Progress Report: June 1, 2014 to November 23, 2014

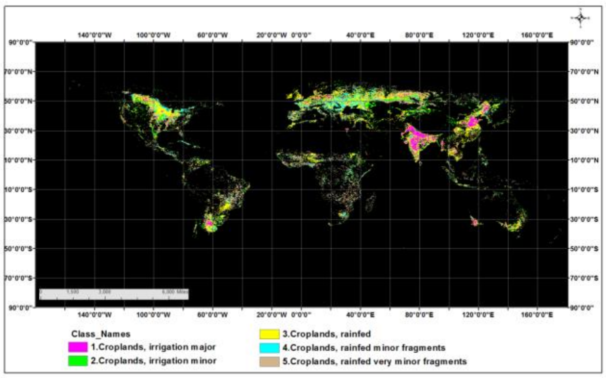
NASA MEaSUREs Project (Project period: June 1, 2013-May 31, 2018)

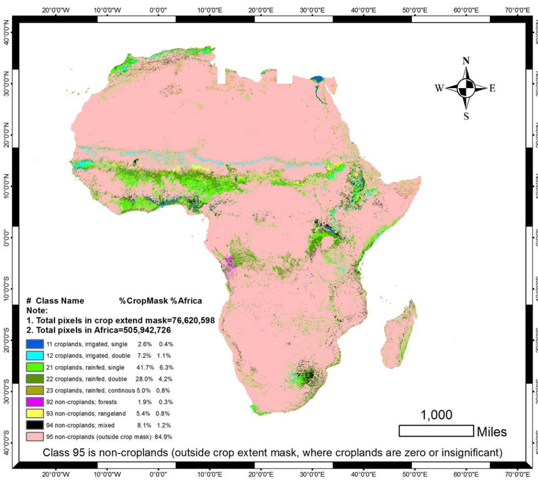
Global Food Security-support Analysis Data @ 30 m (GFSAD30)

(http://geography.wr.usgs.gov/science/croplands/)

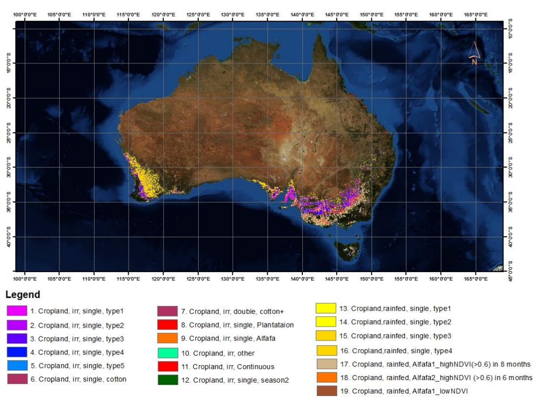
Progress Report: June 1, 2014 to November 23, 2014

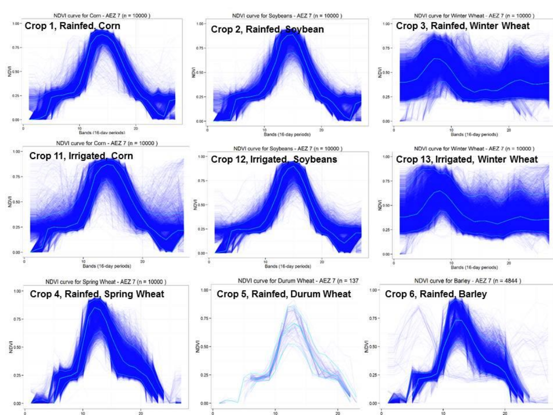
Contact, PI: Prasad S. Thenkabail (pthenkabail@usgs.gov; thenkabail@gmail.com)





GCE 1km Multi-study Crop Mask (aka GCE V1.0) (in press)





<u>Project PI and the USGS Core Team:</u>
Prasad Thenkabail, PI, USGS
Pardhasaradhi Teluguntla, Research Scientist
Jun Xiong, Research Scientist
Justin Poehnelt, IT Expert
<u>Project co-Is</u>:
Cristina Milesi, co-I, NASA AMES\CSUMB
Mutlu Ozdogan, co-I, UW
Russ Congalton, co-I, UNH
Chandra Giri, co-I, USGS EROS
James Tilton, co-I, NASA GSFC
Temuulen Teki, co-I, NAU
<u>PhD Students</u>:
Richard Massey, NAU
Kamini Yadav, UNH
Aparna Phalke, UW

Table of Contents

Progress Report: June 1, 2014 to November 23, 2014

NASA MEaSUREs Project (Project period: June 1, 2013-May 31, 2018)

Global Food Security-support Analysis Data @ 30 m (GFSAD30)

(http://geography.wr.usgs.gov/science/croplands/)

Contact, PI: Prasad S. Thenkabail (pthenkabail@usgs.gov; thenkabail@gmail.com)

EXECUTIVE SUMMARY

Global Food Security-support Analysis Data @ 30 m (GFSAD30) project is funded by NASA MEaSUREs (Making Earth System Data Records for Use in Research Environments), through ROSES solicitation, for a period of 5 years (June 1, 2013- May 31, 2018). The overarching goal of the GFSAD30 project is to produce consistent and unbiased estimates of global agricultural cropland areas, crop types, crop watering method, and cropping intensities using mature cropland mapping algorithms (CMAs) at nominal 30 m spatial resolution. During the process, GFSAD30 project will also develop and release 1 km and 250 m cropland products. This report provides an overview of the progress made till date, specifically for the period of June 1, 2014 to November 23, 2014.

Current achievements include provisional release of GFSAD30 products @ nominal 1km through the Land Processes Distributed Active Archive Center (***LP DAAC***):
Global Cropland Extent (GCE) 1km Crop Dominance (aka GCE V0.0) [see section 2.1.1]
https://earthengine.google.org/#detail/USGS%2FGFSAD1000_V0
Global Cropland Extent (GCE) 1km Multi-study Crop Mask (aka GCE V1.0) [see section 2.1.2]
https://earthengine.google.org/#detail/USGS%2FGFSAD1000_V1

Currently, cropland mapping algorithms (CMAs) are developed and tested to produce cropland products (cropland extent\areas, irrigation versus rainfed, cropping intensity, and crop type) using MODIS 250m time-series data, Landsat 30 m time-series data, numerous secondary data, ground data, and very high spatial resolution imagery (sub-meter to 5 m). Algorithms include spectral matching techniques, parallel k-means computing, automated cropland classification algorithms, linear discriminant analysis, linear spectral mixture model, and hierarchical segmentation *(*HSEG*)*. Extensive ground data is crowd sourced and organized in croplands.org. Also, ~130,000 ground data points are collected and\or sourced. Large number of very high spatial resolution imagery (sub-meter to 5 meter; VHRI) are utilized both in algorithm development and product validation. Ground data, VHRI, and HSEG derived imagery are all used in accuracy assessments.

Project teams are utilizing the computing power of super computers (e.g., NASA NEX), and also coding in Java script and Python in the Google Earth Engine (GEE) Playfield to quickly process massively large quantities of 250m MODIS to 30 m Landsat and numerous other secondary mega-data cubes. Project will organize a 3 day workshop in Reston, VA during January 28-30 to discuss and further advance release of project outputs that include four distinct products, product documentation, peer-reviewed manuscripts, algorithms, Algorithm Theoretical Basis Documents ***(*ATBD's*)*,** and presentations. Project outputs are updated constantly through a common platform: http://geography.wr.usgs.gov/science/croplands/index.html

Progress Report: June 1, 2014 to November 23, 2014

NASA MEaSUREs Project (Project period: June 1, 2013-May 31, 2018)

Global Food Security-support Analysis Data @ 30 m (GFSAD30)

(http://geography.wr.usgs.gov/science/croplands/)

Contact, PI: Prasad S. Thenkabail (pthenkabail@usgs.gov; thenkabail@gmail.com)

1.0 Goal and Objectives of GFSAD30

The overarching goal of GFSAD30 project is to produce consistent and unbiased estimates of global agricultural cropland areas, crop types, crop watering method, and cropping intensities using mature cropland mapping algorithms (CMAs):

There are 6 specific objectives for GFSAD30m project (as in the original proposal):

OBJECTIVE 1: Cropland extent,

OBJECTIVE 2: Crop types (focus on 8 crops that occupy 70% of global croplands),

OBJECTIVE 3: Irrigated vs. rainfed croplands,

OBJECTIVE 4: Cropping intensities\phenology (single, double, triple, continuous cropping),

OBJECTIVE 5: Cropped area computation; and

OBJECTIVE 6: In addition, GCAD four decades will produce continuous data streams at monthly frequency (e.g., illustration for 1 year in Figure 1) from 1982-2017 at 8 km from 1982 to 2000 based on AVHRR GIMMS data and at 250 m from 2001 to 2017 based on MODIS data.

Of these the first 4 are the main products (Figure 1) and the next two are the derived products, once the first 4 products are established.

This progress report provides the work carried out, and the progress made, in meeting the above defined GFSAD30 project goal and objectives to date (June 1, 2013-November 6, 2014; Date Submitted: November 7, 2014).

Global Food Security-support Analysis Data @ 30 m (GFSAD30) Project is a 5-year (June 1, 2013- May 31, 2018) NASA MEaSUREs (Making Earth System Data Records for Use in Research Environments) funded, U. S. Geological Survey (USGS) lead project.

2.0 Overview of GFSAD30 products

GFSAD30 project, will release global cropland products (Figure 1) at 3 resolutions:

- 1 km
- 250 m and
- 30 m

The ultimate goal of the GFSAD30 project is to release global cropland products (Figure 1) at 30 m. However 250m and 1 km products will also be released in the interim. It is not a must for the project to do so. However, interim products at 1 km and 250 m will be invaluable in many ways, especially to help create interest with user communities earlier (rather than wait for 30 m products to come on 4th or 5th year of the project), create interest with wider user groups, and help gather ground data. It is likely that for some areas of the world, we may release 30 m product without releasing 250 m product, especially when some sub-groups conduct studies directly to produce 30 m products.

The specific cropland products delivered at each resolution are as follows:

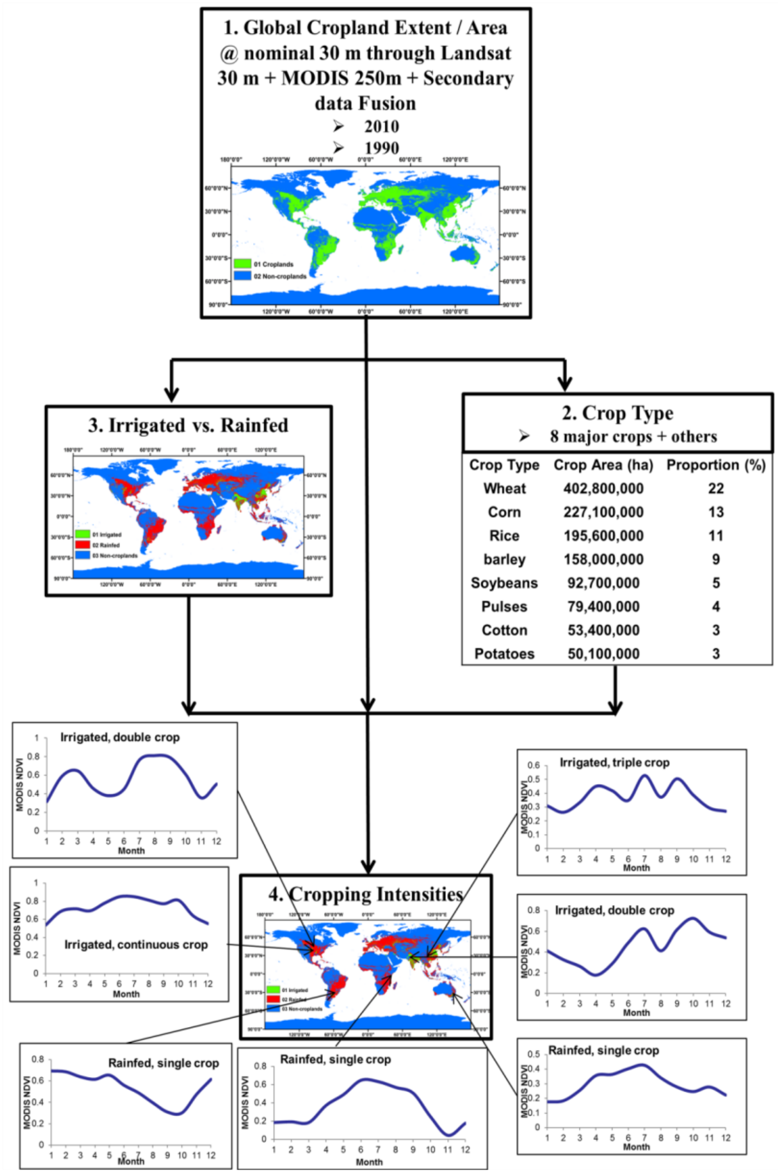


Figure 1. Four main products of global food security-support analysis data @30m (GFSAD30) project. These products are: (a) cropland extent, (b) irrigated vs. rainfed, (c) cropping intensity (single, double, triple, and continuous crop), and (d) crop type.

2.1 1 km products

There are two 1 km cropland products, both of which are already made available at: http://geography.wr.usgs.gov/science/croplands/
These two cropland products are described in section 2.1.1 and section 2.1.2 below.

2.1.1 GCE 1km Crop Dominance (aka GCE V0.0)

This is Global Food Security-support Analysis Data (GFSAD) project's 'GCE' (Global Cropland Extent) 1km Crop Dominance product (e.g., Figure 2a and 2b). Cropland products released at this resolution will be:

- Cropland extent and areas;
- Cropland watering method: irrigation versus rainfed

To a lesser extent

- Crop dominance (not type)

One can download these data from:
http://geography.wr.usgs.gov/science/croplands/

This Global Food Security-support Analysis Data (GFSAD) Project's Global Cropland Extent Product at nominal 1km (GCE 1Km Crop Dominance). The GCE 1KM Crop Dominance provides spatial distribution of the five major global cropland types (wheat, rice, corn, barley and soybeans; which occupy 60% of all global cropland areas) at nominal 1km (GCE 1KM Crop Dominance). The map is produced by overlying the five dominant crops of the world produced by Ramankutty et al. (2008), Monfreda et al. (2008), and Portman et al. (2009) over the remote sensing derived global irrigated and rainfed cropland area map of the International Water Management Institute (IWMI; Thenkabail et al., 2009a, 2009b, 2011).
GCE 1KM Crop Dominance (see Figure below) is an 8 class digital product that provides, at nominal 1 km, information on global:
1. Cropland extent\areas,
2. irrigated versus rainfed cropping,
3. Crop dominance, and
4. Cropping intensity (single, double, triple, and continuous crops).

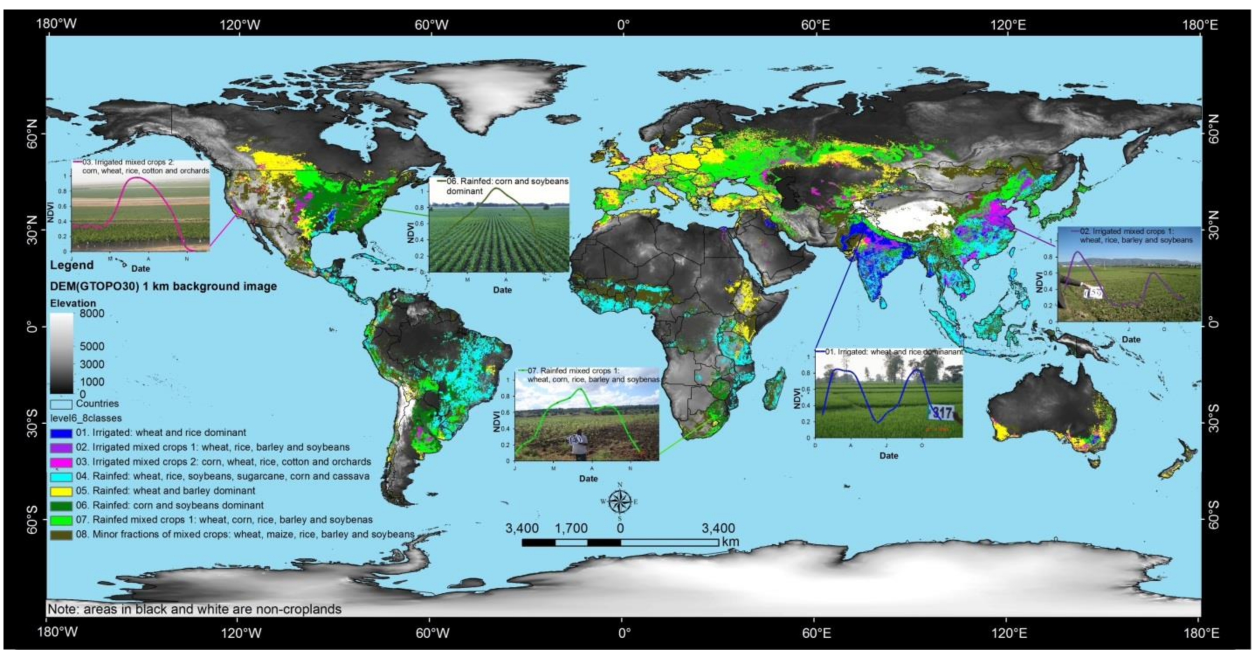


Figure 2a. Spatial distribution of the 5 major global cropland types (wheat, rice, maize, barley and soybeans). The map is produced by overlying the 5 dominant crops (wheat, rice, maize, barley and soybeans; which occupy 60% of all global cropland areas) of the world produced by Ramankutty et al. (2008), Monfreda et al. (2008), and Portman et al. (2009) over the remote sensing derived global irrigated and rainfed cropland area map of the International Water Management Institute (IWMI; Thenkabail et al., 2009a, 2009b, 2011).

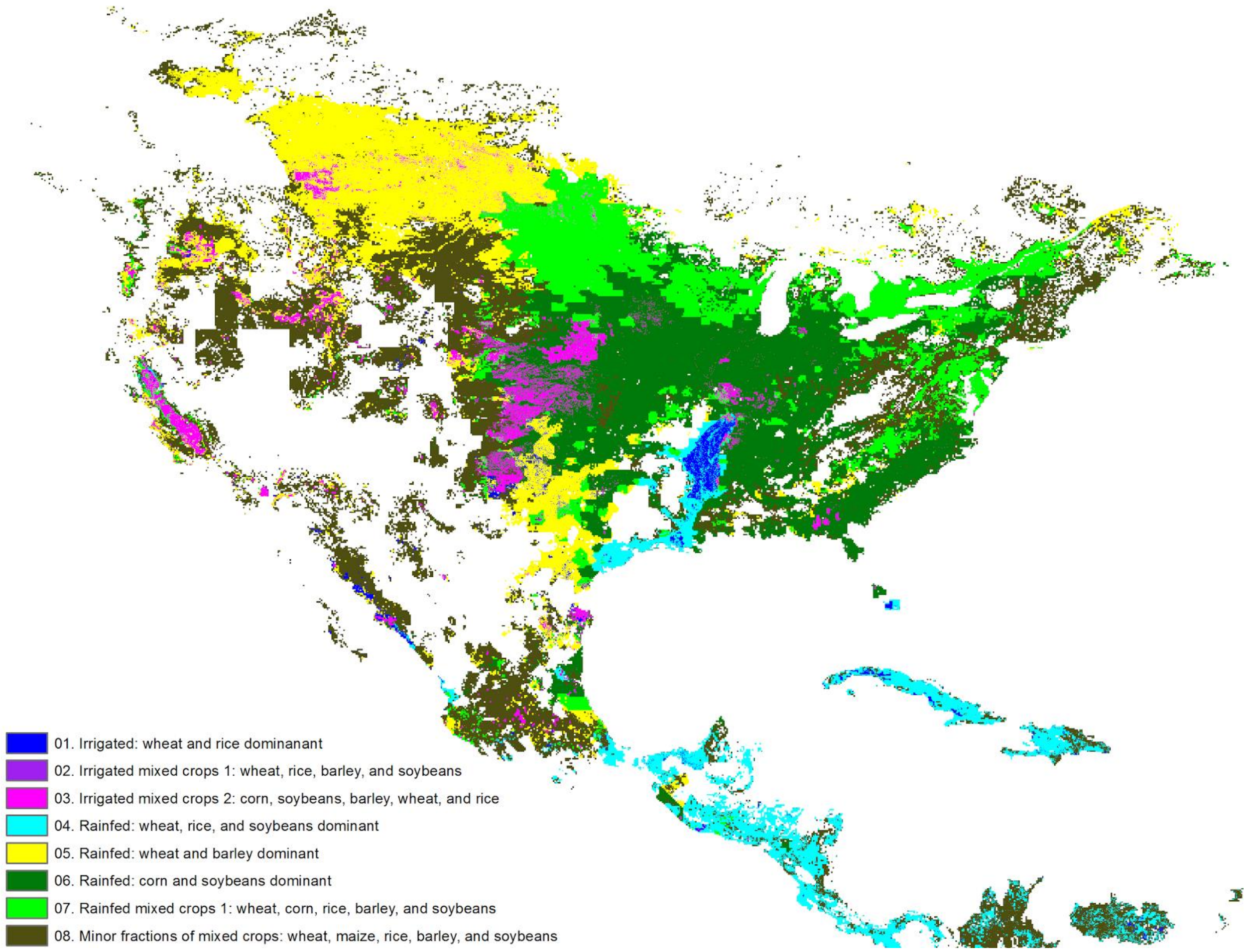


Figure 2b. GCE 1km Crop Dominance (aka GCE V0.0). Figure here shows the USA part of the product. Classes 1 to 3 are croplands that are irrigated. Classes 4 to 7 are croplands that are rainfed. Class 8 is overwhelmingly non-croplands, but have very small fractions of croplands (refer to Thenkabail et al., 2012). Each class has some combination of croplands dominating in them.

2.1.1.1 Product and ATBT References GCE 1km Crop Dominance (aka GCE V0.0):

The GCE 1km Crop Dominance (aka GCE V0.0) is precursor to much more refined 250m and 30 m products. A number of publications pertaining to GCE 1km Crop Dominance (aka GCE V0.0) are given below, in order of importance for the product:

Thenkabail P.S., Knox J.W., Ozdogan, M., Gumma, M.K., Congalton, R.G., Wu, Z., Milesi, C., Finkral, A., Marshall, M., Mariotto, I., You, S. Giri, C. and Nagler, P. 2012. Assessing future risks to agricultural productivity, water resources and food security: how can remote sensing help?. Photogrammetric Engineering and Remote Sensing, August 2012 Special Issue on Global Croplands: Highlight Article. 78(8): 773-782.

Thenkabail, P.S., Hanjra, M.A., Dheeravath, V., Gumma, M. 2011. Book Chapter # 16: Global Croplands and Their Water Use Remote Sensing and Non-Remote Sensing Perspectives. In the Book entitled: "Advances in Environmental Remote Sensing: Sensors, Algorithms, and Applications". Taylor and Francis Edited by Dr. Qihao Weng. Pp. 383-419.

Thenkabail. P., Lyon, G.J., Turral, H., and Biradar, C.M. 2009b. Book entitled: "Remote Sensing of Global Croplands for Food Security" (CRC Press- Taylor and Francis group, Boca Raton, London, New York. Pp. 556 (48 pages in color). Published in June, 2009.
Reviews of this book:

http://www.crcpress.com/product/isbn/9781420090093
http://gfmt.blogspot.com/2011/05/review-remote-sensing-of-global.html

Thenkabail, P.S., GangadharaRao, P., Biggs, T., Krishna, M., and Turral, H., 2007. Spectral Matching Techniques to Determine Historical Land use/Land cover (LULC) and Irrigated Areas using Time-series AVHRR Pathfinder Datasets in the Krishna River Basin, India. Photogrammetric Engineering and Remote Sensing. 73(9): 1029-1040. (Second Place Recipients of the 2008 John I. Davidson ASPRS President's Award for Practical papers).

