ARC Flood Extent Depiction Algorithm Description Document

AFED Version V05R00
Document Revision R00

Atmospheric and Environmental Research, Inc. (AER)
131 Hartwell Avenue
Lexington, MA 02421-3126 USA
Phone: +1-781-761-2288

With contributions by:
John Galantowicz, Jeff Picton, and Ben Root

31 August 2018

Revision History

AFED Version	Document Revision	Date dd/mm/yyyy	Description
V00	R00	12/03/2015	Initial draft version for model outline.
V01	R00	12/05/2015	Updated draft for Model Draft 1
V02	R00	04/09/2015	Updated draft for Model Draft 2
V03	R00	29/03/2016	Updated for Final Historical Model (AFM Phase 1)
V03R01	R00	03/06/2016	Updated for AFM version V03R01.
V03R01	R01	02/08/2016	Added details regarding false positive flash detection step and relative floodability calculation. Renamed document and revised terminology to distinguish AFED from the ARC River Flood Model (AFM-R). Added section on data use policies. Added details and figures regarding iterative downscaling process. Made other minor corrections and additions.
V03R01	R02	15/12/2016	Added details for near real time processing system, including GMI processing. Added data use policy information for NCEP and GFS data. Made other minor corrections and clarifications.
V03R01	R03	23/01/2017	Updated description of near real time operations (Appendix B).
V04R00	R00	18/08/2017	Updated for AFED V04R00 with algorithm changes performed under contract QRSA-040-16 Task Order 1.
V04R01	R00	03/13/2018	Updated for AFED V04R01 with algorithm changes performed under contract QRSA-040-16 Task Order 2, including: new geolocation correction algorithm, revised flooded fraction algorithm with initial flooded fraction ensemble, revised false positive detection algorithm, and changes to some static ancillary datasets.
V05r00	R00	31/08/2018	Updated for AFED V05R00 with algorithm changes performed under contract QRSA-040-16 Task Order 3, including: addition of TMI sensor data, combined use of data from two sensors when available with conditional change from 3- to 2-day flooded fraction averaging, and replacement of the SRTM digital elevation model with the MERIT model with associated changes to relative floodability.

Table of Contents

1 Introduction 1
2 Algorithm high-level description 1
2.1 AFED product definition 1
2.2 Algorithm science background 4
2.3 Inputs 5
2.3.1 Primary microwave sensor inputs 5
2.3.2 Auxiliary microwave sensor inputs 5
2.3.3 Ancillary geographically and temporally varying inputs 6
2.3.4 Ancillary geographically varying static inputs 6
2.3.5 Ancillary lookup table static inputs 7
2.4 Outputs 8
2.4.1 Output products 8
2.4.2 Output quality control metadata definitions 8
2.5 High level flow diagram 9
2.6 High level processing outline 9
2.7 Near real time processing system 11
3 Algorithm components 13
3.1 Download microwave sensor data from archives 13
3.2 Reformat microwave sensor data 16
3.3 Footprint match microwave sensor data 16
3.4 Compute geolocation correction for microwave sensor data 17
3.5 Regrid microwave sensor data 19
3.6 Compute atmospheric correction for microwave sensor data 23
3.7 Compute precipitation and frozen surface conditions flags 25
3.8 Compute flooded fraction 32
3.8.1 Initial flooded fraction 33
3.8.2 Flooded fraction false positive detection 34
3.8.3 Final flooded fraction 35
3.9 Perform downscaling to depict flood extent 36
4 Static ancillary data generation 38
4.1 Footprint matching coefficient sets 38
4.2 Land-water mask (LWM) 39
4.3 Persistent open water area fraction 42
4.4 Tree cover area fraction 42
4.5 Elevation model 42
4.6 Normalized difference vegetation index climatology 42
4.7 CSU SSM/I calibration alignment parameters 43
4.8 AMSR2-to-AMSR-E calibration alignment parameters 43
4.9 GMI-to-AMSR-E calibration alignment parameters 44
4.10 TMI-to-AMSR-E calibration alignment parameters 44
4.11 Persistent open water *Q* end-member 45
4.12 *Q* 31- and 61- day median climatologies 47
4.13 *Q* 61-day historical year-to-year variance 48
4.14 *Q* 61-day median time shift model parameters 48
4.15 *Q* 31-day median time shift model parameter 49

4.16 *Q* tabulated by NDVI and tree cover fraction 49
4.17 Detected seasonal wetlands climatology 50
4.18 Dry-land *Q* end-member climatology 51
4.19 Flooded fraction historical outlier prevalence factor 52
4.20 Minimum detectable flooded fraction 54
4.21 Flooded fraction threshold 54
4.21.1 Relative floodability calculation 54
4.21.2 Flooded fraction threshold calculation 55
5 References 56
A. Data use policies 59
B. Near real time operations 60

1 Introduction

The African Risk Capacity (ARC) flood extent depiction (AFED) is a daily representation of temporarily flooded areas of Africa. The AFED algorithm derives this depiction by processing satellite sensor observations acquired from a number of similar space-borne microwave sensors that together provide a nearly continuous daily record since 1992. The purpose of this algorithm description document (ADD) is to document the relationship between the AFED algorithm theory, mathematical approach, and implementation details. For each separate algorithm process, the ADD summarizes:

- the algorithm theoretical basis;
- inputs that the algorithm uses to sense continuously changing earth surface and atmospheric conditions;
- inputs that the algorithm uses to represent conditions assumed to be unchanging (e.g., topography, persistent water);
- inputs that the algorithm uses to represent expected seasonally changing conditions (e.g., vegetation amount);
- the algorithm outputs;
- the processing procedure and logic used to execute the algorithm.

The ADD describes algorithm logic for every processing step the AFED algorithm applies to transform satellite sensor observations to daily flood depictions. In addition to describing the separate algorithm processes, the ADD also describes the end-to-end processing logic and architecture, system requirements, algorithm data file format and content, and user instructions. A companion document – *ARC Flood Extent Depiction Performance Document* (Galantowicz et al., 2018) – provides AFED performance assessments.

2 Algorithm high-level description

2.1 AFED product definition

The AFED algorithm produces the AFED product. AFED gives a binary (duel) indication of whether an area of land was temporarily flooded or not flooded on a certain date. Each binary indication represents conditions in a geographic area with predefined and unchanging boundaries. All such areas together form a non-overlapping grid of nearly rectangular cells (*grid cells*) covering all land areas of Africa.

AFED indicates any duration of flooding in a grid cell on a certain date; it does not differentiate whether the grid cell was flooded on all or part of the day. Similarly, AFED indicates any area of flooding in the grid cell; it does not differentiate whether all or part of the grid cell was flooded.

AFED uses a land-water mask to distinguish areas of persistent open water from land. AFED may indicate temporary flooding only for grid cells denoted as *land* in the land-water mask. The land-water mask is provided as a separate data product.

Table 1 defines AFED product characteristics.

Table 1: ARC flood extent depiction (AFED) product definition.

Product Characteristic	Definition
Data values:	
Definition	0: Unflooded land or persistent open water 1: Temporarily flooded land
Units	Unitless binary value
Format:	
File data format	Values are stored as georeferenced raster imagery in GeoTIFF (http://trac.osgeo.org/geotiff/) files with one-bit encoding and *deflate* compression.
Single file (tile) coverage	One file covers a 5° latitude x 5° longitude geographic area (tile) described by the coordinates of the upper-left (UL) corner. See Figure 1.
Single file (tile) internal organization	One file includes a 6000 x 6000 pixel raster data array. Each pixel represents a 3-arcsecond latitude x 3-arcsecond longitude geographic area. An arcsecond is 1/60th degree.
File (tile) organization	Data is distributed as 152 5° latitude x 5° longitude tiles per day covering all land areas of Africa. See Figure 1.
Product coverage	All land areas of Africa. See Figure 1.
File naming convention	Template: <product>_<resolution>_<UL latitude>_<UL longitude>_<date>_<version>.<ext> afed_3s_LAT_LONG_YYYYMMDD_vVVrRR.tif Example: afed_3s_05N_010W_20070211_v01r00.tif • afed: ARC flood extent depiction product • 3s: 3 arcseconds latitude x 3 arcseconds longitude resolution • LAT: LLX, where LL is degrees from 00 to 90 and X is "N" or "S" (e.g., 10N) • LONG: LLLX, where LLL is degrees from 000 to 180 and X is "E" or "W" (e.g., 015W) • YYYY: 4-digit year • MM: 2-digit month (01 to 12) • DD: 2-digit day-of-month (01 to 31) • VV: AFED version number (from 01) • RR: Dataset revision number (from 00)
Spatial representation:	
Horizontal datum	World Geodetic System 1984 (WGS84, EPSG:4326[1])
Vertical datum	Not applicable
Coordinate system	Data values represent 3-arcsecond grid cells on a geographic coordinate system grid. Grid cells are edge aligned (see Figure 2).
Temporal representation:	
Reporting frequency	Daily
Reporting time	The nominal reporting time is 2100 UTC[2] of the reporting day.
Time period represented by a report	The nominal time period represented by a report is the 24-hour period prior to the reporting time. See sections 3.4, 3.6, and 3.9 for further details.
Temporal coverage	At this time historical dataset V05R00 covers all dates from 4 January 1992 to 29 August 2018. NRTPS V05R00 covers dates starting from 30 August 2018.

[1]European Petroleum Survey Group (EPSG) Parameter Set code 4326 (http://www.epsg-registry.org/).
[2]Coordinate Universal Time.

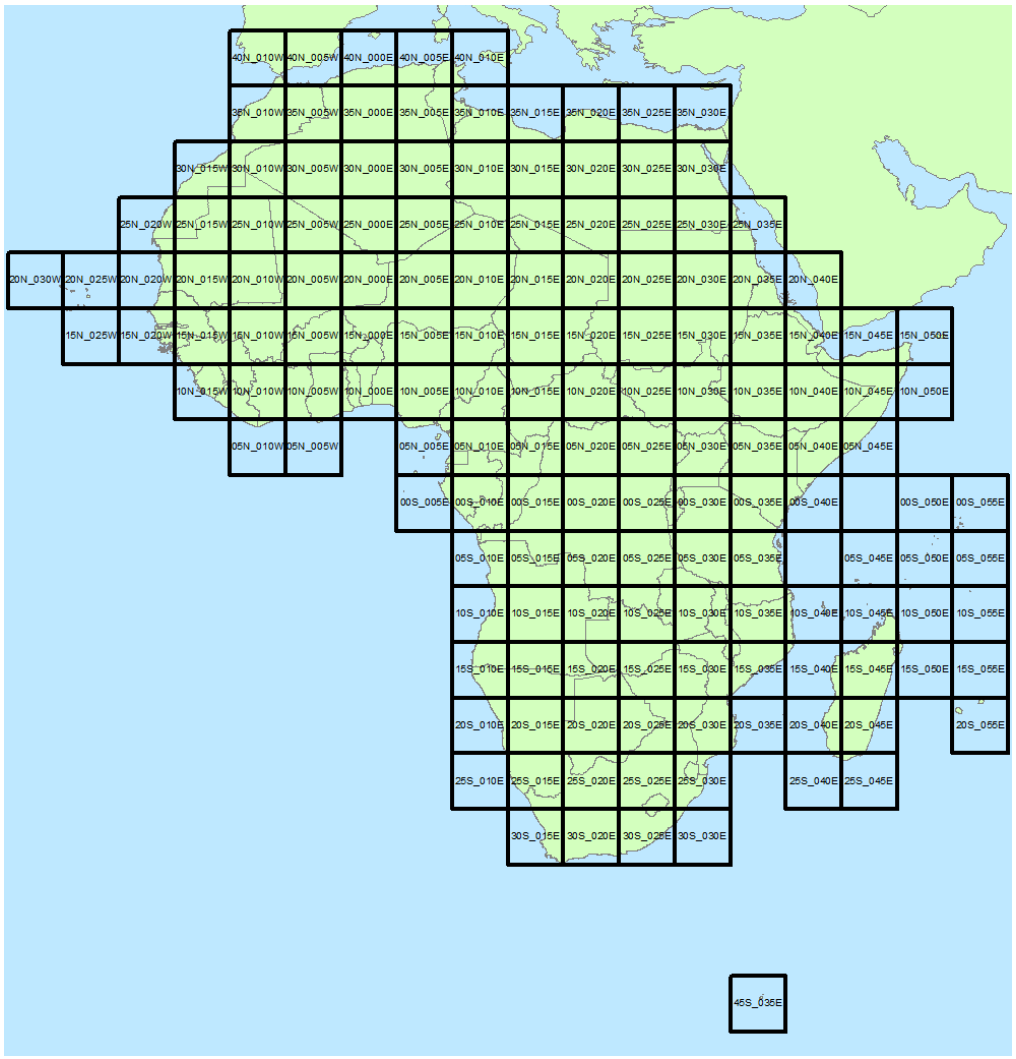

Figure 1: AFED data product tile coverage.

