Remote Sens. **2012**, *4*, 2890-2918; doi:10.3390/rs4102890

OPEN ACCESS

Remote Sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

An Automated Cropland Classification Algorithm (ACCA) for Tajikistan by Combining Landsat, MODIS, and Secondary Data

Prasad S. Thenkabail [1,*] and Zhuoting Wu [1,2]

[1] Flagstaff Science Center, US Geological Survey, Flagstaff, AZ 86001, USA

[2] Merriam-Powell Center for Environmental Research, Northern Arizona University, Flagstaff, AZ 86001, USA; E-Mail: zw27@nau.edu

* Author to whom correspondence should be addressed; E-Mail: pthenkabail@usgs.gov; Tel.: +1-928-556-7221; Fax: +1-928-556-7169.

Received: 23 July 2012; in revised form: 28 August 2012 / Accepted: 12 September 2012 / Published: 25 September 2012

Abstract: The overarching goal of this research was to develop and demonstrate an automated Cropland Classification Algorithm (ACCA) that will rapidly, routinely, and accurately classify agricultural cropland extent, areas, and characteristics (e.g., irrigated *vs.* rainfed) over large areas such as a country or a region through combination of multi-sensor remote sensing and secondary data. In this research, a rule-based ACCA was conceptualized, developed, and demonstrated for the country of Tajikistan using mega file data cubes (MFDCs) involving data from Landsat Global Land Survey (GLS), Landsat Enhanced Thematic Mapper Plus (ETM+) 30 m, Moderate Resolution Imaging Spectroradiometer (MODIS) 250 m time-series, a suite of secondary data (e.g., elevation, slope, precipitation, temperature), and *in situ* data. First, the process involved producing an accurate reference (or truth) cropland layer (TCL), consisting of cropland extent, areas, and irrigated *vs.* rainfed cropland areas, for the entire country of Tajikistan based on MFDC of year 2005 (MFDC2005). The methods involved in producing TCL included using ISOCLASS clustering, Tasseled Cap bi-spectral plots, spectro-temporal characteristics from MODIS 250 m monthly normalized difference vegetation index (NDVI) maximum value composites (MVC) time-series, and textural characteristics of higher resolution imagery. The TCL statistics accurately matched with the national statistics of Tajikistan for irrigated and rainfed croplands, where about 70% of croplands were irrigated and the rest rainfed. Second, a rule-based ACCA was developed to replicate the TCL accurately (~80% producer's and user's accuracies or within 20% quantity disagreement involving about 10 million Landsat 30 m sized cropland pixels of Tajikistan). Development of ACCA was an iterative process involving series of rules that are coded, refined, tweaked, and re-coded till ACCA derived croplands (ACLs) match accurately with TCLs. Third, the ACCA derived cropland layers of Tajikistan were produced for year 2005 (ACL2005), same year as the year used for developing ACCA, using MFDC2005. Fourth, TCL for year 2010 (TCL2010), an independent year, was produced using MFDC2010 using the same methods and approaches as the one used to produce TCL2005. Fifth, the ACCA was applied on MFDC2010 to derive ACL2010. The ACLs were then compared with TCLs (ACL2005 *vs.* TCL2005 and ACL2010 *vs.* TCL2010). The resulting accuracies and errors from error matrices involving about 152 million Landsat (30 m) pixels of the country of Tajikistan (of which about 10 million Landsat size, 30 m, cropland pixels) showed an overall accuracy of 99.6% (k_{hat} = 0.97) for ACL2005 *vs.* TCL2005. For the 3 classes (irrigated, rainfed, and others) mapped in ACL2005, the producer's accuracy was >86.4% and users accuracy was >93.6%. For ACL2010 *vs.* TCL2010, the error matrix showed an overall accuracy on 96.2% (k_{hat} = 0.96). For the 3 classes (irrigated, rainfed, and others) mapped in ACL2010, the producer's and user's accuracies for the irrigated areas were ≥82.9%. Any intermixing was overwhelmingly between irrigated and rainfed croplands, indicating that croplands (irrigated plus rainfed areas) as well as irrigated areas were mapped with high levels of accuracies (~90% or higher) even for the independent year. The ACL2005 and ACL2010, each, were produced using ACCA algorithm in ~30 min using a Dell Precision desktop T7400 computer for the entire country of Tajikistan once the MFDCs for the years were ready. The ACCA algorithm for Tajikistan is made available through US Geological Survey's ScienceBase: http://www.sciencebase.gov/catalog/folder/4f79f1b7e4b0009bd827f548 or at: https://powellcenter.usgs.gov/globalcroplandwater/content/models-algorithms. The research contributes to the efforts of global food security through research on global croplands and their water use (e.g., https://powellcenter.usgs.gov/globalcroplandwater/). The above results clearly demonstrated the ability of a rule-based ACCA to rapidly and accurately produce cropland data layer year after year (hindcast, nowcast, forecast) for the country it was developed using MFDCs that consist of combining multiple sensor data and secondary data. It needs to be noted that the ACCA is applicable to the area (e.g., country, region) for which it is developed. In this case, ACCA is applicable for the Country of Tajikistan to hindcast, nowcast, and forecast agricultural cropland extent, areas, and irrigated *vs.* rainfed. The same fundamental concept of ACCA applies to other areas of the World where ACCA codes need to be modified to suite the area/region of interest. ACCA can also be expanded to compute other crop characteristics such as crop types, cropping intensities, and phenologies.

Keywords: automated cropland classification algorithm (ACCA); truth or reference cropland layer (TCL); ACCA generated cropland layer (ACL); mega file data cube (MFDC); croplands; remote sensing; MODIS; Landsat; classification accuracy; Tajikistan

1. Introduction

Pressure to grow more and more food in a world where population is increasing by about 80 million per year, nutritional demands are swiftly increasing in the developing world, and climate models are predicting that the hottest seasons on record will become the norm by the end of the century—an outcome that bodes ill for feeding the world [1,2]. Yet, cropland areas have either stagnated or even decreased in many parts of the world due to increasing demand of these lands for alternative uses such as bio-fuels, urbanization, and industrialization. Furthermore, ecological and environmental imperatives such as biodiversity conservation and atmospheric carbon sequestration have put a cap on the possible expansion of cropland areas to other lands such as wetlands and forests. Also, of great importance is to take note that food production requires large quantities of water. Indeed, nearly 70–92% of all human water use currently goes towards food production in most Countries [3]. A combination of above issues requires us to produce precise, rapid, and routine mapping of croplands over large areas such as a country, region, and world which in turn will help us accurately assess, allocate, and save water from agriculture for alternative uses such as industrial, recreational, domestic, and environmental. These cropland and water use products will support food security analysis and planning.

There is growing literature on cropland (irrigated and rainfed) mapping across resolutions [4–15]. These methods include: (a) supervised classification methods such as the maximum likelihood classification [16], (b) spectral matching techniques (SMTs) [11], (c) decision tree algorithms [10,17,18], (d) neural network methods[19,20], (e) support vector machine [21], (f) spectral unmixing techniques [22–24], (g) tasseled cap brightness-greenness-wetness [25–28], (h) space-time spiral curves, Change Vector Analysis (CVA)[27], (i) phenology [4,8], (j) fusing climate data with MODIS time-series spectral indices and using algorithms such as decision tree algorithms, image mosaicking using climate data, and sub-pixel calculation of the areas [7,29], and (k) spectral angle mapper [15]. However, most of these methods rely extensively on the human interpretation of spectral signatures, making the process resource-intensive, time-consuming, and difficult to repeat over space and time.

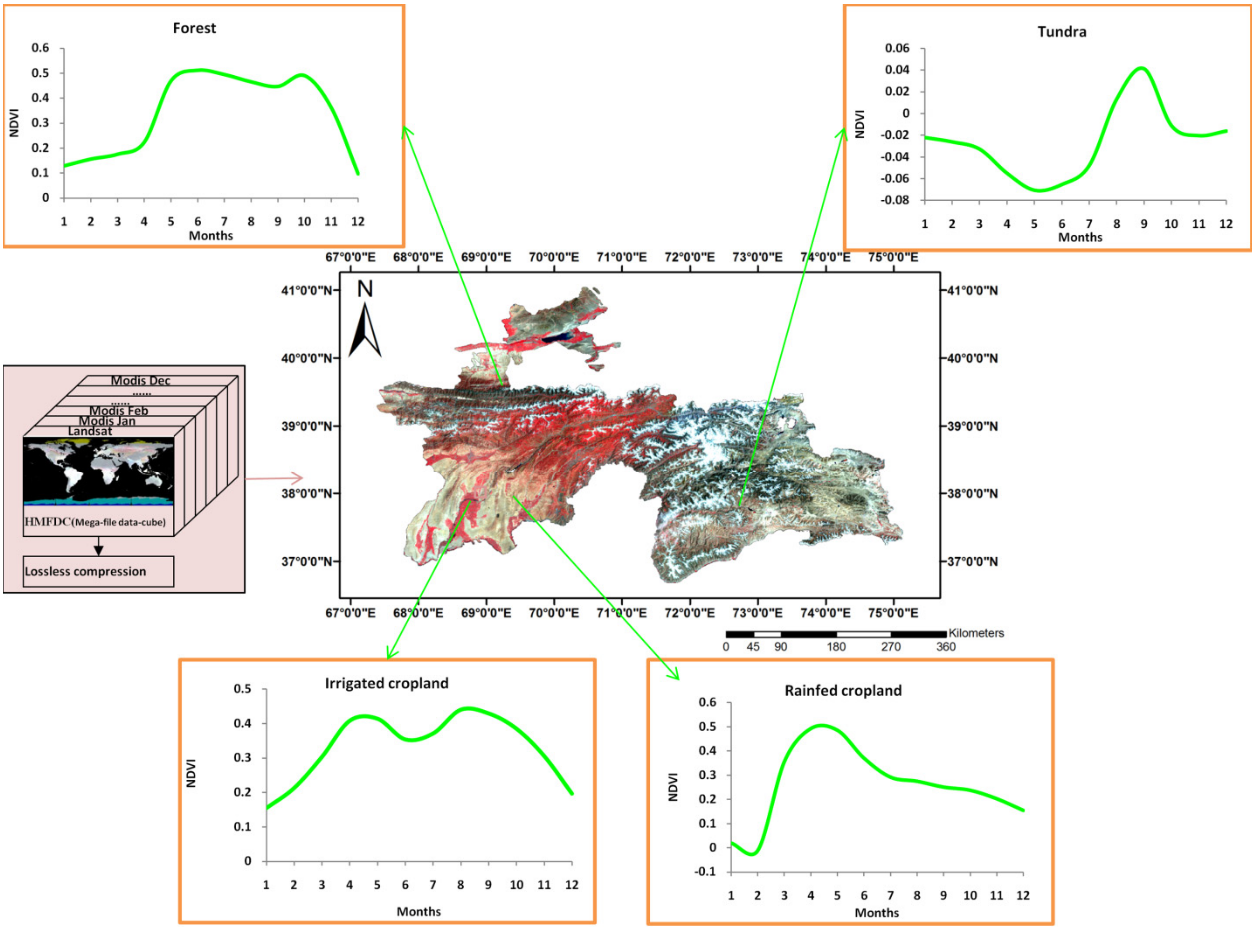
Thereby, the urgent need of the cropland mapping in order to address food and water security scenarios will require the methods to be automated, accurate, and able to provide cropland maps, statistics, and their characteristics (e.g., irrigated *vs.* rainfed, crop types, cropping intensities) rapidly (e.g., producing maps within few hours) year after year (hindcast, nowcast, forecast) over space and time once the MFDCs for the years are ready through automated methods using automated cropland classification algorithms (ACCAs). Fully automated methods do not exist, especially over large areas. The best of existing methods are semi-automated, requiring substantial human interaction, and have large uncertainties when working with independent datasets. These semi-automated methods include: (a) spectral matching techniques (SMTs), (b) ensemble of machine learning algorithms (EMLAs) (e.g., decision trees, neural network), and (c) Classification and Regression Tree (CART). The principle of SMTs [30] is to match the shape, or the magnitude (preferably) both to an ideal or target spectrum (pure class or "end-member"). EMLAs include decision tree algorithms and neural networks [5,31], which are computationally fast to implement. CART is a data mining decision-tree that takes spectral and ancillary data and recursively splits it until ending points or terminal nodes are achieved [32,33]. These methods are powerful, and have shown potential for automation. Nevertheless, the limitations of implementing these algorithms include complexity of methods, inability to demonstrate repeatability of algorithms to produce accurate cropland mapping over time, and the substantial expert interaction required to run the algorithms successfully and accurately when applying them for different time periods such as, for example, for years other than for which the algorithm is developed. Furthermore, the need for new perspectives and concepts for developing simple algorithms, yet powerful and accurate year after year is urgent in the present food security scenario. Given the above background, the overarching goal of this study was to develop an automated cropland classification algorithm (ACCA) that will replicate\reproduce cropland characteristics (e.g., cropland extent, area, geographic specificity, and irrigated *vs.* rainfed using mega file data cube (MFDC) accurately when it is compared with a reference\truth cropland layer (TCL). The goal will lead to accurately map cropland from other land uses and to accurately distinguish between rainfed *vs.* irrigated crop types. For the purpose, the Country of Tajikistan was selected randomly since it was one of the countries of interest for United States Geological Survey (USGS) and United States Agency for International Development (USAID). Two truth\reference cropland layers were produced for Tajikistan: one for year 2005 (TCL2005) and another for the year 2010 (TCL2010). ACCA was developed based on mega file data cube of year 2005 (MFDC2005) which involved multi-sensor remote sensing data and secondary data. The ACCA rules were written based on knowledge base and tweaked till cropland extent, area, geographic specificity of croplands and other required cropland characteristics (e.g., irrigated *vs.* rainfed) in ACL2005 accurately (~80% of producer's and user's accuracies or within 20% quantity disagreement) matched with TCL2005. The strength of ACCA algorithm was tested by applying it on data from independent years to produce cropland extent and their characteristics (e.g., irrigated *vs.* rainfed) year after year (hindcast, nowcast, forecast).