Biradar, C.M., Thenkabail. P.S., Noojipady, P., Li, Y.J., Dheeravath, V., Velpuri, M., Turral, H., Cai, X.L., Gumma, M., Gangalakunta, O.R.P., Schull, M., Alankara, R.D., Gunasinghe, S., and Xiao, X. 2009. Book Chapter 15: Global map of rainfed cropland areas (GMRCA) and statistics using remote sensing. Pp. 357-392. In the book entitled: "Remote Sensing of Global Croplands for Food Security" (CRC Press- Taylor and Francis group, Boca Raton, London, New York. Pp. 475. Published in June, 2009. (Editors: Thenkabail. P., Lyon, G.J., Biradar, C.M., and Turral, H.).

Thenkabail P.S., Wu Z. An Automated Cropland Classification Algorithm (ACCA) for Tajikistan by Combining Landsat, MODIS, and Secondary Data. Remote Sensing. 2012; 4(10):2890-2918.
Download the paper @ this link: http://www.mdpi.com/2072-4292/4/10/2890
ACCA algorithm at this link: https://powellcenter.usgs.gov/globalcroplandwater/content/models-algorithms

Thenkabail, P.S. 2012. Principal Investigator. USGS Global croplands, their water use & food security Web\Data Portal, (https://powellcenter.usgs.gov/globalcroplandwater/)

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., and Dutta, R. 2009a. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. International Journal of Remote Sensing. 30(14): 3679-3733. July, 20, 2009.

Thenkabail, P.S. 2012. Special Issue Foreword. Global Croplands special issue for the August 2012 special issue for Photogrammetric Engineering and Remote Sensing. PE&RS. 78(8): 787-788.

Thenkabail, P.S. 2012. Guest Editor for Global Croplands Special Issue. Photogrammetric Engineering and Remote Sensing. PE&RS. 78(8).

Wu, Z., Thenkabail, P.S., Zakzeski, A., Mueller, R., Melton, F., Rosevelt, C., Dwyer, J., Johnson, J., Verdin, J. P., 2014. Seasonal cultivated and fallow cropland mapping using modis-based automated cropland classification algorithm. J. Appl. Remote Sens. 0001;8(1):083685. doi:10.1117/1.JRS.8.083685.

Wu, Z., Thenkabail, P.S., and Verdin, J. 2014. An automated cropland classification algorithm (ACCA) using Landsat and MODIS data combination for California. Photogrammetric Engineering and Remote Sensing. Vol. 80(1): 81-90.
****

2.1.2 GCE 1km Multi-study Crop Mask (aka GCE V1.0)

This is Global Food Security-support Analysis Data (GFSAD) project's 'GCE' (Global Cropland Extent) 1km Multi-study Crop Mask product (e.g., Figure 3a, 3b). GCE V1.0 product was produced by using 4 baseline products:

1.Thenkabail et al. (Thenkabail et al., 2009b, Biradar et al., 2009, Thenkabail et al., 2011)
   ~1 km global cropland product
2. Pittman et al. (2010)
   250 m global cropland product
3. Yu et al., (2013)
   30 m global cropland product
4. Friedl et al (2010)
   500 m global land use\land cover product

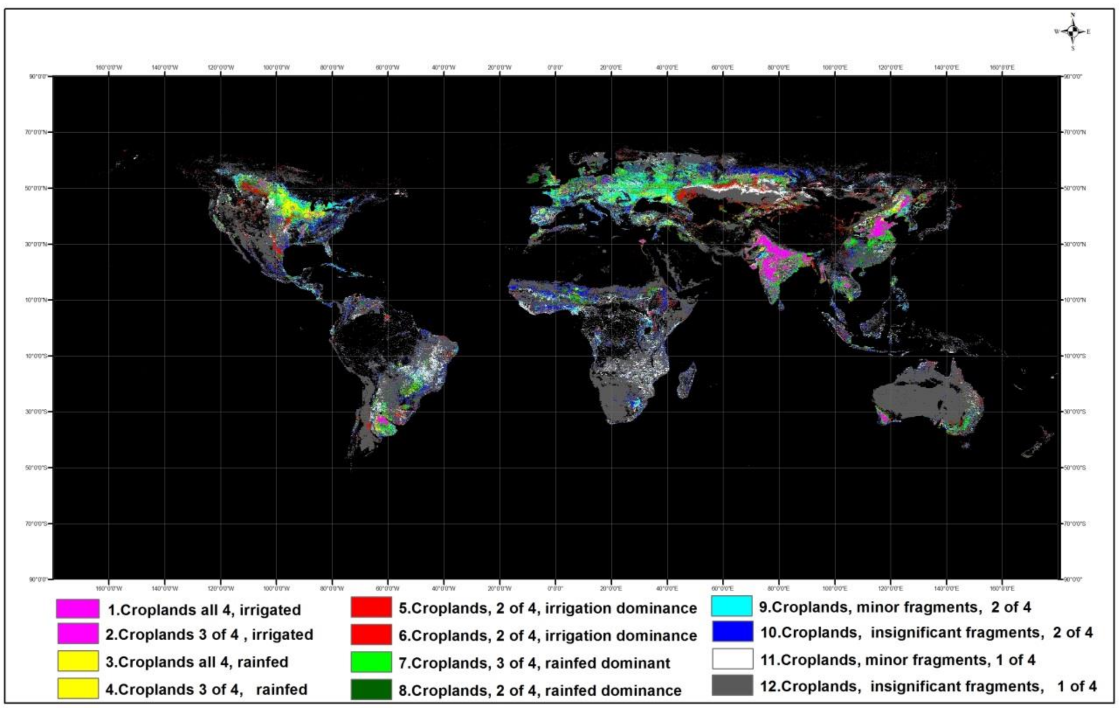
The 'GCE' (Global Cropland Extent) 1km Multi-study Crop Mask product will be:

- Cropland extent and areas;
- Cropland watering method: irrigation versus rainfed

One can download these data from:
http://geography.wr.usgs.gov/science/croplands/

This Global Food Security-support Analysis Data (GFSAD) Project's Global Cropland Extent Product at nominal 1km (GCE 1Km Crop Dominance). This Global Food Security-support Analysis Data (GFSAD) Project's Global Cropland Extent Product at nominal 1km from multi-study crop mask (aka GCE V1.0) provides spatial distribution of a disaggregated five class global cropland extent map derived at nominal 1-km based on four major studies: Thenkabail et al. (2009a, 2011), Pittman et al. (2010), Yu et al. (2013), and Friedl et al. (2010). Classes 1 to Class 5 are cropland classes, that are dominated by irrigated and rainfed agriculture. Class 4 to and Class 5 have minor/very minor fractions of croplands.

GCE 1KM Multi-study Crop Mask (see Figure 2 below) is a 5 class digital product that provides, at nominal 1 km, information on global: 1. cropland extent\areas, 2. irrigated versus rainfed cropping. There is no crop type or crop type dominance information. Cropping intensity (single, double, triple, and continuous crops) can be obtained for every pixel using time-series remote sensing data.



Legend

Figure 3a. A disaggregated twelve class global cropland extent map derived at nominal 1-km based on four major studies: Thenkabail et al. (2009a, 2011), Pittman et al. (2010), Yu et al. (2013), and Friedl et al. (2010). Class 1 to Class 9 are cropland classes, that are dominated by irrigated and rainfed agriculture. Class 10 to and Class 12 have ONLY minor or very minor fractions of croplands. Refer to Teluguntla et al., 2015 for details.

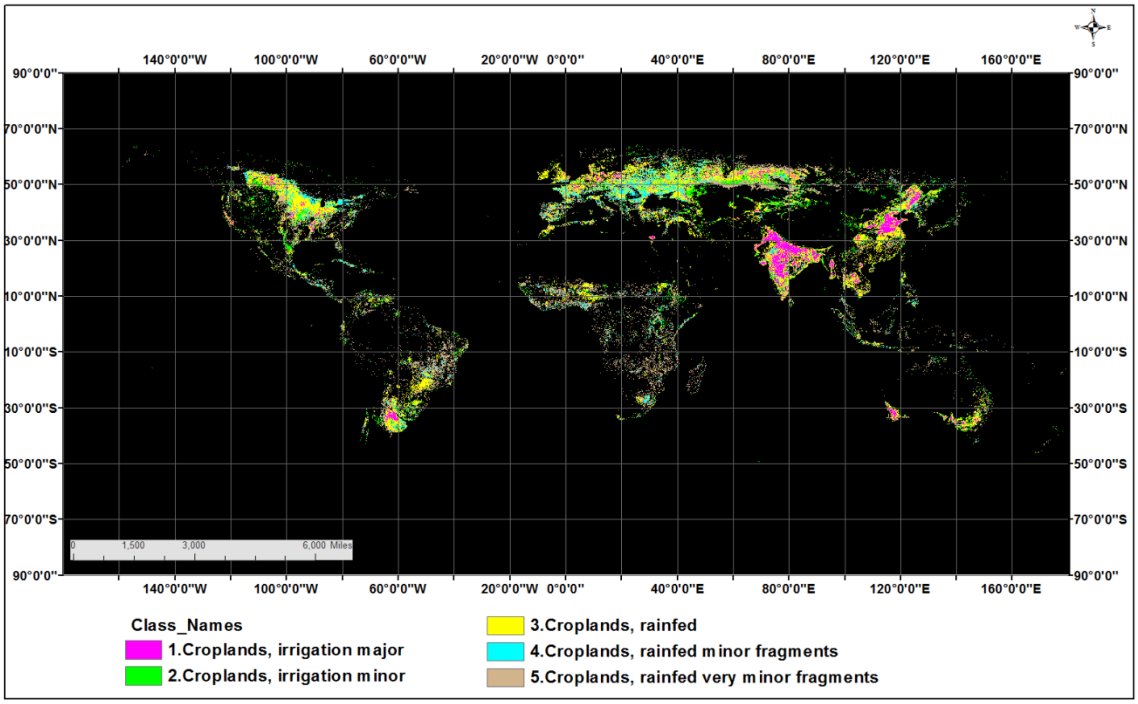


Figure 3b. GCE 1km Multi-study Crop Mask (aka GCE V1.0). Figure here shows the global product. A disaggregated five class global cropland extent map derived at nominal 1-km based on four major studies: Thenkabail et al. (2009a, b, 2011), Pittman et al. (2010), Yu et al. (2013), and Friedl et al. (2010). Classes 1 to Class 5 are cropland classes, that are dominated by irrigated and rainfed agriculture. However, class 4 and Class 5 have ONLY minor or very minor fractions of croplands. Refer to Table 6.7c for cropland statistics of this map. **Note**: **Irrigation major**: areas irrigated by large reservoirs created by large and medium dams, barrages and even large ground water pumping. **Irrigation minor**: areas irrigated by small reservoirs, irrigation tanks, open wells, and other minor irrigation. However, it is very hard to draw a strict boundary between major and minor irrigation and in places there can be significant mixing. So, when major irrigated areas such as the Ganges basin, California's central valley, Nile basin etc. are clearly distinguishable as major irrigation, in other areas major and minor irrigation may inter-mix. Refer to Teluguntla et al., 2015 for details.

2.1.2.1 Product and ATBT References for GCE 1km Multi-study Crop Mask (aka GCE V1.0):

The GCE 1km Crop Dominance (aka GCE V0.0) is precursor to much more refined 250m and 30 m products. A number of publications pertaining to GCE 1km Crop Dominance (aka GCE V0.0) are given below, in order of importance for the product:

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N.S., and Huang, X.M. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *REMOTE SENSING OF ENVIRONMENT*, 114(1), 168-182.

Pittman K., Hansen M.C., Becker-Reshef I., Potapov P.V., Justice C.O. Estimating Global Cropland Extent with Multi-year MODIS Data. Remote Sensing. 2010; 2(7):1844-1863.

Teluguntla, P., Thenkabail, P.S., Xiong, J., Gumma, M.K., Giri, C., Milesi, C., Ozdogan, M., Congalton, R., Tilton, J., Sankey, T.R., Massey, R., Phalke, A., and Yadav, K. 2015. Global Cropland Area Database (GCAD) derived from Remote Sensing in Support of Food Security in the Twenty-first Century: Current Achievements and Future Possibilities. Land Resources: Monitoring, Modelling, and Mapping. Chapter 6, Vol. II., Remote Sensing Handbook. Three Volume Book edited by Prasad S. Thenkabail. Taylor and Francis Inc.\CRC press. Accepted, in Press.

Thenkabail, P.S., Hanjra, M.A., Dheeravath, V., Gumma, M. 2011. Book Chapter # 16: Global Croplands and Their Water Use Remote Sensing and Non-Remote Sensing Perspectives. In the Book entitled: "Advances in Environmental Remote Sensing: Sensors, Algorithms, and Applications". Taylor and Francis Edited by Dr. Qihao Weng. Pp. 383-419.

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., and Dutta, R. 2009a. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. International Journal of Remote Sensing. 30(14): 3679-3733. July, 20, 2009.

Thenkabail, P. S., Lyon, G. J., Turral, H., & Biradar, C. M. (2009b). *Remote Sensing of Global Croplands for Food Security*. Boca Raton, London, New York: CRC Press- Taylor and Francis Group, Published in June,.

Portmann, F., S. Siebert, and P. Döll, 2008. MIRCA2000 – Global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modelling, *Global Biogeochemical Cycles*, *GB0003435*.

Yu, L., Wang, J., Clinton, N., Xin, Q., Zhong, L., Chen, Y., and Gong, P. 2013. International Journal of Digital Earth. FROM-GC: 30 m global cropland extent derived through multisource data integration, *International Journal of Digital Earth* ,DOI:10.1080/17538947.2013.822574.

2.2.1 250 m product

2.2.2 GCE 250m Crop Dominance (aka GCE V2.0)

This is Global Food Security-support Analysis Data (GFSAD) project's 'GCE' (Global Cropland Extent) 250km Crop Type and\or Dominance product. Cropland products released at this resolution will be:

- Cropland extent and areas;
- Cropland watering method: irrigation versus rainfed;
- Cropping intensity;

To a lesser extent

- Crop type and\or dominance

Currently, GCE 250m Crop Dominance (aka GCE V2.0) are produced by various sub-groups, for different continents or large areas (see Figure 4) using a wide array of methods and approaches.

2.2.3 GCE 250m Crop Dominance (aka GCE V2.0) for Australia using: (a) spectral matching techniques (SMTs), and (b) automated cropland classification methods (ACCA) (Pardhasaradhi Teluguntla et al.)

For Australia, extensive field campaign was conducting in October, 2014. This resulted in obtaining ~2500 data points spread across cropland areas of Australia (e.g., Figure 5, 6). These precise locations will be made use to gather ideal spectra based on time-series remote sensing such as MODIS 250 m NDVI (e.g., Figure 7). Then the spectral matching technique (SMT; Thenkabail et al., 2007) will be adopted to determine cropland classes. The SMTs involved matching class spectra (e.g., derived from classification) with ideal spectra (e.g., generated based on true knowledge of the farms) using time-series remote sensing (Figure 7). The two quantitative spectral matching techniques are:

A. Spectral correlation similarity R-square values (SCS R-square values);
B. Spectral similarity value (SSV).

In SCS R-square method, the class spectra are matched with ideal spectra based on magnitude of the spectra. In SSV, the class spectra are matched with ideal spectra based on both magnitude and shape. The resulting output will provide cropland classes (e.g., Figure 8). The Figure 8 is only a preliminary result that requires further development and validation. This will be done using the 2014 field data that is currently organized.

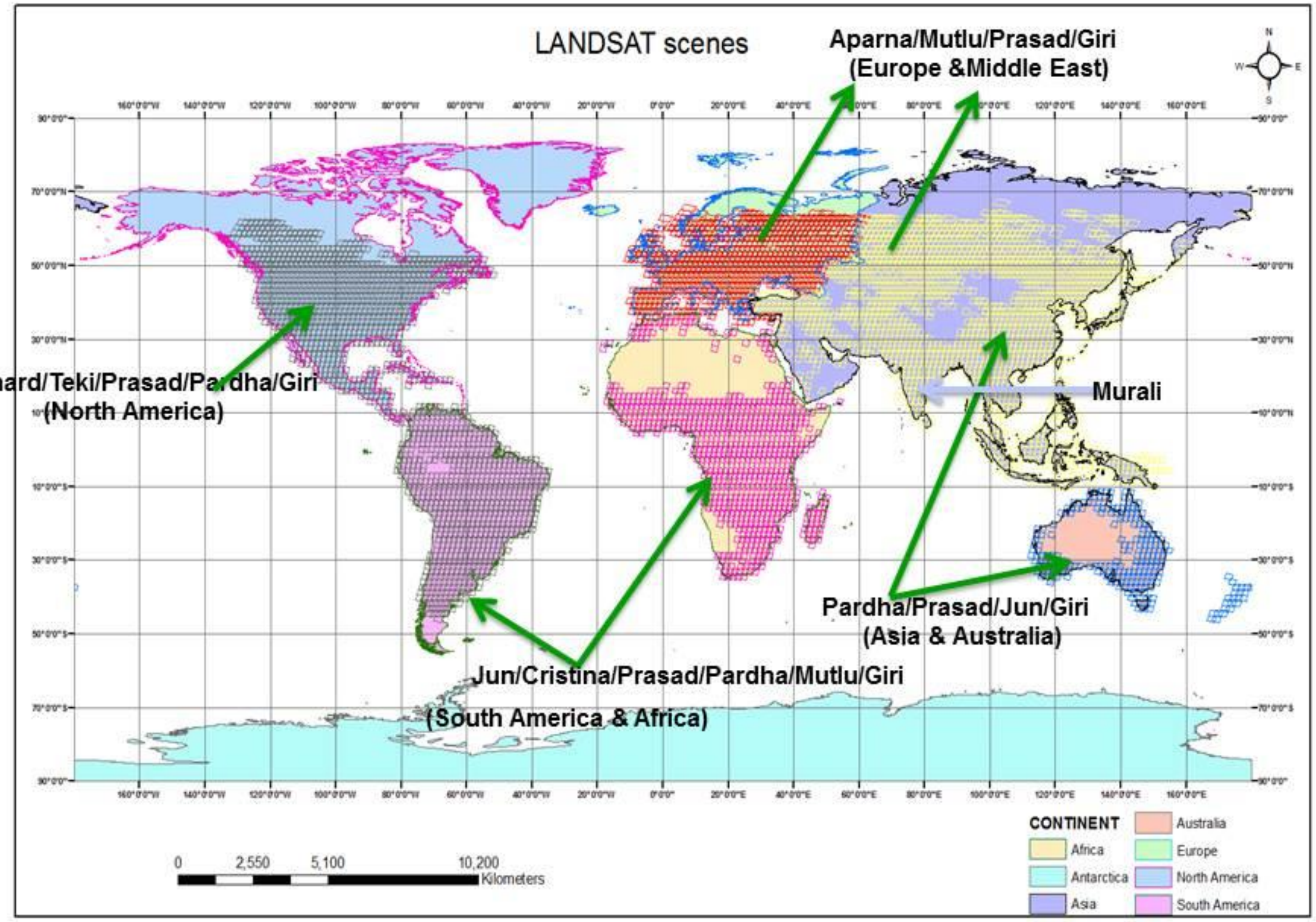


Table 4. Large study areas for which GCE 250m Crop Dominance (aka GCE V2.0) cropland products are produced by various sub-groups, for different continents or large areas using a wide array of methods and approaches.

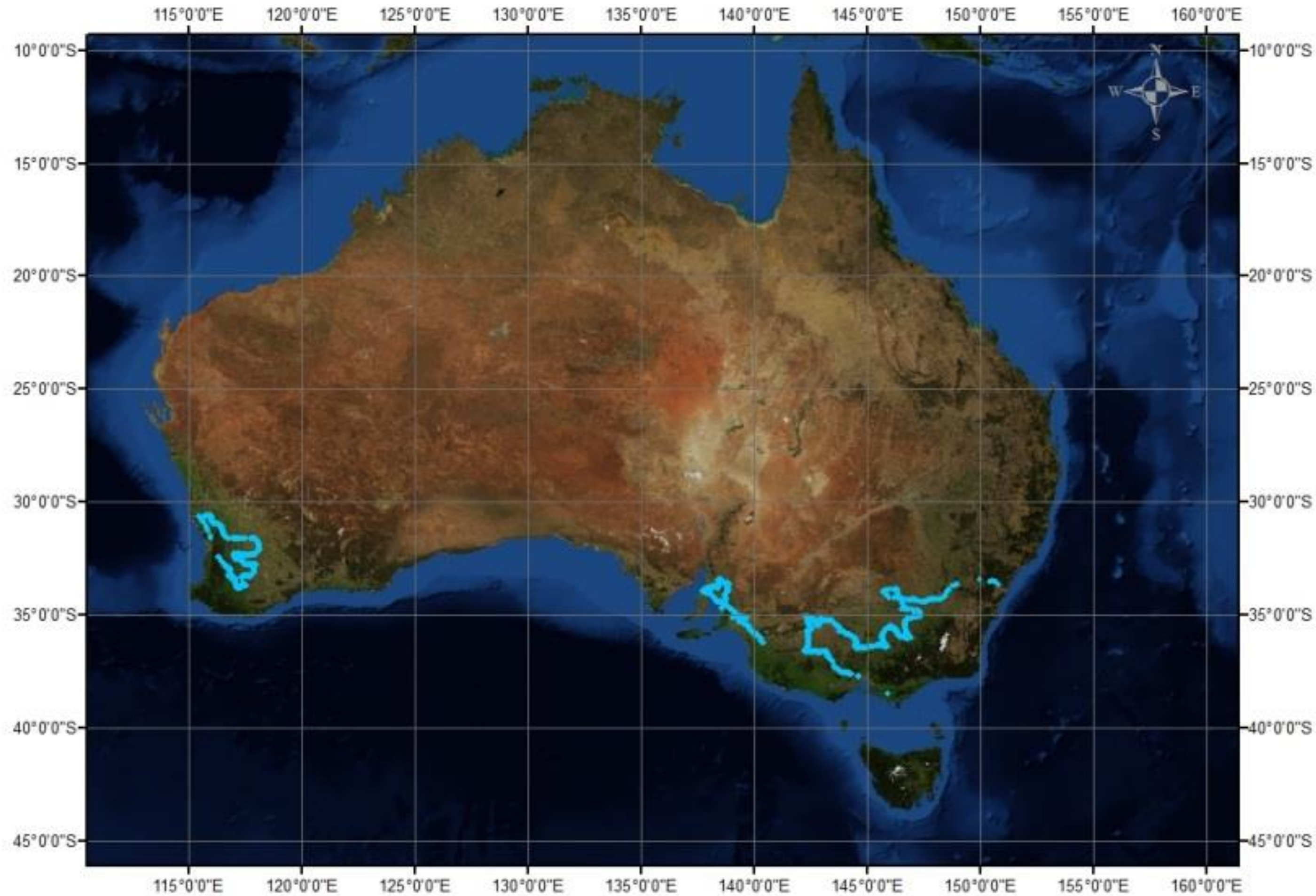


Figure 5. Distribution of ~2500 ground data points for Australia collected in October, 2014. These data points are distributed in the cropland areas of Australia.

