NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) 30-m Cropland Extent Validation (GFSAD30VAL)

User’s Guide
## Document History

<table>
<thead>
<tr>
<th>Document Version</th>
<th>Publication Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>September, 2017</td>
<td>Original</td>
</tr>
<tr>
<td>1.1</td>
<td></td>
<td>Modification made according to USGS reviewer comments</td>
</tr>
</tbody>
</table>
## Contents

Document History ........................................................................................................... 2

### 1.0 Dataset Overview .................................................................................................. 4
#### 1.1 Background ......................................................................................................... 4

### 2.0 Dataset Characteristics ............................................................................................ 5
#### 2.1 Global Food Security Support Analysis Data (GFSAD) Validation 30-m V001 ....... 5
##### 2.1.1 Collection Level .............................................................................................. 5
##### 2.1.2 Granule Level .................................................................................................. 5
##### 2.1.4 Data Layers Classification .............................................................................. 6
##### 2.1.5 Filename Convention ....................................................................................... 6

### 3.0 Dataset Knowledge .................................................................................................. 6
#### 3.1 Frequently Asked Questions ................................................................................ 6

### 4.0 Dataset Access (Applicable Data Tools) ................................................................. 8

### 5.0 Contact Information ................................................................................................ 8

### 6.0 Citation .................................................................................................................... 8
#### 6.1 GFSAD30CEAF ...................................................................................................... 8

### 7.0 Publications .............................................................................................................. 8
#### 7.1 Peer-reviewed publications within the GFSAD project ............................................ 8
#### 7.2 Other relevant publications prior to the GFSAD project ........................................ 10
#### 7.3 Books and Book Chapters ...................................................................................... 11
1.0 Dataset Overview

The goal of the Global Food Security-support Analysis Data @ 30-m (GFSAD30) project was to provide the highest resolution, objective cropland datasets to assist and address global food and water security issues in the twenty-first century. The project developed cropland products using time-series Landsat and Sentinel satellite sensor data, machine learning algorithms, and cloud-based computing. All products produced were assessed for accuracy (validated) using an independent set of reference data and an error matrix approach. The project was funded by the National Aeronautics and Space Administration (NASA) with supplemental funding from the United States Geological Survey (USGS). The project was led by USGS and carried out in collaboration with NASA AMES, University of New Hampshire (UNH), Bay Area Environmental Research Institute (BAERI), University of Wisconsin (UW), NASA GSFC, and Northern Arizona University (NAU). There were a number of International partners, including The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). The validation/accuracy assessment component was led by Dr. Russell G. Congalton at the University of New Hampshire.

This user’s guide provides information about the GFSAD30VAL reference data set generated for the assessment of the cropland extent maps generated by the other GFSAD30 team members. The Coordinate Reference System (CRS) used for the GFSAD30CEAF is a geographic coordinate system (GCS) based on the World Geodetic System 84 (WGS84) reference ellipsoid. The legend is presented in Section 2. Datasets are provided using ESRI vector shapefile format. The year, location, attributes, and file name convention details are provided in section 2.0 of this document.

1.1 Background

Monitoring global croplands is imperative for ensuring sustainable water and food security for people of the world in the twenty-first century. However, the currently available cropland products suffer from major limitations such as: (1) the absence of precise spatial location of the cropped areas; (2) The coarse resolution nature of the map products with significant uncertainties in areas, locations, and detail; (3) The uncertainties in differentiating irrigated areas from rainfed areas; (4) The absence of crop types and cropping intensities; and/or (5) The absence of a dedicated Internet data portal for the dissemination of these cropland products. This project aims to address all these knowledge gaps.