2. Methods

2.1. Study Area

ACCA was developed for Tajikistan in Central Asia (Figure 1). Tajikistan is well suited for ACCA development since the country has minimal outside access, landlocked, about 60% of the country' population depends on agriculture, and irrigation and rainfed agriculture are the backbone of the national economy and livelihood of its people. The total area of Tajikistan is 143,100 Square Kilometers, with elevation ranging from 300 to 7,495 m. The temperature varies between 20–30 °C in spring (March–May) and autumn (September–November), but can exceed 40 °C during summer. The average annual precipitation for most of the republic ranges between 700 and 1,600 mm. The heaviest precipitation falls are at the Fedchenko Glacier, which averages 2,236 mm per year, and the lightest in the eastern Pamirs, which average less than 100 mm per year. Main crops are: cotton, wheat, potatoes, rice, onions, tomatoes, and fruits. Within the United States Famine Early Warning System Network (FEWSNET), Tajikistan is considered as one of the "non-presence" countries, which means that United States does not have an adequate ground presence in the country to understand its systems, especially understanding agriculture. The country is susceptible to famines and is one of the economically poorest nations of the world. In such a scenario, the best and rapid means to gather information for a thorough understanding of food security scenarios is by developing an ACCA applied on fusion of remotely sensed data in order to generate cropland statistics and monitor their progress remotely from space.

Figure 1. Mega-file data cube (MFDC) of Tajikistan using Landsat ETM+, MODIS images, and secondary data as described in Section 2.2. Illustrated for MFDC2005. The time-series NDVI profiles of MODIS for 4 land cover types from typical pixels are illustrated.



2.2. Data Fusion and Mega File Data Cubes (MFDC) Involving MODIS, Landsat, and Secondary Data

Two mega file data cubes (MFDCs), one for year 2005 (MFDC2005) and another for year 2010 (MFDC2010), of Tajikistan were constituted using remote sensing data (Table 1) and other data (Figures 1 and 2). Overall, there are 44 bands of data in each MFDC: MFDC2005 (e.g., Figure 1, Table 1) and MFDC2010 (e.g., Table 1), as explained below. The 44 bands in a single file MFDC2005 for Tajikistan is constituted involving following multiple sources of data, all re-sampled to 30 m resolution. These bands are: (a) Landsat Global Land Survey 2005 (GLS2005) 30 m images for year 2005 (derived from red and near infrared bands: total 2 bands from a single date); (b) MODIS 250 m monthly normalized difference vegetation index (NDVI) maximum value composites (MVCs) [27] (derived from red and near infrared bands: 1 for each month and total 12 bands for the year); (c) MODIS red and near infrared bands (2 bands per month and total 24 bands for the year); (d) Space Shuttle Radar Topographic Mission (SRTM) 90 m elevation and slope (total 2 bands), and

(e) Landsat GLS2005 derived data such as biomass and LAI (inferred from band 4), chlorophyll (band 3), moisture sensitivity (band 5), and thermal emissivity (band 6) (total 4 bands). The MFDC2010 also has 44 bands and was composed using similar approach as described above for MFDC2005, but using images of year 2010. All of these bands were used either in classification or ACCA algorithm development (Section 2.4). The MODIS band 1 and 2 are used as a proxy through computation of NDVI from these bands.

Figure 2. Data layers of Tajikistan in the mega file data cube of independent year 2010 (MFDC2010). Data layers illustrated here are: (**a**) Landsat scaled NDVI, representing biomass and the leaf area index (LAI); (**b**) Landsat band 3 reflectance, representing chlorophyll absorption; (**c**) Landsat band 5 reflectance, representing moisture sensitivity; (**d**) Landsat band 6 digital number (DN), representing thermal emissivity; (**e**) SRTM-derived elevation; (**f**) SRTM-derived slope; and (**g**) MODIS NDVI monthly MVC for the independent year 2010.

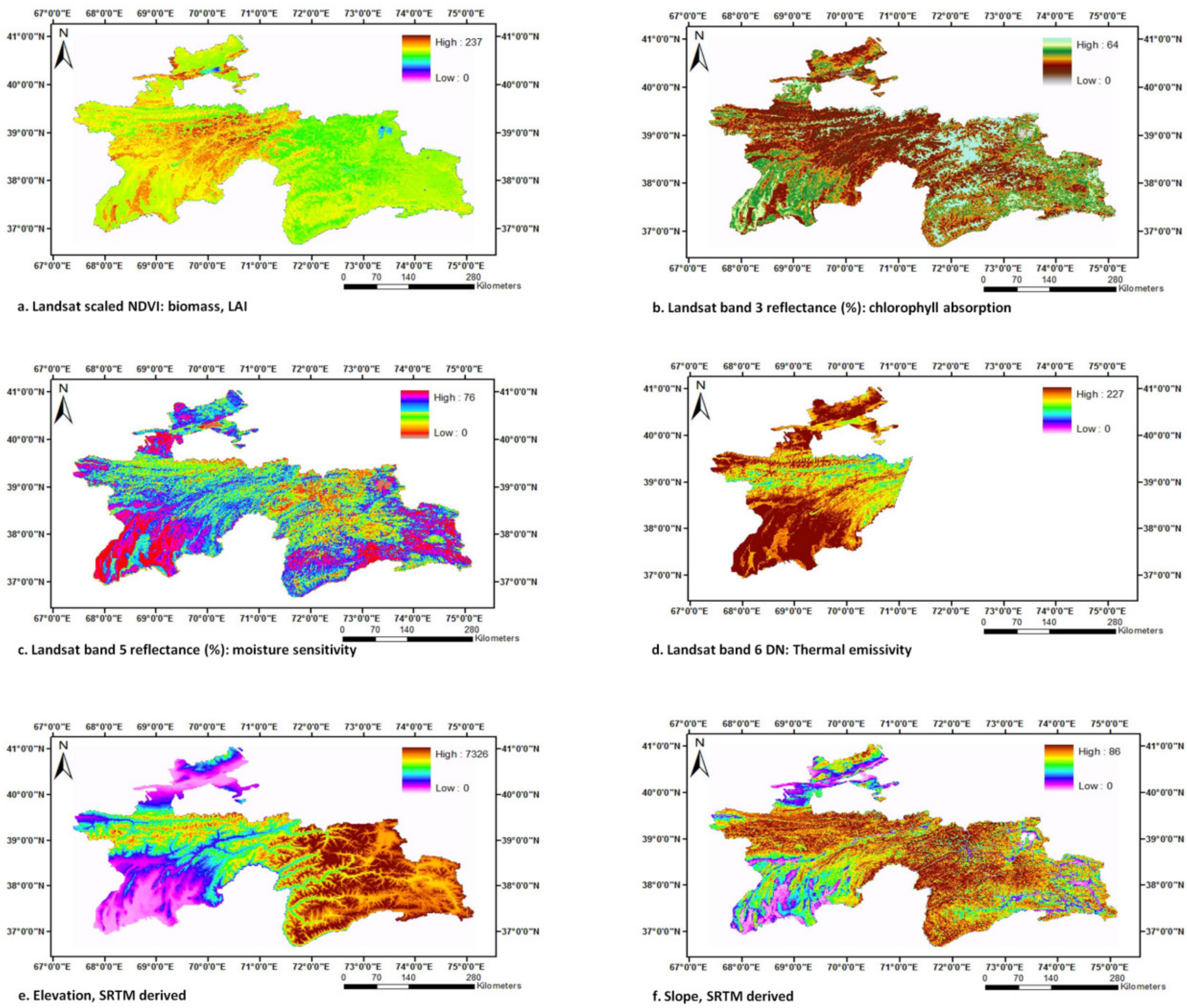


Figure 2. *Cont.*

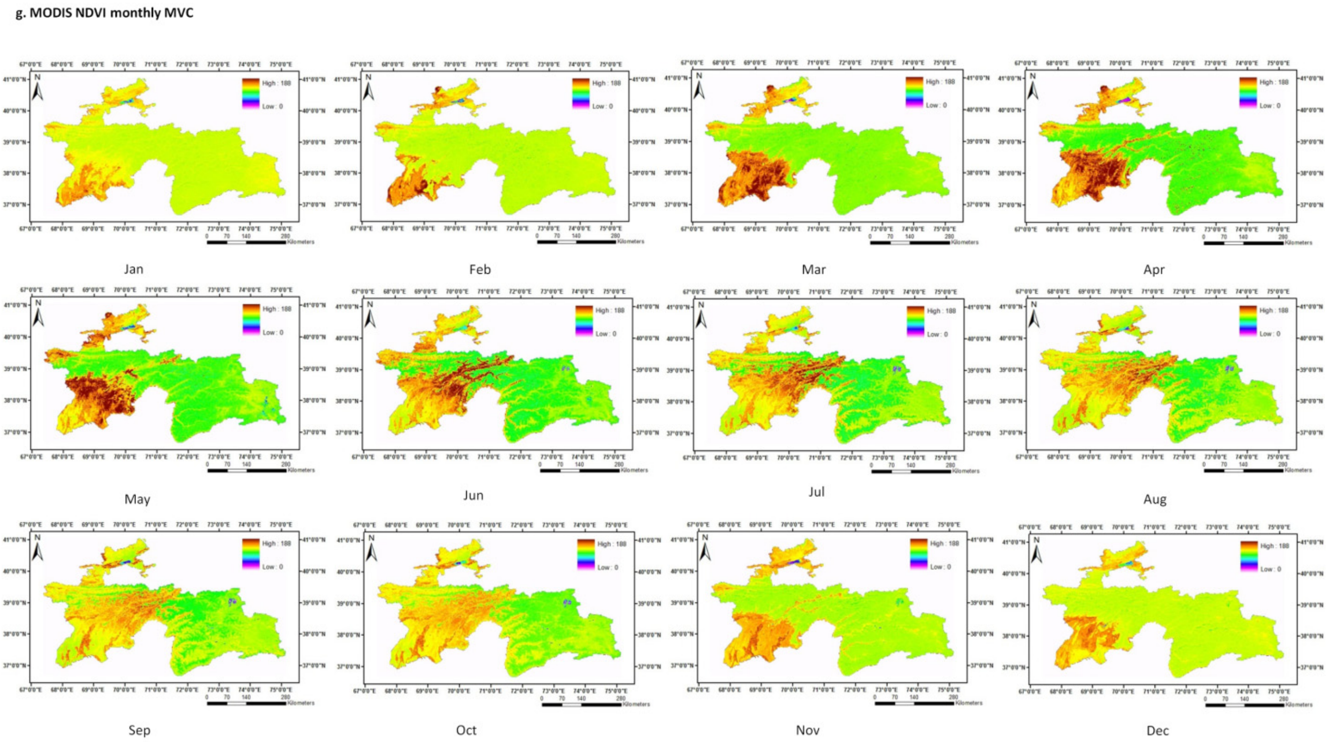


Table 1. Characteristics of image datasets used in creating mega-file data cube for year 2005 (MFDC2005) and for the year 2010 (MFDC2010). The Landsat Global Land Survey (GLS2005) data is fused with MODIS 250m NDVI (derived from band 1 and band 2) monthly maximum value composite (MVC) data for year 2005 (see Figure 1, Section 2.2) in MFDC2005. The Landsat ETM+ 30 m data of 2010 is fused with MODIS 250m NDVI (derived from band 1 and band 2) monthly maximum value composite (MVC) data for year 2010 (see Figure 2, Section 2.2) in MFDC2010.