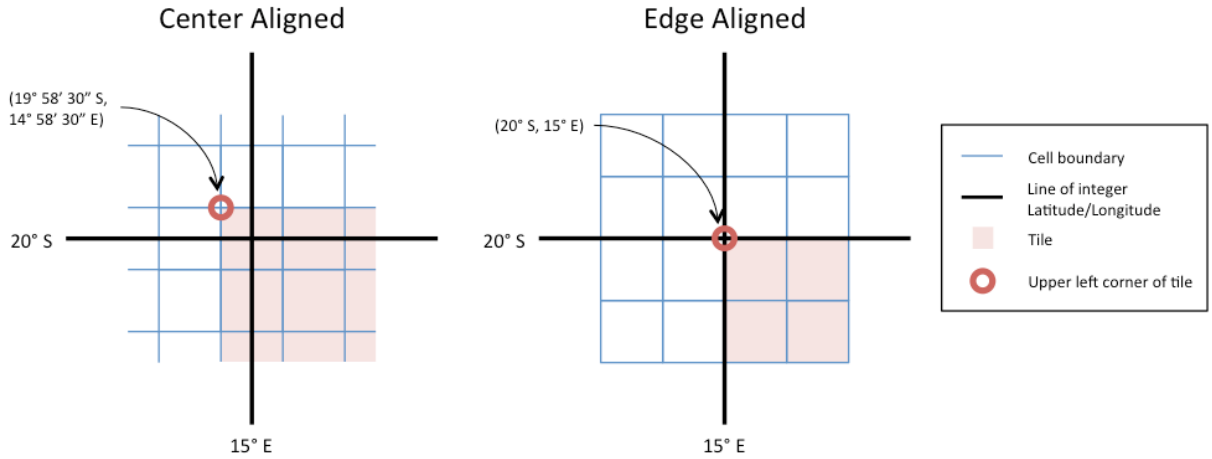

Figure 2: Definition of center-aligned and edge-aligned grid cells.

2.2 Algorithm science background

The primary inputs to the AFED algorithm are measurements made by passive microwave sensors carried on satellites in earth orbit. Table 2 defines key terminology from the field of microwave remote sensing needed to understand the scientific and theoretical basis for the algorithm.

Table 2: Definition of microwave remote sensing terminology used in this document.

Term	Definition
Brightness temperature (herein T_B or TB)	A passive microwave sensor measurement of radiation intensity calibrated to temperature units in kelvin, K
Polarization	The orientation of the measured microwave radiation wave relative to the earth's surface at the point of incidence. Microwave sensors use filters to make radiation measurements representing specific polarizations.
• TB with vertical polarization (V-pol.)	A TB measurement representing radiation intensity isolated to wave oscillations in a plane perpendicular to the earth surface at the point of incidence
• TB with horizontal polarization (H-pol.)	A TB measurement representing radiation intensity isolated to wave oscillations tangent (horizontal) to the earth surface at the point of incidence
Point of incidence	The central geographic location at which the sensor points during a measurement
Footprint	The relative weighting pattern in which geographic areas contribute to a microwave measurement; the peak contribution weight is typically at or near the point of incidence
Field of view (FOV)	The pattern on the earth surface at which the footprint contribution pattern is equal to 1/2 the peak contribution at the point of incidence. The sensor footprint FOV is approximately elliptical; after the footprint matching process the composite footprint FOV is approximately circular. See section 3.3 for further details.
Footprint size or FOV size	The geographic size of the FOV, expressed either as the average FOV diameter or as the distances across the FOV major and minor axes.
Resolution	A relative term referring to the size of the geographic area represented by a sensor measurement or other quantity; FOV size is one measure of resolution and is used as such in this document.
Orbit	The nearly circular path of a satellite around the center of the Earth
Satellite pass direction	The direction of satellite motion in its orbit at the times it passes over the equator: either *ascending* (south-to-north) or *descending* (north-to-south). There are exactly one ascending and one descending pass per orbital period. The AFED algorithm collects and processes ascending and descending pass data in separate groups organized by time of day and resamples them to an earth grid (defined below).
Sun-synchronous orbit	An orbit in which a satellite crosses the equator at roughly the same local solar time for all passes with the same satellite pass direction
Overpass type	Herein, overpass type is identified as *ascending* and *descending* for sun-synchronous satellites or as gridded data file numbered sequentially per day *G1*, *G2*, and *G3* for non-sun-synchronous satellites.
Swath file format	Microwave sensor data in a file organized according to the UTC time at which the data were collected and referenced to the geographic location of the points of incidence by the original data provider.
Earth or map grid	A set of geographic areas with predefined and unchanging boundaries forming a non-overlapping grid of nearly rectangular cells (*grid cells*) covering all or a portion of the Earth. The AFED algorithm uses two grids, one with 3-arcsecond grid cell spacing and one with 5-arcminute spacing.
Frequency or band	Central microwave frequency at which a sensor operates, e.g., 19 GHz, 37 GHz, etc.
Channel	The combined frequency and polarization at which a sensor operates and defining a brightness temperature measurement, e.g., 19 GHz vertical polarization abbreviated as 19 GHz V-pol. or 19V
Flooded fraction	The flooded area of a footprint as a footprint-weighed fraction of the total footprint area
Downscaling	The conversion of coarser-resolution flooded fraction estimates to a flood extent

	depiction at a finer resolution
Flood extent depiction	A set of binary (duel) indications on an earth grid of whether each grid cell was temporarily flooded or not flooded on a certain date

Passive microwave sensors measure radiation emitted naturally from the earth surface and atmosphere. The sensors make several measurement types distinguished by their frequency bands and polarization filters, herein referred to as channels (e.g., 19V, 37H, etc.). The AFED algorithm derives flood depictions using the strong sensitivity of these microwave measurements to flood area within sensor channel footprints. The algorithm uses ancillary data to account for other conditions of the land surface and atmosphere that differ over the earth's surface and/or change over time and have known effects on microwave measurements. From the microwave and ancillary data inputs, the algorithm computes the footprint-weighted flooded fraction, which is a coarser-resolution measure of the total amount of flooding over a relatively large area. The algorithm then employs a downscaling process to convert flooded fraction estimates from many overlapping footprints to a finer-resolution depiction of flooding as a binary flooded/not-flooded map.

2.3 Inputs

2.3.1 *Primary microwave sensor inputs*

The primary microwave sensor input is brightness temperature, TB, in four sensor bands at approximately 19, 37, 22 or 24, and 85 or 89 GHz. The flooded fraction algorithm's primary input is 37 GHz band brightness temperature, which are processed at 22-km resolutionfor the TMI, AMSR-E, AMSR2, and GMI sensors (hereafter AMSRX sensor type) and 50-km resolution for the SSM/I sensors (SSMI sensor type). Section 3.1 describes the sensors and how they are used in the AFED algorithm.

Table 3: Primary microwave sensor inputs

Input	Description
19 GHz V-pol. TB	19 GHz vertical polarization brightness temperatures. Used in algorithms that detect false positives, frozen surface conditions and precipitation.
19 GHz H-pol. TB	19 GHz horizontal polarization brightness temperatures. Used in algorithms that detect false positives, frozen surface conditions and precipitation.
37 GHz V-pol. TB	37 GHz vertical polarization brightness temperatures. Used in flooded fraction algorithm. Used in algorithms that detect frozen surface conditions and precipitation.
37 GHz H-pol. TB	37 GHz horizontal polarization brightness temperatures. Used in flooded fraction algorithm. Used in algorithms that detect frozen surface conditions and precipitation.
22 or 24 GHz V-pol. TB	22 or 24 GHz vertical polarization brightness temperatures. Used in algorithms that detect frozen surface conditions and precipitation.
85 or 89 GHz V-pol. TB	85 or 89 GHz vertical polarization brightness temperatures. Used in algorithms that detect frozen surface conditions and precipitation.
85 or 89 GHz H-pol. TB	85 or 89 GHz horizontal polarization brightness temperatures. Used in algorithms that detect frozen surface conditions and precipitation.

2.3.2 *Auxiliary microwave sensor inputs*

Auxiliary input data vary with the primary microwave sensor inputs.

Table 4: Auxiliary microwave sensor inputs

Input	Description	Units
Latitude	Geographic latitude location of a TB measurement at the point of incidence	Degrees
Longitude	Geographic longitude location of a TB measurement at the point of incidence	Degrees
Observation time	Date and time of a TB measurement	UTC
Data quality flags	Categorical indication of data quality, including data availability	Categorical values

2.3.3 *Ancillary geographically and temporally varying inputs*

The following inputs vary geographically and temporally.

Table 5: Ancillary geographically and temporally varying inputs

Input	Description	Coverage period	Time Sampling	Geographic Grid	Dimensions per Grid Cell
Atmospheric parameters	Characterizes atmospheric conditions (vertical profile of temperature, pressure, water vapor, and clouds) for computation of terms for TB atmospheric correction	30 July 1999 to present	4 times daily: 0000, 0006, 0012, 0018 UTC	1° x 1° global	51 values
		Prior to 30 July 1999	2 times daily: 0000, 0012 UTC	2.5° x 2.5° global	51 values

2.3.4 *Ancillary geographically varying static inputs*

The following inputs are geographically varying and static. They do not change during data processing.