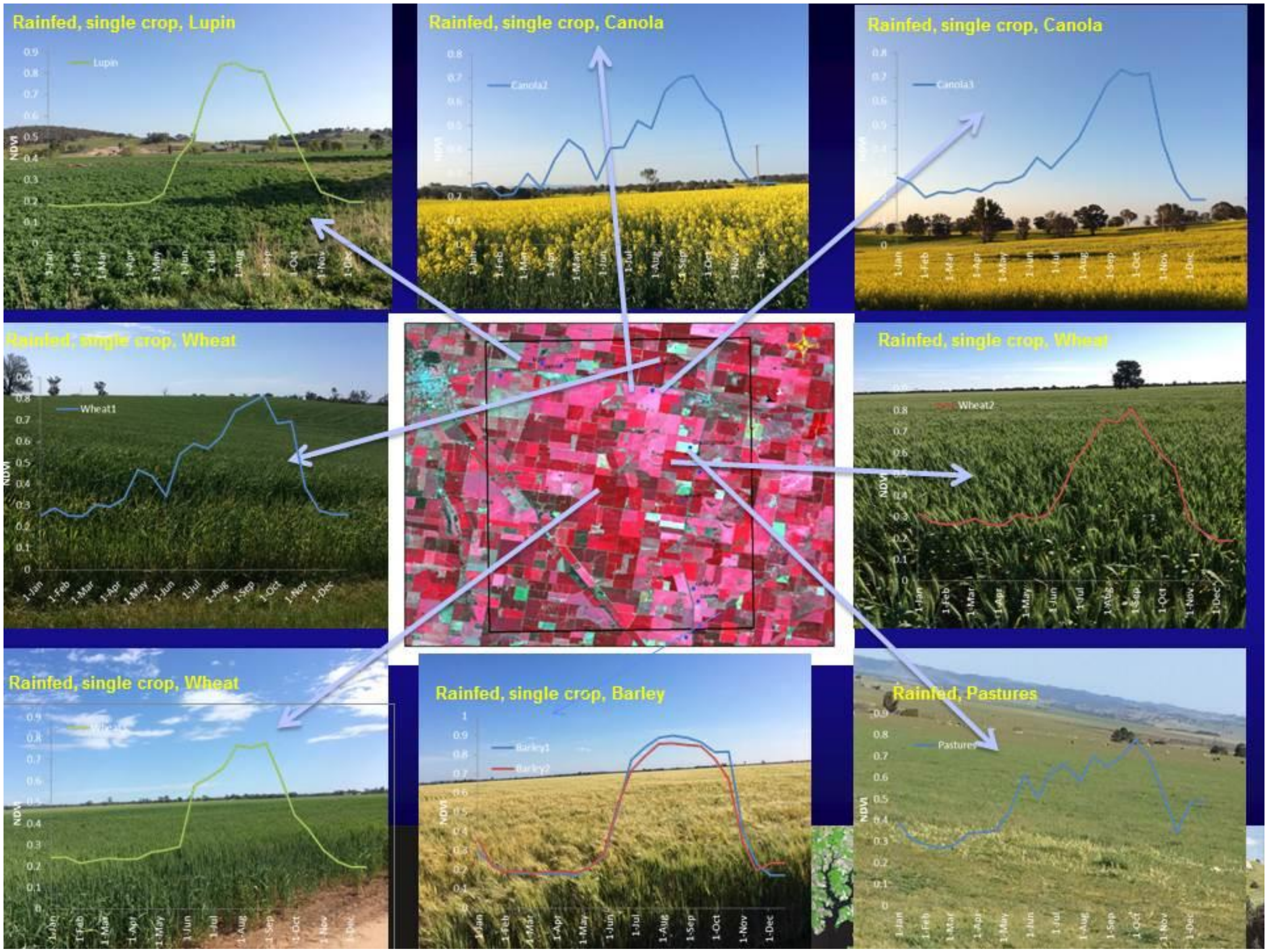


Figure 6. Illustration of few Australian ground data points. The signatures shown are derived from modis data for year 2001. However, we will develop similar signatures for year 2014 so as to correspond with the 2014 ground data.

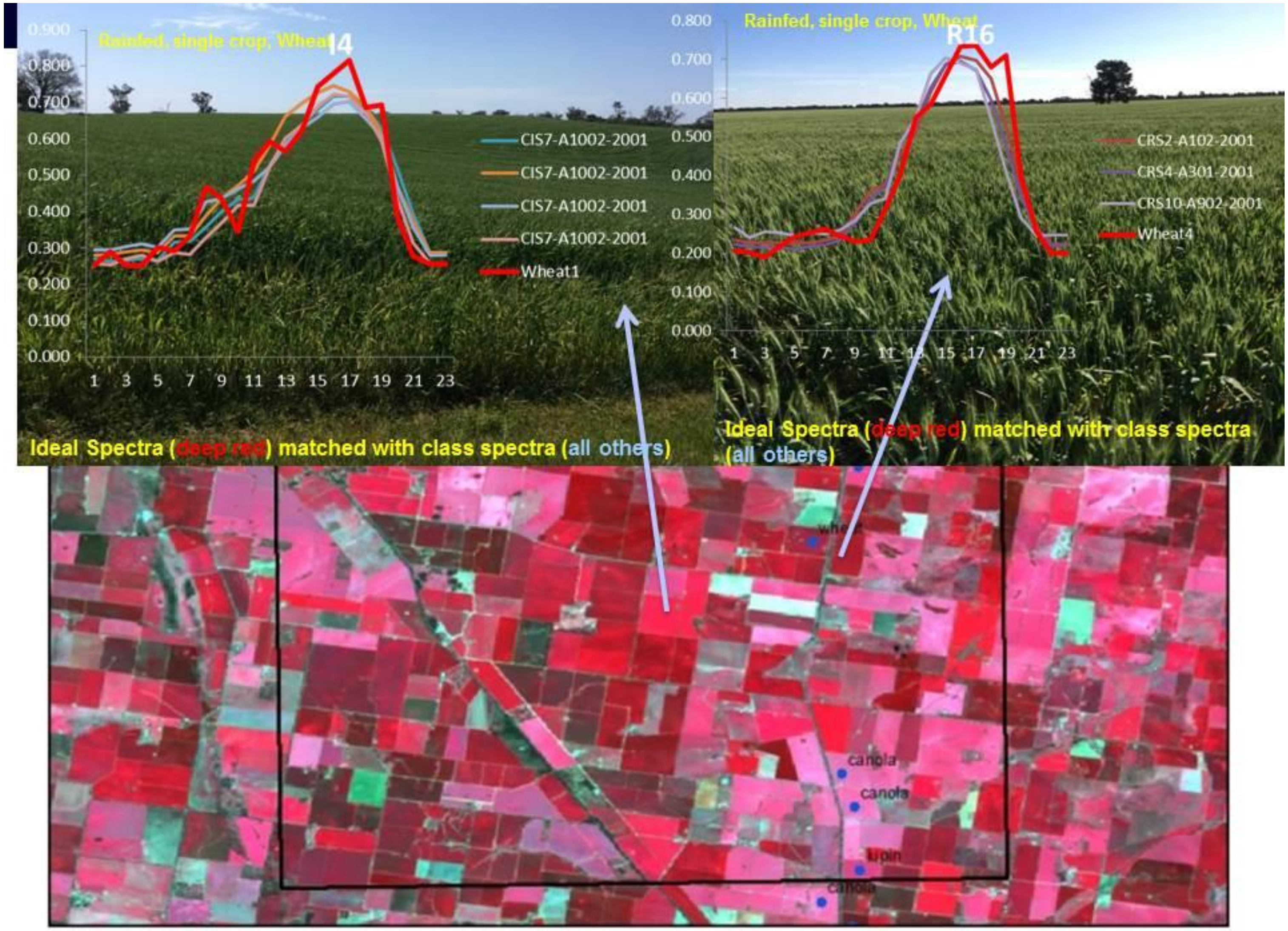


Figure 7. Qualitative spectral matching technique (SMT). Illustration of matching class spectra (all signatures except thick red line) with ideal spectra (thick red lines).

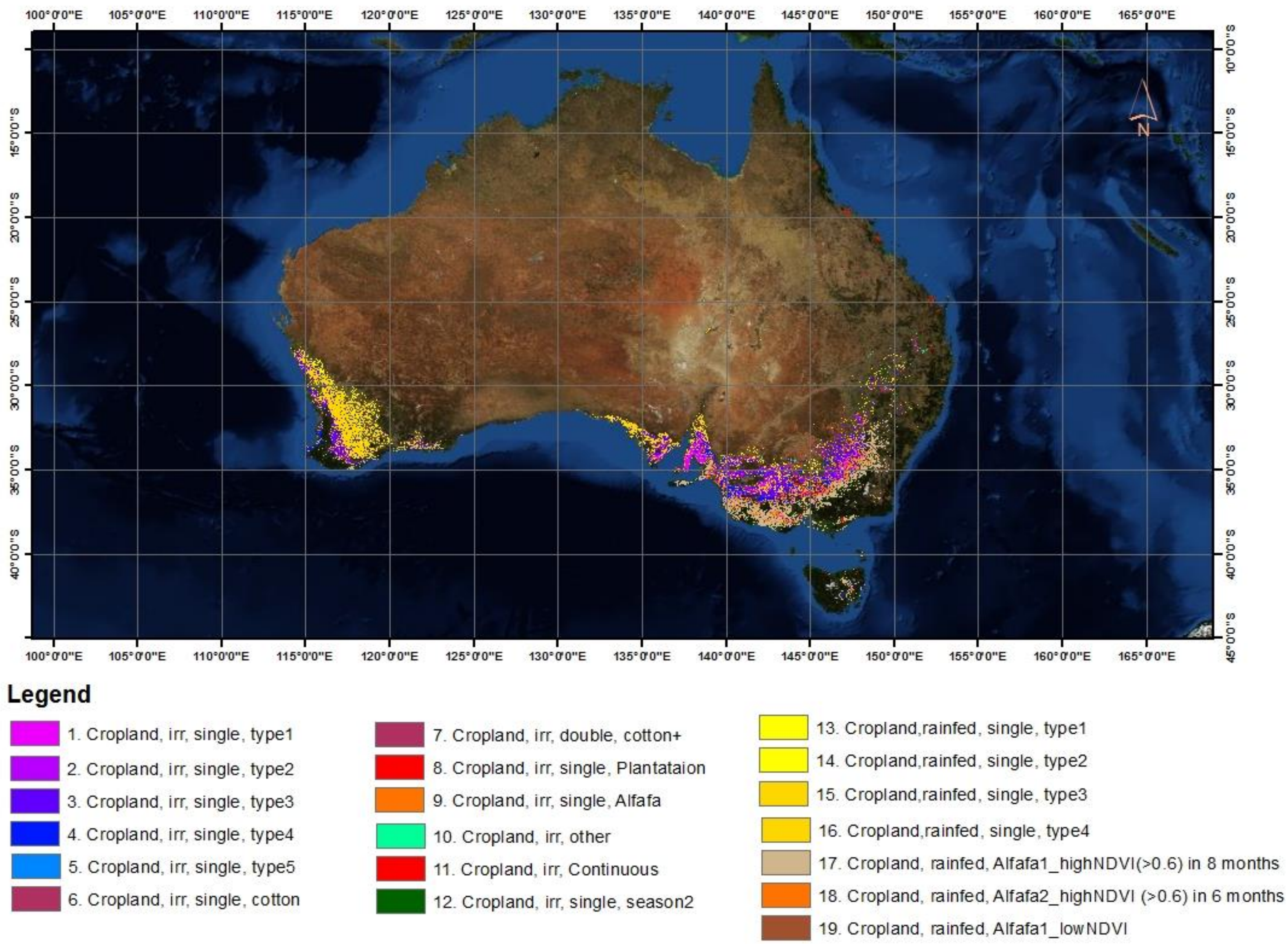


Figure 8. Cropland classes of Australia (preliminary) based on spectral matching technique. This is a preliminary product, currently in development.

2.2.4 **GCE 250m Crop Dominance (aka GCE V2.0) for Africa based on parallel k-means clustering (Jun Xiong et al.)**

Africa cropland products involve parallel computing using k-means clustering. The MODIS 250 m monthly maximum value composite (MVC) data of 11years (2003-2013) was used to derive the GCE 250m Crop Dominance (aka GCE V2.0) for Africa based on parallel k-means clustering. The process is illustrated in a nutshell in Figure 9. The algorithm is run on NASA NEX supercomputer. The resulting 500 cluster classes (Figure 10) are identified and labeled based on extensive ground data and very high resolution (sub-meter to 5 meter) imagery (VHRI) as illustrated in Figure 11 and 12. This leads to cropland products (e.g., Figure 13a, 13b). However, Figure 13a and 13b are only very preliminary products that require substantially more refinement.

The primary challenges for Africa study comes from: 1) Extremely limited local knowledge for training; 2) Highly heterogeneous mixture fields; and 3) Complexities of agriculture system and land management. In this approach, our efforts include FOUR parts: 1). Building standard reference signatures dataset from all-level relevant sources; 2) Stacking and Clustering 253-band MODIS 250m NDVI Time-series Dataset; 3) Identification of generic clusters with signatures using semi-automatically algorithm; 4) Accuracy Assessment of the product.

In step 1, derived reference samples were established based on the interpretation of VHRI (very high resolution images) and temporal profile of MODIS NDVI Data. A web-based system is in development to provide a platform to collect / interpret more reference samples in the entire scientific community rather than limited experts, to form an updating knowledgebase for flexible classifier development. From this method, we've create a reference bank including ~6K samples all over continental Africa, from field work, thematic agriculture atlas, and research publication, and the number of resamples is still increasing.

In step 2, MODIS NDVI Time-series dataset (60GB) in continental Africa was pre-processed. The huge 253-bands dataset was clustered into 500 generic clusters based on the characteristics of NDVI Time Series, after non-crop area was masked out by GCEV1.0 and MODIS land use product (MCD12Q1).

In step 3, all 500 clusters were matched and voted according to the established reference samples bank. 47 groups was semi-automatically generated by the matching and voting algorithm and then 11 classes were identified in the final nominal 2000 GCEV2.3 @250m product.

It is the first time to utilize comprehensive techniques including remote sensing, parallel computing and cloud-deployed GIS service from Google Earth Engine, to a specific field like global cropland mapping. And the workflow designed to absorb coming information is still under developing. Independent accuracy assessment in Step 4 asks for more reference samples in cooperation with other teams like HSeg segmentation and accuracy assessment team. As requirement of project we will release the beta version of GCEV2.3 product for Africa on line, together with research paper, ATDB documentation and accuracy assessment report on schedule.

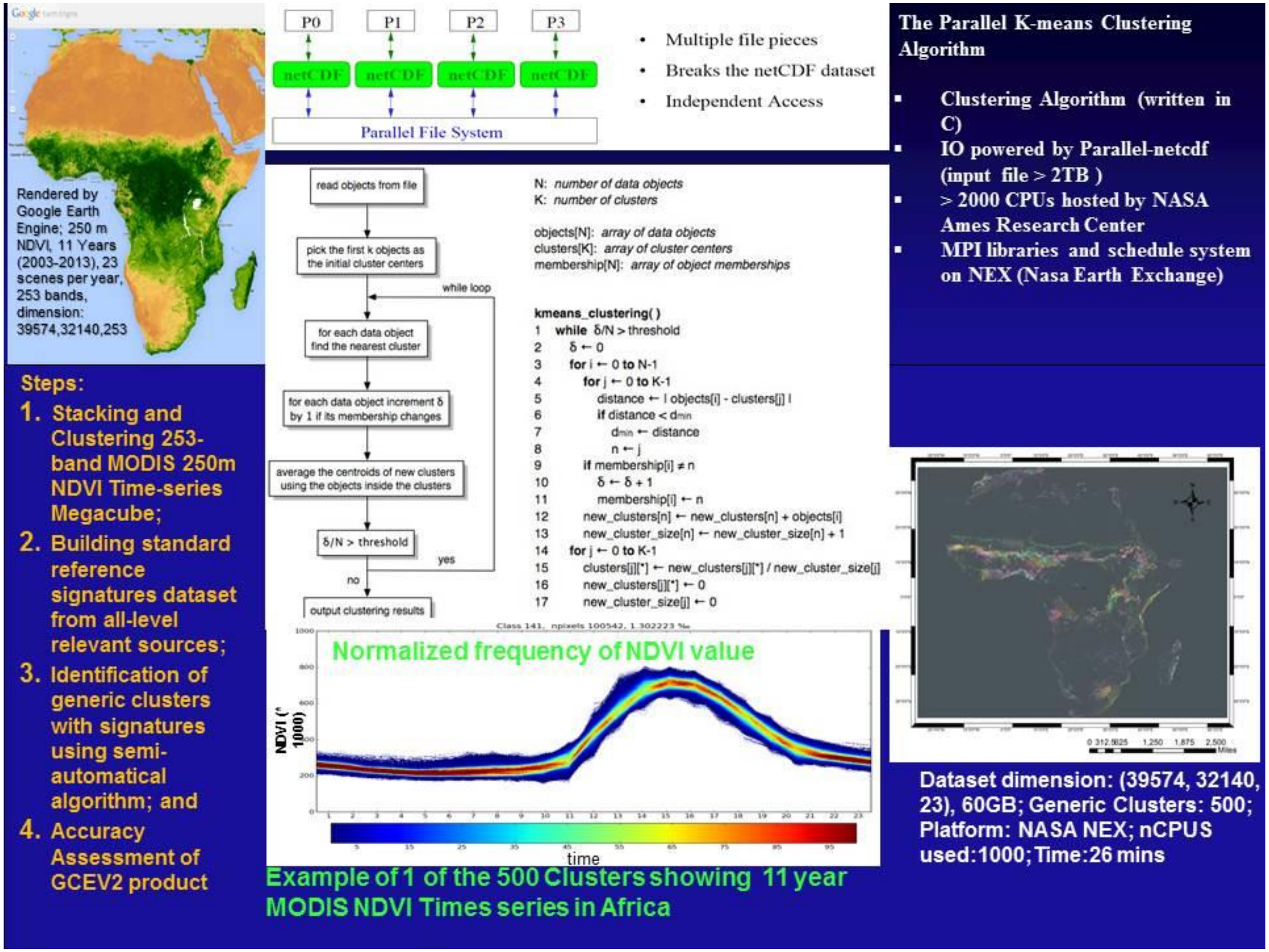


Figure 9. Parallel k-means clustering algorithm running on NASA NEX super computer to determine cropland clusters using time-series MODIS 250m NDVI data rendered by Google Earth Engine (GEE). The use of GEE will help us in interactively generate crop signatures corresponding to ground data points instantly. Also, it will help us view the very high resolution (sub-meter to 5 meter) imagery (VHRI) data in background of each class derived using parallel k-means cluster algorithm.

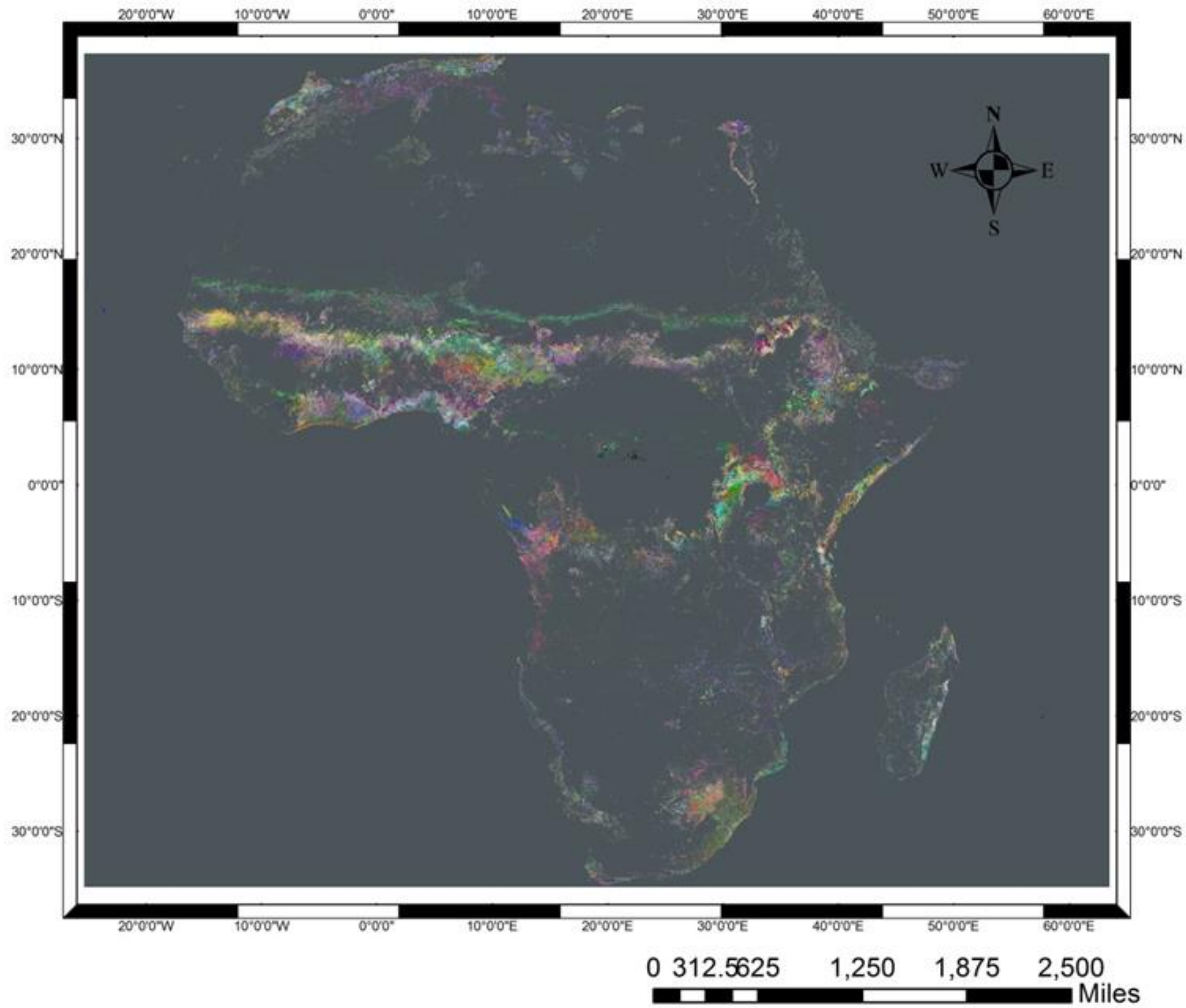
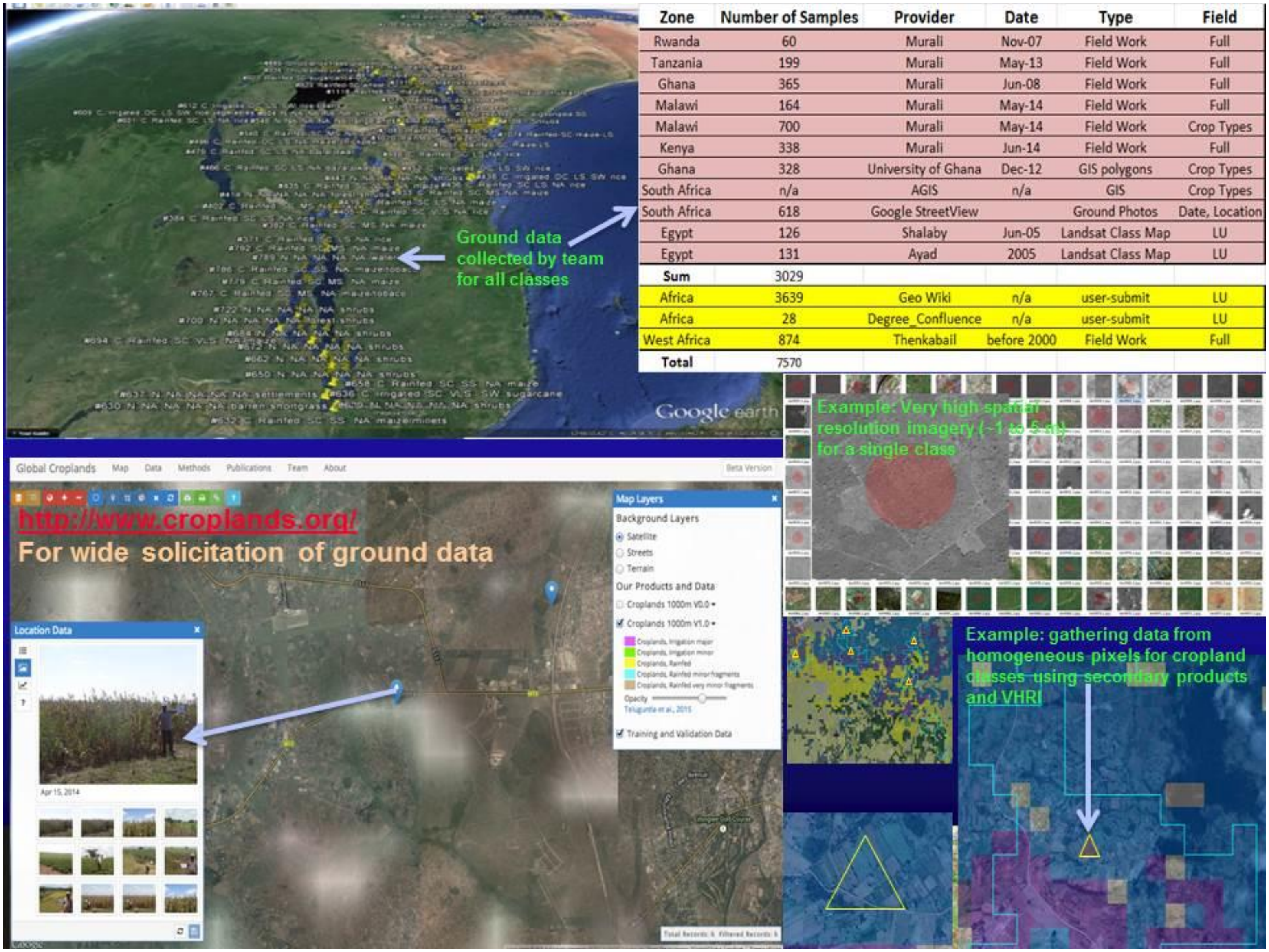


Figure 10. Initial 500 cluster classes of Africa @ 250m resolution based on parallel k-means algorithm running on NASA NEX supercomputer. The clusters are obtained by computing on MODIS 250m monthly maximum value composite (MVC) NDVI data for 11 years (2003-2013).