Satellite	**Sensor**	**Year of Image**	**Spatial**	**Spectral**	**Radiometric**	**Band Range**	**Band Widths**	**Irradiance**	**Data Points**	**Frequency of Revisit**	**Usage**
no unit	no unit	no unit	m	#	bit	μm	μm	$W{\cdot}m^{-2}{\cdot}sr^{-1}{\cdot}\mu m^{-1}$	# per hectares	days	no unit
Landsat	ETM+	GLS2005	30	8	8	0.45–0.52	0.07	1,970	11.1	16	In MFDC2005
						0.52–0.60	0.08	1,843			
						0.63–0.69	0.06	1,555			
						0.75–0.90	0.15	1,047			
						1.55–1.75	0.2	227			
						10.4–12.5	2.1	0			
MODIS	Terra	2005	250	36	12	0.62–0.67	0.05	1,528	0.16	1	In MFDC2005
Landsat	ETM+	2010	30	8	8	Same as GLS2005	Same as GLS2005	Same as GLS2005	11.1	16	In MFDC2010
MODIS	Terra	2010	250	36	12	Same as MODIS2005	Same as MODIS2005	Same as MODIS2005	0.16	1	In MFDC2010

One of the first steps was creating country mosaics and normalizing all the data to surface reflectance. First, GLS2005 and Landsat ETM+ 2010 data were converted to at-sensor reflectance (see [34,35]) using an in-house model written in ERDAS Imagine modeler module, and then to surface reflectance using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) processing system rules [28], followed by mosaicking all the Landsat scenes of GLS2005 and Landsat ETM+ 2010 into a country mosaic (e.g., Figure 1).

2.3. In situ *Data*

In situ field plot data were collected from numerous sources. The main source was very high resolution imagery (VHRI, sub-meter to 5 m, e.g., Figure 3(a)) of the country available from either United States Geological Survey (USGS) or Google Earth sources. USGS has free access to the entire archive of commercial sub-meter to 5 m high resolution imagery (e.g., Quickbird, IKONOS, Geoeye-1) through two US government sources: (a) Commercial Imagery Derived Requirement (CIDR) Database of USGS, and (b) National Geospatial Intelligence Agency (https://warp.nga.mil/). These imagery were used to determine cropland vs. non-croplands, fractional area calculations, detecting irrigation canals, and accuracy assessments. Other sources of in-situ data include nationally produced maps [36], United Nations Food and Agricultural Organization (FAO) reports and maps (http://faostat.fao.org/site/377/default.aspx#ancor) for the country, field photos from specific locations from national sources (e.g., see Figure 3(d)), photos from precise location from the degree confluence project (http://confluence.org/) and geo-wiki (http://www.geo-wiki.org/login.php?ReturnUrl=/index.php, http://waterwiki.net/index.php/Tajikistan). The nationally produced maps [36], and the FAO maps (http://www.fao.org/nr/water/aquastat/maps/index.stm) were used mainly for ensuring that the croplands and their watering source (irrigated or rainfed) mapped by us were geographically consistent, as well as a source of reference. The field photos from different sources provided precise locations of irrigated and rainfed croplands. For example, the VHRI clearly showed irrigated canals which provided additional assistance in detecting irrigated areas along with other datasets mentioned above. From all these data sources, 1,770 data points (Section 2.3, Figures 3 and 4), of irrigated or rainfed croplands from precise locations in Tajikistan were used in producing TCL and assessing its accuracy. The ACCA derived cropland layers (ACL, Section 2.4) were then compared with reference/truth cropland layers (TCLs).

2.4. Algorithm Development Methods

The automated cropland classification algorithm (ACCA) is central to this paper and is described in detail in this section. The goal of ACCA is to rapidly and accurately produce croplands and their characteristics (e.g., irrigated *vs*. rainfed) using fusion of remote sensing and secondary datasets. The ACCA rules are coded with the goal of producing an ACCA generated cropland layer (ACL) that is within 20% quantity disagreement or ~80% user's and producer's accuracies [1,37] when compared with reference/truth cropland layer (TCL).The process of ACCA development involves the following steps:

1. **Generate/obtain truth/reference cropland layer (TCL)**: TCL can be obtained from secondary sources (e.g., from a reliable national institute such as the cropland data layer from United States Department of Agriculture). In absence of such a layer, TCL is generated using remote sensing data (see Section 2.4.1). In this study, TCLs for Tajikistan were produced for the year 2005 (TCL2005) as well as for the year 2010 (TCL2010) as described in detail in Section 2.4.1.

2. **ACCA rules and coding leading to ACCA derived cropland layer (ACL)**: ACCA rules are written for mega file data cube (MFDC, Section 2.2) in order to produce ACCA generated cropland layer (ACL) products that accurately replicate or come very close to replicating TCLs (within 20% quantity disagreement amongst ~80% user's and producer's accuracies). The ACCA was developed for Tajikistan for the year 2005 using MFDC2005 and then applied for the year it was developed (year 2005) as well as for another independent year (year 2010). The ACCA coding process is described in detail in Section 2.4.2.
3. **Difference between the methods for generating TCL and ACL to demonstrate the advantage of ACL**: The process of generating TCL is time consuming. For example, it involves image classification, class identification, and accuracy assessment and so on. Typical methods and approaches of generating TCLs are explained in Thenkabail and others [11]. TCLs can also be obtained from secondary sources, but they too go through time-consuming and resource-intensive processes [11]. In contrast, ACL only requires time and resource to develop the ACCA for the first year. After that the ACCA can be routinely applied for the area for which it is developed year after year to produce ACLs.
4. **Testing ACCA on independent data layers, comparing ACLs *vs.* TCLs**: For Tajikistan, ACCA was developed (Section 2.4.2) using mega file data cube for year 2005 (MFDC2005). This resulted in ACCA generated cropland layer (ACL2005). ACCA is then applied to produce cropland data layers for the year for which it was developed, 2005, and for another independent year, 2010. This resulted in two ACCA generated cropland layers: ACL2005 and ACL2010. The ACLs were then tested for accuracies and errors against reference/truth cropland layers (TCLs) of corresponding years: TCL2005 and TCL2010. First for the year for which it was developed (year 2005) and then for another independent year (year 2010). This process is discussed in Section 2.4.3.

Multiple-resolution data fusion has been used successfully in cropland mapping to improve classification accuracy [38]. The fusion or combination of MODIS time-series, higher spatial resolution Landsat, and secondary data used in this study is considered ideal for cropland classification and assessments [9,22,39–43]. Remote sensing imagery derived vegetation indices such as the Normalized Difference Vegetation Index (NDVI) profile time series are commonly used in indentifying croplands. Because of the distinct crop phenology compared to other land use and land cover types, using temporal signatures can largely improve the accuracy of cropland classification [44]. Overall accuracy of some existing regional cropland classification at 250 m-pixel scale ranged from 84% to 94% [10,43], leaving great potential for producing higher resolution cropland dataset, such as using multi-temporal and multi-sensor remote sensing data via a algorithm.

Some progress has been made in automated land cover classification techniques such as an unsupervised algorithm called independent component analysis [45] and a modified subspace classification method [46]. The advantage of automated algorithm is to avoid time-consuming and laborious manual interpretation and processing. A rule-based decision tree classifier can take advantage of multiple data sources with various resolutions and perform supervised classification in a logical fashion and repeated over time.

Overwhelmingly, the current global agricultural land use information is usually obtained through farmer communications and inspection by authorities (e.g., agricultural extension officers, analysts) through field visits, which may not be accurate or representative at national or global level due to the subjective nature of eye observation involving numerous different field inspectors, and also is labor-extensive, time-consuming and expensive. As a result, few global cropland datasets derived from remote sensing imageries exist [11,47]. The freely available MODIS and Landsat data make it possible to expand the automated algorithms to more accurate and reliable cropland mapping over large areas such as country, regional, and even global scale. The classification outcome can also serve as an independent cropland dataset for national and international agricultural census data, and facilitate food security monitoring and agrarian decision-making.

2.4.1. Producing Truth/Reference Cropland Layer (TCL)

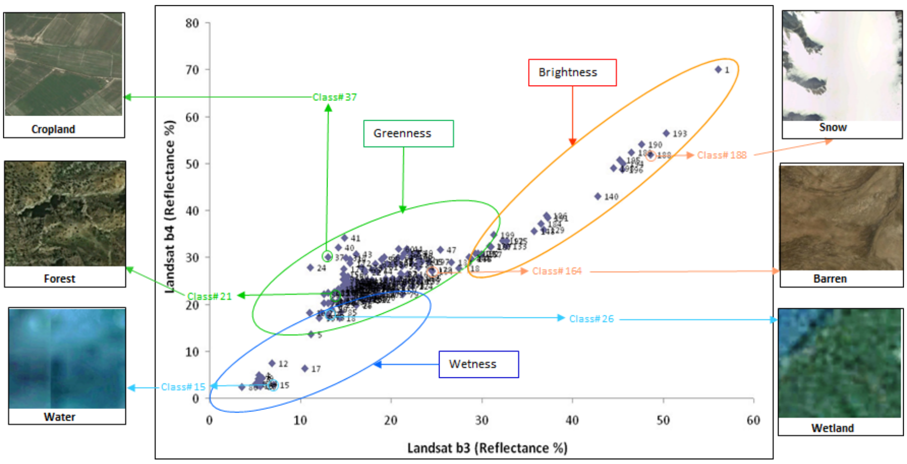
The truth/reference cropland layer (TCL) is the most accurate possible map of croplands (e.g., extent, area) and their characteristics (e.g., irrigated *vs.* rainfed, crop types, cropping intensities) for the area of interest. The TCL can either be obtained from a reliable secondary source (e.g., national sources) or can be produced by us. The TCL is similar to a reference data layer. However, TCL is the "most accurate possible" (MAP). That means that even when a reference data is available, attempt is made to assess its quality, improve it, and bring it to MAP quality.

For Tajikistan, only very coarse resolution cropland products were available [11,12,15], or national statistics and geospatial techniques [36,48]. These products are useful, but do not have sufficient resolution of nominal 30 m required for this study. Thus, a decision was made to first produce a TCL using MFDC2005 (see Section 2.2). As a start an unsupervised classification lead to initial 199 classes (Figure 3(a)). The process of identifying these classes, labeling and producing the TCL involved 5 key steps:

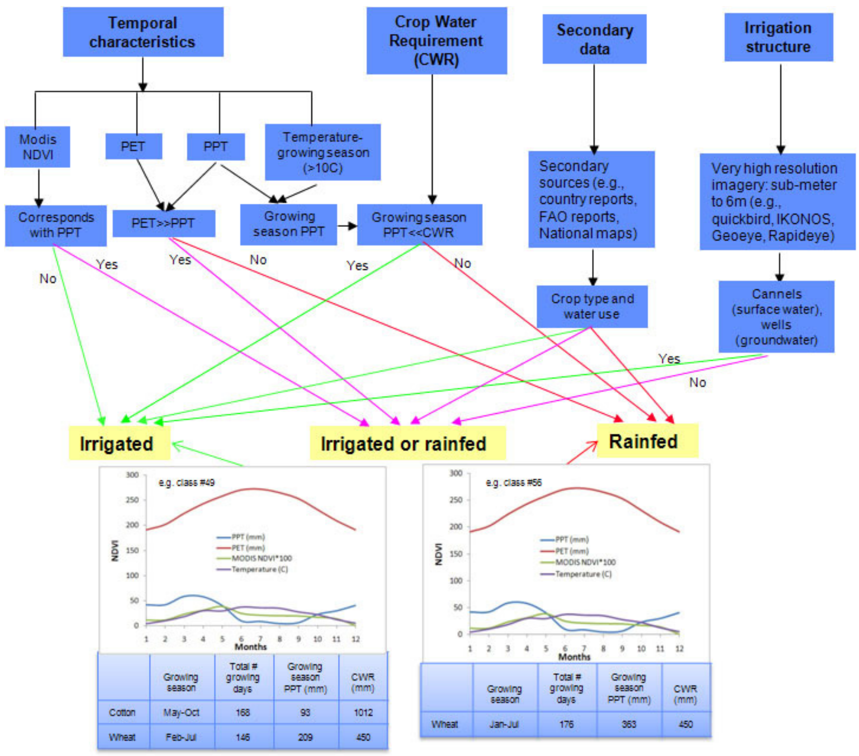
A. bispectral plots (e.g., Figure 3(a)),
B. secondary data and decision trees (e.g., Figure 3(b)),
C. textural characteristics (e.g., Figure 3(c)),
D. spectro-temporal characteristics (e.g., Figure 3(d)), and
E. field-plot data (e.g., Figure 3(d)) and very high resolution imagery (e.g., Figure 3(a)) involving a total of 1,770 points from precise locations (Section 2.3).

The MFDC2005 that has 44 data layers (see Section 2.2) including fusion of Landsat 30 m with MODIS 250 m and numerous secondary data, covering the entire country of Tajikistan. An unsupervised ISOCLASS classification (ERDAS, 2011) was performed on MFDC2005 resulting in 199 classes (Figure 3(a)) for the entire country of Tajikistan. These classes were identified and labeled using well known protocols [27,30,49].Further irrigated areas were distinguished from rainfed croplands using numerous distinct data types (Figure 3(b–d)) and methods described in Thenkabail *et al.* [11]. The process leads to the production of TCL which is discussed in the results section of this paper.