Table 6: Ancillary geographically varying static inputs

Input	Symbol	Description	Geographic Grid	Dimensions per Grid Cell
AFED Land-Water Mask	N/A	Distinguishes persistent open water from land for the purposes of the AFED algorithm	AFED 3-arcsecond grid	1 value
Flooded fraction threshold	f_{f0}	Reference flooded fraction for downscaling 5-arcminute flooded fraction to 3-arcsecond AFED	AFED 3-arcsecond grid	1 value
Persistent open water fraction	f_{pow}	Footprint-weighted fractional coverage of persistent open water as indicated by the AFED Land-Water Mask	AFED 5-arcminute grid	1 value
Dry land Q* end-member seasonal climatology	$Q_{dc}(x,d)$	Specifies the seasonally-varying, historical average dry land Q value for the flooded fraction algorithm. Precalculated per 5-arcminute grid location x and climatological day-of-year d_c for 19 and 37 GHz and separately for ascending and descending passes and for each composite TB resolution per sensor type.	AFED 5-arcminute grid	(365 days) x (2 frequencies) x (2 passes) x (2 resolutions) x (2 sensor types)
Detected seasonal wetlands	N/A	Specifies where and in what season conditions signifying regular annual	AFED 5-arcminute grid	365 days

climatology		flooding were detected in historical *Q* data.		
Q 31-day median time shift model parameter	b_{31}	Provides parameters for the linear model used by the flooded fraction algorithm to compute the expected value of the current 31-day time-centered median from the 31-day retrospective median.	AFED 5-arcminute grid	(365 days) x (2 frequencies) x (2 resolutions) x (2 sensor types)
Q 61-day historical year-to-year variance	$\sigma_{Q61c}^2(d_c,x)$	Provides historical year-to-year variance in 61-day median Q. Used in the flooded fraction algorithm to compute the expected value of the current dry land end-member.	AFED 5-arcminute grid	(365 days) x (2 frequencies) x (2 resolutions) x (2 sensor types)
Q 61-day median time shift model parameters	a_{61}, b_{61}	Provides parameters for the linear model used by the flooded fraction algorithm to compute the expected value of the current 61-day time-centered median from the 61-day retrospective median.	AFED 5-arcminute grid	(365 days) x (2 frequencies) x (2 resolutions) x (2 sensor types) x (2 parameters)
Minimum detectable flooded fraction (MDFF)	N/A	Specifies the flooded fraction limit below which the algorithm sets flooded fraction estimates to zero.	AFED 5-arcminute grid	(365 days) (2 frequencies) x (2 passes) x (2 sensor types)
Coastal mask for MDFF	N/A	Specifies where MDFF application should apply special logic for coastal areas	AFED 5-arcminute grid	1 value
Geolocation correction control points	N/A	Specifies the geographic coordinates of three types of control points for geolocation correction: high land-water gradient, nearby land, and nearby water	AFED 5-arcminute grid	N/A: 3 lists of grid coordinates
Coastal mask for geolocation quality control	N/A	Specifies near-coast grid points for the geolocation correction algorithm to flag as unusable if geolocation correction fails	AFED 5-arcminute grid	1 value
Flooded fraction historical outlier prevalence factor	f_{hop}	Factor on a 0 to 1 scale representing seasonal outlier prevalence in the initial 37-GHz 22-km flooded fraction historical time series	AFED 5-arcminute grid	365 days

* Q is computed as a non-linear function of the polarization ratio index. See section 3.6.

2.3.5 *Ancillary lookup table static inputs*

The following inputs are static lookup tables. They do not change during data processing and they are not graphically varying.

Table 7: Ancillary lookup table static inputs

Input	Symbol	Description	Dimensions
Footprint matching coefficients	a_i	Provides the weighting coefficients for the footprint matching algorithm	(7 channels) x (X scan positions) x (Y coefficients) x (2 resolutions) x (3 sensors)
Persistent open water Q end-member	Q_{pow}	Specifies the persistent open water Q value for the flooded fraction algorithm	(2 frequencies) x (2 passes) x (2 sensor types)

Flooded land Q end-member	Q_f	Specifies the flooded land Q value for the flooded fraction algorithm	(2 frequencies) x (2 passes) x (2 sensor types)
CSU* SSM/I calibration alignment parameters	p_1, p_2	Aligns CSU SSM/I *Q* values with RSS* calibration	(2 frequencies) x (2 passes)
GMI-to-AMSR-E cross-calibration parameters	A_{xGMI}, B_{xGMI}	Aligns GMI *Q* values with AMSR-E	(2 frequencies)
AMSR2-to-AMSR-E cross-calibration parameters	A_{xGMI}, B_{xGMI}	Aligns AMSR2 *Q* values with AMSR-E	(2 frequencies) x (2 passes)
TMI-to-AMSR-E cross-calibration parameters	A_{xTMI}, B_{xTMI}	Aligns TMI *Q* values with AMSR-E	(2 frequencies) x (2 TMI orbital altitudes)

* CSU: Colorado State University. RSS: Remote Sensing Systems.

2.4 Outputs

2.4.1 *Output products*

Table 8: AFED algorithm output products.

Name	Description	Dimensions	Time Sampling	Valid Range	Unit
AFED	ARC flood extent depiction	AFED 3-arcsecond grid	Daily	{0, 1}	Unitless binary value
QC Metadata	Quality control metadata for each grid cell	AFED 5-arcminute grid	Daily	See section 2.4.2	See section 2.4.2

2.4.2 *Output quality control metadata definitions*

The AFED algorithm generates a set of three quality control metadata products daily for each cell on the 5-arcminute AFED grid.

Table 9: AFED algorithm output quality control metadata products.

Name	Description	Units
Total N	Total number of instantaneous estimates used in daily flooded fraction calculation	Unitless integer value in range {0, 15}
Effective Time	Effective time of daily flooded fraction estimate. Computed as the weighted average of the observation times of all instantaneous flooded fraction values used in the day's flooded fraction calculation	Decimal days from 0000 UTC on AFED product date
Reason	Enumerates the primary reason data points were either unavailable from the original source or excluded by the AFED algorithm for each of the gridded sensor data files expected per day for each satellite overpass type. Overpass type is identified as *ascending* and *descending* for sun-synchronous satellites or as gridded data file *G1*, *G2*, and *G3* for non-sun-synchronous satellites.	Categorical values: 0: data point available for use in the algorithm, 1: data point missing from sensor swath, 2: data point available but excluded due to poor quality indicators from data source, 3: data point unavailable because of gaps between sensor swaths or a gap within a sensor swath, 4: data point excluded by frozen-surface detection algorithm, 5: data point excluded by precipitation detection algorithm, 6: data point excluded because atmospheric

		data was unavailable, 7: data point excluded by false positive "flash" detection algorithm.

2.5 High level flow diagram

Figure 3 illustrates the overall AFED algorithm processing flow in two parts. The upper section illustrates the end-to-end processing of microwave sensor data to derive flooded fraction on coarser-resolution footprints and to produce the flood extent depiction by downscaling of the coarser-resolution estimates. The lower section illustrates generation of the static relative floodability and minimum flooded fraction databases that are used in the downscaling process. These databases are generated in an offline process using static data (e.g., topography, microwave measurement footprint shape, persistent water cover) described in section 4.

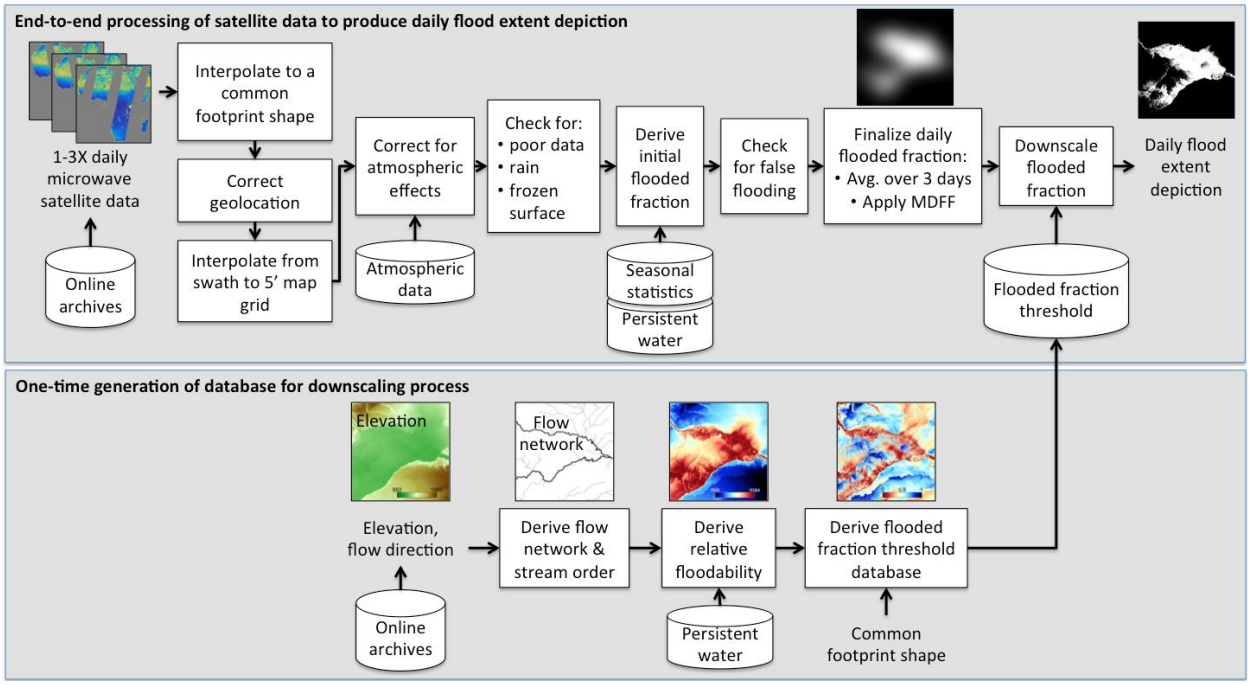

Figure 3: Overall AFED algorithm processing flow

2.6 High level processing outline

The AFED algorithm can be divided into three parts: the microwave (MW) part that derives flooded fraction; the downscaling (DS) part that produces the flood depiction from flooded fraction; and the process management (PM) part. The following table outlines the end-to-end algorithm processing steps and shows how the process and data files are managed by the PM. Processing is parallelized by date and also by sensor in selected periods where sensor operational records overlap. The operator may choose to process one or more days end-to-end or terminate processing at any of the steps leading to intermediate output files. Processing includes separate fine- and coarse-resolution composite brightness temperature datasets (hereafter fine-res CTB and coarse-res CTB, respectively). The fine-res CTB dataset includes all microwave bands and the coarse-res CTB dataset includes only 37-GHz band data.

Table 10: High level processing outline.

Part	Data granules	Processing steps
PM	In: Swath Out: Swath	**Swath data source file identification, downloading, and storage:** 1. Identify microwave sensor **swath data source file(s)** online via ftp meeting satellite, date, and geographic area criteria. 2. For each file: a. If the file is not already stored locally: download the file via ftp. b. Store file locally by date.
PM	In: Swath Out: Swath	**Swath data source file conversion to AFED NetCDF swath file format:** 3. Identify microwave sensor **swath data source file(s)** stored locally meeting satellite, date, and geographic area criteria. 4. For each file: a. If not already done: convert file from its original sensor-specific data format to the common NetCDF format used by the footprint matching process, **AFED NetCDF swath file**. b. Store file locally by date.
PM	Swath	**AFED NetCDF swath file selection for processing:** 5. Identify **AFED NetCDF swath files** stored locally meeting satellite, date, and geographic area criteria. 6. For each identified AFED NetCDF swath file:
MW	In: Swath Out: Swath	**AFED NetCDF swath file conversion to CTB swath file and geolocation correction:** a. If not already done: apply footprint matching interpolation coefficients to TBs in file to produce **coarse-res and fine-res composite TB (CTB) swath files**. b. Store files locally by date c. Apply geolocation correction to CTB swath files, changing file latitude and longitude variables in place.
PM	Swath	**CTB swath file selection for processing:** 7. Identify **CTB swath files** stored locally meeting satellite, date, and geographic area criteria and group by overpass type. 8. For each day and satellite overpass type group:
MW	In: Multiple Swaths Out: 5′ grid	**CTB swath file set conversion to CTB grid file:** a. Loop over coarse-res and fine-res CTB swath data files in time order. i. Find 5′ grid points within swath. ii. For each grid point within swath and not filled by data from a prior swath: 1. Interpolate the CTB data to the grid point. b. Produce one **CTB grid file** per overpass type group and CTB resolution. c. Store files locally by date.
MW	In: 5′ grid Out: 5′ grid	**CTB grid file conversion to CTB grid file with atmospheric correction:** 9. For each lo- and fine-res CTB grid file pair per overpass type group: a. Interpolate atmospheric data to coarse-res CTB observation times, compute coarse-res CTB terms for atmospheric correction, and produce coarse-res and fine-res **CTB grid files with atmospheric correction** using coarse-res atmospheric correction terms for coarse-res and fine-res corrections. b. Store file pair locally by date.
MW	In: 5′ grid Out: 5′ grid	**Precipitation and frozen surface conditions flagging:** 10. For each coarse-res **CTB grid file** per overpass type group: a. Apply algorithms to detect and flag 5′grid points with precipitation or frozen surface and produce **flag grid file**. b. Store file locally by date.
MW	In: 5′ grid Out: 5′ grid	**Flooded fraction computation:** 11. For each coarse-res and fine-res **CTB grid file with atmospheric correction** per overpass type group: a. Flag 5′grid points with missing CTB values to be skipped in processing.

