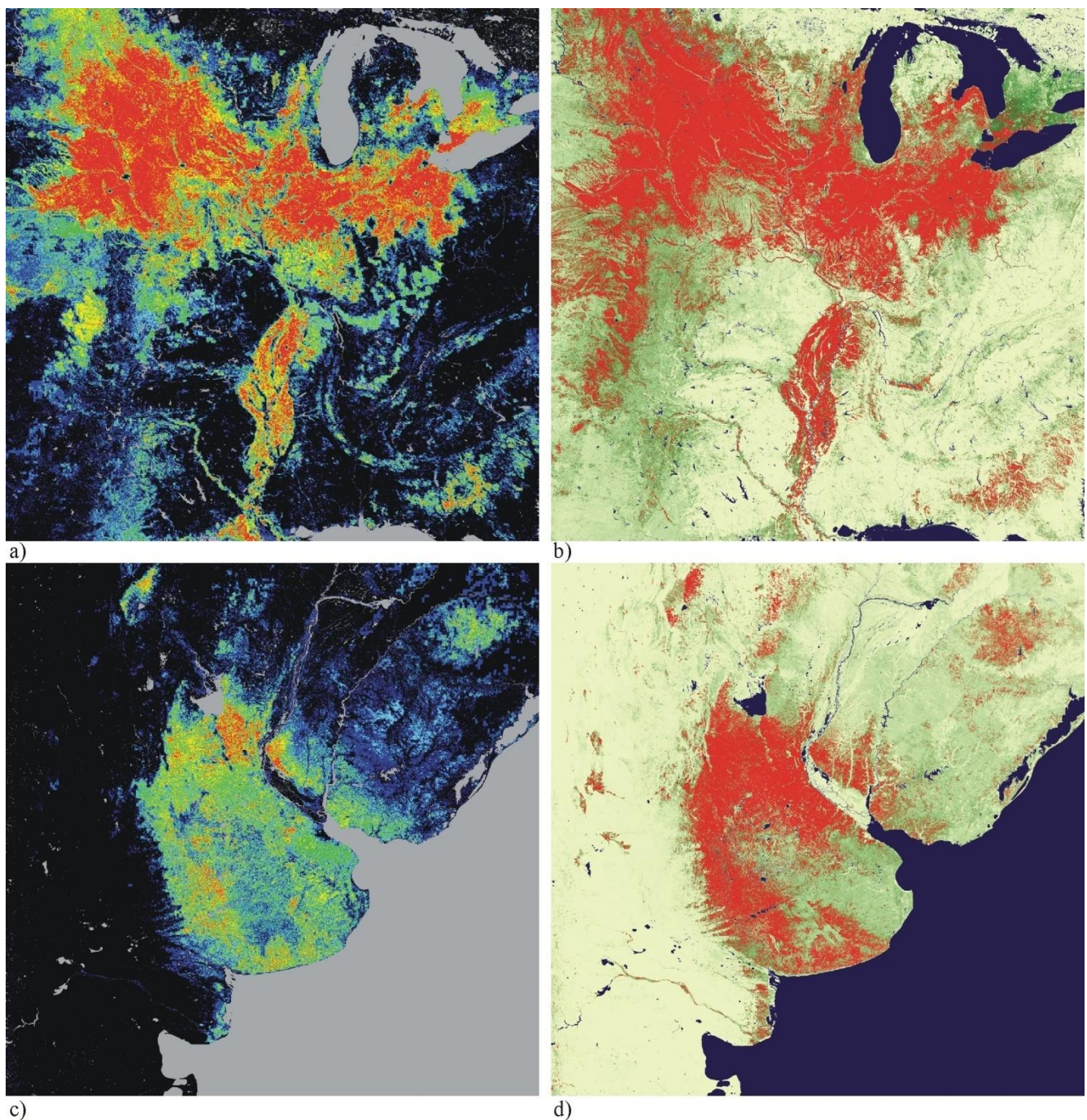