Figure 3. Knowledge capture for producing the truth layer of croplands. (**a**) Step 1: bispectral plots. Cropland *vs.* non-cropland identification and labeling based on tasseled cap bi-spectral plots of unsupervised classes using Landsat red *vs.* near infrared bands and very high resolution imagery (sub-meter to 5 m data) snap-shots as in-situ data. (**b**) Step 2: secondary data and decision trees. Watering method (e.g., irrigated or rainfed) and crop type determination using temporal characteristics of MODIS 250 m NDVI, Crop Water Requirement (CWR), secondary data, and irrigation structure. (**c**) Step 3: textural characteristics using spatial data. Looking at spatial extent to identify and label classes based on large scale (*i.e.*, contiguous) *vs.* small scale (*i.e.*, fragmented) established using very high resolution (sub-meter to 5 m) data. (**d**) Step 4: spectro-temporal characteristics using MODIS time-series. MODIS NDVI temporal profile to identify and label cropping intensities. Illustrated here for irrigated *vs.* rainfed.

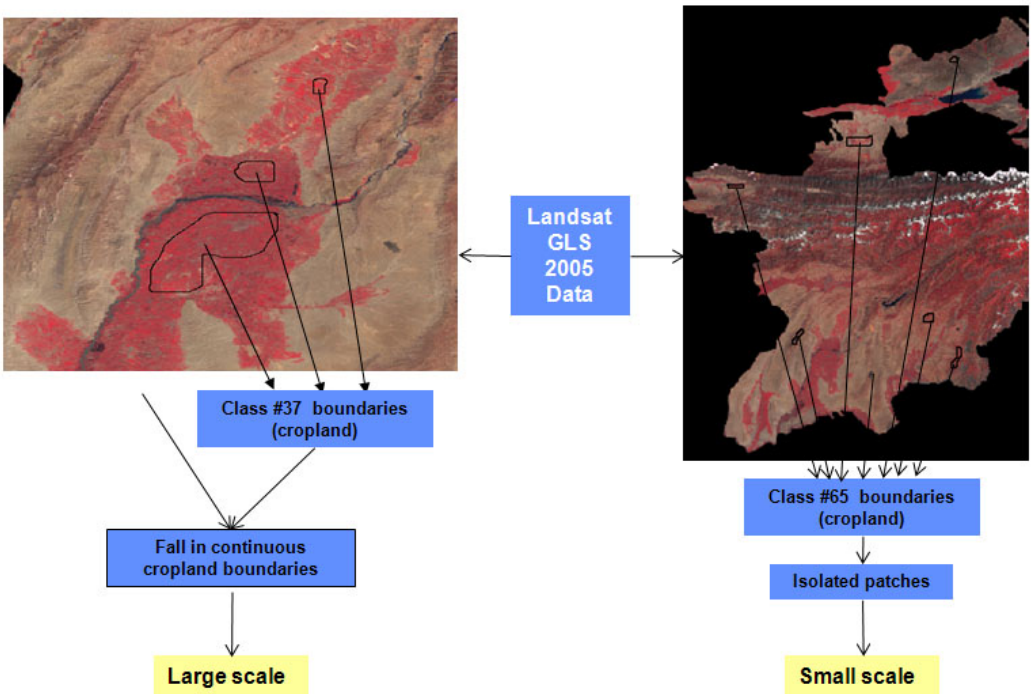


(**a**)

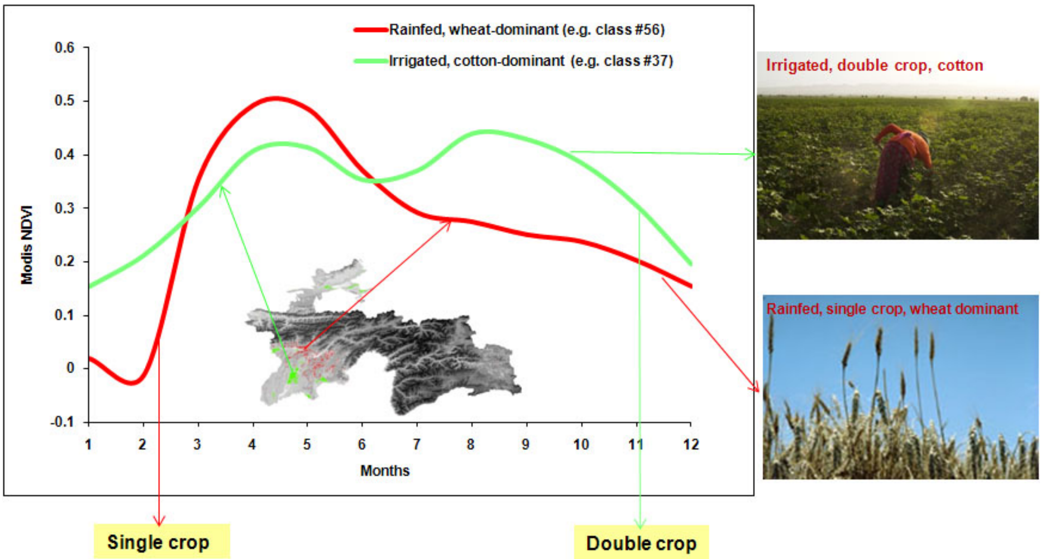


(**b**)

Figure 3. *Cont.*



(**c**)



(**d**)

2.4.2. Automated Cropland Classification Algorithm (ACCA) Derived Cropland Layer (ACL)

The goal of ACCA is to write series of rules or codes based on extensive knowledge of the area (e.g., Section 2.3) that makes use of a reference mega file data cube (MFDC) leading to rapidly and accurately producing a ACCA derive cropland layer (ACL) that accurately (e.g., within 20% quantity disagreement amongst a large number of pixels or ~80% producer's, user's, and overall accuracies) matches the TCL (Section 2.4.1). The process of developing ACCA in Tajikistan involved: (1) making use of a reference mega-file data cube (in this case we used MFDC2005 as reference; Section 2.2) that consists of fusion of Landsat data, MODIS monthly maximum value composite (MVC) NDVI time-series data, and secondary data; (2) intelligently writing rules/codes (e.g., Figure 4 based on

MFDC2005 that will help separate croplands from non-croplands and irrigated from rainfed; (3) segmenting the MFDC based on SRTM derived elevation and/or SRTM derived slope data in coding (Figure 4) to delineate non-croplands from croplands, typically, in conjunction with other remote sensing data in MFDC2005, and (4) highlighting and discerning specific crop features to separate croplands from non-croplands and/or distinguish cropland watering sources (e.g., irrigated *vs*. rainfed) by making use of specific bands such as: (i) chlorophyll absorption (Landsat red band), (ii) moisture sensitivity (Landsat SWIR band), and (iii) irrigated *vs*. rainfed through temperature variability (Landsat thermal band). By writing a series of rules or codes (e.g., Figure 4) through intelligent use of various types and characteristics of MFDC2005 data layers discussed above, we will separate croplands from non-croplands using ACCA.

The process of developing the ACCA will have to go through numerous iterations. It involves writing hundred of simple rules (few illustrated in Figure 4). Typically, every rule such as the one illustrated in Algorithm 1(a,b,c) in Figure 4, captures certain percentage of total cropland area and its characteristics (e.g., irrigated *vs.* rainfed) in TCL. The process is repeated with numerous additional rules (e.g., Algorithms 2–4, Figure 4) applied on MFDC2005 that were written based on all available knowledge (or through trial and error) to capture as much cropland area and as many characteristics as possible. If the rule captures non-croplands then the iteration is repeated by tweaking the rule till we are able to precisely (or near precisely) capture croplands, distinguish them from non-croplands, as well as differentiate irrigated croplands from rainfed croplands.

The key to developing ACCA was multiple fold and iterative. For example, areas with SRTM slopes ≤ 1.5% and cumulative MODIS yearly NDVI ≥ 1,920 (see Algorithm 1(a) in Figure 4) was pure (100%) irrigated areas in Tajikistan as we determined by comparing the ACCA derived result with TCL. The precision in the delineated irrigated areas through a rule or set of rules is established by comparing the ACCA derived croplands through these rules/codes with the TCL (Figure 5(b)). If the ACCA derived cropland did not perfectly match with the TCL then the coding was tweaked till we got the perfect (or nearly perfect) match. Once these pure areas were separated from the rest of the area, further coding to delineate irrigated and rainfed croplands was applied to the remaining area of the country. As the algorithm is further developed, greater complexity in rules/codes and larger number of datasets are involved in further delineating pure cropland areas from non-croplands. For instance, in Algorithm 1(c) (Figure 4), when combination of factors match, for example, when SRTM derived slope > 2.5 but ≤ 7.0 and MODIS yearly cumulative NDVI ≥ 2,000 and elevation ≤ 510 then the areas mapped by these factors are pure irrigated. Algorithm 1(a) through 1(c) (Figure 4) involve SRTM elevation and slope data along with MODIS yearly cumulative NDVI data. Algorithms in Figure 4, involve combinations of Landsat bands, SRTM data, and MODIS monthly data. Coding of Algorithm 1–3 in Figure 4 highlight the rules for mapping irrigated areas and separating them from non-irrigated areas and non-croplands. Coding of Algorithm 4 in Figure 4 highlight the rules for mapping rainfed areas and separating them from non-rainfed areas and non-croplands.

Figure 4. Automated cropland classification algorithm (ACCA). Sample illustrations of irrigated (Algorithm 1(a,b,c), Algorithm 2(a,b,c), and Algorithm 3) and rainfed (Algorithm 4(a,b,c,d,e)) algorithms. Numerous such algorithms are written, based on knowledge capture**,** to finally complete ACCA.