		b. Use data from **flag grid file** to flag 5′grid points with precipitation or frozen surface to be skipped in processing. c. Compute initial flooded fraction estimates for each 5′ grid cell without missing values, precipitation, or frozen surface per each microwave frequency and satellite overpass type group and produce **initial flooded fraction file**. d. Store file locally by date 12. For each day and satellite overpass type group: a. Read or hold in memory coarse-res and fine-res **initial flooded fraction file** data from step 11 for this day and the day before. b. Apply algorithm to detect and flag 5′ grid points with flooded fraction false positive events to be skipped in processing. 13. For each day: a. Take the weighted mean of unflagged fine-res flooded fraction values over all satellite passes for this day and the two previous days as indicated by the historical sensor schedule or NRTPS settings. b. Apply physical limits to flooded fraction values ($0 \leq FF \leq$ land fraction). c. Apply missing value algorithm to fill in for missing values. d. Apply minimum detectable flooded fraction (MDFF) limit to flooded fraction values. e. Produce **daily flooded fraction file**. f. Produce **AFED quality control metadata files**.
DS	In: 5′ grid Out: 3″ grid	Flood extent depiction: 14. For each **daily flooded fraction file**: a. Apply downscaling algorithm to produce binary flooded/not-flooded determination in 3″ grid cells covering all land areas of Africa. b. Produce set of daily **AFED data product files** covering all land areas of Africa.

3″ = 3-arcseconds of latitude and longitude
5′ = 5-arcminutes of latitude and longitude

2.7 Near real time processing system

As of AFED version V04R00, the AFED near real time processing system (NRTPS) executes the same algorithm code as the historical processing system (HPS) with the following modifications and additional components needed to meet ARC's near real time data requirements.

- *Automation*: The NRTPS produces AFED automatically on a daily basis using a *crontab*. The *crontab* (*Cron* table) is a configuration file for the *Cron* job scheduler—a standard utility on Unix-like operating systems—that specifies the NRTPS commands to execute at intervals throughout the day (e.g., once every three hours). Once executed, NRTPS scripts check for the existence of required input files (i.e., from prior steps in the processing chain) and expected output files; processing proceeds only when inputs files are present and output files are not. For automatic recovery from data or system outages, the *crontab* (1) executes scripts to process the current product day and the two prior days such that no step is repeated for the same day unless the prior attempts failed to produce output files and (2) executes scripts to fill in missing data after a waiting period (to be determined, e.g., 24 hours) and proceed with processing. To protect against the generation of incomplete files, most processes include an "atomic save" step: an output file is first created with a randomly generated name that is only later changed to a name recognized by further processing steps after the generating processes has successfully completed.

- *Use GFS data for atmospheric correction*: The NRTPS uses data from the Global Forecast System (GFS) in the atmospheric correction (described in section 3.6) while HPS uses atmospheric data from National Centers for Environmental Prediction (NCEP) retrospective atmospheric analyses. The NRTPS uses GFS data for two reasons: GFS is intended for consumption by users depending on routine, reliable, and timely data and it includes forecast data that NRTPS stores and accesses as a contingency if results from one or more GFS analysis times are unavailable when needed. The NRTPS includes an independent processing stream called the ARC GFS Provider (AGP) that accesses GFS data from public servers, reformats it for compatibility with the NRTPS atmospheric correction step, and stages the resulting files for NRTPS retrieval. To protect against the temporary loss of data flow from GFS, the AGP produces files from GFS forecast results for future times and replaces them with analysis results when they become available.
- *Product delivery:* When new results are produced, NRTPS automatically sends the AFED and metadata files to an FTP site for ARC access and sends a notification email to designated recipients containing coverage statistics for the microwave data used in the products and metadata files (*reason* and *total N*, described in Table 9) and flooded fraction graphics as attachments.
- *Use of cloud computing*: The NRTPS, AGP, and FTP site run on three virtual servers hosted on the Amazon Web Services (AWS) Amazon Elastic Compute Cloud (EC2). Establishment of an AWS virtual server (or virtual machine, VM) entails creation of a VM image (or Amazon Machine Image, AMI), which is the data construct stored in the AWS library that fully describes the VM's configuration, operating system, and other software prior to its instantiation as a server. After instantiating a VM, we upload all the static data and prerequisite intermediate product files from prior days needed by the NRTPS to begin processing for the current date and activate the *crontab*.
- *GMI processing*: Automated NRTPS includes two data processing streams: one for AMSR2 (Advanced Microwave Scanning Radiometer 2, Global Change Observation Mission 1st-Water satellite) and one for GMI (GPM Microwave Imager, Global Precipitation Measurement satellite). GMI's satellite is in a non-sun-synchronous orbit, which means that swaths cannot be grouped and placed on a common grid by time of day, as they are for all other sensors, which are on sun-synchronous satellites. NRTPS uses a satellite orbit prediction model to catalog GMI swaths expected to cross Africa, associate them with grid files, and match incoming swath data files to the catalog for processing by the regridding algorithm. Further NRTPS steps use the database to determine the type and number of GMI gridded data files falling on each day. NRTPS enumerates GMI grid files as G1, G2, or G3, and there may be one, two or three data possible files per day. The NRTPS uses the orbit prediction model to designate the grid files to be recognized by the system each day, meaning that in the event of a gap in data availability, the NRTPS will skip over a grid with no data but maintain its place in the predicted grid sequence. This logic is consistent with that implemented for sun-synchronous satellites by the HPS, which implicitly expected two grids per day (designated as the descending and ascending pass direction grids), and for processing consistency the NRTPS uses the revised grid logic for AMSR2 as well as GMI, with the exception that AMSR2 grid files continue to be identified by pass direction.

As of AFED version V04R00, all HPS processing steps use the same Python language algorithm code as the NRTPS. This approach eliminates concerns that HPS-NRTPS code differences might cause numerical inconsistencies between historical and NRT AFED products.

3 Algorithm components

Table 11 lists the separate algorithm components described in detail in the following sections.

Table 11: AFED algorithm component summary

Name	ADD section	Algorithm Component Description
Download microwave sensor data from online archives	3.1	Downloads and organizes microwave sensor swath data files in HDF format for a given satellite, date range, and geographic area
Reformat microwave sensor data	3.2	Creates custom NetCDF-format microwave sensor swath data files for footprint matching algorithm
Footprint match microwave sensor data	3.3	Applies footprint matching interpolation coefficients to TBs in NetCDF-format microwave sensor swath data files to produce composite TBs (CTBs)
Compute geolocation correction for microwave sensor data	3.4	Adjusts CTB geographic coordinates to optimize CTB alignment with coastlines
Regrid microwave sensor data	3.5	Interpolates CTBs from their swath locations to points on a 5-arcminute fixed earth grid
Compute atmospheric correction for microwave sensor data	3.6	Uses atmospheric data, CTBs, and CTB observation time and location data to compute atmospheric transmission and emission terms associated with each CTB data point
Compute precipitation and frozen surface conditions flags	3.7	Computes flags indicating precipitation or frozen surface conditions from top-of-atmosphere CTBs on the 5-arcminute earth grid.
Compute flooded fraction	3.8	Computes flooded fraction from atmospherically-corrected CTBs on the 5-arcminute earth grid
Perform downscaling to depict flood extent	3.9	Produces the 3-arcsecond AFC flood extent depiction (AFED) product from flooded fraction inputs on the 5-arcminute earth grid

3.1 Download microwave sensor data from archives

End-to-end AFED processing begins with the acquisition of microwave sensor data from online public archives using automated scripts or manual download. Table 12 lists the baseline and alternative AFED algorithm microwave sensors.

Table 12: Microwave sensors and satellite platforms carrying them

Sensor	Platform	Operational	Footprint Resolution		AFED usage
			19 GHz	37 GHz	
SSM/I[1]	DMSP F11	12/1991-5/2000	69 x 43 km	37 x 28 km	Sensor for 1992-1997 historical product period
	DMSP F13	5/1995-11/2009			
	DMSP F14	5/1997-8/2008			
	DMSP F15	12/1999-present			
TMI[2]	TRMM	12/1997-9/2014	30 x 18 km / 35 x 21 km*	16 x 9 km / 18 x 10 km*	- Sensor for 1998-2014 historical product period - Covers 38°S – 38°N latitude
AMSR-E[3]	NASA Aqua	6/2002-10/2011	27 x 17 km	14 x 10 km	Sensor for 2002-2011 historical product period
AMSR2[4]	Jaxa GCOM-W1	7/2012-present	22 x 14 km	12 x 7 km	Sensor for 2012-present historical product period and near real time AFED process
GMI[5]	GPM	3/2014-present	18 x 11 km	16 x 9 km	- Sensor for 2014-present historical product period and near real time AFED process - Covers 65°S – 65°N latitude
Alternate sensors for near real time AFED process					
WindSat[6]	Coriolis	1/2003-present	27 x 16 km	13 x 8 km	Data not readily available

* TMI resolution increased in August 2001 when the TRMM satellite orbital altitude was boosted from 350 to 400 km.

The AFED algorithm uses the following data sources for sensor data:

- **RSS-calibrated SSM/I (1992-2011)**: National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC) Operational Climate Data Record (CDR) Program, *SSMI(S) Brightness Temperatures – RSS, Version 7* (Hilburn and Wentz, 2008). This version of the SSM/I data has been calibrated by Remote Sensing Systems (RSS) to the same standards across all SSM/I series sensors listed below as well as AMSR-E.
  Home page: http://www.ncdc.noaa.gov/cdr/operationalcdrs.html
  Bulk data access:
  http://www.ncdc.noaa.gov/has/HAS.FileAppRouter?datasetname=SSMIRSS&subqueryby=STATION&applname=&outdest=FILE
  FTP (script access): N/A
- **CSU-calibrated SSM/I (2012)**: National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC) Operational Climate Data Record (CDR) Program, *SSMI(S) Brightness Temperature - CSU* (Berg et al., 2013; Kummerow et al., 2013). Prior to AFED version V05R00, CSU-calibrated SSM/I data provided coverage for 2012, which is not covered by the RSS SSM/I dataset used for other periods. As of V05R00, AFED processing does not use SSM/I data after

[1] Special Sensor Microwave/Imager (http://www.ncdc.noaa.gov/oa/rsad/ssmi/swath/index.html)
[2] Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (http://trmm.gsfc.nasa.gov/)
[3] Advanced Microwave Scanning Radiometer-EOS (http://nsidc.org/data/amsre)
[4] Advanced Microwave Scanning Radiometer 2, Global Change Observation Mission 1st-Water (http://www.jaxa.jp/projects/sat/gcom_w/index_e.html)
[5] GPM Microwave Imager, Global Precipitation Measurement (http://pmm.nasa.gov/)
[6] http://www.nrl.navy.mil/WindSat/, http://www.cpi.com/twiki/bin/view/WindSat/WebHome

1997/1/11.
Home page: http://www.ncdc.noaa.gov/cdr/operationalcdrs.html
Bulk data access:
http://www.ncdc.noaa.gov/has/HAS.FileAppRouter?datasetname=CSU_SSMIS&subqueryby=STATION&applname=&outdest=FILE
FTP (script access): N/A