Zone	Number of Samples	Provider	Date	Type	Field
Rwanda	60	Murali	Nov-07	Field Work	Full
Tanzania	199	Murali	May-13	Field Work	Full
Ghana	365	Murali	Jun-08	Field Work	Full
Malawi	164	Murali	May-14	Field Work	Full
Malawi	700	Murali	May-14	Field Work	Crop Types
Kenya	338	Murali	Jun-14	Field Work	Full
Ghana	328	University of Ghana	Dec-12	GIS polygons	Crop Types
South Africa	n/a	AGIS	n/a	GIS	Crop Types
South Africa	618	Google StreetView		Ground Photos	Date, Location
Egypt	126	Shalaby	Jun-05	Landsat Class Map	LU
Egypt	131	Ayad	2005	Landsat Class Map	LU
Sum	3029				
Africa	3639	Geo Wiki	n/a	user-submit	LU
Africa	28	Degree_Confluence	n/a	user-submit	LU
West Africa	874	Thenkabail	before 2000	Field Work	Full
Total	7570				

Figure 11. Process of class identification. Involves the use of ground data, very high resolution (sub-meter to 5 meter) imagery (VHRI), and time-series MODIS 250 m NDVI signatures of each cluster.

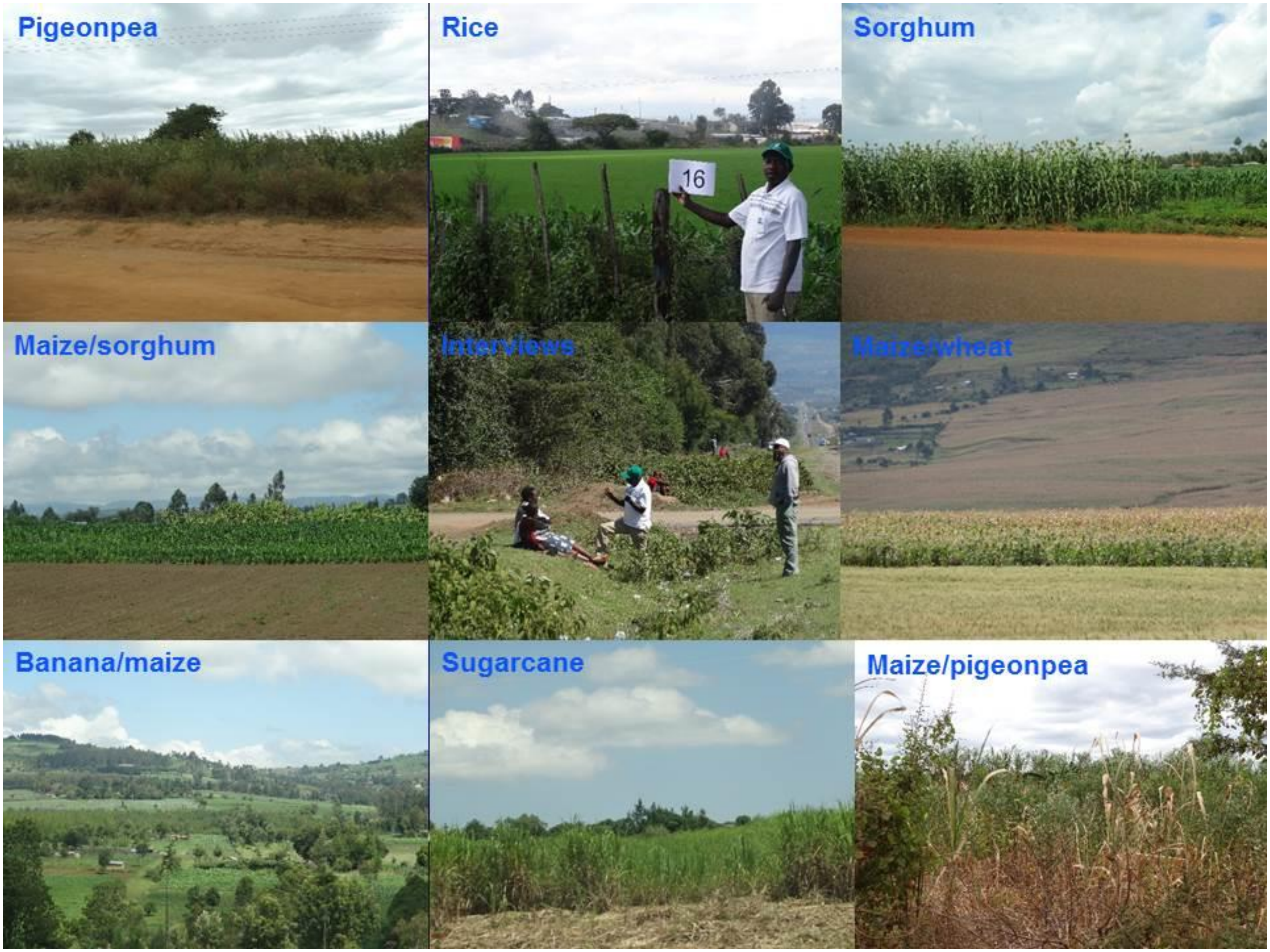


Figure 12. Typical ground data used for Africa. This data is now built into croplands.org.

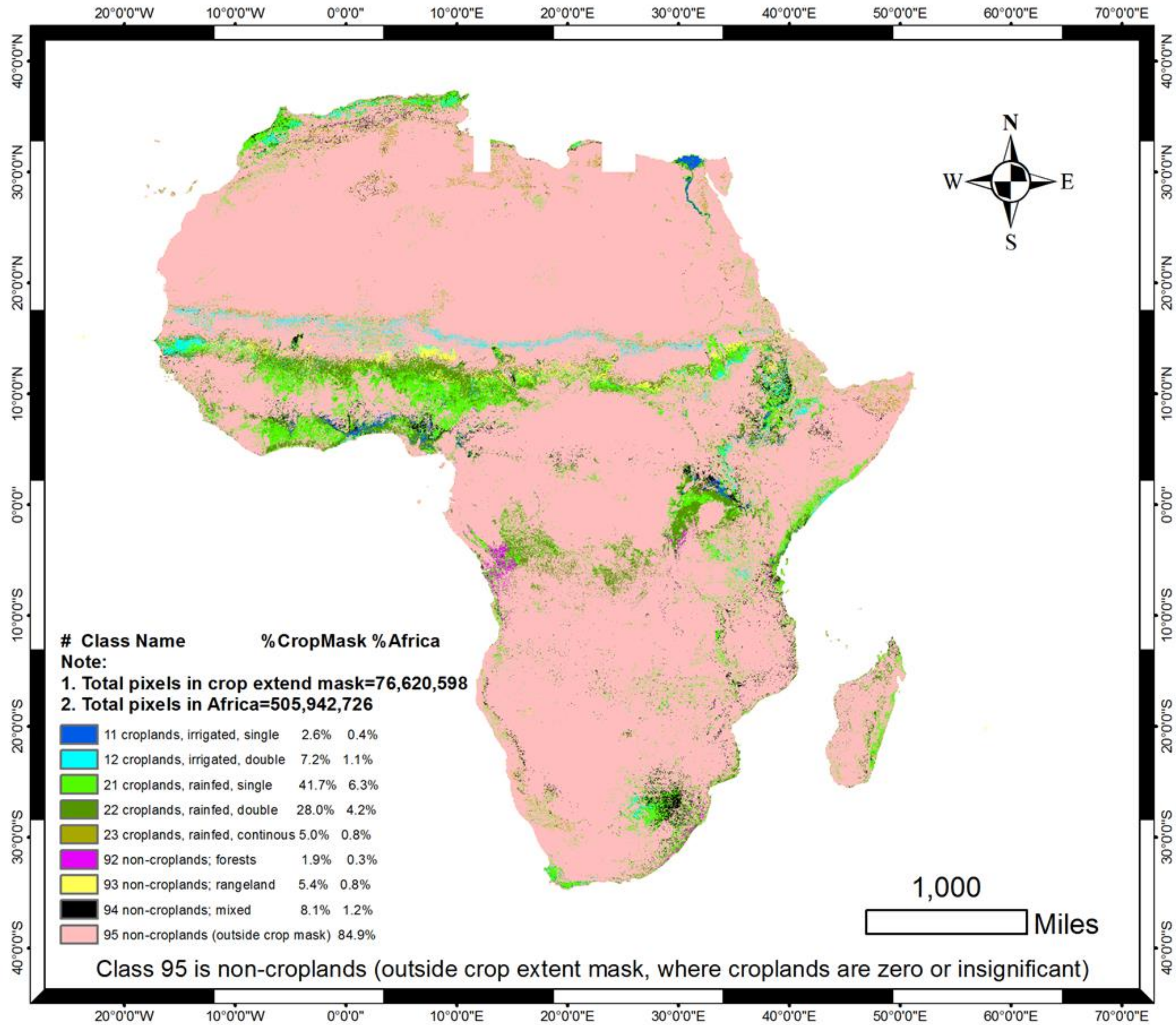


Figure 13a. Preliminary disaggregated cropland product of Africa using MODIS time-series 250 m spatial resolution data. This product is expected to significantly change with further analysis.

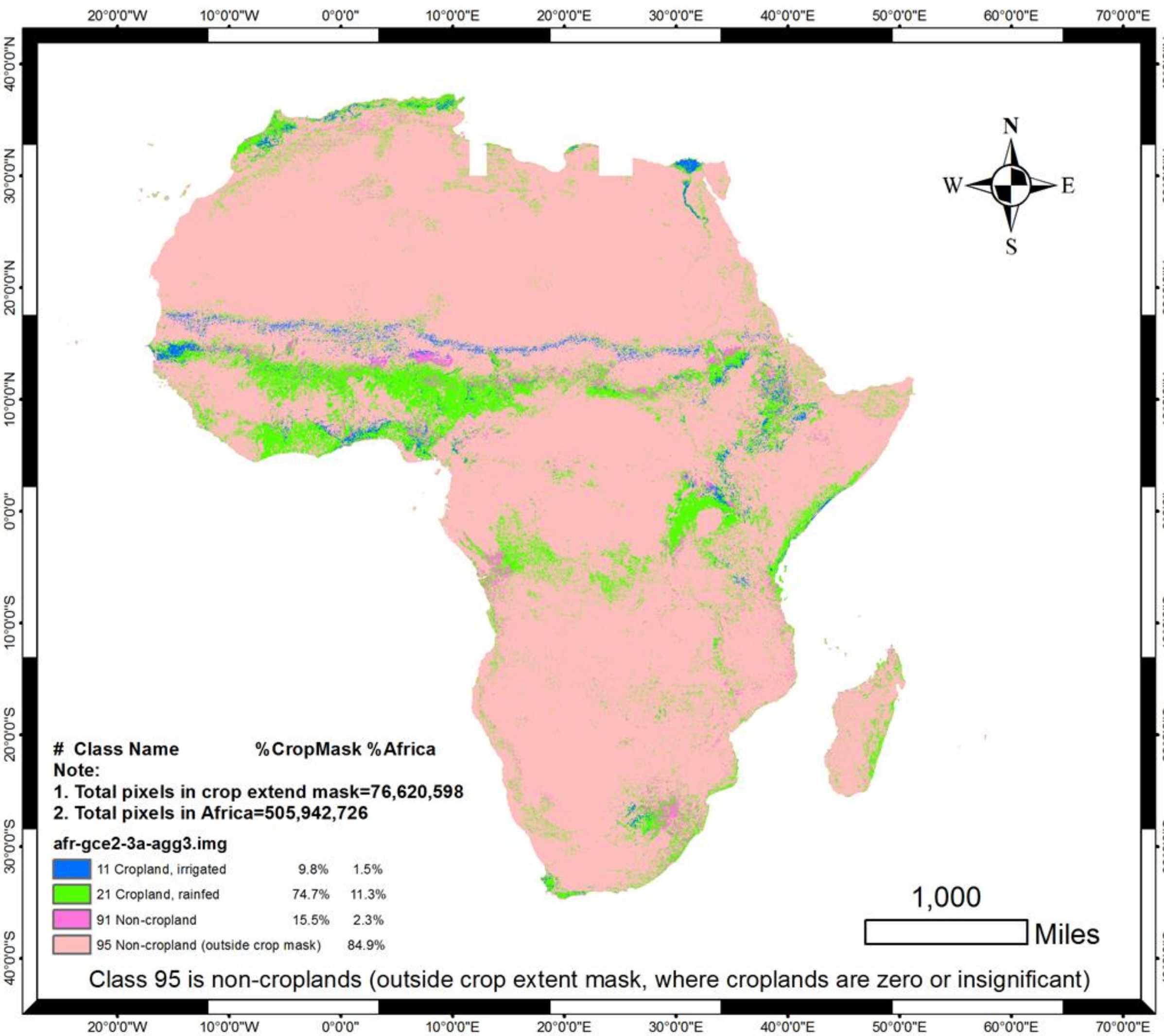


Figure 13b. Preliminary aggregated cropland product of Africa using MODIS time-series 250 m spatial resolution data. This product is expected to significantly change with further analysis.

2.2.5 GFSAD30 30 meter approach to cropland extent for Africa using Linear Spectral Mixture Model (LSMM) (Cristina Milesi et al.)

Characterizing cropland extent over Africa at 30m is challenging for two main reasons: 1) cropland areas are very fragmented and heterogeneous since few industrial farms exists and small farmers tend to grow a variety of crops at a very small scale; 2) the amount of good quality data from Landsat is limited because of persistent cloud cover over the tropical portion of the continent, the lack of recent Landsat 5 TM data and the data quality degradation of Landsat 7 data caused by SLC-off.

In this approach to 30m identification of cropland areas in Africa we exploit existing Landsat time series for the 2010 epoch to tease out the seasonal variability of agricultural fields and contrast it to the less variable surrounding natural covers. Cycles of fallow, emergence, growth to full canopy and harvest are typical of cropped fields and are responsible for the high inter-annual variability in the balance between the amounts of vegetation and substrate. Accurate measures of fractional cover of vegetation, bright and dark substrates can be obtained from unmixing the spectral response of Landsat data. The technique of unmixing satellite data relies on identifying spectral end members from pure pixels and then inverting a mixture model (e.g., Figure 14). While the identification of end members is often done on an image specific basis, here we use a set of global Landsat spectral end members so that a standardized spectral mixture model can be applied consistently across Landsat scenes and across time series. The goal is to characterize the seasonal variability in vegetation and substrate cover for the continent of Africa (e.g., Figure 14). This can be achieved by un-mixing time series of Landsat data for pre-emergence, emergence, peak of growing season and post-harvest images.

We test the approach both on time series from the archive database of Landsat data and from the early releases of global WELD data available on NEX. We show that the main combinations in which substrate and canopy cover over Africa are: 1) high vegetation/soil variability, 2) Fallow: high substrate with low variability (vegetation fraction close to zero); 3) Rangeland: low vegetation and low substrate fractions, low variability; 4) Forest, wetlands: high vegetation, high dark fractions, and low variability. It is expected that the identification of cropland extent from 30 m will be able to take advantage of the unique spectral signature of soil and substrate seasonal variation.

As part of this project we have developed platform independent codes for linear spectral un-mixing that can be deployed in parallel on the NASA Earth Exchange.

Next steps will include training of the WELD DATA with ground targets identified with the HSeg segmentation software and define a decision tree model that can efficiently separate cropland areas from surrounding non-cropped pixels.

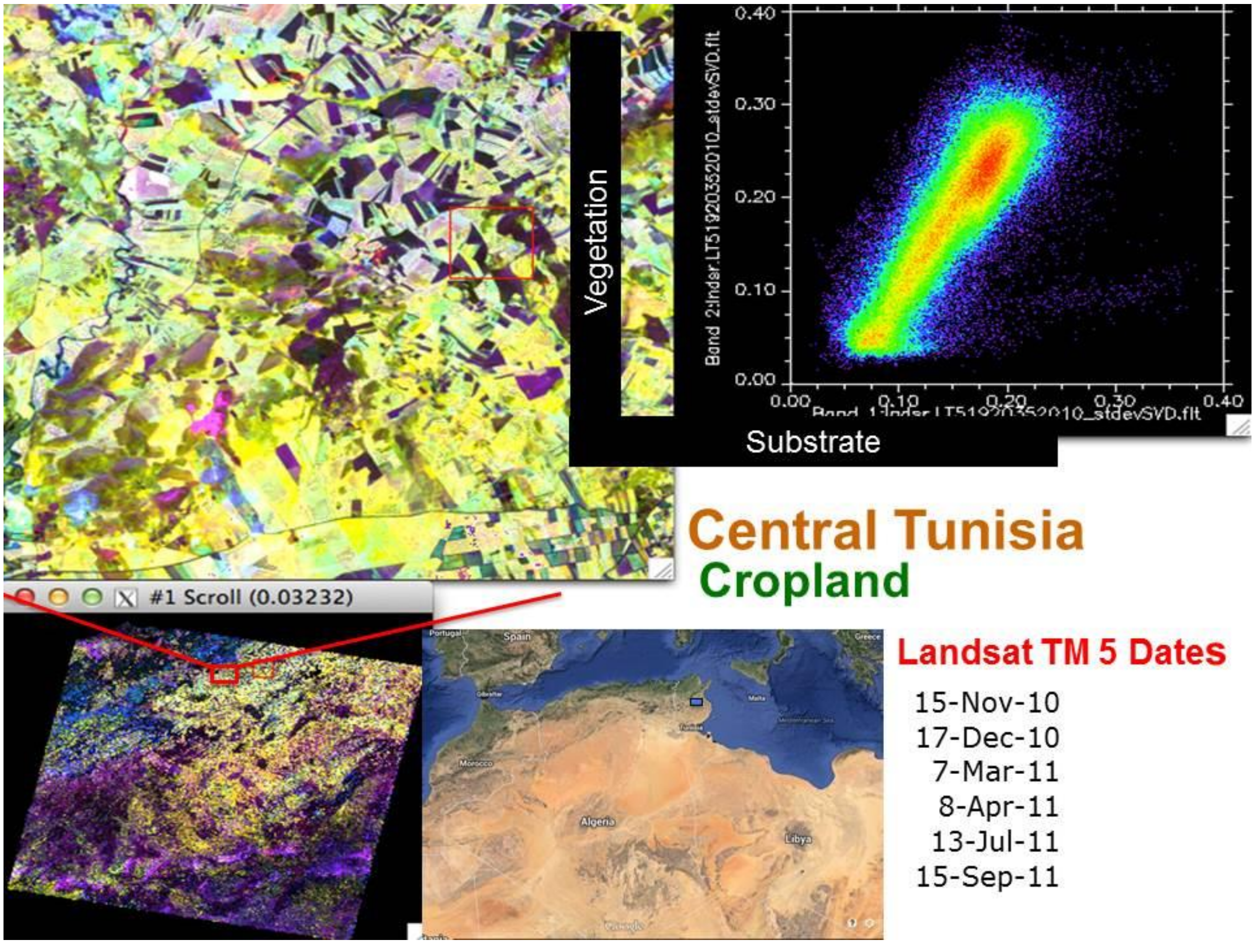
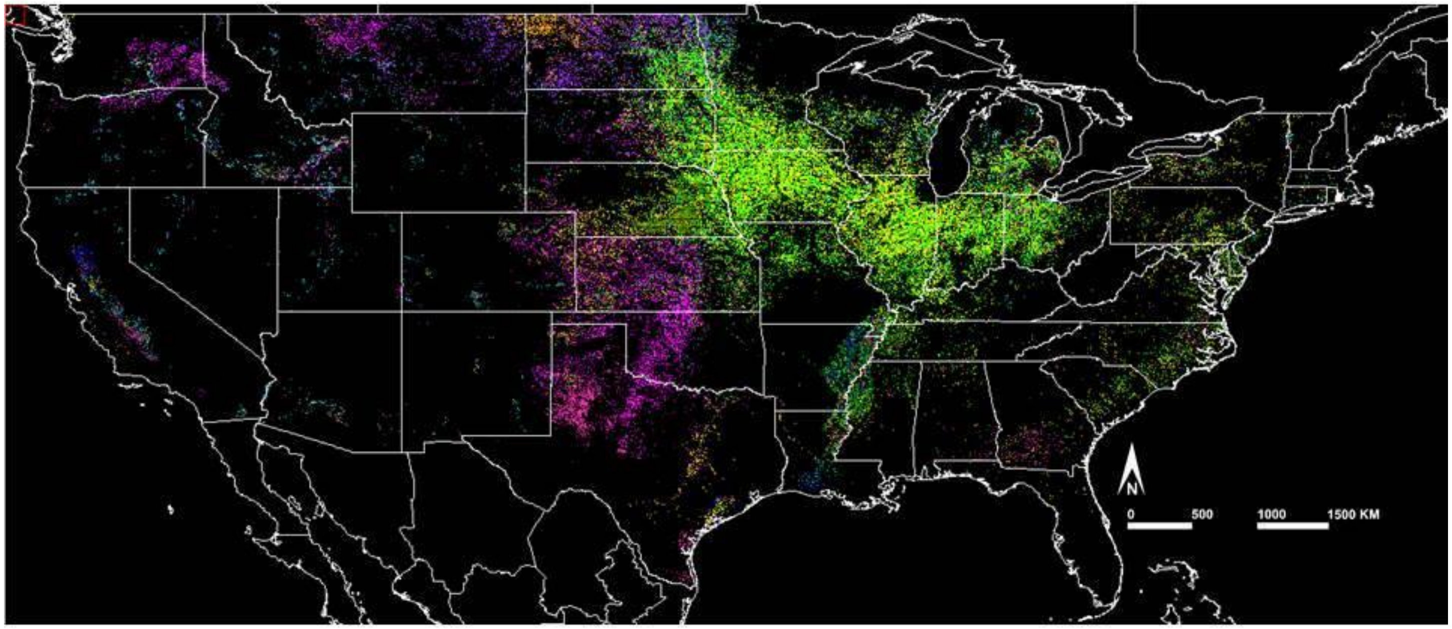


Figure 14. Illustration of inverting a standardized linear Spectral mixture model (LSMM) that converts the Landsat reflectance to sub-pixel fractions estimates of Substrate (S), Vegetation (V) and Dark (D) materials.