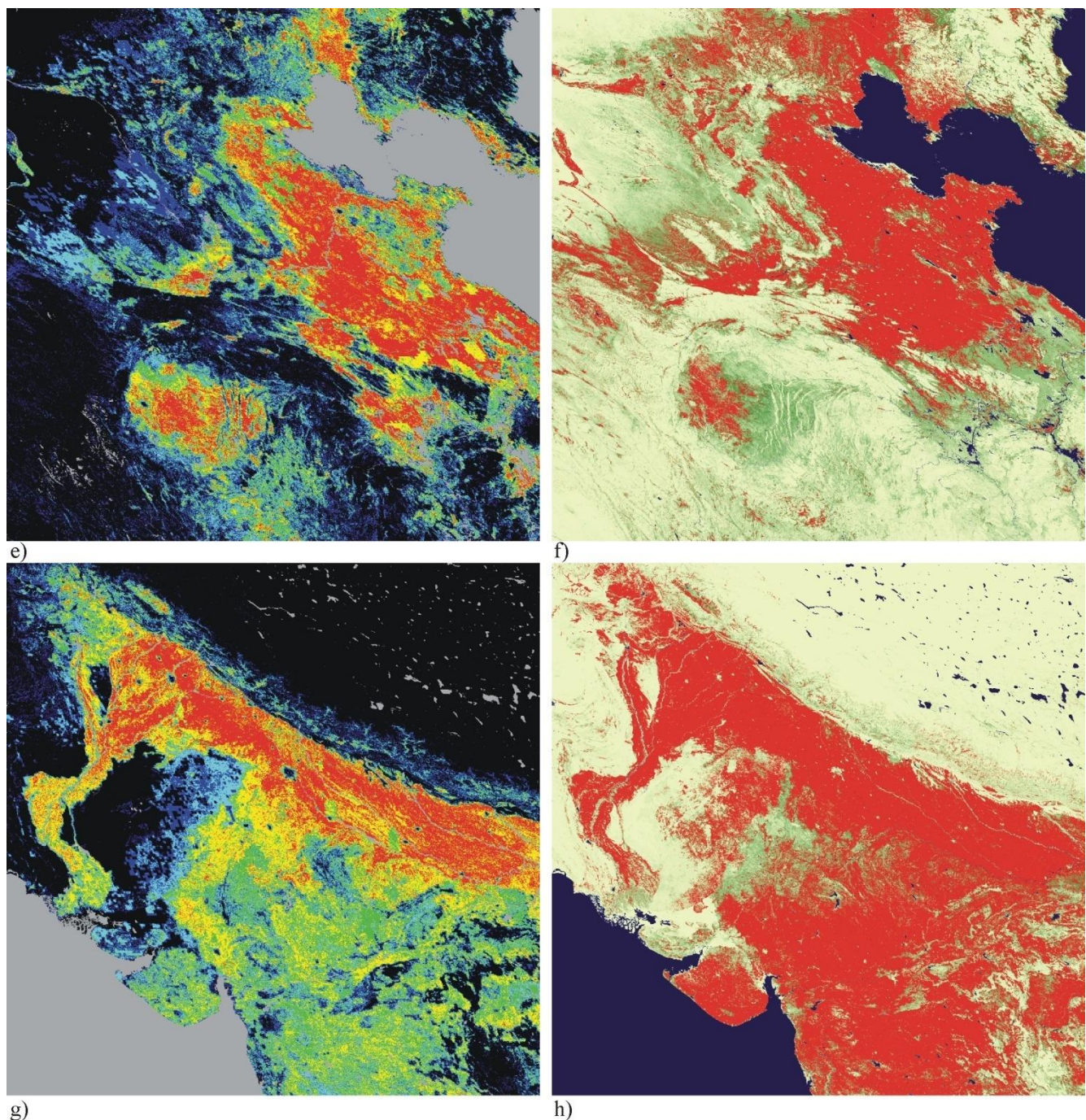
Figure 6. *Cont.*