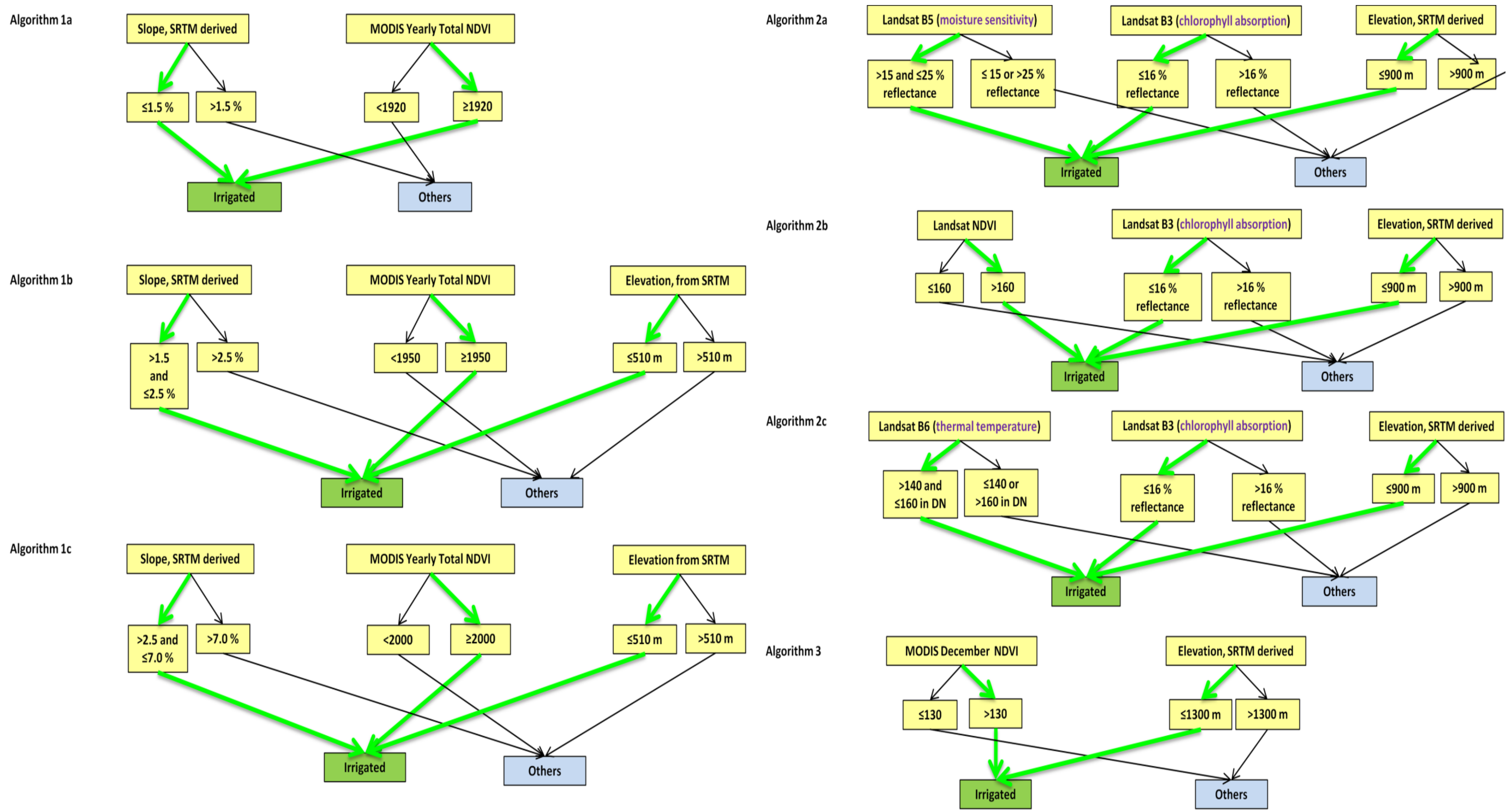


Figure 4. *Cont.*

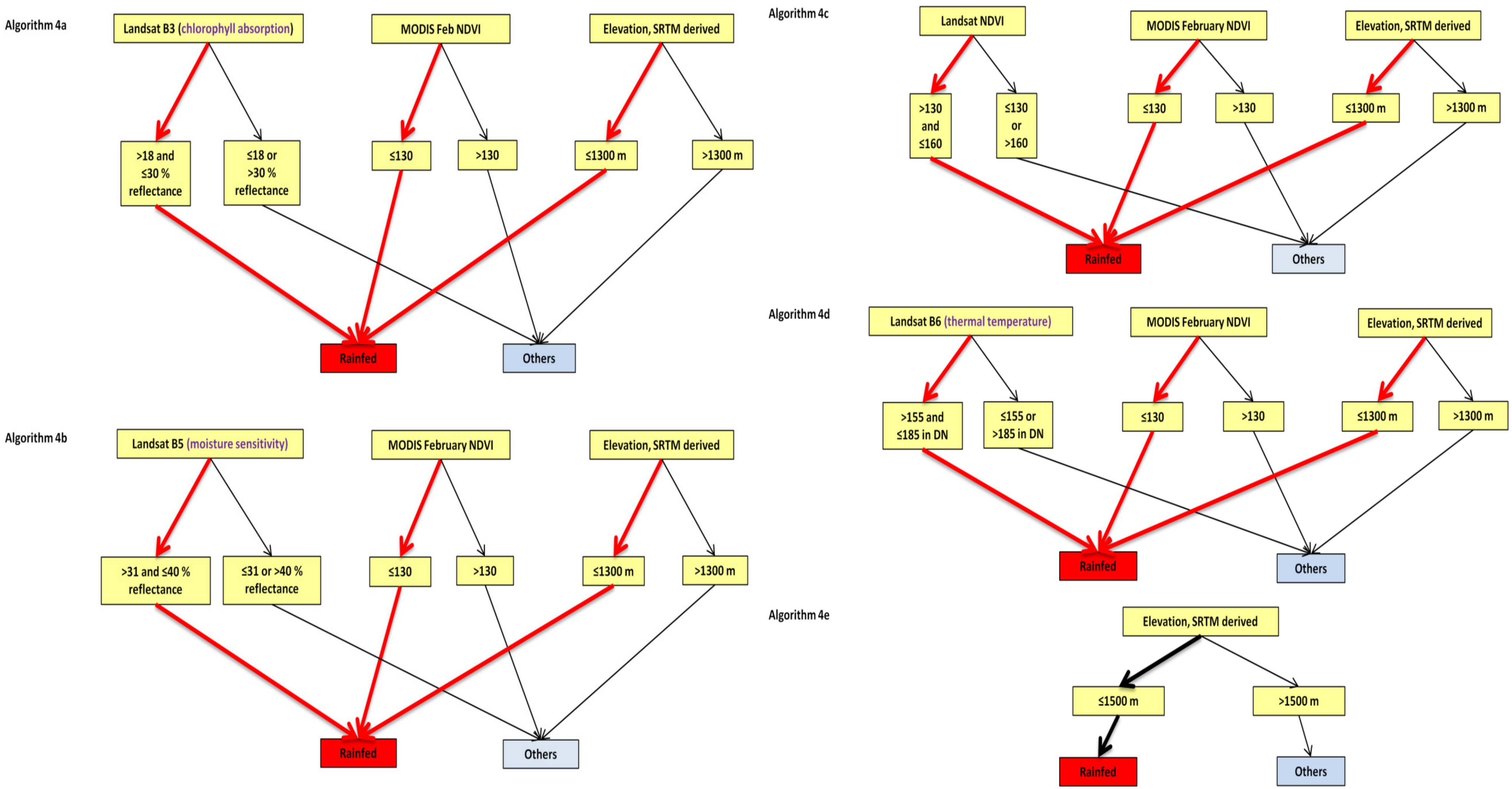
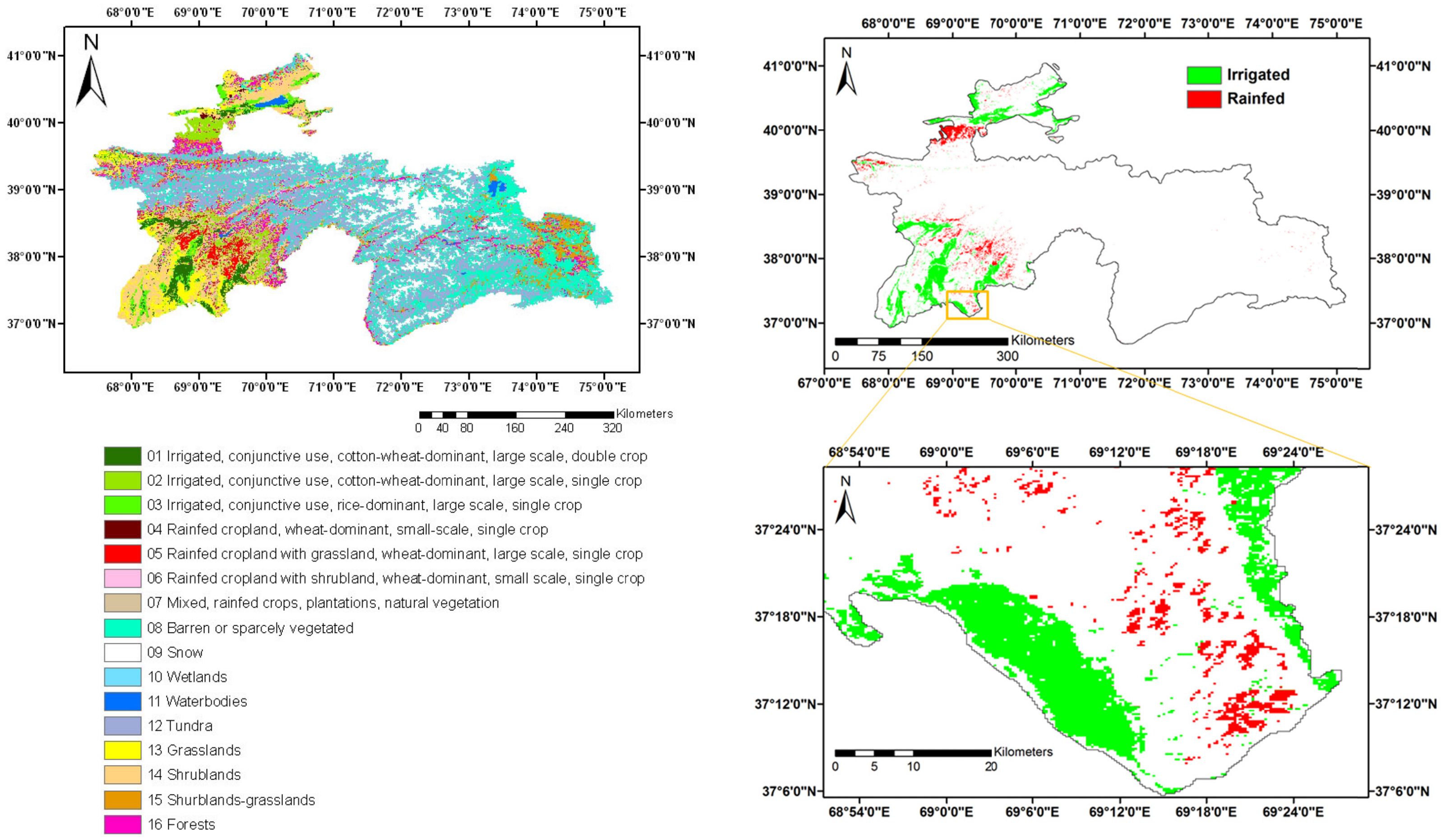


Figure 5. Land cover classification (16 classes, (**a**)) and **truth cropland layer (2 classes, (b)) for Tajikistan for year 2005 (TCL2005)** generated using mega-file data cube of year 2005 (MFDC2005, Figure 1, Table 1). Detailed image interpretation techniques involved bispectral plots (e.g., Figure 3(a)), secondary data (e.g., Figure 3(b)), decision trees (e.g., Figure 3(b)), textural characteristics (e.g., Figure 3(c)), spectro-temporal characteristics (e.g., Figure 3(d)), and field-plot data (e.g., Figure 3(d)).



(**a**) (**b**)

The process was continued until when numerous algorithms put together to produce an ACCA derived cropland layer of Tajikistan for year 2005 (ACL2005) accurately (within 20% quantity disagreement amongst a large number of pixels or ~80% producer's, user's, and overall accuracies) matched with TCL of Tajikistan for year 2005 (TCL2005). The ACL2005 was compared with TCL2005 to assess accuracies and errors. The ACL2010 was then compared with TCL2005 to assess accuracies and errors. In Tajikistan, accuracies and errors involved about 10 million cropland pixels of Landsat (30 m) resolution. The ACCA algorithm and the associated MFDC2005 data are made available on US Geological Survey's (USGS) Powell Center Global Croplands (https://powellcenter.usgs.gov/globalcroplandwater/) ScienceBase for others to download and test: http://www.sciencebase.gov/catalog/folder/4f79f1b7e4b0009bd827f548.

2.4.3. Applying ACCA on Independent Years

The ACCA developed for Tajikistan in this study was based on the MFDC2005. Then the ACCA was applied on mega file data cube of Tajikistan for the year 2010 (MFDC2010). The MFDC2010 had similar data layers as MFDC2005 (Section 2.2), except that the years of image data were different. The ACCA generated cropland layer for year 2010 (ACL2010) was then compared with TCL for year 2010 (TCL2010).

3. Results and Discussions

First, we present and discuss the production of TCL2005 for the country of Tajikistan. The process involved in producing TCL2010 is similar to TCL2005 and hence won't be discussed. Second, we present and discuss the results of the automated cropland classification algorithm (ACCA) derived croplands (ACL) of Tajikistan for the year 2005 (ACL2005) and the year 2010 (ACL2010). Third, we compare ACL2005 with TCL2005 and ACL2010 with TCL2010, and assess accuracies and errors of ACLs relative to TCLs.

3.1. Truth/Reference Layer of Croplands for Tajikistan Based on Data for Year 2005

First, the truth/reference cropland layer (TCL) of the entire country of Tajikistan was produced for year 2005 (TCL2005; Figure 5(a,b)) based on the datasets and methods described in Section 2.4.1. Figure 5(a) has 6 cropland classes, of which classes 1 to 3 are irrigated croplands and classes 4 to 6 are rainfed croplands. The rest are non-cropland classes. The irrigated classes are grouped together as one irrigated class, and so are the rainfed classes, with the resulting map (Figure 5(b)) showing only cropland areas of Tajikistan, masking out all non-cropland areas. The cropland and non-cropland statistics are provided in Table 2. The total country area is 14,218,169 ha, of which the irrigated areas (as of year 2005) was 711,000 ha (~5% of the country area) and rainfed croplands 270,000 ha (~1.9% of the country area). Therefore, the total cropland area is 981,000 ha (Table 2). This is almost the same as 984,000 ha reported by the State Statistical Committee of the Republic of Tajikistan [36]. The same source reports irrigated areas as 724,000 ha in year 2006, again almost the same as determined in this study. Overall, the irrigated *vs.* rainfed areas of Tajikistan is roughly 70% *vs*. 30%, respectively. This is true for year 2010 and 2005 (http://www.stat.tj/en/, http://www.mongabay.com/history/tajikistan/

tajikistan-climate.html, http://www.fao.org/nr/water/aquastat/maps/index.stm). The results in Table 2, provide statistics for year 2005.