2.2.6GCE 250m Crop Dominance (aka GCE V2.0) for N. America\USA based on automated cropland classification algorithm (Richard Massey, Teki Sankey et al.)

The North American product is first developed taking the United States of America. The method involves development of an automated cropland classification algorithm (ACCA) using the USDA cropland data layer (CDL) as reference. The process begins by creating a knowledge base of US croplands (e.g., Figure 15, Table 1). Through this process, it was determined that the 10 irrigated and 10 rainfed cropland classes occupy ~95% of all US cropland areas (Figure 15, Table 1). So, all our studies will be limited to these. In order to establish maximum separability between classes, we determined that analyzing classes taking the FAO's agroecological zones (AEZs) will be most appropriate. Based on this analogy, the 10 irrigated and the 10 rainfed classes are segmented for each AEZ (e.g., Figure 16 for AEZ7). MODIS 250 m NDVI time-series spectral characteristics are determined for each of the classes taking ~10,000 random sample locations, in most cases (e.g., Figure 17). Once the knowledge is understood, the process of building algorithm will focus on separating maximum number of classes with greatest possible accuracies. Currently we are developing such an algorithm for year 2008, which will be rigorously tested for accuracies using USDA CDL reference data. The emphasis is then to ensure that the algorithm works automatically for independent years.



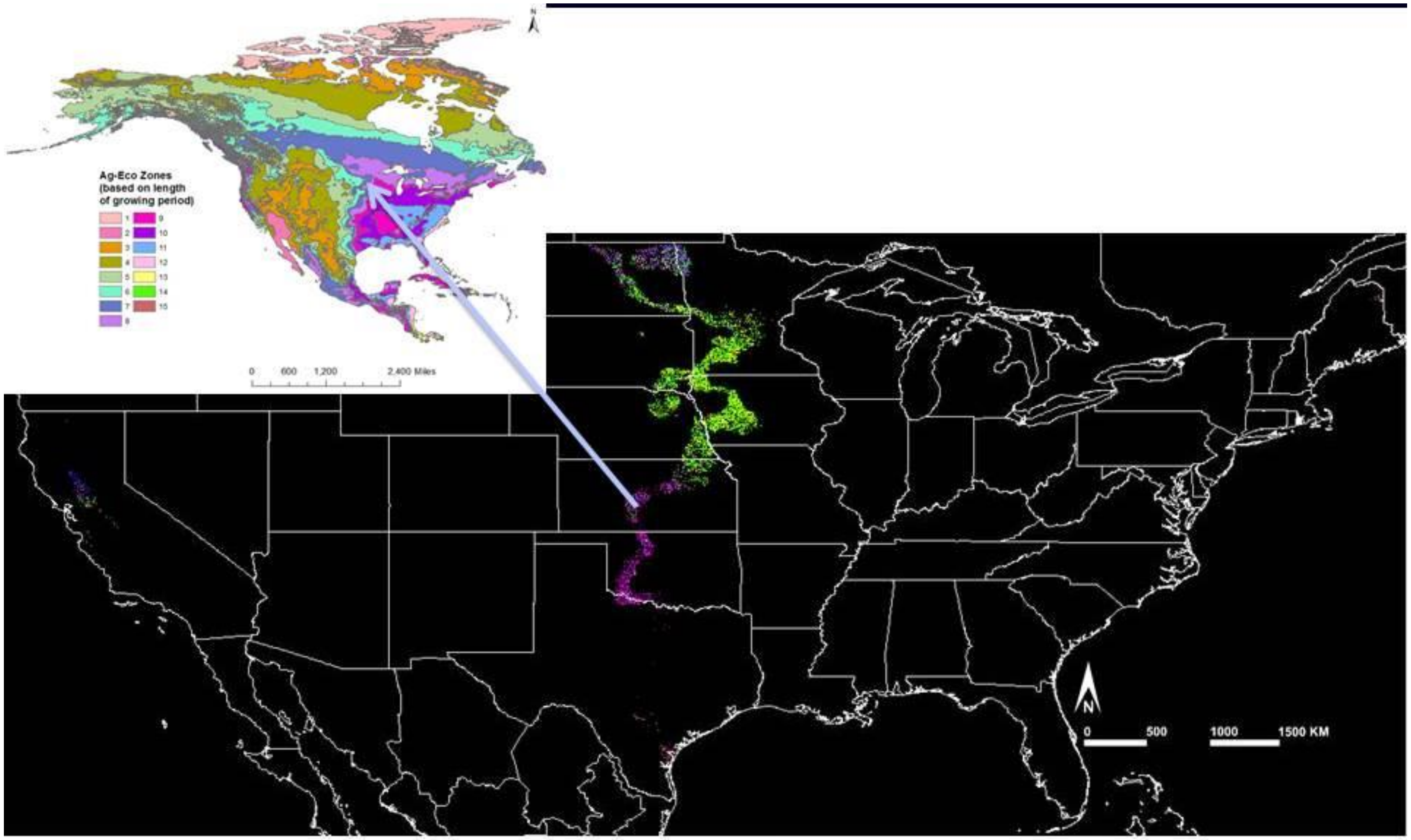
Class Rain-fed	%Maj Crop	% all crop US	Class Irrigated	%Maj Crop	% all crop US
1 Corn	31.5	29.6	11 Corn	6.4	6.0
2 Soybeans	23.8	22.4	12 Soybeans	3.9	3.6
3 Wheat (w)	13.3	12.5	13 Wheat (w)	2.8	2.6
4 Wheat (s)	6.4	6.0	14 Wheat (s)	0.4	0.4
5 Wheat (d)	0.8	0.8	15 Wheat (d)	0.1	0.1
6 Barley	0.7	0.7	16 Barley	0.2	0.2
7 Potatoes	0.2	0.2	17 Potatoes	0.2	0.2
8 Alfalfa	2.6	2.5	18 Alfalfa	1.1	1.0
9 Cotton	2.6	2.5	19 Cotton	1.5	1.4
10 Rice	0.5	0.4	20 Rice	1.0	1.0

Figure 15. Croplands of USA [Source: USDA CDL]. The area statistics of these 20 classes (10 irrigated and 10 rainfed) are shown in Table 1.

Table 1. Cropland distribution in the United States of America during the Year 2008 (Source: USDA, 2008)

Crop Class Number	Class name	GFSAD30 Code #	GFSAD30 Irrigated or rainfed	Major Crops Names	Pixels #	%age of Major I&R Crops	%age of All crops
1	Croplands, rainfed, corn	2	R	Corn	116305056	31.5	29.6
2	Croplands, rainfed, soybeans	5	R	Soybeans	87964380	23.8	22.4
3	Croplands, rainfed, wheat-winter	11	R	Wheat (w)	49229460	13.3	12.5
4	Croplands, rainfed, wheat-spring	12	R	Wheat (s)	23718114	6.4	6.0
5	Croplands, rainfed, wheat-durum	13	R	Wheat (d)	3138723	0.8	0.8
6	Croplands, rainfed, barley	6	R	Barley	2605932	0.7	0.7
7	Croplands, rainfed, potatoes	8	R	Potatoes	668367	0.2	0.2
8	Croplands, rainfed, alfalfa	99	R	Alfalfa	9734760	2.6	2.5
9	Croplands, rainfed, cotton	71	R	Cotton	9734760	2.6	2.5
10	Croplands, rainfed, rice	3	R	Rice	1748817	0.5	0.4
11	Croplands, irrigated, corn	2	I	Corn	23713641	6.4	6.0
12	Croplands, irrigated, soybeans	5	I	Soybeans	14309064	3.9	3.6
13	Croplands, irrigated, wheat-winter	11	I	Wheat (w)	10271142	2.8	2.6
14	Croplands, irrigated, wheat-spring	12	I	Wheat (s)	1437912	0.4	0.4
15	Croplands, irrigated, wheat-durum	13	I	Wheat (d)	330687	0.1	0.1
16	Croplands, irrigated, barley	6	I	Barley	616392	0.2	0.2
17	Croplands, irrigated, potatoes	8	I	Potatoes	771309	0.2	0.2
18	Croplands, irrigated, alfalfa	99	I	Alfalfa	4069548	1.1	1.0
19	Croplands, irrigated, cotton	72	I	Cotton	5439546	1.5	1.4
20	Croplands, irrigated, rice	3	I	Rice	3759966	1.0	1.0
Total area of Major crops in USA (Landsat Pixels)					369567576	**100.0**	**94.0**
Total area of all crops in USA (pixels)					393225483		
% rainfed croplands (of 20 major)						**82.5**	
% irrigated croplands (of 20 major)						**17.5**	

Note:
1. I = irrigated, R= rainfed
2. pixels are Landsat 30 m pixels. 1 pixel = 0.09 hectares
3. area proportions are determined by pixels. They are not re-projected and actual areas not calculated.
4. Irrigated/non-irrigated classes are generated using MODIS irrigated map (Ozdogan et al., 2009)



Class Rain-fed	% AEZ	% all crop US	Class Irrigated	% AEZ	% all crop US
1 Corn	33.1	3.9	11 Corn	3.2	0.4
2 Soybeans	30.1	3.6	12 Soybeans	2.1	0.2
3 Wheat (w)	12.2	1.4	13 Wheat (w)	1.7	0.2
4 Wheat (s)	12.5	1.5	14 Wheat (s)	0.1	0.0
5 Wheat (d)	0.0	0.0	15 Wheat (d)	0.0	0.0
6 Barley	0.7	0.1	16 Barley	0.0	0.0
7 Potatoes	0.1	0.0	17 Potatoes	0.0	0.0
8 Alfalfa	0.7	0.1	18 Alfalfa	0.5	0.1
9 Cotton	0.9	0.1	19 Cotton	0.2	0.0
10 Rice	0.3	0.0	20 Rice	1.4	0.2

Figure 16. Classes within the FAO agroecological zone 7 (AEZ7) for USA derived from USDA cropland layer for year 2008.

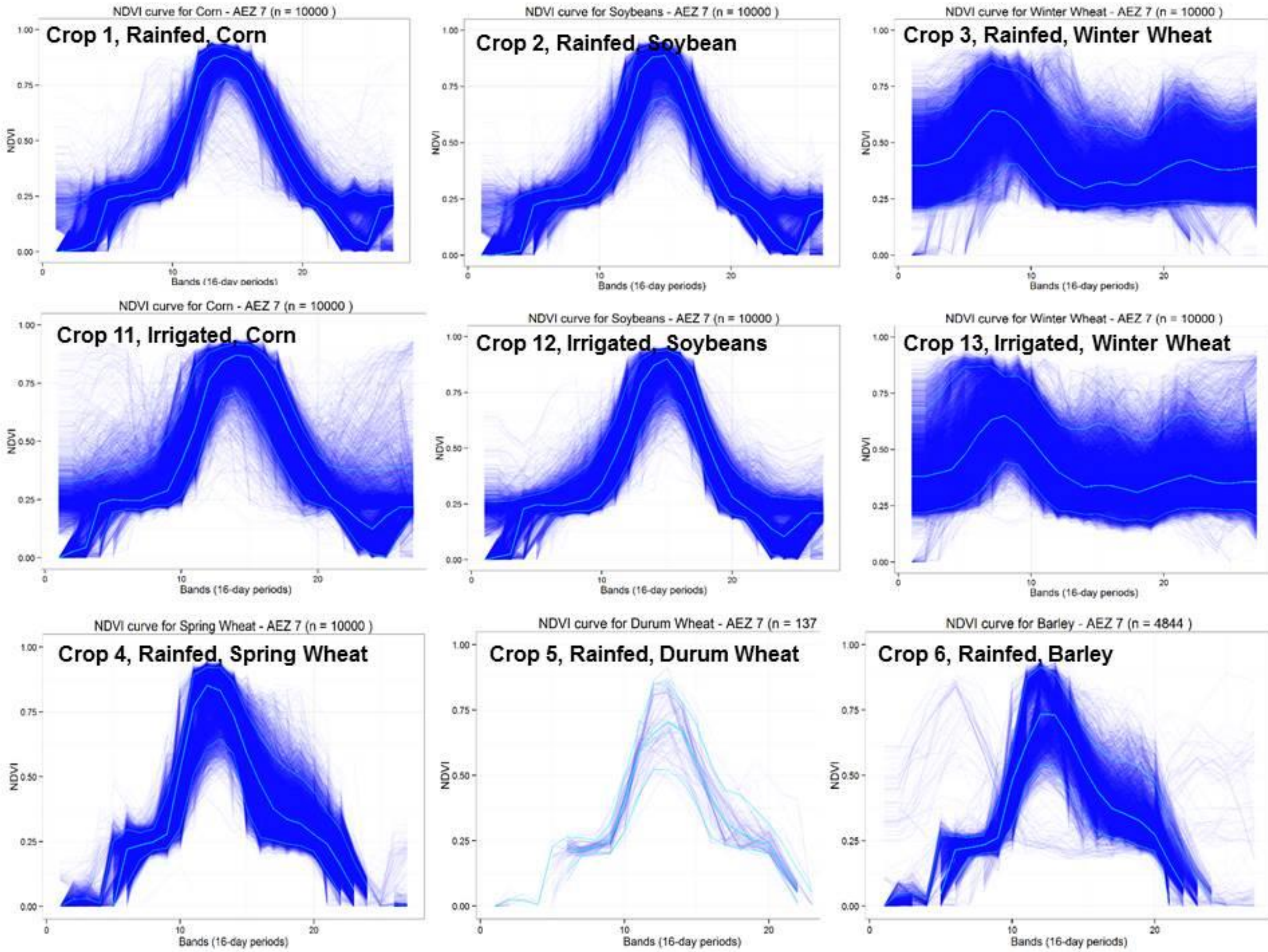


Figure 17. Derived MODIS 250 m NDVI time-series signatures for the certain rainfed and irrigated crops of USA for FAO AEZ7. The spatial distribution of these crops are shown in Figure 15.

2.2.7 Linear Discriminate Analysis (LDA) for distinguishing croplands from non-croplands of Europe (Aparna Phalke, Mutlu Ozdogan et al.)

Time-series Landsat imagery is used to train a Linear Discriminate Analysis (LDA) algorithm to separate croplands from non-croplands in Europe. The process involves organizing time-series Landsat images for different zones of Europe ((e.g., Figure 18) and then classifying croplands versus non-croplands based on LDA equations into mutually exclusive and exhaustive classes based on a set of measurable croplands versus non-croplands features (e.g., mean EVI, standard deviation of EVI, maximum EVI, Minimum EVI, and so on; e.g., Figure 19) derived from time-series Landsat images. They are planning to do this at 3 distinct levels:

1. Footprint/scene level: The LDA model is derived based on randomly generated training data of croplands vs. non-croplands of a footprint/scene (e.g., a single Landsat scene or few scenes in a homogeneous area) and applied to the same footprint\scene or, at the most, to nearby footprints\scenes to classify and separate croplands from non-croplands;
2. Zonal level: The LDA model is derived based on randomly generated training data of the croplands vs. non-croplands of a zone (e.g., a agroecological zone) and applied to all scenes within the zone to classify and separate croplands from non-croplands;
3. Regional level: The LDA model is derived based on randomly generated training data of croplands vs. non-croplands of a whole region (e.g., entire Europe) and applied to entire region to classify and separate croplands from non-croplands.

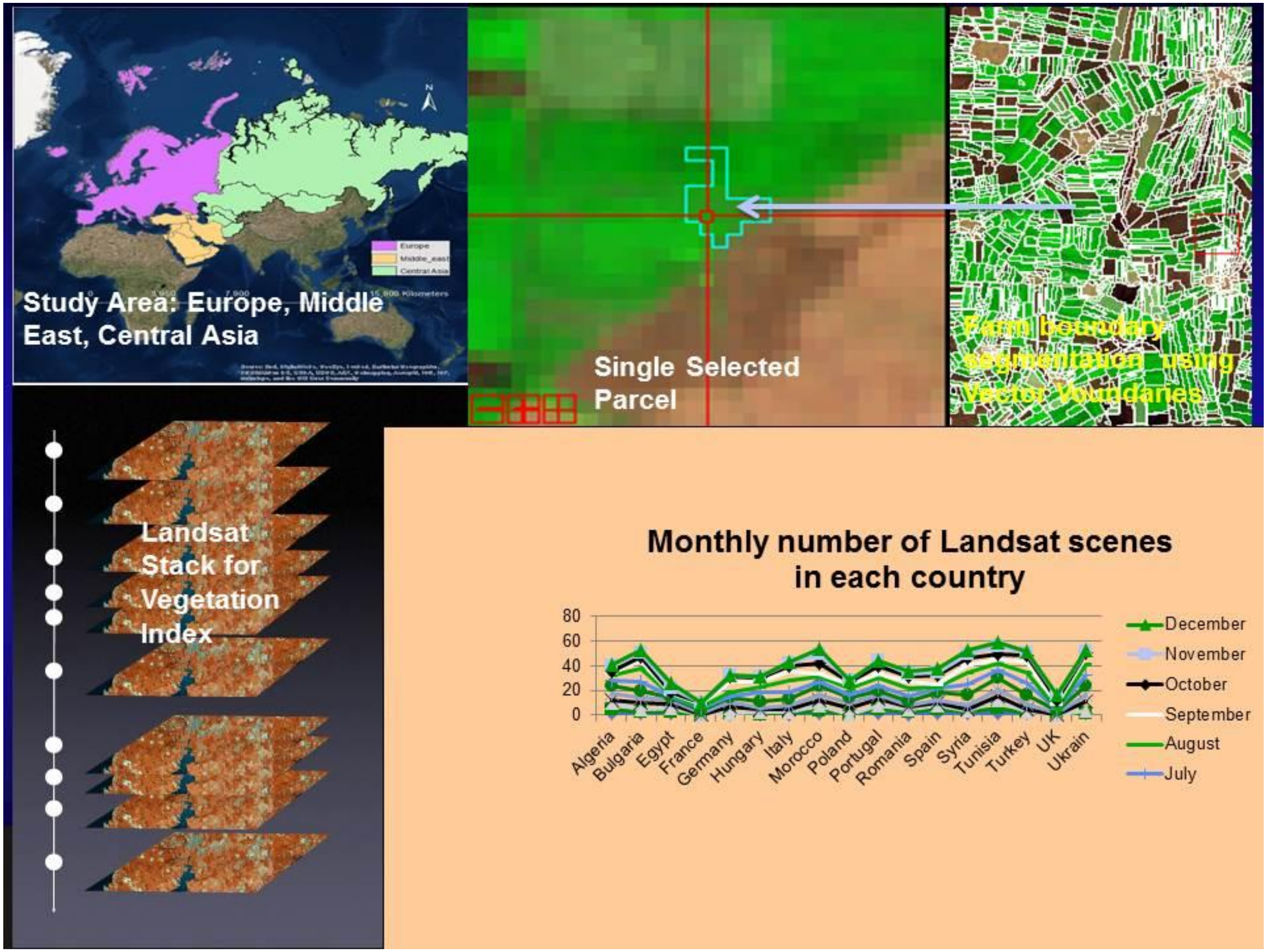


Figure 18. Process of developing linear discriminant analysis (LDA) algorithm using time-series Landsat data for Europe.

	Variable	Definition
1	Mean of EVI	It is the mean value of EVI of the particular footprint Landsat time-series
2	Standard deviation of EVI	It is the standard deviation value of EVI of the particular footprint Landsat time-series
3	Maximum of EVI	It is the maximum value of EVI of the particular footprint Landsat time-series
4	Minimum of EVI	It is the minimum value of EVI of the particular footprint Landsat time-series
5	Variance of EVI	It is the variance value of EVI of the particular footprint Landsat time-series
6	Range of EVI	It is the range value of EVI of the particular footprint Landsat time-series
7	Count of EVI	It is the number of observation recorded in each date of acquired Landsat scene
8	Slope	It is the topographic variable consisting of the inclination of the surface.
9	Elevation	It is the height of the surface above sea level.

Coefficients of linear discriminants:

	LD1	LD2	LD3	LD4
turkey$mean	1.321835e-03	3.013637e-03	4.663059e-04	5.776202e-04
turkey$sd	-2.032111e-03	-2.596891e-03	-7.681568e-03	5.523297e-03
turkey$max	-7.091220e-05	2.275280e-04	3.524008e-04	-5.721943e-04
turkey$min	7.036666e-06	4.361454e-04	-1.638853e-04	5.913712e-06
turkey$variance	1.377855e-07	-2.916269e-07	1.653012e-06	-1.103393e-06
turkey$range	-6.870982e-05	1.922650e-04	3.493801e-04	-5.512965e-04
turkey$number.of.observations	1.803958e-02	-3.554534e-02	-8.661410e-02	2.443064e-02
turkey$slope	9.375192e-02	3.713403e-02	-1.343249e-01	-1.378963e-01
turkey$elevation	2.675338e-03	-1.740076e-03	1.321873e-03	8.454065e-04

Proportion of trace:

LD1	LD2	LD3	LD4
0.8875	0.0774	0.0298	0.0053

LD1+ LD2 + LD 3 +LD4 =100%

Linear discriminant 1 (LD1)= 1.321835e-03*mean + (-2.032111e-03)*sd + (-7.091220e-05)*max + 7.036666e-06*min + 1.377855e-07*variance + (-6.870982e-05)range + 1.803958e-02*NumberOfObservation + 9.375192e-02*slope + 2.675338e-03*elevation

Figure 19. A typical linear discriminant analysis (LDA) equation used to separate croplands from non-croplands.

3.0 Landsat 30 m cropland products (Jun Xiong, Pardhasaradhi Teluguntla, et al.)
Ultimately, the 30 m products will be produced by various sub-groups based on study areas (Figure 20) where different co-Is have expertise and\or interest. Even though there are ~9500 Landsat images covering the terrestrial area, ~50% of these images (see Table 2) are required to cover areas where cropland currently exist and where they can potentially exist in future or where they existed in the past. These images were processed, standardized (Table 3,4) and mosaicked (Figure 21a through 21g).