- **AMSR-E**: National Snow and Ice Data Center (NSIDC, http://nsidc.org) *AMSR-E/Aqua L2A Global Swath Spatially-Resampled Brightness Temperatures, Version 3*. Calibrated to same standard as RSS SSM/I data.
  Home page: http://nsidc.org/data/amsre/
  Bulk data access: http://reverb.echo.nasa.gov/reverb
  FTP (script access): ftp://n5eil01u.ecs.nsidc.org/SAN/AMSA/AE_L2A.003/
- **AMSR2**: Jaxa (Japan Aerospace Exploration Agency), GCOM-W1 Data Providing Service, *GCOM-W1 AMSR2 Brightness Temperature L1B and L1R, Version 1.1* (or later).
  Home page: http://gcom-w1.jaxa.jp/index.html (free registration required)
  Bulk data access: http://gcom-w1.jaxa.jp/searchsat.html
  HTTP data browse: http://gcom-w1.jaxa.jp/product-download.html
  SFTP (script access): sftp -oPort=2051 *username*@gcom-w1.jaxa.jp
- **GMI:** Goddard Earth Science Data and Information Services Center (GES DISC, http://daac.gsfc.nasa.gov)
  Home Page:
  https://daac.gsfc.nasa.gov/datasets/GPM_1BGMI_V05/summary?keywords=gmi
  Bulk data access: http://reverb.echo.nasa.gov/reverb
  FTP (script access): ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata
- **TMI:** Goddard Earth Science Data and Information Services Center (GES DISC, http://daac.gsfc.nasa.gov)
  Home Page:
  https://daac.gsfc.nasa.gov/datasets/GPM_1BTMI_V05/summary?keywords=tmi
  Bulk data access: http://reverb.echo.nasa.gov/reverb
  FTP (script access): ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata

The algorithm uses a common script for TMI, AMSR-E, AMSR2, and GMI downloads via FTP/SFTP. The script uses a satellite orbit prediction model to determine which data granules (i.e., full- or half-orbit segments) on a given day include data from a selected region, specified either as a shapefile polygon (as it is for AFED) or latitude-longitude box. The script downloads selected granules as needed for AFED algorithm processing, reducing the local storage burden.

For historical processing, we downloaded global RSS and CSU SSM/I full-orbit data granules manually using NOAA's Hierarchical Data Storage System (HDSS) Access System (HAS). A script computed the degree to which each granule overlapped the AFED Africa grid domain, set aside granules with no overlapped, and, on days when data from more than one DMSP platform were available, selected the platform to use for further processing at each overpass time.

A comma-separate-variable, CSV, file (e.g., *afm_sensor_schedule_v05r00.csv*) lists the sensors, platform, and flooded fraction resolution used in AFED processing per each day through

2018-08-31-24. As of AFED V05R00, the algorithm uses all overpass types available from the sensor or sensors listed each day.

3.2 Reformat microwave sensor data

Scripts convert downloaded microwave sensor data files in sensor-specific data formats to a common NetCDF format. This eliminates unneeded data fields present in the downloaded data and further reduces storage space for data granules to be retained.

3.3 Footprint match microwave sensor data

Footprint matching is a well-developed process for normalizing the spatial sampling of multi-frequency microwave sensor measurements and controlling spatial sampling during earth gridding (Backus and Gilbert, 1970; Stogryn, 1978; Poe, 1990; Galantowicz and England, 1991; Galantowicz et al., 2003; Galantowicz, 2004; Gu and England, 2007). Microwave sensors measure weighted averages of brightness temperature over elliptical footprints with two-dimensional Gaussian-like horizontal weighting functions. Typically, lower-frequency channels have coarser resolution footprints than higher-frequency channels (e.g., Table 12). Footprint matching is a linear process that resamples neighboring footprints in the sensor swath reference frame to form composite footprints optimized to match a defined reference shape common among all channels of interest. This process is summarized in the equation:

$$T_{Bc} = \sum_{i=1}^{N} a_i T_{Bi} \tag{1}$$

where a_i are the coefficients (summing to 1) weighting N brightness temperature measurements, T_{Bi}, to form the *composite brightness temperature* sample, T_{Bc}. The coefficient set depends on the position in the swath of the composite sample because the relative geometries of the measurements shift with the sensor scan position. Derivation of coefficients is described in section 4.1.

The footprint matching algorithm creates footprint-matched (composite) brightness temperatures (CTBs) in the swath reference frame. The algorithm computes CTB samples in the portions of each half-orbit granule where the scan lines intersect land. The CTB are created at predetermined points within the sensor swath geometry that are intentionally spaced to provide oversampled data (i.e., sample ground spacing is less than half the CTB resolution). The compositing process also calculates CTB geographic coordinates by interpolation of the input TB latitude and longitude data in the along-track/along-scan swath geometry.

The AFED algorithm computes AMSR-E, AMSR2, and GMI CTBs at two resolution levels – 22 km (37 GHz channels) and 27 km (19, 24, 37, 89 GHz channels) –TMI CTBs at 22 km (37 GHz) and 36 km (19, 22, 37, 85 GHz), and SSM/I CTBs at 50 km (37 GHz) and 69 km (19, 22, 37, 85 GHz). The defined reference shapes for 27-, 36-, and 69-km CTBs have a circular Gaussian weighting function:

$$W(x) = W_0 \exp\left(-a_2 x^2 / D^2\right) \tag{2}$$

where x is distance from the footprint center, D is the resolution diameter, W_0 is the weight at x=0, and a_2 is a constant defined such that $W(D/2) = 0.5$. The defined reference shapes for 22- and 50-km CTBs have a modified weighting function that concentrates more total weight within the area defined by $x<D/2$ without changing the feature that $W(D/2) = 0.5$:

$$W(x) = W_0 \exp\left(-a_{2.4}\, x^{2.4} / D^{2.4}\right). \tag{3}$$

The term *reference footprint shape* (RFS) encompasses the concepts of resolution and weighting function. Table 13 summarizes how the four RFS types are used in the AFED algorithm.

Table 13: Reference footprint shapes (RFS) for composite TBs (CTB)

Reference footprint shape	Resolution [km]	Weighting function	Sensors	Bands [GHz]
22D	22	Modified Circular Gaussian	AMSR-E, AMSR2, GMI	37
27C	27	Circular Gaussian	SSM/I	19, 24, 37, 89
36C	36	Circular Gaussian	TMI	19, 22, 37, 85
50D	50	Modified Circular Gaussian	AMSR-E, AMSR2, GMI	37
69C	69	Circular Gaussian	SSM/I	19, 22, 37, 85

3.4 Compute geolocation correction for microwave sensor data

The AFED algorithm relies on accurate geographic coordinates (i.e., latitude and longitude pairs) provided as auxiliary data in the input microwave sensor swath data files for each TB measurement. Geolocation data from all microwave sensors (including AMSR-E, AMSR2, GMI, and SSM/I) include errors that may be significant fractions of the resampled footprint size (e.g., 1 or 2 km out of 22-km composite footprints). Along coastlines and near large water bodies or flooding, geolocation error is a source of CTB variance because small dislocations perpendicular to land-water boundaries can make a large difference in the amount of water within CTB footprints. (Geolocation is a lesser error source away from water because dry-land TB varies more gradually in every direction.) High CTB variance increases MDFF and reduces the 3-day flooded fraction average, which reduces AFED flood extents and the likelihood that the algorithm will detect smaller floods.

The AFED geolocation correction algorithm replaces latitude and longitude coordinate data in CTB swath files with adjusted values. The algorithm preserves the original coordinate data in the file under different variable names. The algorithm uses the high correlation between the *Q*-transformed CTBs—defined in Section 3.8—and persistent water fraction (f_{pow}, e.g., Figure 18) to solve for the coordinate correction that minimizes a cost function with normalized *Q* and f_{pow} terms:

$$\sum_{i=1}^{N}\left[\frac{Q_i - Q_{land,i}}{Q_{water,i} - Q_{land,i}} - \frac{f_{pow}(\lambda'_i, \phi'_i) - f_{pow}(\lambda'_{land,i}, \phi'_{land,i})}{f_{pow}(\lambda'_{water,i}, \phi'_{water,i}) - f_{pow}(\lambda'_{land,i}, \phi'_{land,i})}\right]^2 \quad (4)$$

where Q_i is one of *N* selected *Q* optimization points in each CTB swath originally located at latitude λ_i and longitude ϕ_i; $Q_{land,i}$ and $Q_{water,i}$ are land and water normalization points near Q_i; $f_{pow}(\lambda'_i, \phi'_i)$ is precomputed, gridded f_{pow} data interpolated to the corrected Q_i coordinates, (λ'_i, ϕ'_i); and the other terms are f_{pow} likewise interpolated to the corrected coordinates of $Q_{land,i}$ and $Q_{water,i}$. The *Q* optimization points in each CTB swath are the set of points nearest predetermined control locations with high f_{pow} gradient. We tabulated high-gradient control locations along with coordinates for the nearest land and water normalization points per each CTB resolution (22 to 69 km). Figure 4 shows a subset of the 22-km resolution high-gradient, water, and land control points over part of Africa. The control location tables are provided to the algorithm as ancillary static input data.

After assembling and normalizing *N* Q_i data points, the algorithm executes iterative cost function minimization to solved for two parameters per swath that together define a coordinate

shift: the geographic arc length of the shift (in kilometers) and the earth azimuth angle of the shift relative to the satellite ground track direction earth azimuth angle. The satellite ground track azimuth varies along the satellite path. Defining the shift direction relative to the ground track azimuth allows the algorithm to correct for uncertainties in the sensor-to-earth geometry that vary slowly through the satellite orbit. The correction cannot correct for other types of geolocation error due to factors such as earth elevation or sensor sample timing uncertainties. If the swath has an insufficient number of *Q* optimization points ($N<50$), the algorithm does not attempt geolocation correction and instead flags coastal CTBs to be left out of AFED processing.

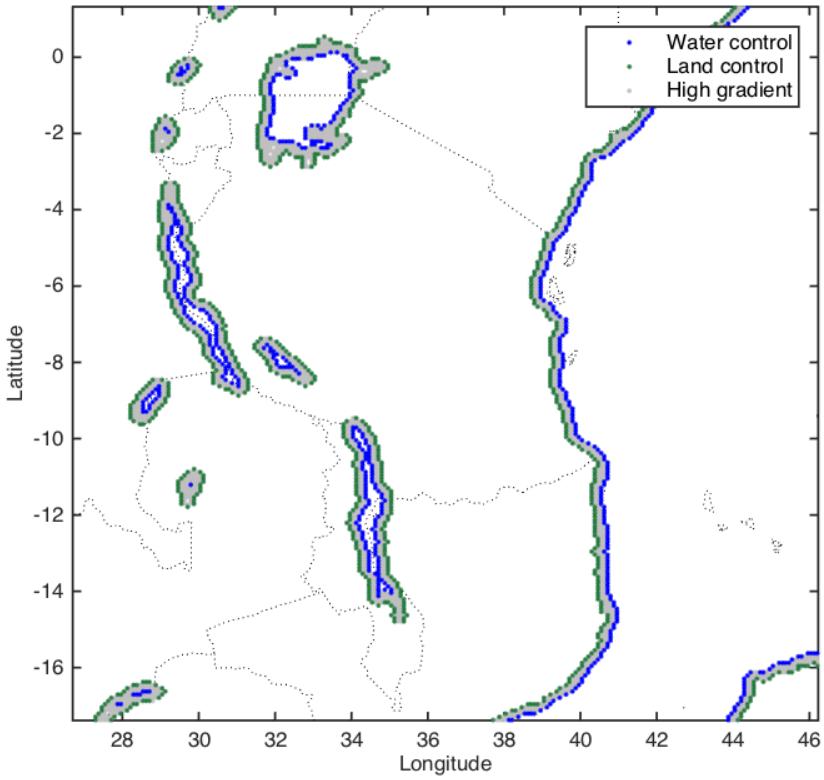

Figure 4: Examples of land, water, and high-gradient control points used in geolocation algorithm for 22-km resolution data.