e)

f)

g)

h)

In addition to geographic regions, countries were also grouped into regions representing one of four dominant crop types: corn, rice, soybeans and wheat and analyzed with respect to the cropland probability layer. In order to be included in a dominant crop region, a country had to have a minimum of 1 million hectares of production cropland according to FAS, and at least 25% of that cropland must be allocated to a single production crop. Countries with more than one dominant crop were included in all applicable regions. The selected countries represent 52.3% of global corn production, 68.7% of rice production, 58.9% of wheat production and 74.9% of soybean production. Table 5 shows the countries included in each of these dominant crop regions.

Table 5. Countries with at least 25% of cropland allocated to a single crop for the four dominant food production crops.

Corn	Rice	Soybeans	Wheat
Angola	Bangladesh	Argentina	Afghanistan
Benin	Burma	Bolivia	Algeria
Brazil	Cambodia	Brazil	Australia
Colombia	China	Paraguay	Canada
Congo (Kinshasa)	Colombia	United States	European Union
Cote d'Ivoire	Guinea		Egypt
Ethiopia	India		Iran
Ghana	Indonesia		Iraq
Kenya	Japan		Kazakhstan
North Korea	North Korea		Moldova
Malawi	Madagascar		Morocco
Mexico	Nepal		Pakistan
Moldova	Peru		Russia
Mozambique	Philippines		Syria
Nepal	Thailand		Tunisia
Peru	Vietnam		Turkey
Philippines	South Korea		Turkmenistan
Serbia			Ukraine
South Africa			Uzbekistan
Tanzania			
Togo			
Uganda			
United States			
Zambia			
Zimbabwe			

Table 6 shows the median cropland probability for areas with differing levels of comparative agreement among the heritage cropland data sets. No median probability reflects a single crop type, as all countries have some level of crop diversification. However, consistently higher cropland probabilities are found for corn and soybean-dominant countries, reflecting a more generic spectral-temporal signature for these crop types as compared to rice and wheat, which are less unambiguously mapped. However, if crop dominance is defined more stringently, for example if one crop accounts for over 60% of national/regional crop production, wheat regions are more readily identified than rice-growing areas. Ten countries matched this definition for rice dominance, the largest of which are Indonesia, Bangladesh and Burma. Six countries met the definition for wheat dominance, the largest being Australia and Kazakhstan. Table 7 shows the comparison between rice and wheat in these exceedingly single-crop dominant countries.

Table 6. MODIS 250 m median crop probability for countries with 25% or more of cropland area in corn, rice, soybeans or wheat, as a function of heritage global cropland map agreement.

Level of Agreement	Corn	Rice	Soybeans	Wheat
5 of 5	85.89	58.02	86.75	62.65
At Least 4 of 5	66.93	46.58	69.24	55.52
At Least 3 of 5	50.57	40.50	55.51	49.65
At Least 2 of 5	33.10	34.22	41.70	42.92
At Least 1 of 5	19.91	25.26	27.50	32.78

Table 7. MODIS 250 m median crop probability for countries with 60% or more of cropland area in rice, or wheat as a function of function of heritage global cropland map agreement.

Level of Agreement	Rice	Wheat
5 of 5	20.84	73.79
At Least 4 of 5	19.93	57.19
At Least 3 of 5	15.69	41.84
At Least 2 of 5	10.49	26.25
At Least 1 of 5	4.90	10.29

The primary driver behind the significant decline in rice performance between the 25% and 60% dominance thresholds is the presence of India and China on the first list and their absence in the second. Though they are by far the world's two largest rice producers, both also have fairly diverse types of crop production with both being large producers of corn, wheat and soybeans. The mixture of crops within China and India contributes to an overstatement of the strength of rice cropland characterization in Table 5.

4. Conclusion

The benefits of the global probability layer are evident in the production of the final MODIS 250 m cropland layer. Since cropland spectral phenologies vary significantly at the global scale, it is difficult to create a generic set of decision rules that can account for such variation. By using a statistical database, in this case the FAS PSD, a per country cropland extent map was derived from the cropland probability layer. Such an approach works as long as within any sub-region (nation), actual croplands are assigned relatively higher probabilities than non-croplands. This appears to have occurred in South/East Asia, where a low probability threshold yielded a degree of concurrence with heritage cropland data sets. Such a flexible procedure is not needed for other, more stable land cover targets such as forests, where spectral consistency through space and time is greater.

The cropland extent map presented here will be incorporated into the USDA Global agriculture monitoring project (GLAM) as a reference for FAS analysts in assessing crop condition using MODIS NDVI phenological profiles [33]. However, static depictions of global cropland extent such as the one presented in this study can be improved upon. Annual estimates of cropland and crop type are possible with MODIS data [34,35], and targeting specific crop types rather than cropland areas in aggregate

could lead to improved performance, especially in rice dominant areas. This study indicates that MODIS is appropriate for mapping the primary broadleaf crop production regions of the globe. Annual MODIS crop indicator maps for these regions could be used in novel ways, such as to stratify crop production regions for targeted sampling of higher-resolution data sets for area estimation. Moving from static to dynamic cropland monitoring applications is the next step in global cropland mapping with MODIS data.