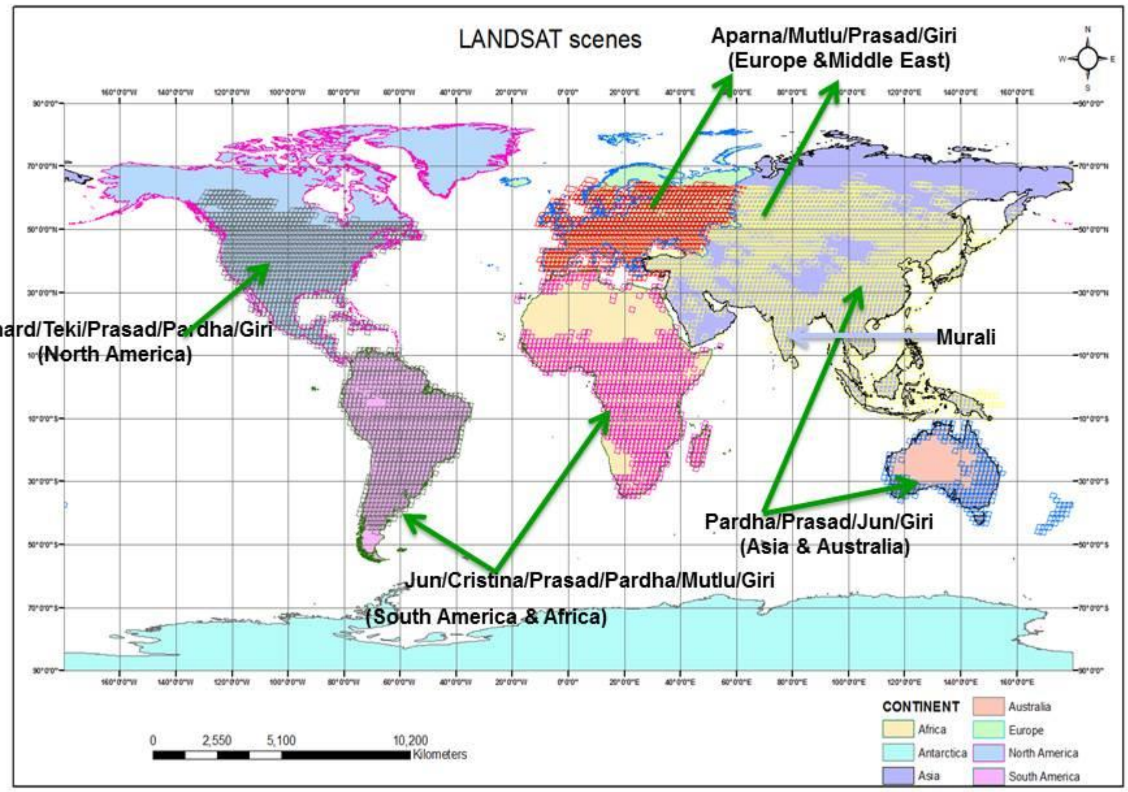


Figure 20. Shows the 7 regions and the Landsat tiles over these regions over cropland areas and\or potential cropland areas. Overall, there are 4990 Landsat tiles over croplands\potential croplands (above figure) of ~9000 Landsat tiles (see Table 1) covering the entire terrestrial Earth. The areas where there is currently zero croplands and\or their chances of occurring in future are about zero (e.g., Sahara desert, Antarctica), no Landsat images are selected to avoid processing unnecessary images for cropland studies.

Global Landsat BDI for 4 epochs (1975, 1990, 2000, 2005, and 2010) is summarized in Table 1. The Table 1 shows the number of images, sensor from which they are acquired, and the total storage volume required these images will be used in GFSAD30 project. The GLS scenes are band separate, in UTM coordinates, WGS-84 datum, are distributed in GeoTIFF format, and are compressed using tar and gzip / bZip. Collectively, these datasets provide consistent observations of global, orthorectified, leaf-on, “cloud free” data (Gutman, et al., 2008).

Epoch	Scenes	Size*	ETM+	TM	MSS	ALI
2010	8453	1.5TB	3719	4734	n/a	n/a
2005	9375	1.64TB	7087	2288	n/a	n/a
2000	8755	2.18TB	8755	n/a	n/a	n/a
1990	7371	975GB	n/a	7371	n/a	none
1975	7592	250GB	n/a	n/a	7592	n/a

Table 2. Global coverage of the Landsat GLS Data for epoch 2010, 2005, 2000, 1990, 1975. The table shows the sensor from which the data are acquired and the required storage volume.

Region Name	Sensor	Scenes	Format	Single-Band	Total Size
Africa	ETM	891	Erdas img	144 GB	720 GB
Asia	ETM	1651	Erdas img	366 GB	1830 GB
Australia	ETM	224	Erdas img	39 GB	195 GB
Europe	ETM	640	Erdas img	67 GB	335 GB
Middle-east	ETM	193	Erdas img	37 GB	185 GB
North-America	ETM	900	Erdas img	133 GB	665 GB
South-America	ETM	709	Erdas img	79 GB	395 GB

Table 3. Distribution of Landsat 2000 images in each of the 7 regions in cropland areas.

Sensor	Data Sets	Units	Scale factor	Data Type	Fill Value	Valid Range
TM/ETM+	Band 3 RED	TOA reflectance (-)	0.001	int16	-9999	0 - 1000
	Band 4 NIR	TOA reflectance (-)	0.001	int16	-9999	0 - 1000
	Band 5 SWIR	TOA reflectance (-)	0.001	int16	-9999	0 - 1000
	Band 6 THM*	Radiometric temperature (K)	0.1	int16	-9999	0 - 1000
ALL	NDVI	Normalized Difference Vegetation Index	0.001	int16	-9999	-1000 - 1000

Table 4. Landsat data normalization and scaling.

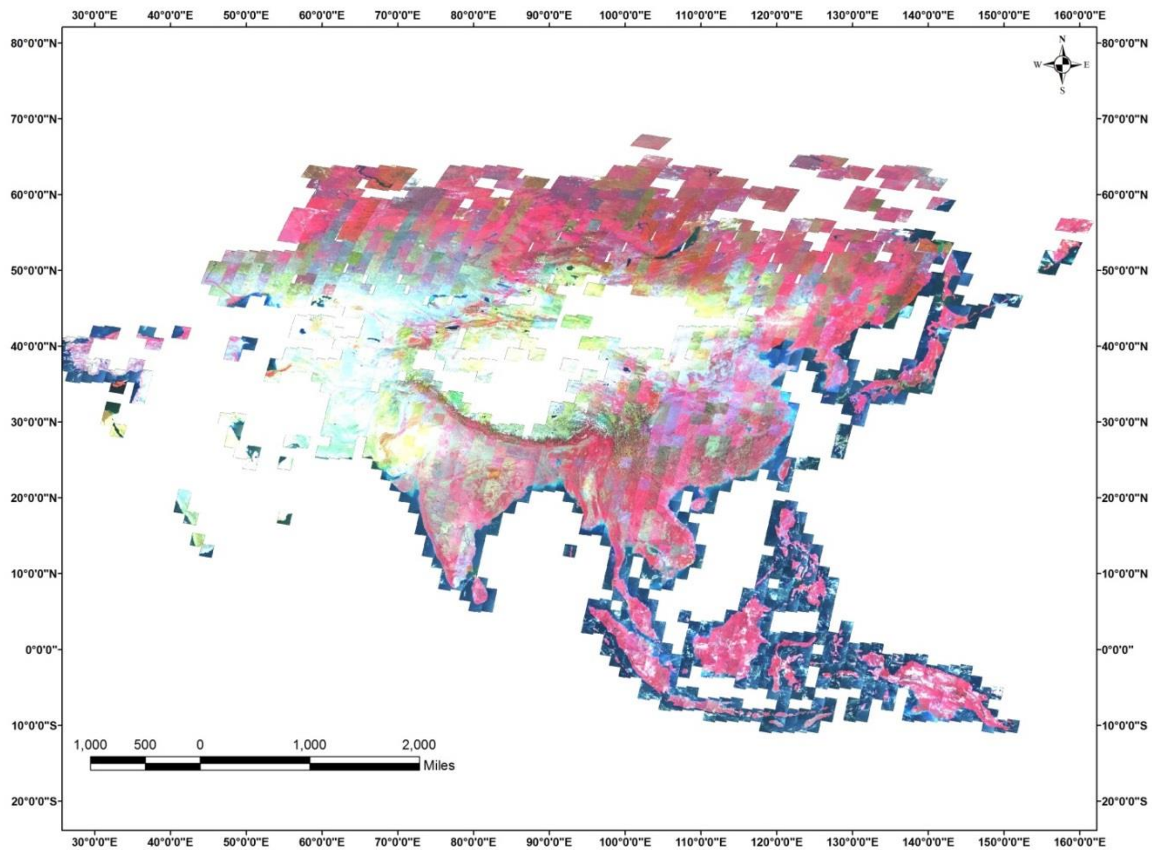


Figure 21a. Landsat Global Land Survey 2000 (GLS2000) mosaic of Asia. A total of 1651 images of croplands or where even a fraction of croplands exist have been mosaicked. Depicted here as False color composite (FCC) RGB bands 4,3,5. These data are in top of the atmosphere reflectance (TOA) expressed in %. The black areas show areas where zero croplands exist and considered sheer waste of time and resources to analyze for cropland characteristics. Data is processed and mosaicked on NASA AMES NEX supercomputer.

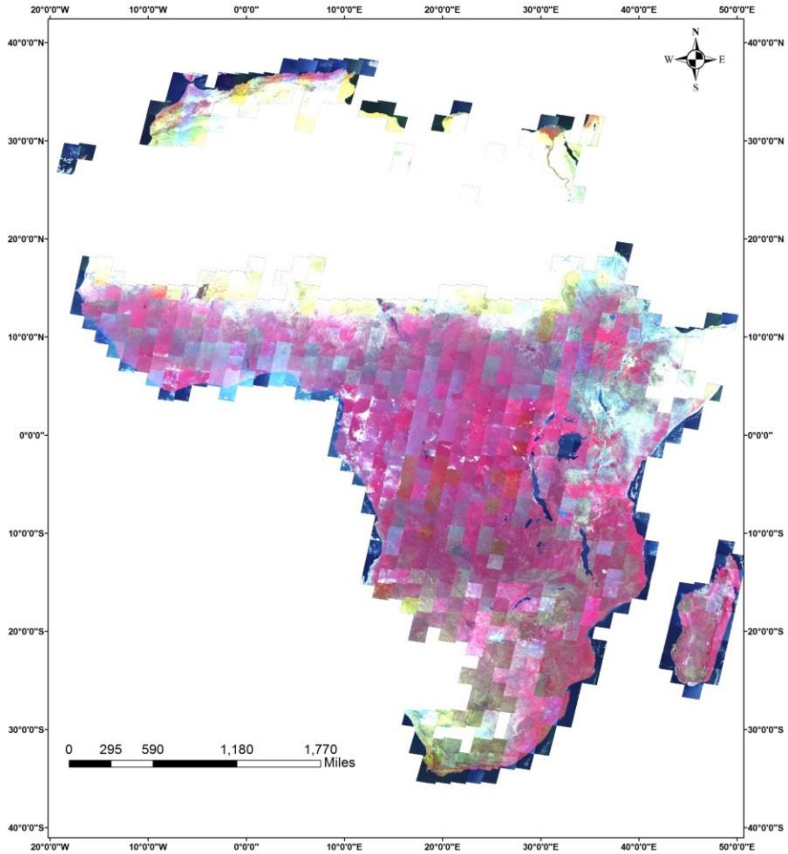


Figure 21b. Landsat Global Land Survey 2000 (GLS2000) mosaic of Africa. A total of 891 images of croplands or where even a fraction of croplands exist have been mosaicked. Depicted here as False color composite (FCC) RGB bands 4,3,5. These data are in top of the atmosphere reflectance (TOA) expressed in %. The black areas show areas where zero croplands exist and considered sheer waste of time and resources to analyze for cropland characteristics. Data is processed and mosaicked on NASA AMES NEX supercomputer.

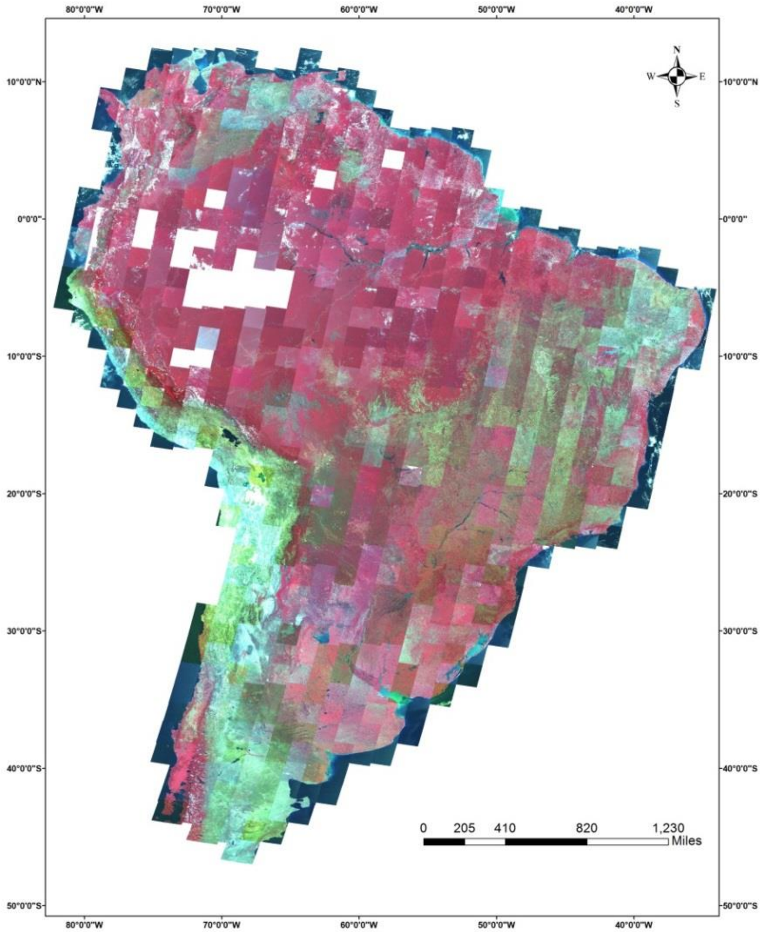


Figure 21c. Landsat Global Land Survey 2000 (GLS2000) mosaic of S. America. A total of 809 images of croplands or where even a fraction of croplands exist have been mosaicked. Depicted here as False color composite (FCC) RGB bands 4,3,5. These data are in top of the atmosphere reflectance (TOA) expressed in %. The black areas show areas where zero croplands exist and considered sheer waste of time and resources to analyze for cropland characteristics. Data is processed and mosaicked on NASA AMES NEX supercomputer.

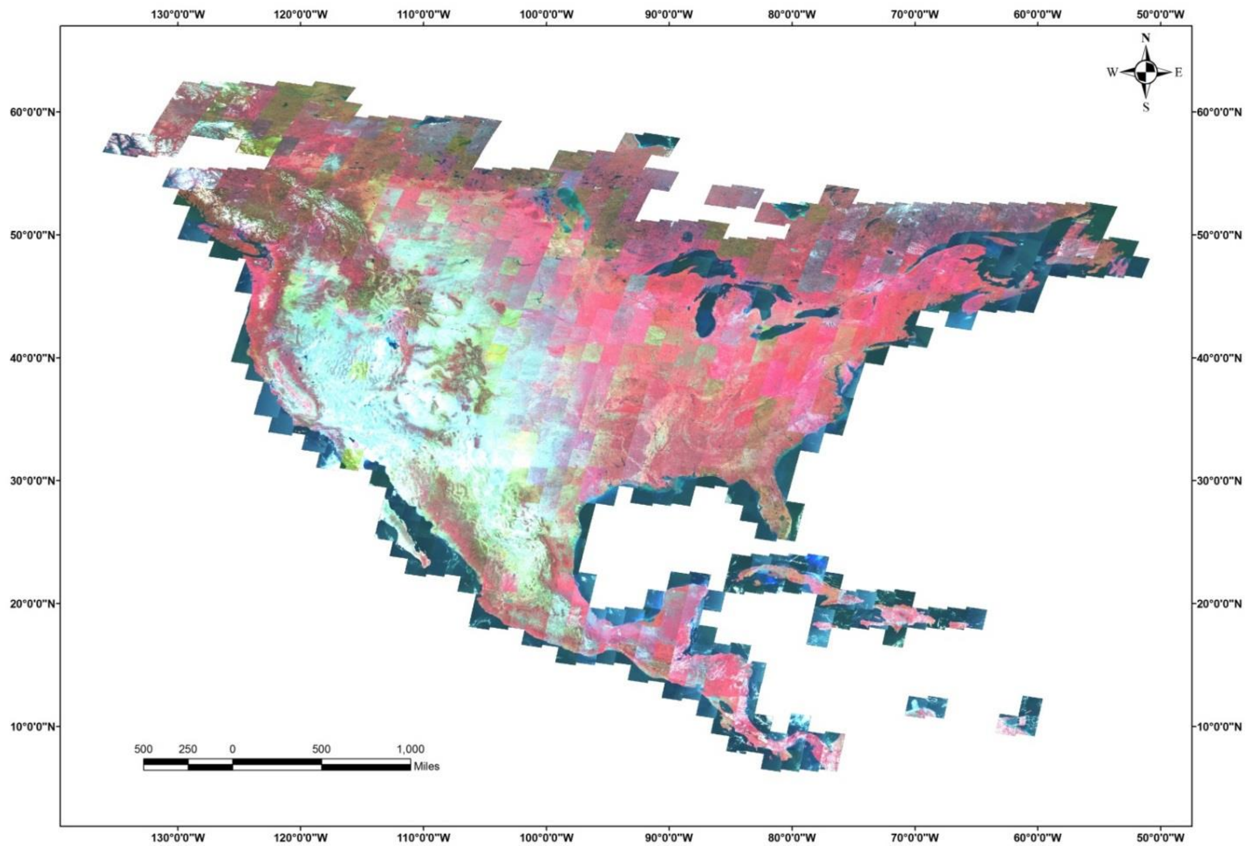


Figure 21d. Landsat Global Land Survey 2000 (GLS2000) mosaic of N. America. A total of 900 images of croplands or where even a fraction of croplands exist have been mosaicked. Depicted here as False color composite (FCC) RGB bands 4,3,5. These data are in top of the atmosphere reflectance (TOA) expressed in %. The black areas show areas where zero croplands exist and considered sheer waste of time and resources to analyze for cropland characteristics. Data is processed and mosaicked on NASA AMES NEX supercomputer.

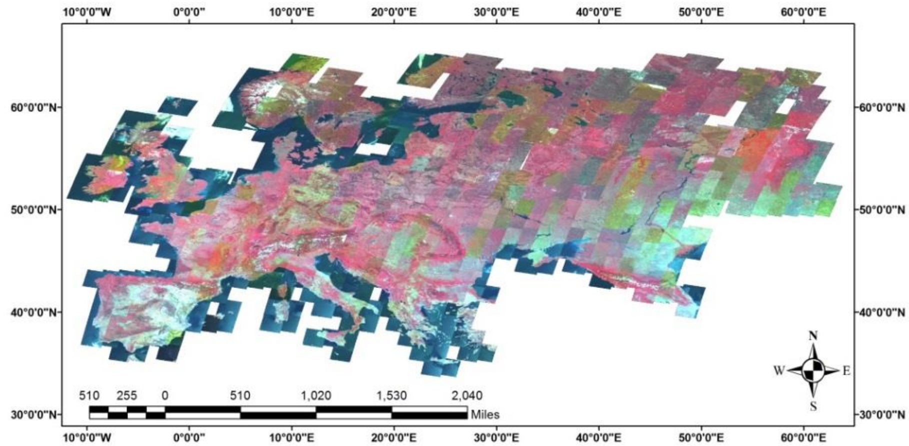


Figure 21e. Landsat Global Land Survey 2000 (GLS2000) mosaic of Europe. A total of 640 images of croplands or where even a fraction of croplands exist have been mosaicked. Depicted here as False color composite (FCC) RGB bands 4,3,5. These data are in top of the atmosphere reflectance (TOA) expressed in %. The black areas show areas where zero croplands exist and considered sheer waste of time and resources to analyze for cropland characteristics. Data is processed and mosaicked on NASA AMES NEX supercomputer.

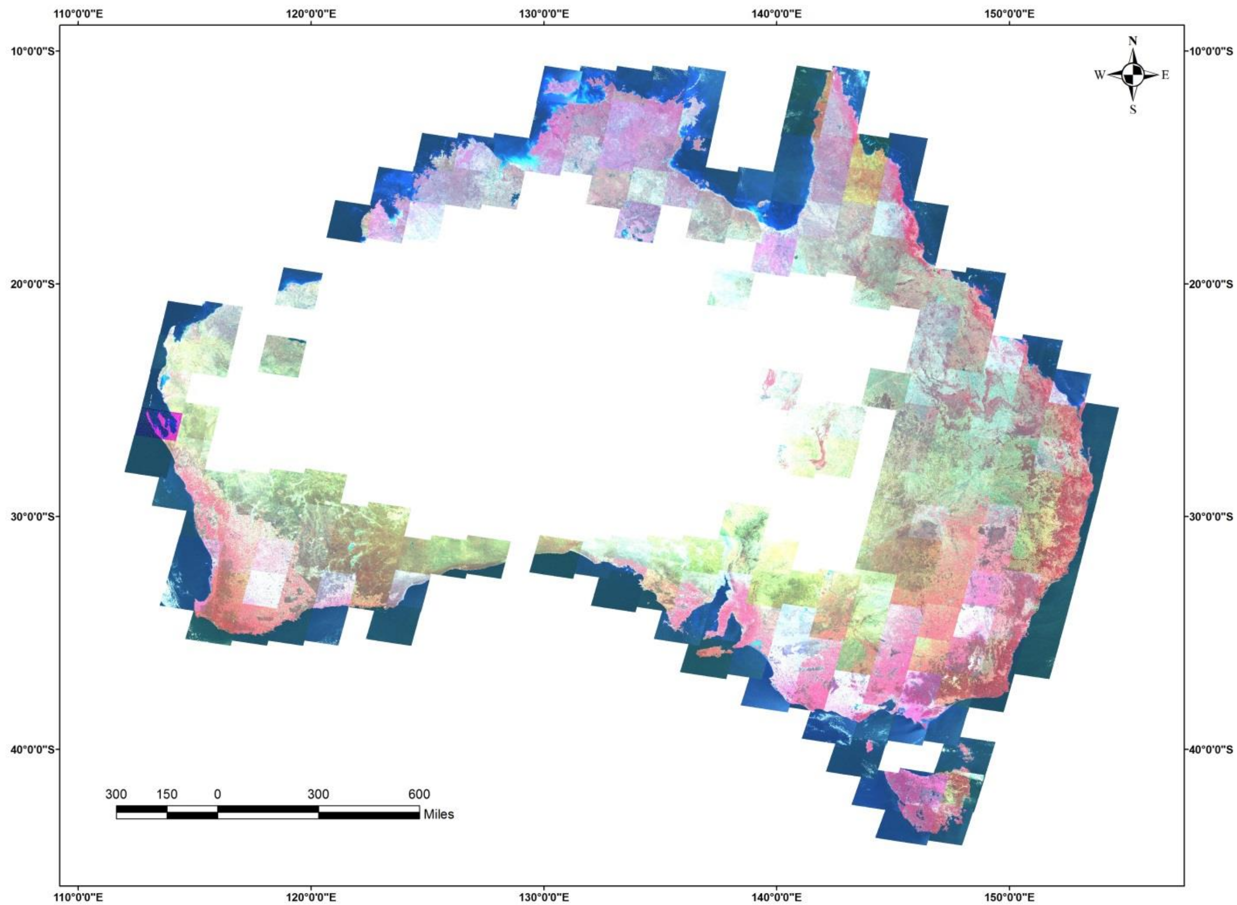


Figure 21f. Landsat Global Land Survey 2000 (GLS2000) mosaic of Australia. A total of 224 images of croplands or where even a fraction of croplands exist have been mosaicked. Depicted here as False color composite (FCC) RGB bands 4,3,5. These data are in top of the atmosphere reflectance (TOA) expressed in %. The black areas show areas where zero croplands exist and considered sheer waste of time and resources to analyze for cropland characteristics. Data is processed and mosaicked on NASA AMES NEX supercomputer.