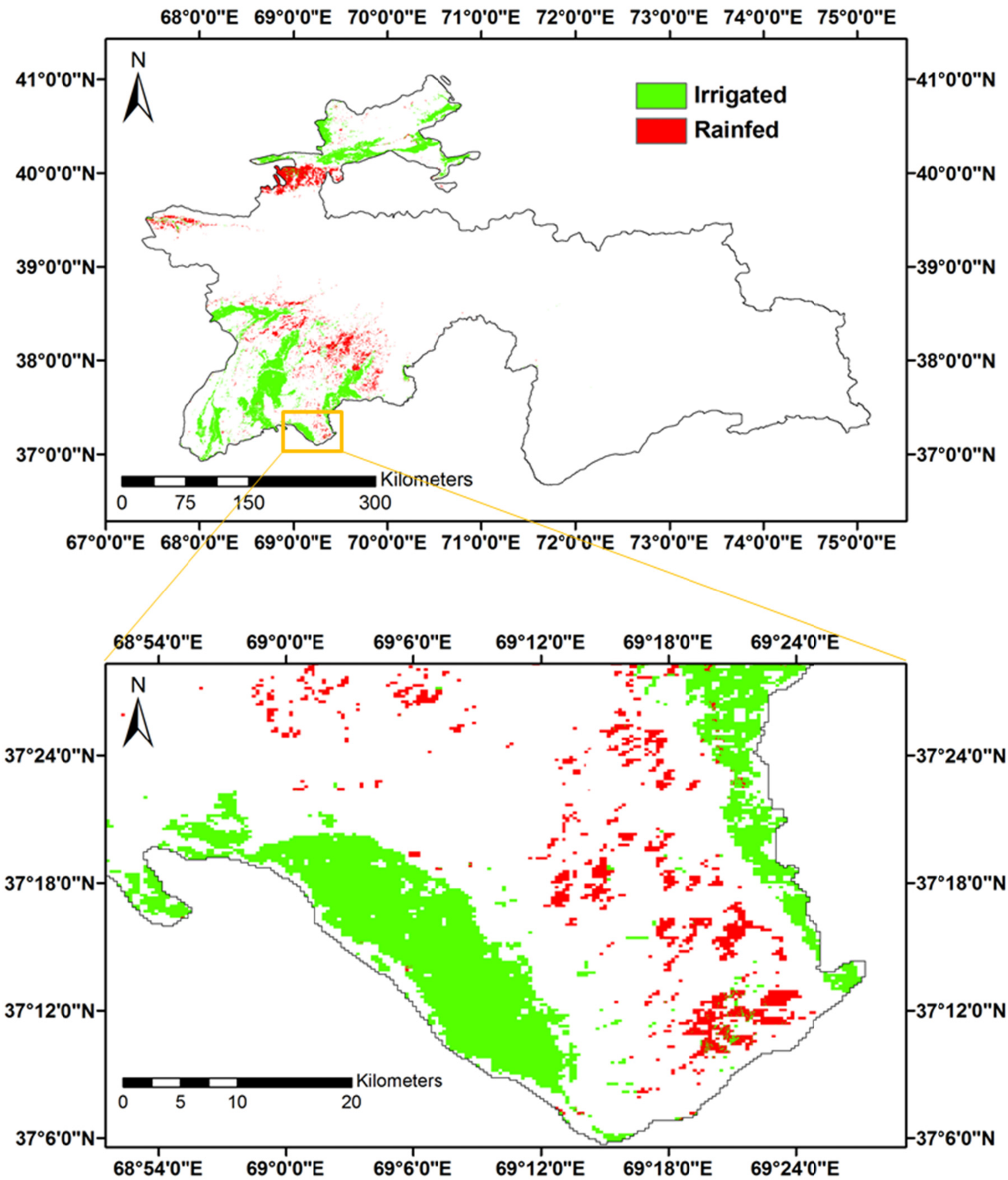
Table 2. Truth cropland layer (TCL) of Tajikistan statistics (see related maps in Figure 5(a,b)). Class characteristics include the full pixel areas (FPAs), Landsat GLS 2005 and MODIS 2005 NDVI.

Class #	Class Name	Area	% Total Area	Landsat NDVI	MODIS NDVI Profile											
					1	2	3	4	5	6	7	8	9	10	11	12
		ha	no units	no units	no units	no units	no units	no units	no units	no units	no units	no units	no units	no units	no units	no units
1	01 Irrigated, conjunctive use, cotton-wheat-rice-dominant, large scale, double crop	710,166	5	0.36	0.09	0.19	0.28	0.36	0.37	0.33	0.36	0.41	0.39	0.35	0.28	0.15
2	02 Rainfed, wheat-barley-dominant, large scale, single crop	273,188	1.9	0.18	0.05	0.00	0.31	0.42	0.47	0.33	0.26	0.24	0.22	0.20	0.18	0.03
3	03 Shrub/rangeland dominate with rainfed croplands	470,037	3.3	0.20	0.01	−0.01	0.22	0.38	0.39	0.37	0.29	0.26	0.23	0.21	0.09	0.01
4	04 Shrublands-grasslands	3,384,129	23.6	0.16	0.02	0.03	0.12	0.24	0.24	0.24	0.21	0.20	0.18	0.16	0.07	0.03
5	05 Mixed, shrublands, grasslands, urban built-up	70,554	0.5	0.21	0.01	0.02	0.20	0.33	0.34	0.34	0.29	0.26	0.24	0.22	0.16	0.09
6	06 Forest	835,732	5.8	0.20	0.00	−0.01	0.02	0.12	0.16	0.28	0.28	0.25	0.22	0.18	0.02	−0.01
7	07 Tundra	3,596,562	25.1	0.20	−0.02	−0.03	−0.04	−0.05	−0.05	0.01	0.14	0.20	0.17	0.09	−0.02	−0.02
8	08 Wetlands	116,386	0.8	0.21	0.01	0.02	0.02	−0.05	0.02	0.15	0.26	0.25	0.23	0.20	0.12	0.00
9	09 Barren or sparcely vegetated	3,380,265	23.6	0.11	−0.01	−0.03	−0.02	−0.01	−0.02	0.02	0.08	0.10	0.09	0.05	−0.01	−0.01
10	10 Snow	1,381,150	9.7	0.05	−0.03	−0.03	−0.04	−0.14	−0.25	−0.08	−0.08	−0.05	−0.04	−0.03	−0.03	−0.03
11	11 Waterbodies	91,831	0.6	−0.11	−0.15	−0.17	−0.18	−0.23	−0.11	−0.06	−0.05	−0.07	−0.12	−0.13	−0.23	−0.17
Total		14,218,169	100													

3.2. ACCA Generated Croplands (ACLs) for Tajikistan for the Same Year (2005) as That of the Truth Cropland Layer (TCL, Year 2005)

Figure 6 shows the ACCA generated cropland data layer of Tajikistan for the year 2005 (ACL2005). Once the mega file data cube (Section 2.2) for year 2005 (MFDC2005) was ready, ACL2005 (Figure 6) for Tajikistan took ~30 min to generate on a Dell Precision desktop T7400 computer using ACCA algorithm (Section 2.4.2).

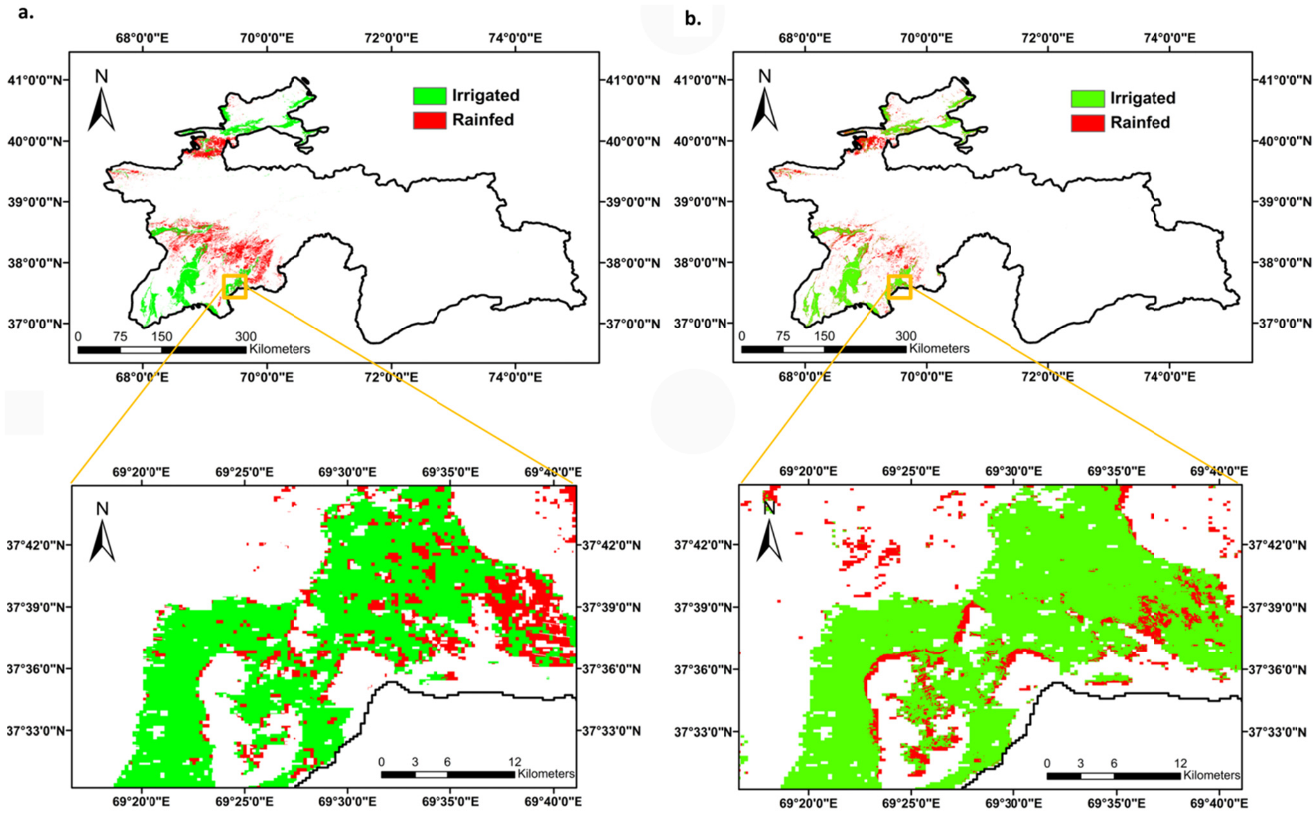
Figure 6. Automated Cropland Classification Algorithm (ACCA) derived croplands of Tajikistan for year 2005 **(ACL2005)** using the same data (MFDC2005). This is compared with CTL (Figure 5). The result matrix is in Table 3(a).



3.3. Truth/Reference Layer of Croplands for Tajikistan Based on Data for Year 2010

TCL2010 (Figure 7(a)) for Tajikistan was generated using exactly the same approach and methods used to produce TCL2005 (Section 3.1). The only difference being that the mega file data cube for year 2010 (MFDC2010) was used to produce TCL2010 instead of MFDC2005.

Figure 7. TCL2010 (a) *vs.* ACL2010 (b). Automated Cropland Classification Algorithm (ACCA) derived croplands of Tajikistan for year 2010 (**ACL2010**, (b)) is compared with the truth cropland layer for Tajikistan for year 2010 (**TCL2010**, (a)). TCL2010 is generated using the independent data layer (MFDC2010). ACL2010 is produced by applying ACCA algorithm on MFDC2010. The result matrix is in Table 3(b).



3.4. ACCA Generated Croplands for Tajikistan for the Independent Year of 2010

ACL2010 (Figure 7(b)) for Tajikistan was produced by applying ACCA on data from an independent year (year 2010) using MFDC2010. Once the mega file data cube (Section 2.2) for year 2010 (MFDC2010) was ready, ACL2010 (Figure 6) for Tajikistan took ~30 min to generate, using a Dell Precision desktop T7400 computer.

3.5. Accuracies and Errors

The accuracies and errors were established by creating an error matrix of Tajikistan involving:

1. TCL2005 (Figure 5(b)) *vs.* ACL2005 (Figure 6); and
2. TCL2010 (Figure 7(a)) *vs.* ACL2010 (Figure 7(b)).

The results are presented in Table 3(a,b). The overall accuracy of the ACCA derived cropland layer for 2005 (ACL2005) was 99.6% (k_{hat} = 0.97) (Table 3(a); Figure 5(b), Figure 6). With about 152 million pixels (each of 30 m resolution) for the country of Tajikistan, and such high level of accuracy shows the very high degree of performance of the ACCA algorithm. For the 3 classes (irrigated, rainfed, and others) mapped, the producer's accuracy was > 86.4% and users accuracy was > 93.6%. Errors of omissions and commissions for irrigated areas were negligible. The overall accuracy of the

ACCA derived cropland layer for independent year 2010 (ACL2010) was 96.2% (k_{hat} = 0.96) (Table 3(b); Figure 7(a,b)). Again, it showed very high degree of accuracy even for the independent year. For the irrigated areas the producer's accuracy was 90.8% and user's accuracy was 82.9%. The ACL2005 *vs.* TCL2005 and ACL2010 *vs.* TCL2010 are shown in Figure 8. However, the rainfed croplands had significant uncertainties. This implies that it is feasible to classify total croplands (irrigated + rainfed) as well as irrigated croplands accurately using ACCA algorithm, but uncertainties for rainfed croplands were significant and needs further investigation. In all likelihood, resolving rainfed croplands will require greater knowledge base through field visits and a better understanding of the rainfed system. This would potentially lead to a need for additional data layers such as evapotranspiration and additional temporal high resolution imagery.