Satellite-derived cropland extent maps at high spatial resolution are necessary for food and water security analysis. Therefore, the GFSAD30 team produced cropland extent maps of the continental regions of the world at 30-m resolution for the nominal year 2015 using Sentinel-2 and Landsat-8 time-series imagery. These cropland extent maps are part of a global data release, whereby each continental region will be made publically available. Global cropland extent maps, indicating cropland and non-cropland areas, provide a working baseline data set to develop higher-level products such as crop watering method (irrigated or rainfed), cropping intensities (e.g., single, double, or continuous cropping), crop type mapping, cropland fallow, as well as the assessment of cropland productivity (productivity per unit of land), and crop water productivity (productivity per unit of water or “crop per drop”). Uncertainties associated with cropland extent maps have a cascading effect on all these higher-level cropland products.

Cloud-based geo-spatial computing platforms and satellite imagery offer opportunities for producing precise and accurate maps of cropland extent and area that meet the spatial and temporal requirements for a broad set of applications. Such maps can be a significant improvement compared to existing products, which tend to be coarser resolution, are often not representative of regions with highly dynamic change, and have a fixed set of cover classes. Cloud-based computing platforms such as
Google Earth Engine and new earth-observing satellites like the Sentinel constellation have brought significant improvements to land use/land cover (LULC) mapping and agriculture monitoring.

Producing the GFSAD30 cropland extent maps would not be complete without a thorough and independent accuracy assessment. The biggest limiting factor in conducting such an assessment is the lack of sufficient and appropriate reference data to generate valid error matrices. Therefore, significant effort was expended to collect and/or generate sufficient reference data using existing sources, ground-based field campaigns, and interpretation of very high resolution imagery (VHRI) to complete this assessment. The reference dataset (GFSAD30VAL) generated by the team at the University of New Hampshire is described in this User's Guide.

For a very detailed description of the entire accuracy assessment process including collection of other reference data, appropriate sampling, error matrix generation, and difference image creation, please refer to the algorithm theoretical basis document (ATBD) of GFSAD30VAL.

### 2.0 Dataset Characteristics

The global food security-support analysis data @ 30-m cropland extent validation (GFSAD30VAL) dataset and characteristics are described below.

#### 2.1 Global Food Security Support Analysis Data (GFSAD) 30-m Validation V001

**2.1.1 Collection Level**

<table>
<thead>
<tr>
<th>Short name</th>
<th>GFSAD30VAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Granularity</td>
<td>Static</td>
</tr>
<tr>
<td>Temporal Extent</td>
<td>2015, nominal</td>
</tr>
<tr>
<td>Spatial Extent</td>
<td>Global</td>
</tr>
<tr>
<td>File size</td>
<td>~31 MB</td>
</tr>
<tr>
<td>Coordinate System</td>
<td>Geographic</td>
</tr>
<tr>
<td>Datum</td>
<td>WGS84</td>
</tr>
<tr>
<td>File Format</td>
<td>ESRI Shapefile</td>
</tr>
</tbody>
</table>

**2.1.2 Granule Level**

<table>
<thead>
<tr>
<th>Number of Layers</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Samples</td>
<td>19171</td>
</tr>
<tr>
<td>Sample Unit Size</td>
<td>~90 m (3x3 30 m pixels)</td>
</tr>
</tbody>
</table>
### 2.1.4 Data Layers Classification

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Cropland</td>
<td>Non-Cropland areas</td>
</tr>
<tr>
<td>Cropland</td>
<td>Cropland and Fallow-land</td>
</tr>
</tbody>
</table>

### 2.1.5 Filename Convention

GFSAD30VAL_2015_001_2017001121526.tif

GFSAD30VAL = Product Short name
30 = 30 m Spatial Resolution*
VAL = Validation
2015 = Nominal Year
001 = Version
20170526 = Processing Date in YYYYJJJHHMMSS

*The spatial resolution of the sample unit is ~90 m

### 3.0 Dataset Knowledge

The following questions address the user information regarding the GFSAD30VAL collection.