Figure 5 shows the effect of geolocation correction on *Q* statistics compiled over January 2013 from ascending and descending pass AMSR2 data. V04R01 data include the geolocation correction and V03R00 data do not. The geolocation correction reduces monthly *Q* standard deviations (right hand map) in coastal zones and many inland areas. A few areas have higher standard deviations, which may be a result of geolocation bias correction from areas with lower natural variability to those with higher variability. The geolocation correction also changes the monthly *Q* mean—notably along coastlines, wetlands, and rivers—which is another indication that the algorithm corrects for systematic geolocation bias in addition to day-to-day random geolocation error.

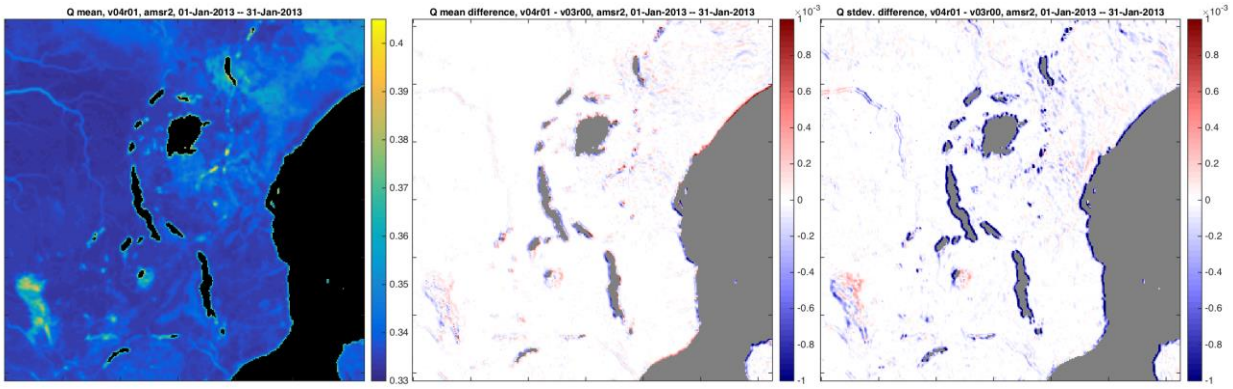

Figure 5: Change in *Q* statistics with (V04R01) and without (V03R00) geolocation correction, AMSR2 22-km data, 2013-01-01 to 2013-12-31. Left: V04R01 *Q* mean. Middle: *Q* mean difference, V04R01-V03R00. Right: *Q* standard deviation difference, V04R01-V03R00.

3.5 Regrid microwave sensor data

The algorithm interpolates CTBs from their swath locations to points on a 5-arcminute fixed earth grid using two-dimensional cubic spline interpolation (Press et al., 1986). Two-dimensional spline interpolation requires that the data to be interpolated are associated with two independent variables. The algorithm finds the location of an earth coordinate grid point in the sensor sampling coordinate system of a swath, which is defined by the following two dimensions: (1) the direction of progressive sensor scans and (2) the position of sensor samples along each scan. Here, we will refer to the sensor sampling coordinate system dimensions as *scan* and *position*. The method then applies the two-dimensional spline interpolation technique to the data in the scan and position coordinate system.

Figure 6 shows how the *scan-position* sensor sampling coordinate system geometry relates to latitude and longitude for a location in the middle of the swath. The "x" symbol denotes an example earth grid location to which CTB data are to be interpolated. The nearest swath data point to the grid point is denoted by its scan number and position number in the scan ("(scan,pos)"). In this case, the grid point is located between the nearest scan and the previous scan ("scan-1") and between the nearest scan position and the previous scan position ("pos-1"). Note that the *scan-position* coordinate axes are not at right angles to each other. This is more evident at the edge of the CTB swath, as shown in Figure 7. Here, the example grid point is also located between the nearest scan and the previous scan and between the nearest scan position and the previous scan position, but the curvature of the scans means that the angle between the scan and position coordinate axes is more oblique than it is at the middle of the swath.

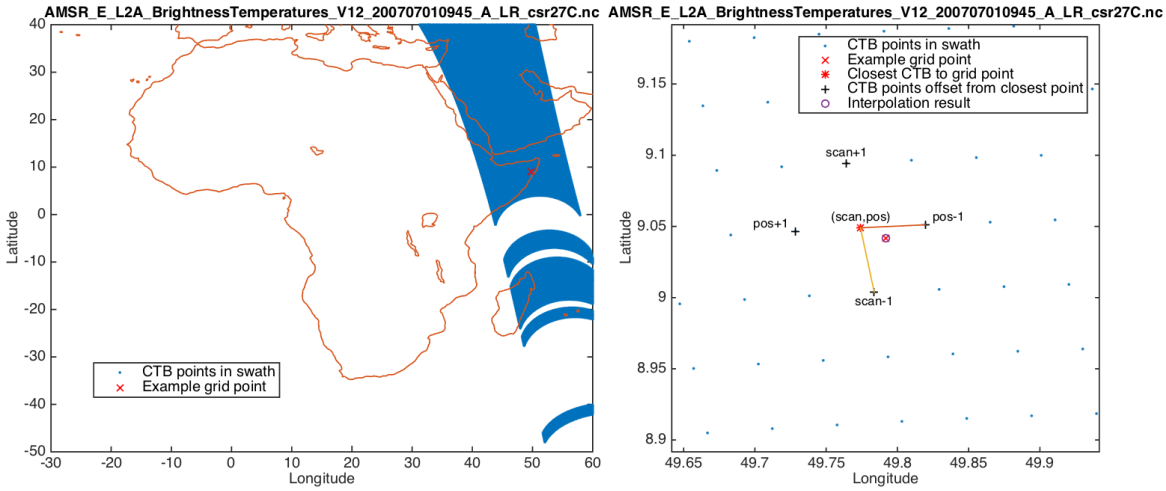

Figure 6: Geometry for spline interpolation of CTB data from swath points to grid points for an example grid point "x". Left: Locations of CTB points in a single swath and an example grid point. Right: Detailed view of the area around the example grid point. Progressive sensor scans pass through the points labeled "scan-1", "(scan,pos)", and "scan+1". Progressive CTB positions along "scan" progress through the points labeled "pos-1", "(scan,pos)", and "pos+1."

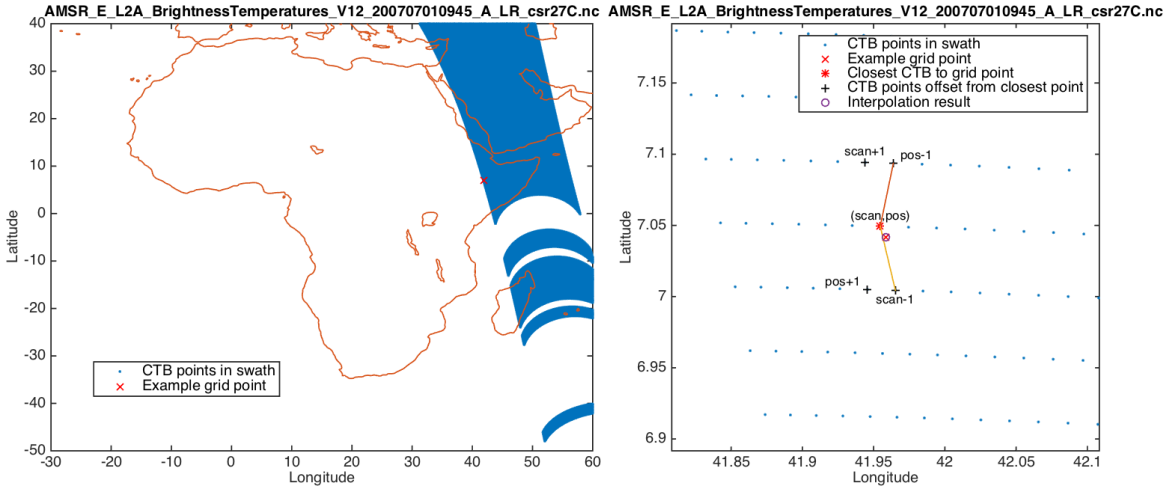

Figure 7: Like Figure 6 but for a grid point near the edge of the CTB data swath.

Figure 6 and Figure 7 also illustrate the results of the spline interpolation process applied not to CTB data but to the latitude and longitude coordinates of the CTB data. This test verifies that the methodology produces the expected results when the correct results are known exactly. In this case, we know the correct results—the geographic coordinates of the grid point, which are indicated by the "*" symbol in the figures—because we specified them in advance. We found that the interpolation results—indicated by "o" symbols—were well-aligned with the grid point coordinates ("x"), providing evidence that the method is mathematically sound and implemented correctly.

The microwave regridding algorithm may be expected to produce one to three gridded data files per day depending on the satellite orbit. For sun-synchronous satellites (those carrying AMSR-E, AMSR2, and SSM/I), two grid files per day are expected, one each for data from the satellite's ascending and descending pass direction swaths. For non-sun-synchronous satellites (those carrying TMI and GMI), there may be one to three grid files per day depending on the

timing of satellite passes over Africa. At grid points where multiple swaths with the same pass direction overlap, the algorithm retains data from the earliest swath of the day and discards the overlapping data from later swaths. The AFED PM executes the regridding algorithm separately for each RFS.

In V03R01 historical processing, the algorithm avoided circumstances where swaths crossing Africa might be divided between successive days by shifting the start of day during the regridding process. For example, for AMSR-E the descending pass equator crossing time is about 0130 local solar time, meaning that some descending swath data in a daily group covering Africa may fall on the prior UTC day. The algorithm shifted the start of day for AMSR-E data collections to 2100 UTC on the prior day so that the successive series of passes covering all of Africa are grouped together in a single gridded data file assigned to the day on which most of the data were observed. In the NRTPS, this functionality is replaced by a swath database and logic to group swaths that cover Africa successively with a temporal gap less than the duration of a full orbit. The database approach accommodates data from sensors on non-sun-synchronous satellites (like GMI) by assigning the grid file to the day corresponding to the predicted average time of swath data to be combined in the file. For data from sun-synchronous satellite, this approach is equivalent to that achieved in V03R01 and V04R00 historical processing. As of V04R01, the algorithm began using the swath database approach for all processing. Orbital modeling parameters are not available for parts of the SSM/I record so we built an SSM/I swath database by analysis SSM/I swath data source files.

Figure 8 and Figure 9 illustrate regridding algorithm results for 27C RFS data from the eight AMSR-E microwave sensor channels acquired on 2 June 2002 during descending satellite passes. (Note that descending pass swaths slant from the northeast to southwest because of west-to-east earth rotation as the satellite moves in a nearly due-south direction; similarly, ascending passes slant from southeast to northwest.) The swaths do not overlap, which is typical of AMSR-E at these latitudes and means that algorithm logic to handle overlaps is not activated for AMSR-E; nonetheless, the algorithm will continue to include the logic for potential use with other sensors.