Figure 8. Comparisons between truth cropland layers (TCL) of 2005 (**a**) / 2010 (**b**) and ACCA-derived cropland layers (ACL) of 2005 (**c**) / 2010 (**d**). Specifically, comparisons of **TCL2005 (a)** ***vs.*** **ACL2005 (c)**, and **TCL2010 (b)** ***vs.*** **ACL2010 (d)**. Automated Cropland Classification Algorithm (ACCA) derived croplands of Tajikistan for years 2005 (**ACL2005**) and 2010 (**ACL2010**) are compared with the truth cropland layers for Tajikistan for year 2005 (**TCL2005**) and year 2010 (**TCL2010**). The result matrices are in Table 3(a,b).

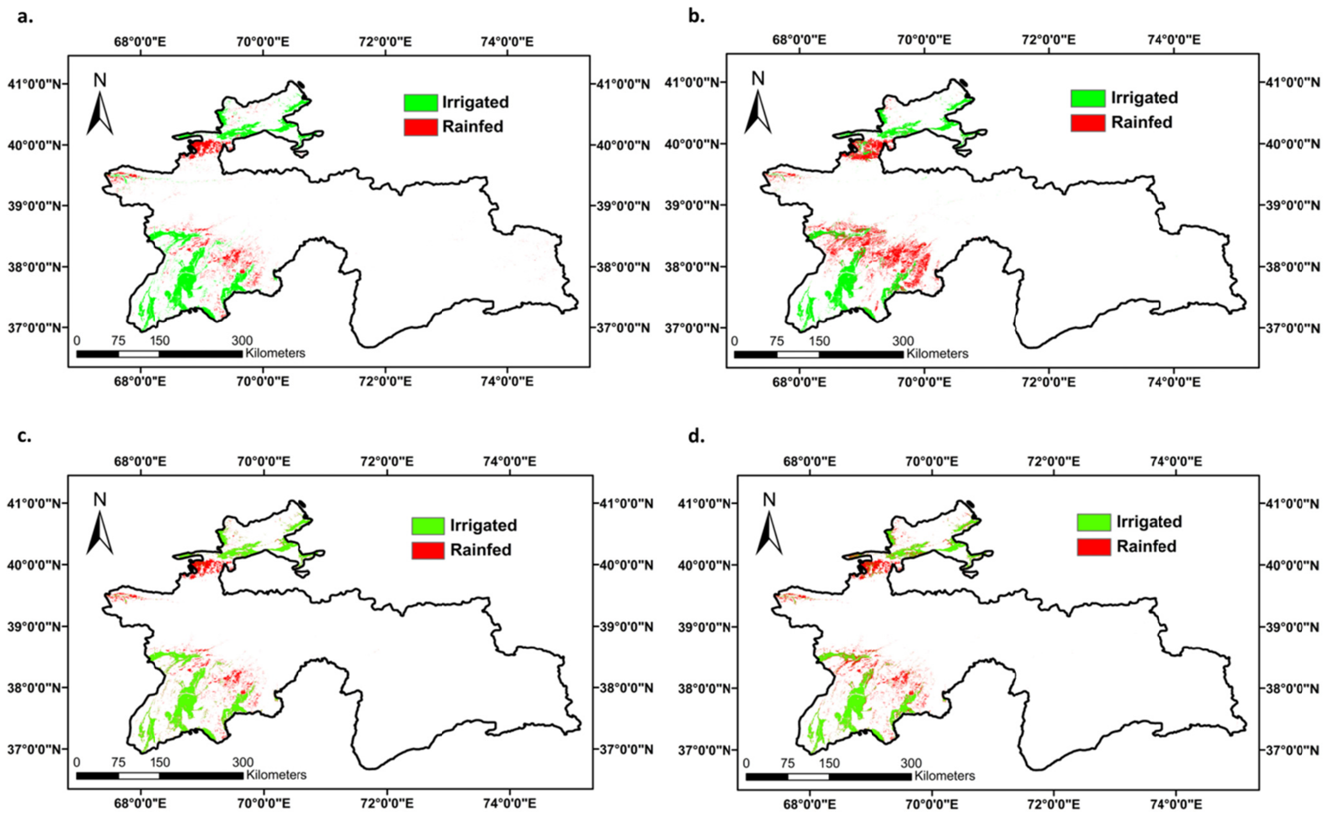


Table 3. Accuracies and errors of irrigated and rainfed classes of Tajikistan established through an error matrix by comparing the cropland truth layer (CTL; Figure 5(b)) (x-axis) with: (**a**) ACCA derived cropland layer (Figure 6) using mega file data cube for year 2005 (MFDC2005), and (**b**) ACCA derived cropland layer (Figure 7) using mega file data cube for year 2010 (MFDC2010).

a. ACCA Algorithm Derived Data for Year 2005 (TRUTH LAYER YR 2005)

	Irrigated areas	Rainfed areas	All other LCLU classes	Row total	Producer's accuracy	Errors of Omissions
Irrigated areas	7,398,009	152,082	30,326	7,580,417	97.6	2.4
Rainfed areas	143,585	2,519,546	252,914	2,916,045	86.4	13.6
All other LCLU classes	24,215	20,577	142,205,696	142,250,488	99.97	0.03
Column total	7,565,809	2,692,205	142,488,936	152,123,251		
User's accuracy	97.8	93.6	99.8	152,746,950		
Errors of Commission	2.2	6.4	0.2			
			Overall accuracy	99.6		
			K_{hat}	0.97		

b. ACCA Algorithm Derived Data for Year 2010 (Landsat ETM+ 2010 and MODIS 2010) (TRUTH LAYER YR 2010)

	Irrigated areas	Rainfed areas	All other LCLU classes	Row total	Producer's accuracy	Errors of Omissions
Irrigated areas	7,258,443	412,788	325,116	7,996,347	90.8	9.2
Rainfed areas	882,856	2719,506	3,917,696	7,520,058	36.2	63.8
All other LCLU classes	618,103	1,255,142	178,375,106	180,248,351	99.0	1.0
Column total	8,759,402	4,387,436	182,617,918	188,353,055		
User's accuracy	82.9	62.0	97.7	195,764,756		
Errors of Commission	17.1	38.0	2.3			
			Overall accuracy	96.2		
			K_{hat}	0.96		

4. Uniqueness, Importance, Impact and Limitations of ACCA

4.1. The Uniqueness of ACCA Algorithm

The uniqueness of ACCA is 3 fold. First, The ACCA algorithm is able to retain the spatial location of ~7 million irrigated cropland pixels and ~3 million rainfed cropland pixels within 20% quantity disagreement (~80 producer's and user's accuracies) for the year for which the model was developed (Table 3(a); Figures 5(b) and 6) as well as for the independent year of 2010 (Table 3(b); Figure 7) clearly implies the robustness of the model. The ACL2005 *vs.* TCL2005 and ACL2010 *vs.* TCL2010 are shown in Figure 8. The ACCA algorithm concept is also adopted in a separate study for the state of California by Wu and Thenkabail [50], where we successfully tested ACCA for 3 independent years and obtained an accuracy of within 10% quantity disagreement (~90% producer's and user's accuracies). Comparing ACL with TCL also means comparing ACL with other classification methods because TCL by nature involves other classification methods as espoused in Section 2.4.1.

Second, the concept of ACCA is ideally suited for development and application over large areas such as a country, or a region or a state or a specific agroecosystem. Knowledge is captured and automated in the algorithm and is primed to be used on a standardized MFDC for any year, requiring very minimal human interaction. ACCA is automated to rapidly compute cropland areas year after year once the mega file data cube (MFDC) for the year in question is set up. For example, ACL2010 was then produced for Tajikistan by ACCA within ~30 min using Dell Precision T7400 computer once the MFDC2010 was ready.

Third, ACCA algorithm is built based on mega file data cube (MFDC) concept involving multi-sensor data fusion along with secondary data. In the case of Tajikistan, we used 44 bands (see Section 2.2) in the development of ACCA involving Landsat ETM+ bands, MODIS 250 m monthly MVC NDVI time-series, bands 1 and 2, and SRTM elevation and slope. Other powerful classification methods such as the spectral matching techniques (SMTs), ensemble of machine learning algorithms (EMLAs) (e.g., decision trees, neural network), and Classification and Regression Tree (CART) all require substantial human interactions and\or significantly greater training data in order to develop robust cropland algorithms that can successfully run on independent data sets and are not automated like ACCA.

4.2. Limitations of ACCA and the Way forward to Developing a Global ACCA

There is one significant limitation of ACCA that needs to be noted. The concept of ACCA for cross site application remains the same. However, specific regions will have to have their own ACCA algorithms to account for unique cropland characteristics of the regions. Further, the ACCA algorithm works on specific MFDC data types used in coding. When these data types are changed (e.g., data for other sensors, additional secondary data), ACCA requires to be modified to take into consideration these additional datasets. The ACCA developed here contributes towards that goal by implementing such a vision for a country. Given the complexity of global agriculture, there will not be a single global algorithm at such high spatial resolution like 30 m that can produce croplands accurately year after year, using multi-sensor remote sensing data fusion along with data integration from other sources, leading to standard mega file data cubes (MFDCs) discussed in this paper.

However, the way forward is to develop ACCA-like algorithms for individual countries, or sub-nations in case of very large nations with complex agro-ecosystems. One can then integrate these different ACCA algorithms, each working on specific areas, and run them together to get the croplands of the entire larger area.

4.3. Implications and Applications of ACCA in Global Cropland Mapping

Even though cropland mapping using remote sensing has been in existence for several decades now [3,29], none of them are fully automated. There are some semi-automated algorithms [5,15,31,32]. However, the need for automated cropland classification algorithm like ACCA presented in this paper is more urgent given the need to address the global food security issue in the twenty-first century. The importance of this need is highlighted by the efforts of Group on Earth Observing (GEO) agricultural initiative (http://www.earthobservations.org/cop_ag_gams.shtml) such as GEO global agricultural monitoring (GEO GLAM). Remote sensing data fusion and automated cropland classifications are a must to address the bleak scenario of global food and water security [1,2].

Uncertainty in global cropland mapping and global cropland water use assessments continues to be high [3]. Reducing this uncertainty will require Earth Observation data from multiple sensors routinely and frequently during the crop growing season. Acceptable overall accuracies [51] should be greater than 85% along with equally high levels of producer's and user's accuracies. Further, global food security analysis will require multiple cropland parameters: cropland areas, cropland watering method (irrigated *vs.* rainfed), cropping intensities, and crop types. In addition, the cropland characteristics of one year are best studies when they are compared with long-term means. Continuous satellite sensor data records are now available, for example, from 1972 onwards from Landsat, 1982 onwards from AVHRR, and 2000 onwards from MODIS to enable such comparisons.

5. Conclusions and Way Forward

The paper presents the development and implementation of an Automated Cropland Classification Algorithm (ACCA) for a country. Specifically, the study demonstrated the ability to compute ACCA algorithm-derived cropland layers (ACLs) for the Country of Tajikistan automatically and accurately once the mega file data cubes (MFDCs), involving combination of Landsat ETM+ 30 m, MODIS 250 m monthly NDVI maximum value composite time-series, and secondary data, were ready. This lead to ACCA computed cropland layer of Tajikistan for the year: (A). 2005 (ACL2005), a non-independent year, derived using MFDC for the year 2005 (MFDC2005), and (B). 2010 (ACL2010), an independent year, derived using MFDC2010. The ACLs were then compared with the truth\reference cropland layers (TCLs). When compared with: 1. TCL2005, the ACL2005 provided an overall accuracy of 99.6% (k_{hat} = 0.97); and 2. TCL2010, the ACL2010 provided an overall accuracy of 96.2% (k_{hat} = 0.96). The producer's and user's accuracies for the total croplands (irrigated plus rainfed) as well as for irrigated classes were above 82.9%, but typically over 90% in most cases. The ACCA algorithm and associated files for Tajikistan are made available through US Geological Survey's (USGS) Sciencebase: http://www.sciencebase.gov/catalog/folder/4f79f1b7e4b0009bd827f548 and the USGS Powell Center: https://powellcenter.usgs.gov/globalcroplandwater/content/models-algorithms.

The results clearly demonstrated the ability of ACCA algorithm to compute total cropland areas as well as irrigated cropland areas consistently, as illustrated for the country of Tajikistan in this study, rapidly and accurately, year after year. Thus, ACCA has the ability to hindcast, nowcast, and futurecast total cropland areas and irrigated areas accurately and automatically for the region for which it was developed. Further development of the algorithm should consider generating rainfed croplands with greater certainty as well as generating crop types and cropping intensities. The ACCA developed in this study is applicable to the area for which it was developed (*i.e.*, Tajikistan). For other cropland areas of the World, the ACCA concept remains the same. However, the codes (Figure 4) need to be modified to suite the regions of interest. Also, the ACCA concept can be further expanded to derive other crop characteristics (e.g., crop types, cropping intensities, and crop phonologies). The MFDC data requirements for other regions will vary depending on the complexity of the agricultural systems of the regions of interest. For example, additional remote sensing (e.g., non-optical, thermal) or secondary (e.g., soils) data may be required to better code and train ACCA to accurately compute croplands extent, areas, and characteristics in complex agroecosystems of the World. However, once the ACCA algorithm is developed for a region taking all the needed datasets and factors into consideration, it can then be applied to hindcast, nowcast and futurecast over the region for which it is developed. The ultimate goal is to combine these series of regional ACCA algorithms into a single global ACCA algorithm.