### 3.1 Frequently Asked Questions

**What are the accuracies of the global cropland extent product?**

For the entire world, the global cropland extent product had an overall accuracy of 91.7%. For the cropland class, the producer’s accuracy was 83.4% (errors of omission of 16.6%) and user’s accuracies of 78.3% (errors of commissions of 21.7%). This means, for the entire global product, 16.6% of the croplands are missing and and 21.7% of non-croplands get mapped as croplands. However, we performed accuracies in 72 distinct zones of the world and provide error matrix for each of these zones. For those interested in these details, please refer to ATBD of the GFSAD30VAL as well as ATBD’s of each of the 7 zones of the world, where these details are provided.
Is this the only reference dataset used to assess the GFSAD30 cropland extent maps?
No, reference data used in this project were collected in a number of ways: 1- any existing data sets that were appropriate were collected including those from the US and Canada, 2- some ground-based field campaigns were conducted, and 3- interpretation of very high resolution imagery (VHRI). The reference data collected from VHRI was performed by two different groups in this project. Reference data collected by USGS team members for the world were divided into two parts. One part (60% of the data) was made available to the mapping teams to use for training and testing of their algorithms used to generate the cropland extent maps. The other part (40% of the data) was put aside and restricted for use in the independent validation conducted by the accuracy assessment team. All these data exist online at Croplands.org and are available through the LPDAAC. The other database, described in this User’s Guide (GFSAD30VAL) was also generated from VHRI and has remained independent from all the other reference data.

What does the GFSAD30VAL product contain?
This database contains sample locations (geographic coordinates), a label (either cropland or non-cropland), and zone (as described in GFSAD30VAL ATBD document) for randomly selected areas all over the globe (see Section 2.4.1)

What's the definition of the crop extent?
For the entire Global Food Security-Support Analysis Data project at 30-m (GFSAD30) project, cropland extent was defined as: “lands cultivated with plants harvested for food, feed, and fiber, include both seasonal crops (e.g., wheat, rice, corn, soybeans, cotton) and continuous plantations (e.g., coffee, tea, rubber, cocoa, oil palms). Cropland fallow are lands uncultivated during a season or a year but are farmlands and are equipped for cultivation, including plantations (e.g., orchards, vineyards, coffee, tea, rubber)” (Teluguntla et al., 2015). Cropland extent includes all planted crops and fallow lands. Non-croplands include all other land cover classes other than croplands and cropland fallow.

How can the dataset be obtained?
All the GFSAD30 products are downloadable through the Land Processes Distributed Active Archive Center (LP DAAC). GFSAD30VAL is among them. You can also visualize these data @: croplands.org by going to the “products” drop-down menu there.

Can anyone use these data?
Yes. The reference data and all products associated with the GFSAD30 project are freely available to anyone who would like to use them.

4.0 Dataset Access (Applicable Data Tools)

The GFSAD30VAL dataset is available through the LP DAAC Data Pool and NASA Earthdata Search. GFSAD data visualization and information can also be found at Global Croplands Website.

5.0 Contact Information

LP DAAC User Services  
U.S. Geological Survey (USGS)  
Center for Earth Resources Observation and Science (EROS)  
47914 252nd Street  
Sioux Falls, SD 57198-0001

Phone Number: 605-594-6116  
Toll Free: 866-573-3222 (866-LPE-DAAC)  
Fax: 605-594-6963

Email: lpdaac@usgs.gov  
Web: https://lpdaac.usgs.gov

For the Principal Investigators, feel free to write to:

Russell G. Congalton at russ.congalton@unh.edu  
Prasad S. Thenkabail at pthenkabail@usgs.gov

More details about the GFSAD project and products can be found at: globalcroplands.org

6.0 Citation

6.1 GFSAD30VAL


7.0 Publications

The following publications are related to the development of the above croplands products:

7.1 Peer-reviewed publications within GFSAD project


Web sites and Data portals:
http://croplands.org (30-m global croplands visualization tool)
http://geography.wr.usgs.gov/science/croplands/products.html#LPDAAC (dissemination on LP DAAC)
croplands.org (crowdsourcing global croplands data)
7.2 Other relevant past publications prior to GFSAD project


Thenkabail, P.S., Gangadhararao Rao, P., Biggs, T., Krishna, M., and Turral, H., 2007b. 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).


### 7.3 Books and Book Chapters