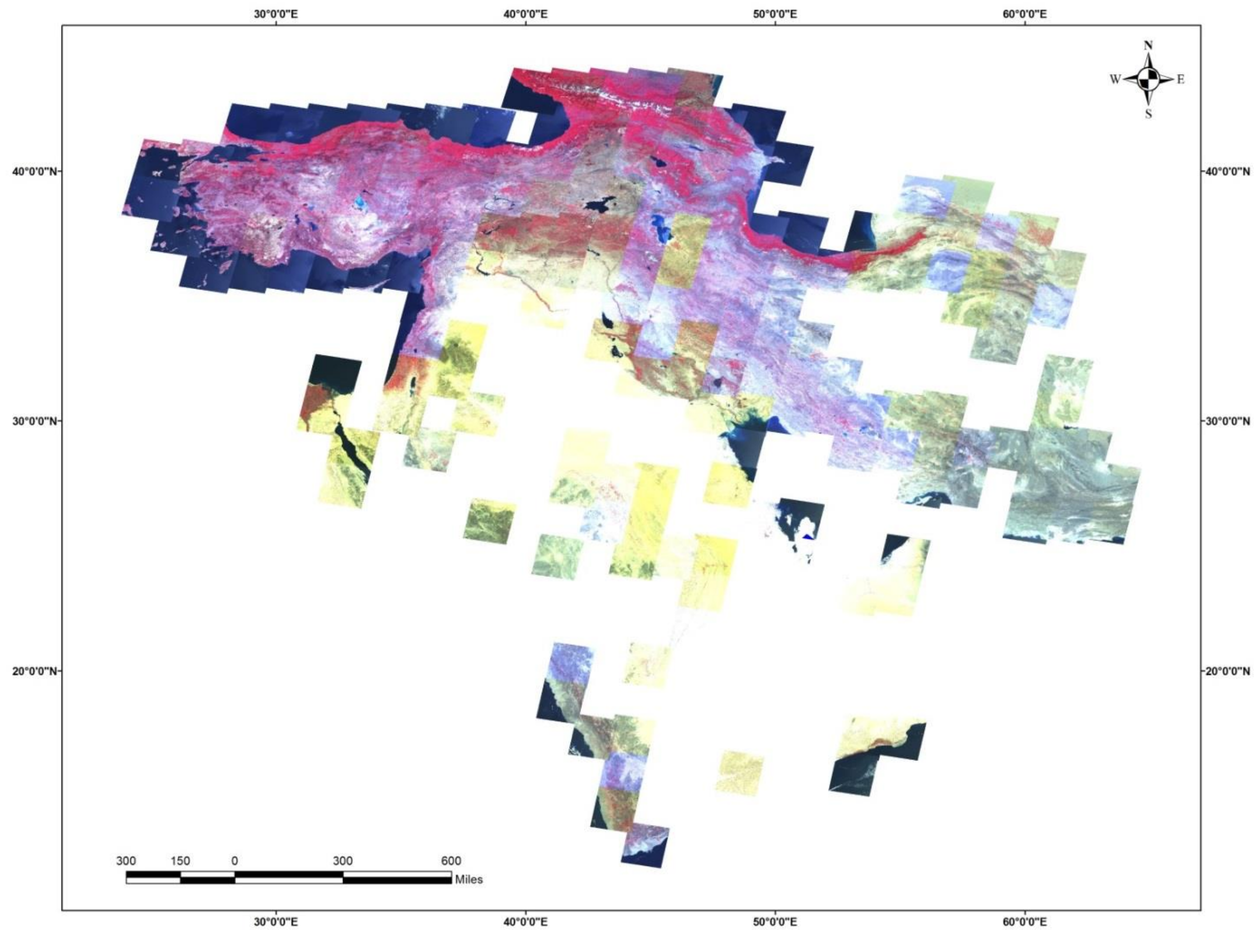


Figure 21g. Landsat Global Land Survey 2000 (GLS2000) mosaic of Middle East. A total of 193 images of croplands or where even a fraction of croplands exist have been mosaicked. Depicted here as False color composite (FCC) RGB bands 4,3,5. These data are in top of the atmosphere reflectance (TOA) expressed in %. The black areas show areas where zero croplands exist and considered sheer waste of time and resources to analyze for cropland characteristics. Data is processed and mosaicked on NASA AMES NEX supercomputer.

4.0 Global ground data (Mutlu Ozdogan, Justin Poehnelt, Jun Xiong, Pardhasaradhi Teluguntla et al.)

Currently ~130,000 ground data points (Figure 22) have been collected and\or sources and organized into a database. During the project period, we are collecting ground data from various locations of the world (e.g., Figure 23a to 23d). These data will be shared during the project period through croplands.org and other mechanisms.

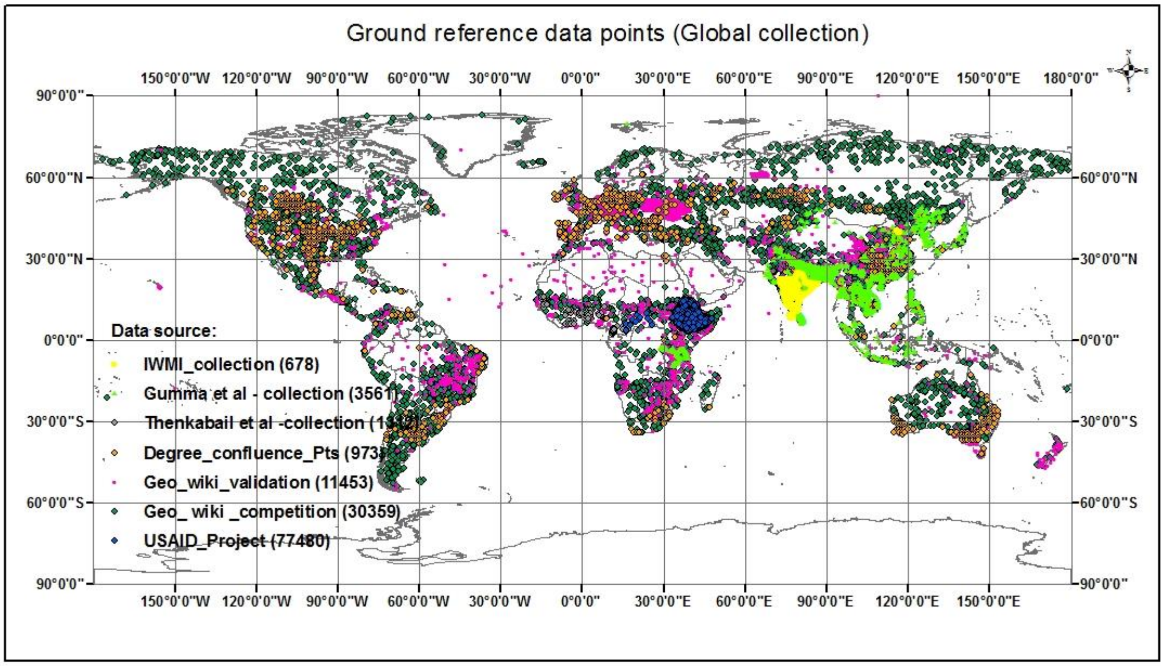


Figure 22. Global ground data base built by University of Wisconsin team. Over 125,000+ data points are available. This is ongoing work and the point so far organized is shown above. Each point has location, digital images\s, and cropland and other land use characteristics. Most of these data is collected by Dr. Murali Krishna Gumma and Prasad Thenkabail's earlier team at the International Water Management Institute.

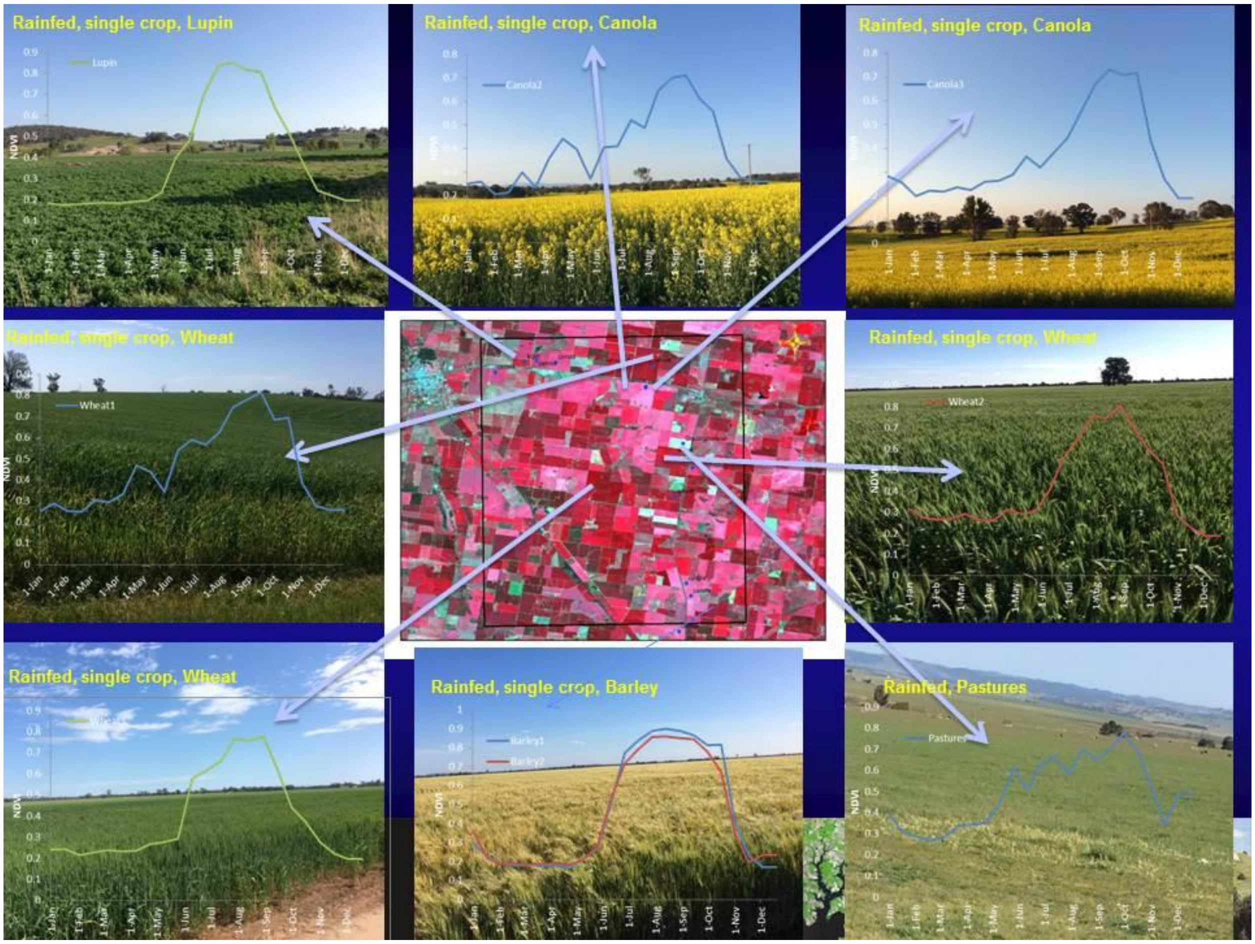


Figure 23a. Typical recent ground data collected in Australia.

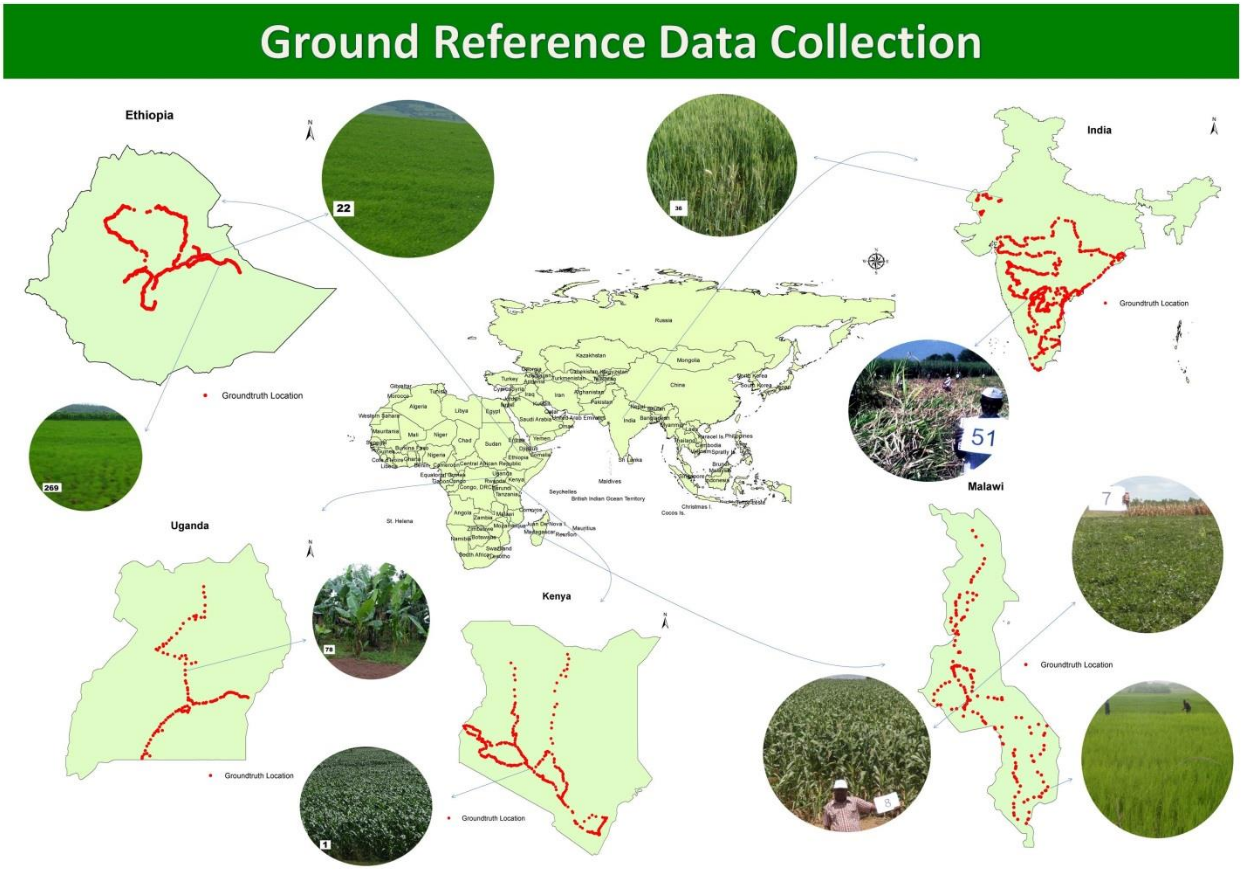


Figure 23b. Typical recent ground data collected in Africa and Asia.

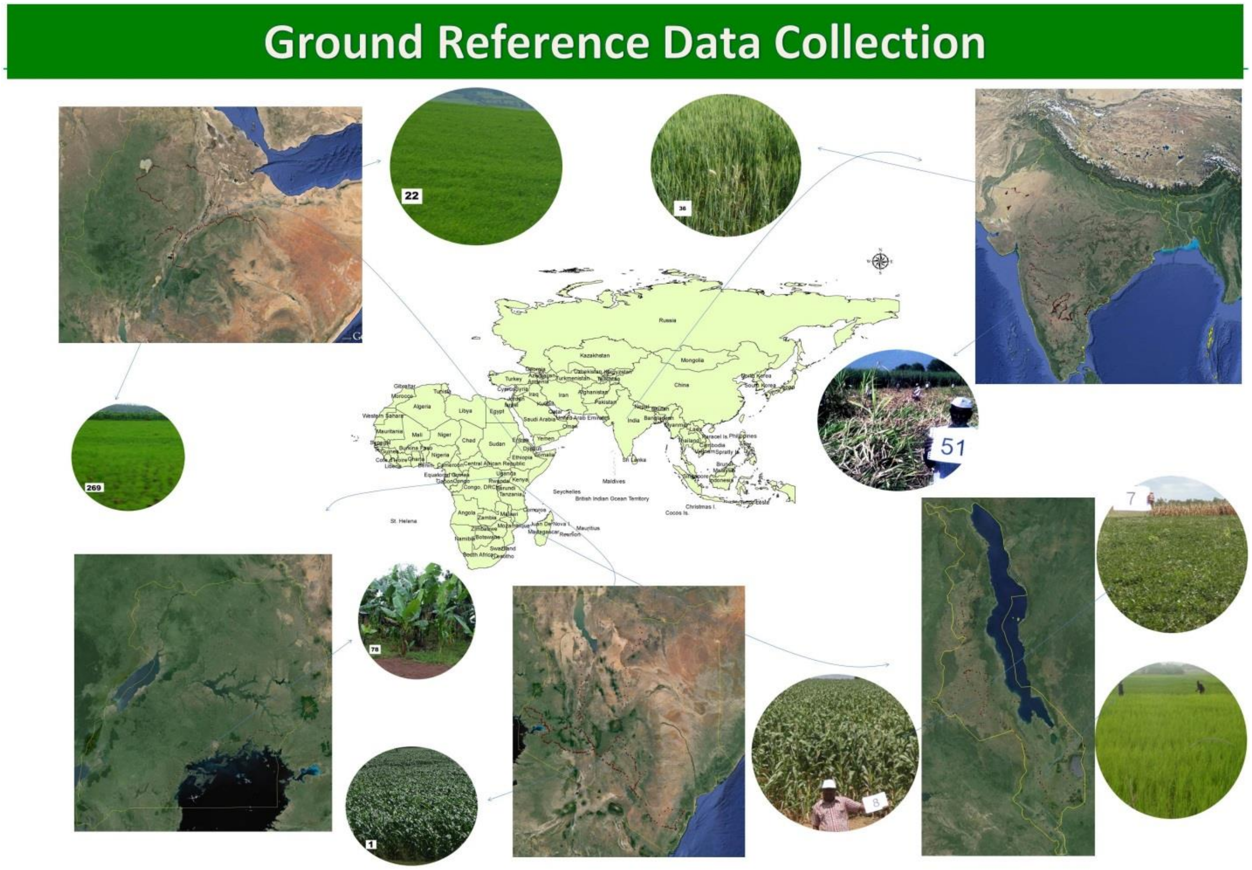


Figure 23c. Typical recent ground data collected in Africa and Asia.

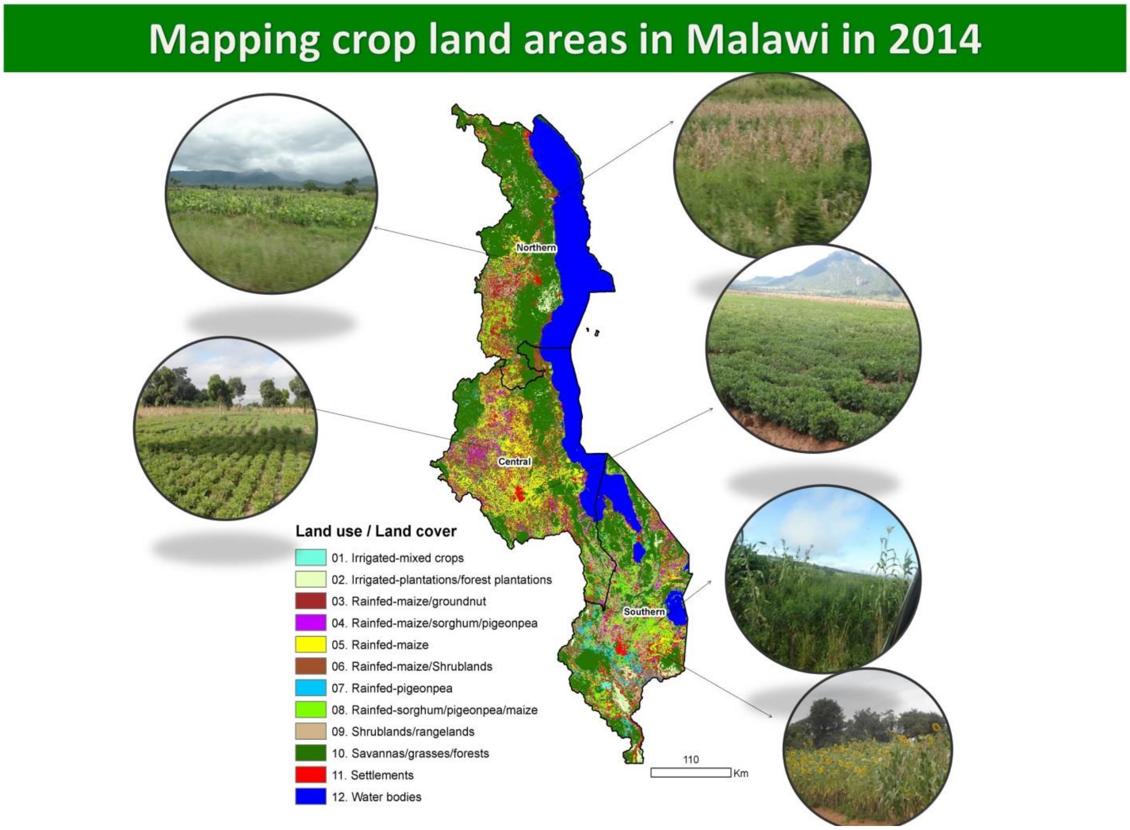


Figure 23d. Typical recent ground data collected in Africa and Asia.

5.0 Accuracy, Uncertainty, and Reference Data issues @ University of New Hampshire
(**Russell G. Congalton** and Kamini Yadav et al.)

The team at the University of New Hampshire has specific tasks as part of this GFSAD30 project to work on the aspects of spatial uncertainty, accuracy assessment, and reference data collection protocols for mapping global croplands. To that end, in the last year, the team has accomplished the following:

1- Conducted a detailed review of previous global land cover and cropland mapping projects. As part of this effort a review paper was written centered around compiling information about past global land cover mapping projects and an uncertainty analysis to show where most errors occur in these projects. This paper is has been submitted to the journal Remote Sensing and is currently being revised based on reviewer comments. The current version of the paper is attached..

2- Developed a field reference data collection procedure document for the entire team. This document includes a detailed field form. The procedure was presented to the entire team and vetted during our January Team Meeting in Menlo Park, CA and used to facilitate reference data collection in Australia. A full copy of the current version of this document is attached.

3- Enhanced the computer program written the previous year in R (e.g., Figure 24), an open source statistics package, to compute an error matrix and the associated descriptive statistics. Since R is open source, this program can be distributed freely to whomever on the team or elsewhere that might wish to use it. A user's guide to go along with the program was also written and is currently in review (attached). A slide showing the program output is attached.