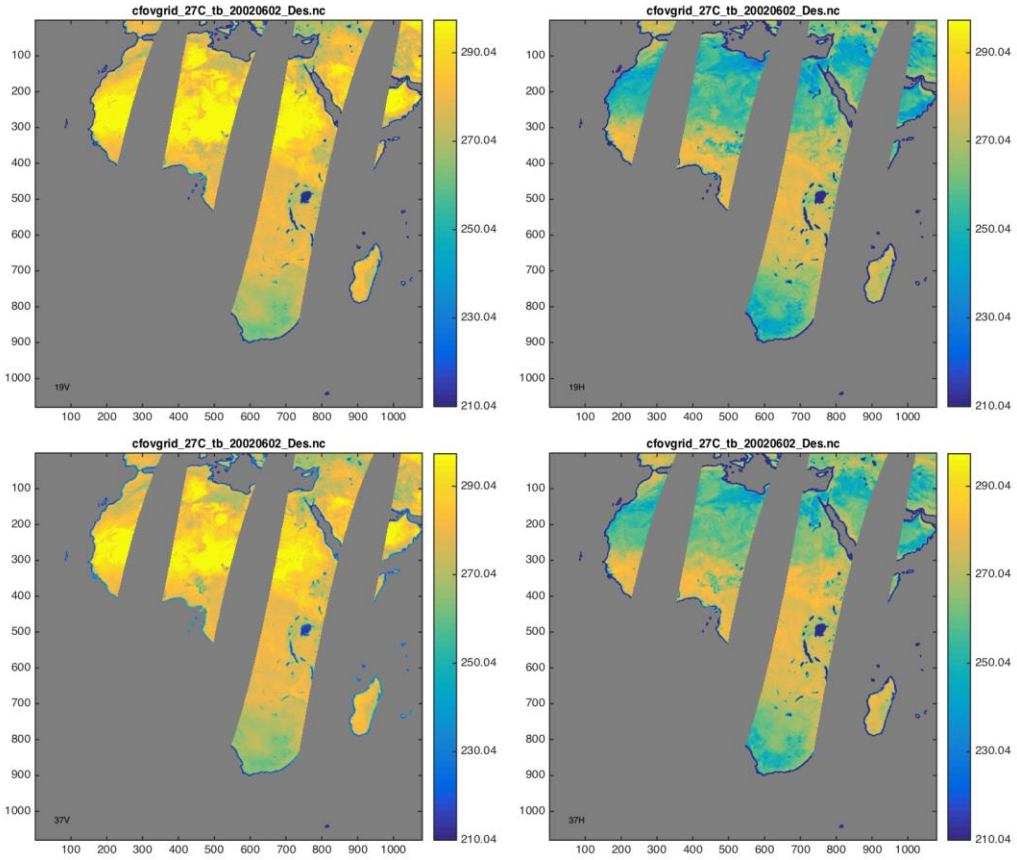

Figure 8: Gridded AMSR-E 27C RFS microwave brightness temperature data [K] on the AFED 5-arcminute grid for descending passes on 2 June 2002. From top left: 19 GHz vertical polarization (19V), 19 GHz horizontal polarization (19H), 37V, and 37H. Axes indicate grid cell row and column number.

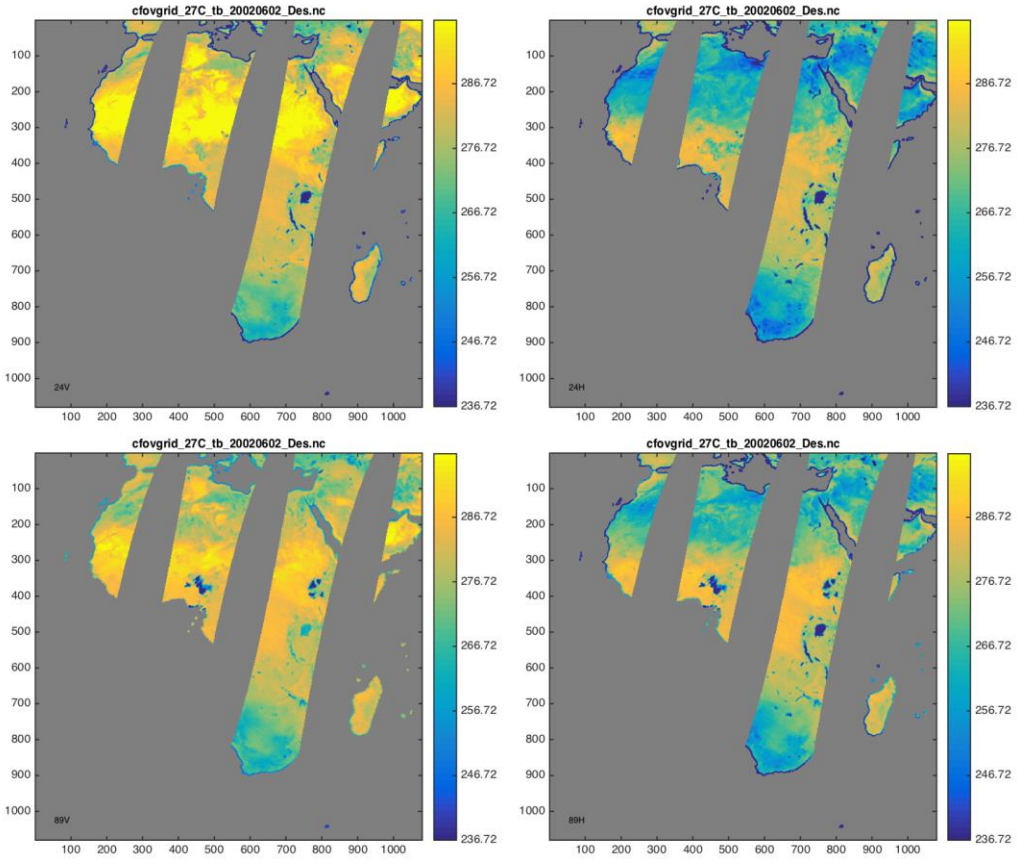

Figure 9: Like Figure 8 but for 24V, 24H, 89V, and 89H data.

3.6 Compute atmospheric correction for microwave sensor data

The AFED atmospheric correction algorithm uses a physics-based microwave radiative transfer model (Moncet et al., 2011). The model is formulated as if to estimate the brightness temperature (TB) measured by a satellite microwave sensor at a microwave frequency ν and polarization p in non-scattering atmospheres:

$$T_{\nu,p}^{B} = T_{\nu}^{\uparrow} + \tau_{\nu}\left[\varepsilon_{\nu,p}T_{e,\nu} + \left(1-\varepsilon_{\nu,p}\right)T_{\nu}^{\downarrow}\right] \quad (5)$$

$$= \left(T_{\nu}^{\uparrow} + \tau_{\nu}T_{\nu}^{\downarrow}\right) + \varepsilon_{\nu,p}\left(\tau_{\nu}T_{e,\nu} - \tau_{\nu}T_{\nu}^{\downarrow}\right) \quad (6)$$

$$T_{\nu,p}^{B} = A_{\nu} + \varepsilon_{\nu,p}\left(B_{\nu}T_{e,\nu} + C_{\nu}\right) \quad (7)$$

where τ_{ν} is the total atmospheric transmittance along the sensor line of sight to the surface, $T_{\nu}^{\uparrow}$ and $T_{\nu}^{\downarrow}$ represent the upwelling and downwelling atmospheric emission, respectively, $T_{e,\nu}$ is the effective emission temperature of the surface, and $\varepsilon_{\nu,p}$ is the surface emissivity. The attenuated cosmic background is included in the downwelling emission term. The *A*, *B* and *C* terms in (5) are abbreviations of the corresponding terms in (6). Since we are dealing exclusively with conically scanning imagers, the dependence of these variables on sensor viewing angle (which is approximately constant) is omitted from our notation.

The AFED atmospheric correction algorithm estimates the earth-leaving (EL) brightness temperature as:

$$T_{n,p}^{EL} = \frac{A_n - C_n}{B_n} \quad (8)$$

The earth-leaving TB algorithm corrects for two components of the atmospheric radiative transfer that reduce land-water contrast: the total atmospheric transmittance and the upwelling atmospheric emission. The algorithm does not correct for reflected downwelling atmospheric emission or for the effective emission temperature of the surface because to do so the algorithm would need a surface emissivity estimate. Surface emissivity cannot be accurately estimated when land surface conditions are unpredictable, as they are during flood events. The flooded fraction algorithm (section 3.8) compensates for the lack of surface temperature information by using the polarization ratio index, which approximately cancels the surface temperature effect because it appears in all terms in the ratio's numerator and denominator.

The AFED algorithm relies on global numerical weather prediction model (NWP) atmospheric analyses (*Kanamitsu*, 1989; *Kalnay et al*., 1990) to supply atmospheric data to estimate the A, B, and C terms in eq. 8. The algorithm uses atmospheric surface temperature and pressure and vertical profiles of temperature and water vapor from the datasets listed in Table 14. For historical processing, we acquired NCEP data from the University Corporation for Atmospheric Research (UCAR) Research Data Archive (RDA); the NRTPS acquires GFS data from the NOAA National Operational Model Archive and Distribution System (NOMADS). The algorithm interpolates the 6- or 12-hourly NWP products to the time and footprint center location of each microwave sensor observation.

Table 14: Atmospheric data sources

Temporal Range	Usage Period	Dataset Name, Number, and Link	Time Sampling	Geographic Grid
01/07/1976 – 31/03/1997	05/12/1991 – 31/03/1997	NCEP FNL Operational Model Global Tropospheric Analyses, July 1976 to April 1997[1], ds082.0, http://rda.ucar.edu/datasets/ds082.0/	2 times daily: 0000, 0012 UTC	2.5° x 2.5° global
04/01/1997 – 30/06/2007	04/01/1997 – 29/07/1999	NCEP FNL Operational Model Global Tropospheric Analyses, April 1997 through June 2007[2], ds083.0, http://rda.ucar.edu/datasets/ds083.0	2 times daily: 0000, 0012 UTC	2.5° x 2.5° global
30/07/1999 – present	30/07/1999 – 31/12/2015	NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999[3], ds083.2, http://rda.ucar.edu/datasets/ds083.2/	4 times daily: 0000, 0006, 0012, 0018 UTC	1° x 1° global
02/03/2004 – present	All NRTPS processing	GFS-ANL or GFS, 003 (1º) – Domain, https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs	4 times daily: 0000, 0006, 0012, 0018 UTC	1° x 1° global

[1]National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 1980.
[2]National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 1997.
[3]National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 2000, updated daily.

There were occasional gaps in the atmospheric data records during which the historical processing system could not compute earth-leaving TBs. The dates on which both the ascending and descending passes were affected are listed in Table 15; all affected data are noted in the quality control *reason* metadata files (Table 9). The algorithm did not update AFED during these periods, which means that AFED remains constant and equal to the last valid AFED value prior to the period. It may be preferable for ARC to use AFED results derived from the non-atmospherically corrected TBs (i.e., top-of-atmosphere TBs) during some of these periods. An AFED version designated v03r01_toa has been produced from top-of-atmosphere SSM/I TBs for the periods 1992 to 2002 and October 2011 through August 2012. The NRTPS is protected against future gaps in atmospheric data flow from GFS by the temporary storage of GFS forecast data as a contingency to be used in the event that GFS analysis data are not available when needed.

Table 15: Dates without atmospheric data

Date Range	Sensor
05-Mar-1992 27-Mar-1993 01-May-1993 08-Jul-1993 11-Aug-1995 02-Apr-1996 18-Mar-1997 – 22-Mar-1997 04-Apr-1997 21-Apr-1997 – 27-Apr-1997 29-Apr-1997 02-Oct-1997	SSM/I
29-Feb-2012	SSM/I

The radiative transfer model used for computing A, B, and C is described by *Moncet et al.* (2011). The AFED PM executes the model once per each satellite ascending and descending pass per day. The atmospheric correction algorithm uses the same A, B, and C values for the fine- (22D and 50D) and coarse- (27C, 36C and 69C) resolution CTB data.

3.7 Compute precipitation and frozen surface conditions flags

The AFED algorithm precipitation and frozen surface conditions detection algorithm uses top-of-atmosphere coarse-res (27C, 36C, or 69C) RFS CTB data at 19, 24 (or 22), 37, and 89 (or 85) GHz. The flooded fraction algorithm uses these flags to indicate microwave data points that should not be used for flood mapping because the presence of rain or frozen surface conditions may lead to errors in flooded fraction retrieval. We based the algorithm on published models and made several modifications to make the algorithm more sensitive to rain to reduce the incidence of AFED algorithm errors.