Acknowledgements

References

1. Boatwright, G.O.; Whitehead, V.S. Early warning and crop condition assessment research. *IEEE Trans. Geosci. Remote Sens*. **1986**, *24*, 54-64.
2. Group on Earth Observations. *Report from the Workshop on Developing a Strategy for Global Agricultural Monitoring in the framework of Group on Earth Observations (GEO), 16-18 July 2007, FAO, Rome*; GEO: Geneva, Switzerland, 2007.
3. National Agricultural Statistics Service. *Cropland Data Layer*; USDA-NASS: Washington, DC, USA, 2010; Available online: www.nass.usda.gov/research/Cropland/SARS1a.htm (accessed on 6 January 2010).
4. Wolfe, R.; Roy, D.; Vermote, E. The MODIS land data storage, gridding and compositing methodology: L2 Grid. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1324-1338.
5. Vermote, E.F.; Saleous, N.El; Justice, C.O.; Kaufman, Y.J.; Privette, J.L.; Remer, L.; Roger, J.C.; Tanre, D. Atmospheric correction of visible to middle-infrared EOS-MODIS data over land surfaces: Background, operational algorithm and validation. *J. Geophys. Res.* **1997**, *102*, 17131-17141.
6. Ramankutty, N.; Foley, J.A. Characterizing patterns of global land use: An analysis of global croplands data. *Glob. Biogeochem. Cycles* **1998**, *12*, 667-685.
7. Ramankutty, N.; Evan, A.T.; Monfreda, C.; Foley, J.A. Farming the Planet: Geographic distribution of global agriculture in the year 2000. *Glob. Biogeochem. Cycles* **2008**, *22*, B1003.
8. Loveland, T.R.; Reed, B.C.; Brown, J.F.; Ohlen, D.O.; Zhu, Z.; Yang, L.; Merchant, J.W. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1303-1330.
9. Hansen, M.C.; DeFries, R.S.; Townshend, J.R.G.; Sohlberg, R. Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sens.* **2000**, *21*, 1331-1364.
10. Friedl, M.A.; McIver, D. K.; Hodges, J.C.F.; Zhang, X.Y.; Muchoney, D.; Strahler, A.H.; Woodcock, C.E.; Gopal, S.; Schneider, A.; Cooper, A.; Baccini, A.; Gao, F.; Schaaf, C. Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens. Environ.* **2002**, *83*, 287-302.

11. Bartholomé, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959-1977.

12. Arino, O.; Leroy, M.; Ranera, F.; Gross, D.; Bicheron, P.; Nino, F.; Brockman, C.; Defourny, P.; Vancutsem, C.; Achard, F.; Durieux, L.; Bourg, L.; Latham, J.; Di Gregorio, A.; Witt, R.; Herold, M.; Sambale, J.; Plummer, S.; Weber, J-L.; Goryl, P.; Houghton. N. GlobCover—A global land cover service with MERIS. In *Proceedings of 'Envisat Symposium 2007'*, Montreux, Switzerland, April 2007.

13. Thenkabail, P.S.; Biradar C.M.; Noojipady, P.; Dheeravath, V.; Li, Y.J.; Velpuri, M.; Gumma, M.; Reddy, G.P.O.; Turral, H.; Cai, X.L.; Vithanage, J.; Schull, M.; Dutta, R. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. *Int. J. Remote Sens.* **2009**, *30*, 3679-3733..

14. Chandrashekhar, M.; Thenkabail, P.S.; Noojipady, P.; Li, Y.; Dheeravath, V.; Turral, H.; Velpuri, M.; Gumma, M.K.; Gangalakunta, O.B.P.; Cai, X.L.; Xiao, X.; Schull, M.A.; Alankara, R.D.; Gunasinghe, S.; Mohideen, S. A global map of rainfed cropland areas (GMRCA) at the end of the last millennium using remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 114-129.

15. Tucker, C.J.; Grant, D.; Dykstra, J. NASA's global orthorectified Landsat data set. *Photogramm. Eng. Remote Sensing* **2004**, *70*, 313-322.

16. Africover. *Land Cover Mapping Based on Satellite Remote Sensing*; Food and Agriculture Organization of the United Nations: Roma, Italy, 2002; Available online: http://www.africover.org/download/documents/Short_Project_description_en.pdf (accessed on 1 March 2010).