4- Ms. Kamini Yadav spent 2 weeks (July 20 – August 2, 2014) at USGS Flagstaff, AZ working with PI Dr. Prasad Thenkabail and his team on field sampling and protocols for reference data collection. In addition, she interacted with the team on the mapping components of this project to further understand their needs for accuracy methodology.

5- Work continued to obtain high-resolution imagery from NGA without much success. However, team member Dr. James Tilton (NASA) was able to set up a process by which he could request imagery for the entire team. In addition, Dr. Chandra Giri (team member from USGS EROS) provided 500 high spatial resolution images from around the globe that were used for another project, but that show great promise for use in this project.

6- Given this new access to high resolution imagery, we began some analysis of this imagery using the HSeg Program written by team member, Dr. James Tilton (e.g., Figure 25). This program provides segmentation and classification of high resolution imagery that may prove useful as reference data. This work is ongoing and Ms. Yadav will spend a week in December at NASA Goddard working with Dr. Tilton on this analysis. A slide of our current analysis is provided.

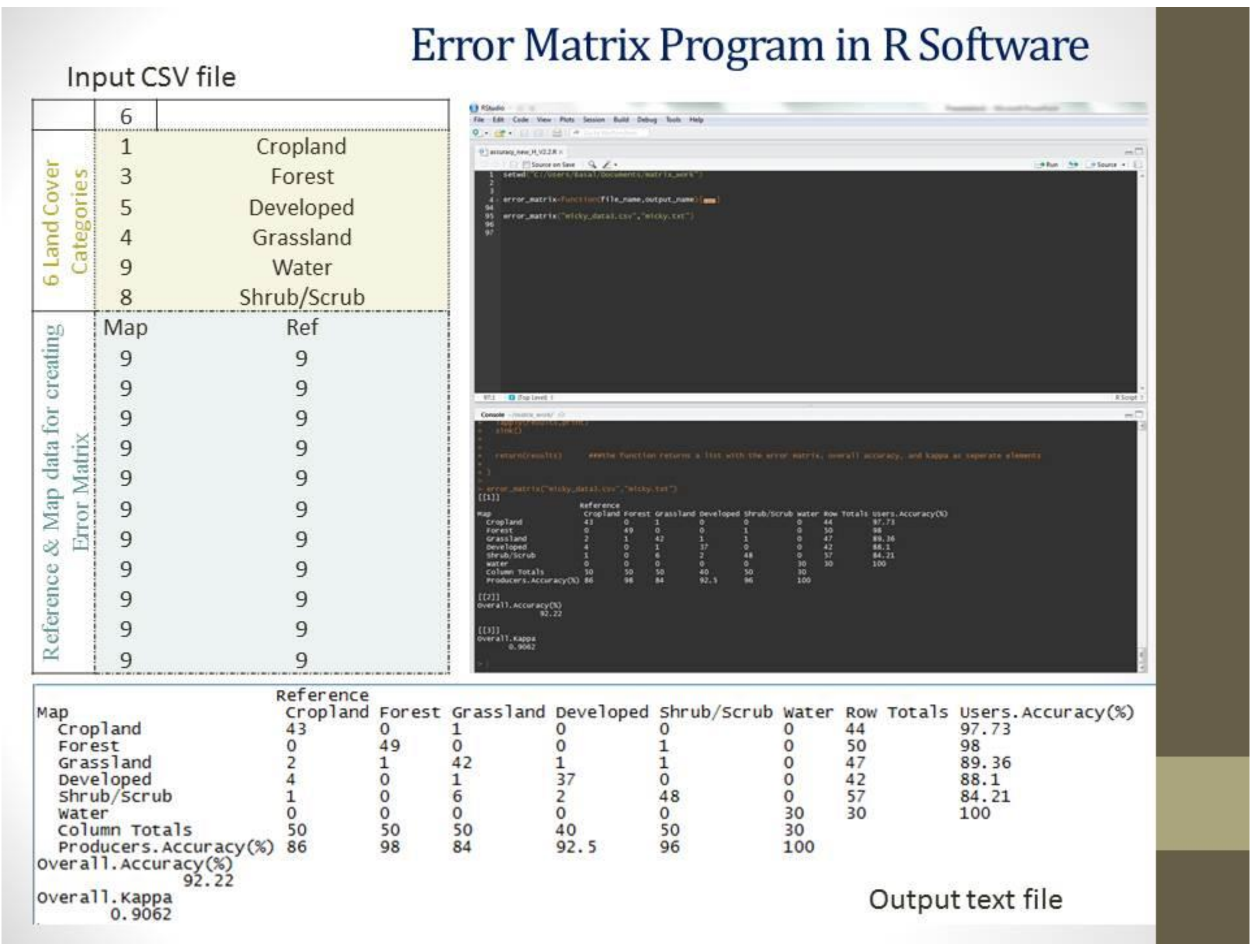


Figure 24. Error matrix in R software.

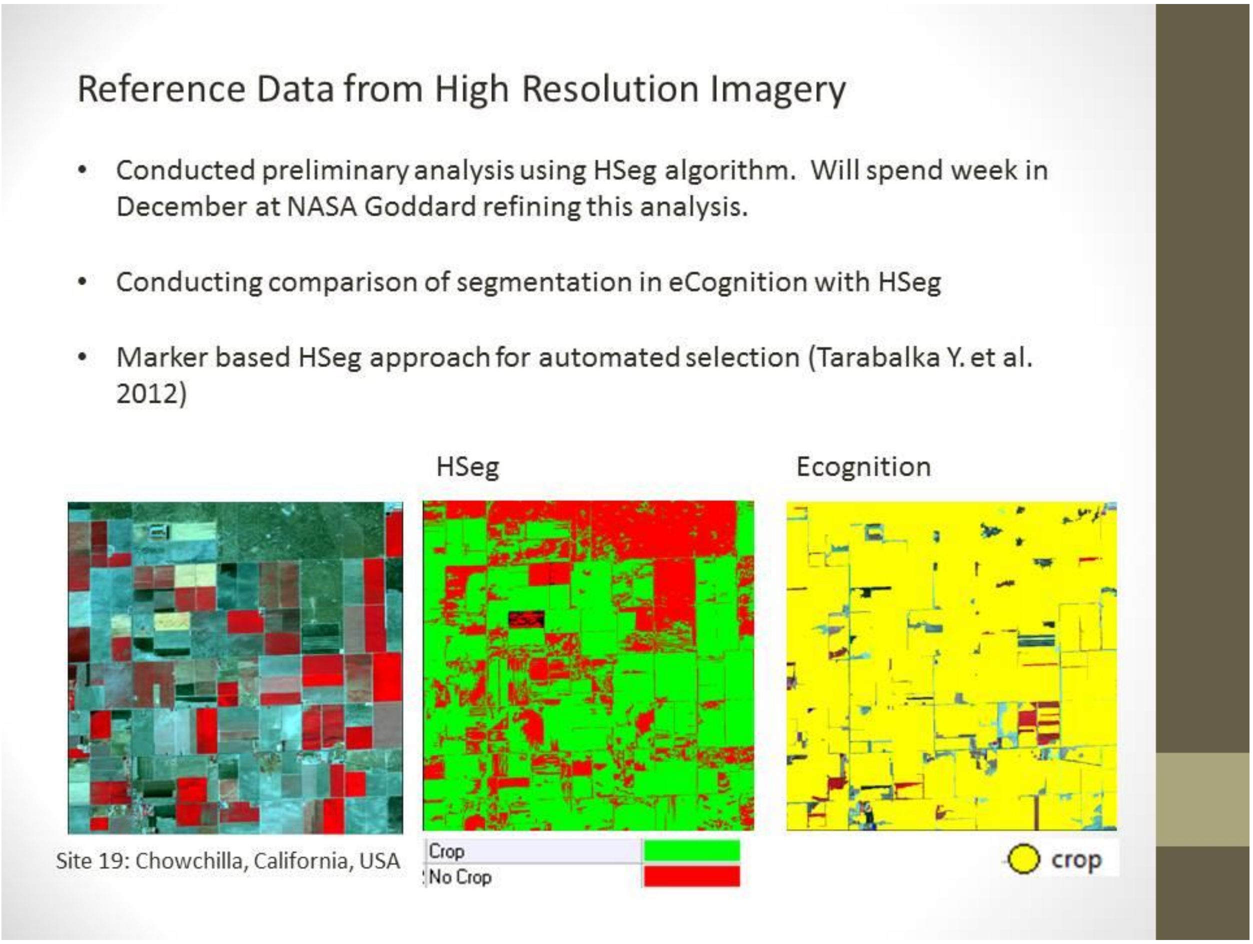
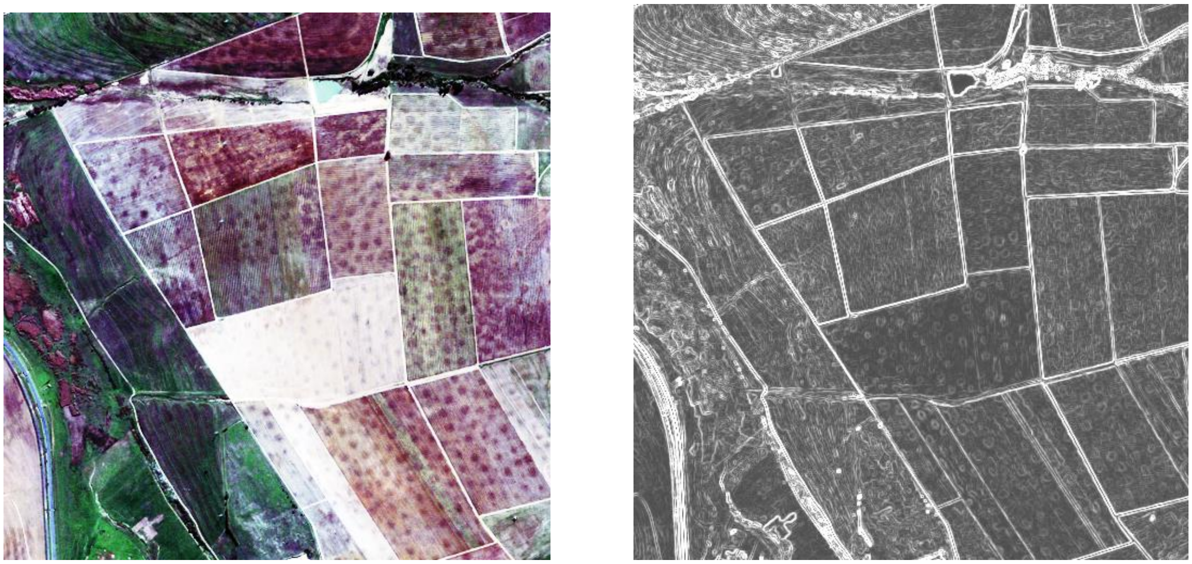


Figure 25. Reference data generated through HSeG and eCognition software using very high spatial resolution imagery (VHRI).

6.0 Hierarchical Segmentation (HSeg) algorithm for cropland products: work @ NASA GSFC (James Tilton et al.)

We have previously developed a best merge region-growing approach that integrates nonadjacent region object aggregation with the neighboring region merge process usually employed in region growing segmentation approaches (e.g., Figure 26). This approach has been named HSeg, because it provides a hierarchical set of image segmentation results. Up to this point, HSeg considered only global region feature information in the region growing decision process. We present here three new versions of HSeg that include local edge information into the region growing decision process at different levels of rigor. We then compare the effectiveness and processing times of these new versions HSeg with each other and with the original version of HSeg. Figure 26 shows the subset of the image data (Figure 26a) along with the subset of the edge image (Figure 26b) produced by the Frei-Chen edge operator (maximum of the four spectral bands). Figure 26 shows a 512 by 512 pixel subset of this image to more clearly illustrate the results.



(a) (b)

Figure 26. (a) Bands 3, 2, 1 of a 512x512 pixel subset of a Quickbird image over South Africa displayed as RGB. (b) A 512x512 pixel subset of the Frei-Chen edge operator for this image, maximum over spectral bands.

Currently, we have proposed three alternate approaches for incorporating the edge information into HSeg image segmentation approach (Figure 27, 28). Some quantitative results from a plurality vote classification approach provide mixed results concerning the effectiveness of the new implementations as compared to the previous version of HSeg. However, we have noted that large homogeneous areas are merged into one region much earlier in the region growing process with the new versions, as was desired (e.g., Figure 29). We will continue to evaluate and compare these versions of HSeg on other data sets, noting the tradeoffs between computation time and segmentation quality.

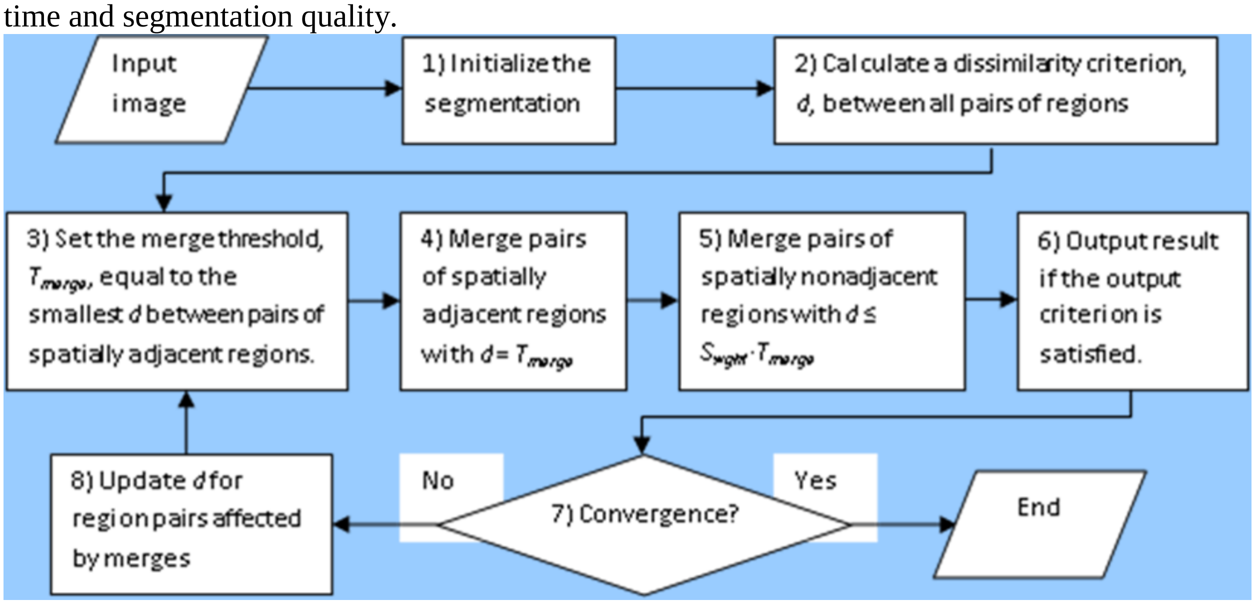


Figure 27. HSeg illustrating alternate approaches for incorporating the edge information into HSeg image segmentation approach.

$$G_1 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & \sqrt{2} & 1 \\ 0 & 0 & 0 \\ -1 & -\sqrt{2} & -1 \end{bmatrix} \qquad G_2 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & 0 & -1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & 0 & -1 \end{bmatrix} \qquad G_3 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 0 & -1 & \sqrt{2} \\ 1 & 0 & -1 \\ -\sqrt{2} & 1 & 0 \end{bmatrix}$$

$$G_4 = \frac{1}{2\sqrt{2}}\begin{bmatrix} \sqrt{2} & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -\sqrt{2} \end{bmatrix} \qquad G_5 = \frac{1}{2}\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \qquad G_6 = \frac{1}{2}\begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

$$G_7 = \frac{1}{6}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \qquad G_8 = \frac{1}{6}\begin{bmatrix} -2 & 1 & -2 \\ 1 & 4 & 1 \\ -2 & 1 & -2 \end{bmatrix} \qquad G_9 = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Figure 28. HSeg illustrating alternate approaches for incorporating the edge information into HSeg image segmentation approach.

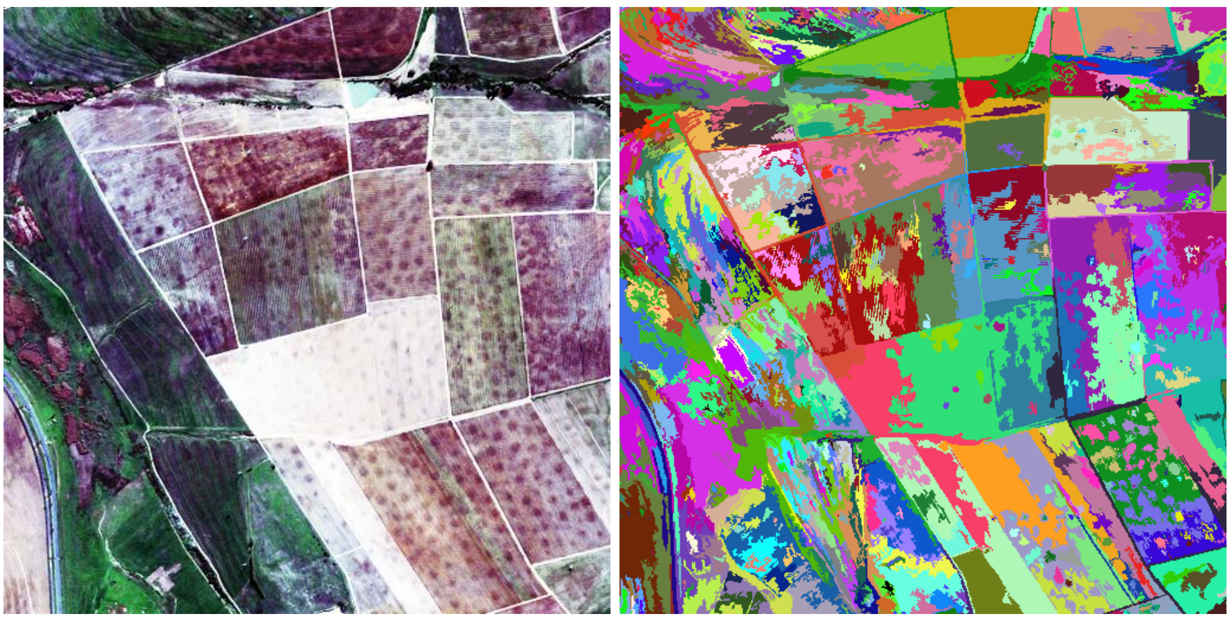


Figure 29. Large homogeneous areas are merged into one region much earlier in the region growing process with the new versions of Hseg, as was desired.

7.0 Google Earth Engine (GEE) Computing (Justin Poehnelt, Jun Xiong et al.)

One of the goals of this project is to use the power of Google Earth Engine (GEE) to power and process massively large data volumes. GEE houses all the Landsat and MODIS archives, and can be used to perform parallel processing by scripting in GEE playfield using Java Script and Python. Currently, we have on GEE, the following GFSAD30 data:

Global Cropland Extent (GCE) 1km Crop Dominance (aka GCE V0.0) [see section 2.1.1]
https://earthengine.google.org/#detail/USGS%2FGFSAD1000_V0
Global Cropland Extent (GCE) 1km Multi-study Crop Mask (aka GCE V1.0) [see section 2.1.2]
https://earthengine.google.org/#detail/USGS%2FGFSAD1000_V1

The GEE will allow to interactively zoom in to any are of the world, instantly process and classify data based on java script and python coding in play field, and save data.

8.0 Crowdsourcing global ground data through Croplands.org (Justin Poehnelt et al.)

The main objectives of crowdsourcing global ground data through Croplands.org are:

A. Create an online ground data mechanism that is live, interactive, and analytical;
B. mechanism for others to contribute their own knowledge on a location to the project;
C. Allows our team to review and modify ground data using high resolution satellite imagery, NDVI time series and other data sources as available.

Visit croplands.org site at:
http://www.croplands.org/

9.0 LP DAAC (Pardhasaradhi Teluguntla et al.)

All GFSAD30 data will be disseminated through the Land Processes Distributed Active Archive Center (**LP DAAC**), A NASA Earth Observing System Data and Information System (EOSDIS). Currently, we have provisional distribution of global cropland products:

Global Cropland Extent (GCE) 1km Crop Dominance (aka GCE V0.0) [see section 2.1.1]
http://e4ftl01.cr.usgs.gov/GFSAD/GFSADCD1KM/
Global Cropland Extent (GCE) 1km Multi-study Crop Mask (aka GCE V1.0) [see section 2.1.2]
http://e4ftl01.cr.usgs.gov/GFSAD/GFSADCM1KM/

10.0 GFSAD30 web portal (Justin Poehnelt et al.)

The global food security-support analysis data @ 30 m (GFSAD30) project web site provides every information on the project that includes project goal and objectives, project progress, link to appropriate product releases and all other necessary information. This information can be obtained through the links, here:
http://geography.wr.usgs.gov/science/croplands/
http://geography.wr.usgs.gov/science/climateLCC.html

11.0 Computing Resources allocation on the NASA Earth Exchange (NEX)

We upgraded the space allocated to the GFSAD30 project on the NASA Earth Exchange (https://nex.nasa.gov) from 25TB to 50TB. This additional space was required to allow prototyping parallelized classifiers that can take advantage of parallel computing capabilities of Pleiades Supercomputer and, at the same time, explore the use of dense time stacks of Landsat data, both from the Landsat archive and from WELD CONUS data. We also started experimenting with the processing of large Landsat composites with Endeavor, the latest supercomputer acquired by NASA's Advanced Supercomputer Facility. Endeavor (http://www.nas.nasa.gov/hecc/resources/endeavour.html) is a large shared memory system (6 TB of total memory) which should increase the performance of unsupervised clustering algorithm applied to large datasets (i.e., continental Landsat composites). Currently, any computational job requiring more than 252 GB will not run on the Pleiades supercomputer. Clustering techniques based on unsupervised algorithms are inherently difficult to parallelize and require large memory resources. We are in the process of testing the performance of unsupervised classifiers on continental Landsat composites with the large shared memory of Endeavor.

12.0 GFSAD30 Team composition

GFSAD30 Project Team names and affiliation (Project Team)

Prasad Thenkabail, PI, USGS
Cristina Milesi, co-I, NASA AMES\CSUMB
Mutlu Ozdogan, co-I, UW
Russ Congalton, co-I, UNH
Chandra Giri, co-I, USGS EROS
James Tilton, co-I, NASA GSFC
Temuulen Teki, co-I, NAU
Pardhasaradhi Teluguntla, Research Scientist, BAERI\USGS
Jun Xiong, Post doc, NAU\USGS
Justin Poehnelt, IT expert, NAU\USGS
Murali Krishna Gumma, ICRISAT
Richard Massey, PhD student, NAU\USGS
Aparna Phalke, PhD student, UW
Kamini Yadav, PhD student, UNH
Gu Jianyu, PhD student, UNH

LP DAAC (web portal, data portal, web map)

Dave Meyer, U. S. Geological Survey
Stacie Doman Bennett, U. S. Geological Survey

Web Master (web portal, data portal, web map)

Justin Poehnelt, IT expert, NAU\USGS
Jeff Peters, U. S. Geological Survey

Google Earth Engine: (Python and Java scripts on Playfield);

Justin Poehnelt, IT expert, NAU\USGS
Jun Xiong, Post doc, NAU\USGS
David Thau, Google

IT Support

Mr. Miguel Velasco, U. S. Geological Survey
IT Unit of Astrogeology, Flagstaff, USGS

14.0 Workshops and Meetings

The fourth workshop of the GFSAD30 was held on June 24-26, 2014 at the USGS in Sioux Falls, SD. The fifth workshop will be held in Reston, VA during January 28-30, 2015.