This research is expected to make significant contributions to global cropland mapping efforts such as the one proposed by the Group on Earth Observation Global Agricultural Monitoring (GEO GLAM) and other similar regional and national initiatives where accurate, reliable, consistent, rapid, and routine cropland mapping is expected year after year, which in turn will contribute to food and water security analysis and decision making. The study will also contribute to the efforts of global food security through research on global croplands and their water use (e.g., https://powellcenter.usgs.gov/globalcroplandwater/).

Acknowledgements

This work is supported by the US Geological Survey's (USGS) WaterSMART (Sustain and Manage America's Resources for Tomorrow) project and Famine Early Warning Network (FEWSNET) project. Further, there was continued financial support for this research from the USGS Geographic Analysis and Monitoring (GAM) and Land Remote Sensing (LRS) programs which come under the USGS Core Science System of Climate and Land Use Change. This support is gratefully acknowledged. Special thanks to James Verdin and James Rowland of USGS. Authors are grateful for continued support and encouragement from Susan Benjamin, Director of the USGS Western Geographic Science Center and Edwin Pfeifer, USGS Southwest Geographic Team Chief. Finally, authors would also like to thank USGS John Wesley Powell Center for Analysis and Synthesis for funding the Working group on Global Croplands (WGGC). Our special thanks to Powell Center Directors: Jill Baron and Marty Goldhaber. Inputs from WGGC team members (http://powellcenter.usgs.gov/current_projects.php#GlobalCroplandMembers) are acknowledged. The WGGC web site (https://powellcenter.usgs.gov/globalcroplandwater/) support provided by Megan Eberhardt Frank, Gail A. Montgomery, Tim Kern and others is deeply appreciated.

References

1. Van den Bergh, F.; Wessles, K.J.; Miteff, S.; van Zyl, T.L.; Gazendam, A.D.; Bachoo, A.K. HiTempo: A platform for time-series analysis of remote-sensing satellite data in a high-performance computing environment. *Int. J. Remote Sens.* **2012**, *33*, 4720–4740.
2. Tilman, D.; Balzer, C.; Hill, J.; Befort, B.L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20260–20264.
3. Thenkabail, P.S.; Hanjra, M.A.; Dheeravath, V.; Gumma, M. A holistic view of global croplands and their water use for ensuring global food security in the 21st century through advanced remote sensing and non-remote sensing approaches. *Remote Sens.* **2010**, *2*, 211–261.
4. 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.
5. 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.; *et al*. Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens. Environ.* **2002**, *83*, 287–302.
6. Hansen, M.C.; Defries, R.S.; Townshend, J.R.G.; Sohlberg, R.; Dimiceli, C.; Carroll, M. Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS data. *Remote Sens. Environ.* **2002**, *83*, 303–319.
7. Ozdogan, M.; Woodcock, C.E. Resolution dependent errors in remote sensing of cultivated areas. *Remote Sens. Environ.* **2006**, *103*, 203–217.
8. Wardlow, B.D.; Kastens, J.H.; Egbert, S.L. Using USDA crop progress data for the evaluation of greenup onset date calculated from MODIS 250-meter data. *Photogramm. Eng. Remote Sensing* **2006**, *72*, 1225–1234.
9. Wardlow, B.D.; Egbert, S.L.; Kastens, J.H. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sens. Environ.* **2007**, *108*, 290–310.
10. Wardlow, B.D.; Egbert, S.L. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the US Central Great Plains. *Remote Sens. Environ.* **2008**, *112*, 1096–1116.
11. Thenkabail, P.S., Lyon, G.J., Turral, H., Biradar, C.M., Eds. *Remote Sensing of Global Croplands for Food Security*; CRC Press: Boca Raton, FL, USA, 2009.
12. Thenkabail, P.S.; Biradar, C.M.; Noojipady, P.; Dheeravath, V.; Li, Y.; Velpuri, M.; Gumma, M.; Gangalakunta, O.R.P.; Turral, H.; Cai, X.; *et al*. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. *Int. J. Remote Sens.* **2009**, *30*, 3679–3733.
13. Xiao, X.; Boles, S.; Frolking, S.; Li, C.; Babu, J.Y.; Salas, W.; Moore, B., III. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. *Remote Sens. Environ.* **2006**, *100*, 95–113.
14. Gumma, M.K.; Nelson, A.; Thenkabail, P.S.; Singh, A.N. Mapping rice areas of South Asia using MODIS multitemporal data. *J. Appl. Remote Sens.* **2011**, *5*, 053547–053547-26.
15. Thenkabail, P.S., Lyon, G.J., Huete, A., Eds. *Hyperspectral Remote Sensing of Vegetation*; CRC Press: Boca Raton, FL, USA, 2011.

16. EL-Magd, I.A.; Tanton, T.W. Improvements in land use mapping for irrigated agriculture from satellite sensor data using a multi-stage maximum likelihood classification. *Int. J. Remote Sens.* **2003**, *24*, 4197–4206.
17. De Fries, R.S.; Hansen, M.; Townshend, J.R.G.; Sohlberg, R. Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *Int. J. Remote Sens.* **1998**, *19*, 3141–3168.
18. Pittman, K.; Hansen, M.C.; Becker-Reshef, I.; Potapov, P.V.; Justice, C.O. Estimating global cropland extent with multi-year MODIS data. *Remote Sens.* **2010**, *2*, 1844–1863.
19. Liu, J.; Shao, G.; Zhu, H.; Liu, S. A neural network approach for enhancing information extraction from multispectral image data. *Can. J. Remote Sens.* **2005**, *31*, 432–438.
20. Atzberger, C.; Rembold, F. Estimating Sub-Pixel to Regional Winter Crop Areas Using Neural Nets. In *Proceedings of ISPRS TC VII Symposium: 100 Years ISPRS–Advancing Remote Sensing Science*, Vienna, Austria, 5–7 July 2010.
21. Mathur, A.; Foody, G.M. Crop classification by support vector machine with intelligently selected training data for an operational application. *Int. J. Remote Sens.* **2008**, *29*, 2227–2240.
22. Lobell, D.B.; Asner, G.P. Cropland distributions from temporal unmixing of MODIS data. *Remote Sens. Environ.* **2004**, *93*, 412–422.
23. Yang, C.; Everitt, J.H.; Bradford, J.M. Airborne hyperspectral imagery and linear spectral unmixing for mapping variation in crop yield. *Precis. Agric.* **2007**, *8*, 279–296.
24. Chen, Z.; Li, S.; Ren, J.; Gong, P.; Zhang, M.; Wang, L.; Xiao, S.; Jiang, D. Agricultural Applications. In *Advances in Land Remote Sensing*: *System*, *Modeling*, *Inversion and Application*; Liang, S., Ed.; Springer: New York, NY, USA, 2008; pp. 397–421.
25. Crist, E.P.; Cicone, R.C. Application of the tasseled cap concept to simulated thematic mapper data. *Photogramm. Eng. Remote Sensing* **1984**, *50*, 343–352.
26. Cohen, W.; Goward, S. Landsat's role in ecological applications of remote sensing. *Bioscience* **2004**, *54*, 535–545.
27. Thenkabail, P.S.; Schull, M.; Turral, H. Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data. *Remote Sens. Environ.* **2005**, *95*, 317–341.
28. Masek, J.G.; Huang, C.; Wolfe, R.; Cohen, W.; Hall, F.; Kutler, J.; Nelson, P. North American forest disturbance mapped from a decadal Landsat record. *Remote Sens. Environ.* **2008**, *112*, 2914–2926.
29. Ozdogan, M.; Gutmanm, G. A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: An application example in the continental US. *Remote Sens. Environ.* **2008**, *112*, 3520–3537.
30. Thenkaball, P.S.; GangadharaRao, P.; Biggs, T.W.; Krishna, M.; Turral, H. Spectral matching techniques to determine historical Land-use/Land-cover (LULC) and irrigated areas using time-series 0.1-degree AVHRR pathfinder datasets. *Photogramm. Eng. Remote Sensing* **2007**, *73*, 1029–1040.
31. Chan, J.C.; Paelinckx, D. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens. Environ.* **2008**, *112*, 2999–3011.

32. Goel, P.K.; Prasher, S.O.; Patel, R.M.; Landry, J.A.; Bonnell, R.B.; Viau, A.A. Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Comput. Electron. Agric.* **2003**, *39*, 67–93.
33. Zheng, H.; Chen, L.; Han, X.; Zhao, X.; Ma, Y. Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: The importance of phosphorus application rates under drought conditions. *Agr. Ecosyst. Environ.* **2009**, *132*, 98–105.
34. Thenkabail, P.S. Inter-sensor relationships between IKONOS and Landsat-7 ETM+ NDVI data in three ecoregions of Africa. *Int. J. Remote Sens.* **2004**, *25*, 389–408.
35. Chander, G.; Markham, B.L.; Helder, D.L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+ and EO-1 ALI sensors. *Remote Sens. Environ.* **2009**, *113*, 893–903.
36. State Statistical Committee of the Republic of Tajikistan. Agriculture in Tajikistan. In *Statistical Yearbook*; Statistical Agency under President of the Republic of Tajikistan: Dushanbe, Tajikistan, 2007.
37. Pontius, R.G., Jr.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429.
38. Watts, J.D.; Powell, S.L.; Lawrence, R.L.; Hilker, T. Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. *Remote Sens. Environ.* **2011**, *115*, 66–75.
39. Biggs, T.W.; Thenkabail, P.S.; Gumma, M.K.; Scott, C.A.; Parthasaradhi, G.R.; Turral, H.N. Irrigated area mapping in heterogeneous landscapes with MODIS time series, ground truth and census data; Krishna Basin, India. *Int. J. Remote Sens.* **2006**, *27*, 4245–4266.
40. Simonneaux, V.; Duchemin, B.; Helson, D.; Er-Raki, S.; Olioso, A.; Chehbouni, A.G. The use of high-resolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central Morocco. *Int. J. Remote Sens.* **2008**, *29*, 95–116.
41. Dheeravath, V.; Thenkabail, P.S.; Chandrakantha, G.; Noojipady, P.; Reddy, G.P.O.; Biradar, C.M.; Gumma, M.K.; Velpuri, M. Irrigated areas of India derived using MODIS 500 m time series for the years 2001–2003. *ISPRS J. Photogramm.* **2010**, *65*, 42–59.
42. Lv, T.; Liu, C. Study on extraction of crop information using time-series MODIS data in the Chao Phraya Basin of Thailand. *Adv. Space Res.* **2010**, *45*, 775–784.
43. Shao, Y.; Lunetta, R.S.; Ediriwickrema, J.; Liames, J. Mapping cropland and major crop types across the Great Lakes Basin using MODIS-NDVI Data. *Photogramm. Eng. Remote Sensing* **2010**, *76*, 73–84.
44. Serra, P.; Pons, X. Monitoring farmers' decisions on Mediterranean irrigated crops using satellite image time series. *Int. J. Remote Sens.* **2008**, *29*, 2293–2316.
45. Ozdogan, M. The spatial distribution of crop types from MODIS data: Temporal unmixing using Independent Component Analysis. *Remote Sens. Environ.* **2010**, *114*, 1190–1204.
46. Bagan, H.; Yamagata, Y. Improved Subspace classification method for multispectral remote sensing image classification. *Photogramm. Eng. Remote Sensing* **2010**, *76*, 1239–1251.
47. Ramankutty, N.; Evan, A.T.; Monfreda, C.; Foley, J.A. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Glob. Biogeochem. Cy.* **2008**, doi:10.1029/2007GB002952.

48. Siebert, S.; Hoogeveen, J.; Frenken, K. *Irrigation in Africa; Europe and Latin America—Update of the Digital Global Map of Irrigation Areas to Version 4. Frankfurt Hydrology Paper 05*; University of Frankfurt: Frankfurt am Main, Germany, 2006.
49. Biradar, C.M.; Thenkabial, P.S.; Islam, M.A.; Anputhas, M.; Tharme, R.; Vithanage, J.; Alankara, R.; Gunasinghe, S. Establishing the best spectral bands and timing of imagery for land use-land cover (LULC) class separability using Landsat ETM+ and Terra MODIS data. *Can. J. Remote Sens.* **2007**, *33*, 431–444.
50. Wu, Z.; Thenkabail, P.S. An automated cropland classification algorithm (ACCA) by combining MODIS, Landsat, and Secondary Data for the State of California. *Photogramm*. *Eng*. *Remote Sensing* **2012**, in review.
51. Congalton, R., Green, K., Eds. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed.; CRC/Taylor & Francis: Boca Raton, FL, USA, 2009; p. 183.