17. Homer, C.; Dewitz, J.; Fry, J.; Coan, M.; Hossain, N.; Larson, C.; Herold, N.; McKerrow, A.; VanDriel, J.N.; Wickham, J. Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogramm. Eng. Remote Sensing* **2007**, *73*, 337-341.

18. Fisette, T.; Chenier, R.; Maloley, M.; Gasser, PY.; Huffman, T.; White, L.; Ogston, R.; Elgarawany, A. Methodology for a Canadian agricultural land cover classification. In *Proceedings of 1st International Conference on Object-based Image Analysis*, Salzburg, Austria, July 2006; In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2006; Volume XXXVI 4/C42.

19. Fairbanks, D.H.K.; Thompson, M.W.; Vink, D.E.; Newby, T.S.; van den Berg, H.M.; Everard, D.A. The South African land-cover characteristics database: a synopsis of the landscape. *South Afr. J. Sci.* **2000**, *96*, 69-82.

20. Buttner, G.; Feranec J.; Jaffrain, G.; Mari, L.; Maucha, G.; Soukup, T. The CORINE Land Cover 2000 Project. *EARSeL eProceedings* **2004**, *3*, 331-346.

21. Vermote, E.F.; Saleous, N.El; Justice, C.O. Atmospheric correction of MODIS data in the visible to middle infrared: First results. *Remote Sens. Environ.* **2002**, *83*, 97-111.

22. Carroll, M.; Townshend, J.; Hansen, M.; DiMiceli, C.; Sohlberg, R.; Wurster, K. Vegetative cover conversion and vegetation continuous Fields. In *Land Remote Sensing and Global Environmental Change*; NASA's Earth Observing System and the Science of ASTER and MODIS Series; Ramachandran, B., Justice, C.O., Abrams, M., Eds.; Springer-Verlag: New York, NY, USA, 2010.

23. Holben, B. Characteristics of maximum-value composite images from temporal AVHRR data. *Int. J. Remote Sens.* **1986**, *7*, 1417-1434.

24. Hansen, M.C.; DeFries, R.S.; Townshend, J.R.G.; Carroll, M.; Dimiceli, C.; Sohlberg, R.A. Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions* **2003**, *7*, 1-15.

25. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S-Plus*; Springer-Verlag: New York, NY, USA, 1994.

26. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123-140.

27. Reynolds, C.A. *Input Data Sources, Climate Normals, Crop Models, and Data Extraction Routines Utilized by PECAD*; USDA: Washington, DC, USA, 2001; Available online: www.pecad.fas.usda.gov/cropexplorer/datasources.cfm (accessed on 14 October 2009).

28. FAOSTAT. *FAO Statistical Databases*; Food and Agricultural Organization of the United Nations: Rome, Italy, 2001.

29. Galford, G.L.; Mustard, J.F.; Melillo, J.; Gendrin, A.; Cerri, C.C.; Cerri, E.P. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sens. Environ.* **2008**, *112*, 576-587.

30. Carroll, M.L.; Townshend, J.R.; DiMiceli, C.M.; Noojipady, P.; Sohlberg, R.A. A new global rater water mask at 250m resolution. *Int. J. Dig. Earth* **2009**, *2*, 291-308.

31. Ioffe, G.; Nefedova, T.; Zaslavsky, I. From spatial continuity to fragmentation: The case of Russian farming. *Ann. Assoc. Amer. Geographers* **2004**, *94*, 913-943.

32. ROSSTAT. *Regions of Russia: Social and Economic Indicators*; Russian Federal State Statistics Service: Moscow, Russian, 2008.

33. Becker-Reshef, I.; Justice, C.O.; Sullivan, M.; Vermote, E.; Tucker, C.; Anyamba, A.; Small, J.; Pak, E.; Masouka, E.; Schmaltz, J.; Hansen, M.; Pittman, K.; Birkett, C.; Willians, D.; Reynolds, C.; Doorn, B. Monitoring global croplands with coarse resolution earth observations: The Global Agriculture Monitoring (GLAM) project. **2010**, *2*, 1589-1609.

34. Wardlow, B.D.; Egbert, S.L.; Kastens, J.H. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sens. Environ.* **2007**, *108*, 290-310.

35. Chang, J.; Hansen, M.C.; Pittman, K.; Carroll, M.; DiMiceli, C. Corn and soybean mapping in the United States using MODIS time-series data sets. *Agronomy J.* **2007**, *99*, 1654-1664.