The algorithm produces flags for precipitation and frozen surface conditions (snow or frozen ground) per each sensor footprint in a two-step process. In the first step, the algorithm uses a modified version of the decision-tree detection method of Ferraro et al. (1996, hereafter *Ferraro1996*) to flag snow and frozen ground and make primary flags for precipitation. In the second step, the algorithm uses image processing techniques to dilate the primary precipitation flags to areas within a 7- or 9-grid cell radius; the algorithm checks areas covered by the dilation a second time with a more sensitive precipitation test than that used for the primary flag. The

decision tree and image processing steps treat land and water areas separately because the decision tree can provide a stricter and more accurate test for rain over water, which has more predicable conditions. However, detection tests that assume a water background frequently produce false positives when there is a significant land fraction in the footprint. As a result, as of AFED V04R00 the algorithm uses land decisions wherever footprints include less than 95% water.

Step 1: Primary flags

For flagging frozen surface conditions (snow or frozen ground), the algorithm applies the decision tree given in *Ferraro1996* Fig. 1. This decision tree also tests for rain but we instead used the more rigorous rain detection method discussed below. In some cases, rain and snow flags may occur simultaneously but the AFED algorithm uses the flags as independent indicators of data quality so there is no logical conflict.

For rain, we modified the *Ferraro1996* method in several ways:

1. Scattering conditions test: The *Ferraro1996* decision tree for rain detection (*Ferraro1996*, Fig. 9) begins with a test to identify scattering conditions: SI>10, where SI is defined in *Ferraro1996* eq. 6. We conducted tests that indicated that a constant threshold (of 10) was leading to rain detection false positives in arid regions of Africa. We modified the algorithm to replace the constant threshold with a local climatology, $SI_0(d_c,x)$, where d_c, is climatological day-of-year and x is grid cell location. We computed SI_0 for TMI, AMSR-E, AMSR2, and GMI using 2003-2011 AMSR-E data and for SSM/I using 1994-2001 SSM/I data. First, for each day-of-year, we computed the 15th percentile over all eight years of the 61-day median SI value, SI_{61}, for ascending (day) and descending (night) pass data. We used the 15th percentile value to reduce as much as possible the effect of rain in the SI_0 calculation. Second, we computed descending-ascending SI_{61} difference, ΔSI_{61}, which is an indicator of arid conditions. Then we computed $SI_0 = 10 + 2max(\Delta SI_{61},0) + max(SI_{61}-7,-1)$. Relative to the original algorithm (threshold of 10), the first term will reduce rain false positives in arid areas (where $\Delta SI_{61}>0$) and the second term will reduce false positives where $SI_{61}>7$ and increase true positives where $6<SI_{61}<7$.
2. Snow flag in precipitation test: The *Ferraro1996* decision tree for rain detection (*Ferraro1996*, Fig. 9) includes an *Identify Snow* test; when there is a positive snow test the precipitation test is always negative. We have concluded that the snow test incorrectly uses an "or" instead of an "and" in the statement "$TB_{22V}<265$ or $TB_{22V}<175+0.49TB_{85V}$", where TB_{22V} and TB_{85V} indicate brightness temperature measurements made in the 22 and 85 GHz bands at vertical polarization. Our conclusion that an "and" is correct in this context – to indicate a *positive* snow test – is based on empirical results and is supported by *Ferraro1996* Fig. 1, where an "or" is used to indicate a *negative* snow decision. We also changed the first term in this test to "$TB_{22V}<264$"; the effect of this change is to slightly increase rain flagging in frozen surface conditions.
3. Desert flag in precipitation test: The *Ferraro1996* decision tree for rain detection (*Ferraro1996*, Fig. 9) includes an *Identify Desert* test, $TB_{19V}- TB_{19H}>20$; when there is a positive desert test the precipitation test is always negative. We change the threshold in this test to 22 from 20 to slightly increase rain flagging in some conditions (e.g., over flooded land).

4. Arid flag in precipitation test: The *Ferraro1996* decision tree for rain detection (*Ferraro1996*, Fig. 9) includes an *Identify Arid Soil* test, $TB_{85V}>253$ and $TB_{19V}-TB_{19H}>7$; when there is a positive arid soil test the precipitation test is always negative. We changed this test to a two-part test. In the first part, there is a positive test for arid soil where "$SI < SI_0+5$ and $TB_{85V}>253$ and $TB_{19V}- TB_{19H}>7$". This change helps increase rain flagging in some conditions by restricting arid soil flags to conditions with lower SI values. In the second part, there is a positive test for arid soil where "$SI \geq SI_0+5$ and $TB_{85V}>255$ and $TB_{19V}- TB_{19H}>15$". This restores arid flagging for some cases with higher SI values and extreme surface conditions.

Figure 10 and Figure 11 show the results of the changes in the primary flags algorithm for year 2003. The AFED and *Ferraro1996* methods produce similar patterns in the number of rain flags per year (Figure 10), although the AFED algorithm has notably higher flagging rates in several areas including along the Congo River, the Barotse Plain wetlands straddling Angola and Zambia, and in Southern Africa. In each of these areas, the additional flags appear to be consistent with the surrounding areas (e.g., the Congo and Barotse Plain in Figure 10) and correspond to the expected rainfall seasonality (not shown). Figure 11 shows that the net effect of the AFED algorithm is more often more rain flagging than less and the number of additional flags per year is rarely more than 10.

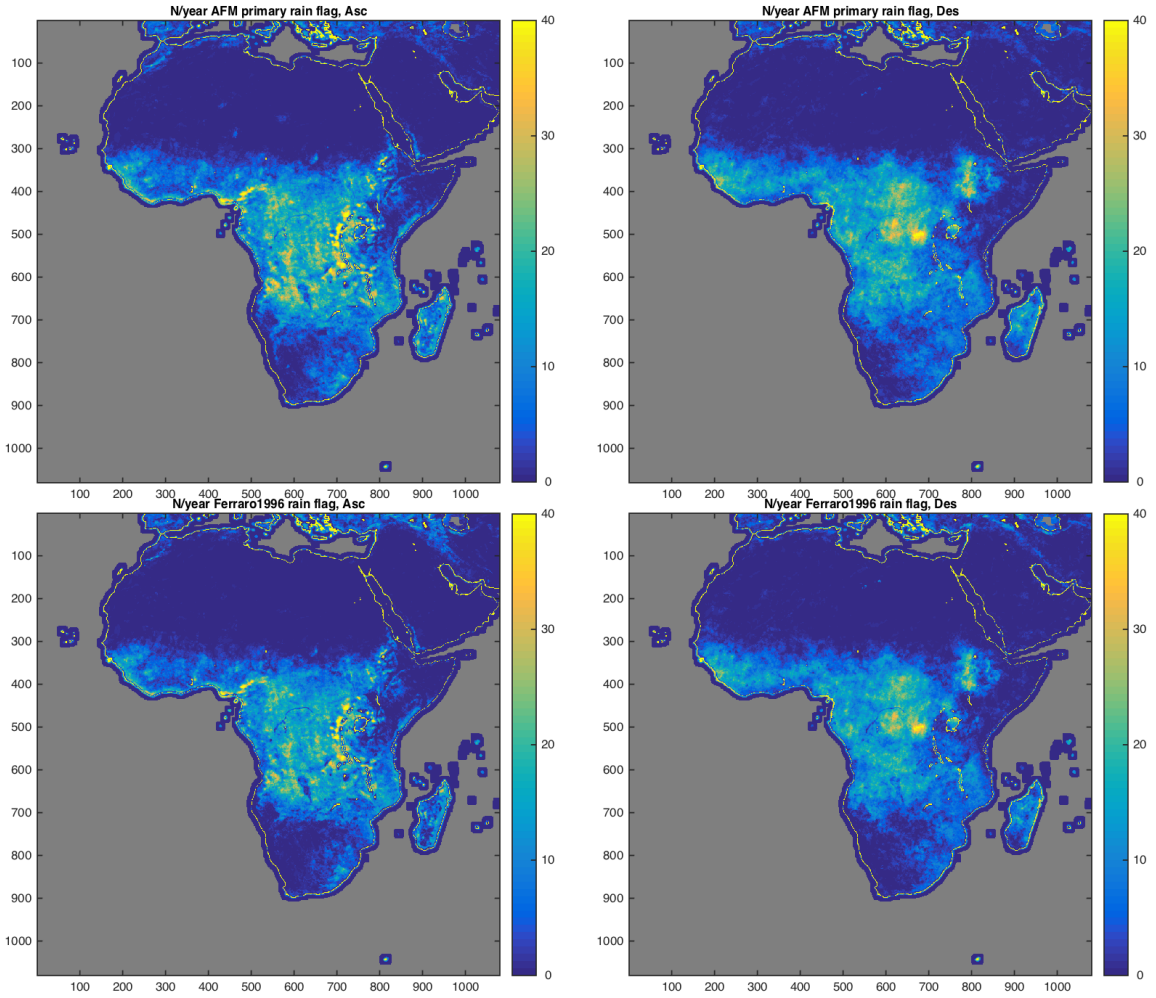

Figure 10: Number of primary rain flags per year 2003 for AMSR-E. Upper left: AFED algorithm, ascending pass. Upper right: AFED algorithm, descending. Lower left: *Ferraro1996*, ascending. Lower right: *Ferraro1996*, descending.

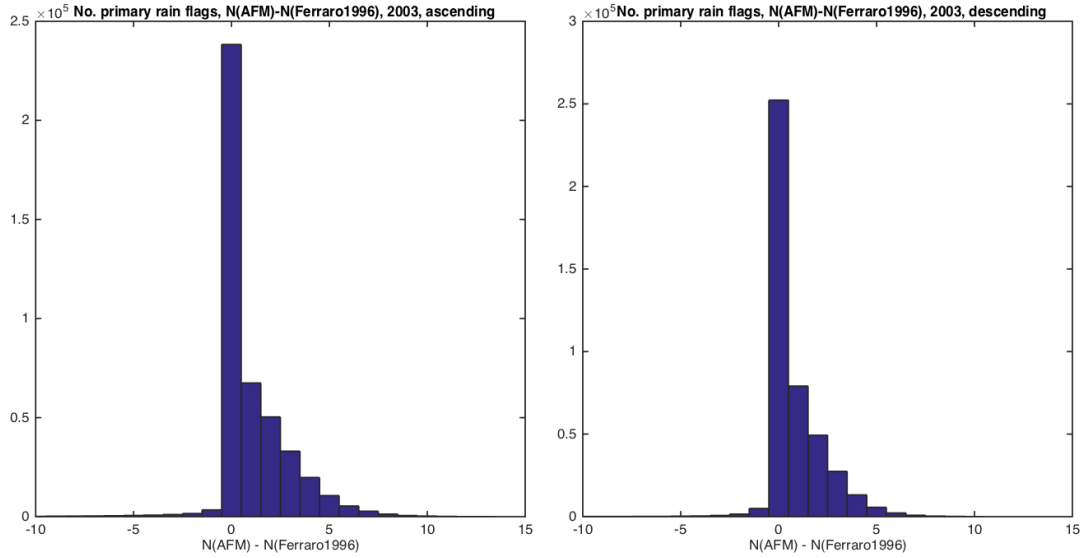

Figure 11: Difference in number of primary flags per year 2003, AFED algorithm vs. *Ferraro1996* model, excluding areas where persistent water cover fraction is greater than 0.2. Left: Ascending pass data. Right: Descending pass.

Step 2: Flag dilation

The secondary precipitation test expands the primary flag to surrounding areas:

1. Dilation: The algorithm creates an image in which the primary flags from step 1 are dilated, or expanded, by a radius of 7-grid cells for 27-km or 36-km data and 9-grid cells for 69-km data to the surrounding areas. Flags from land areas (less than 95% water) are only dilated to other land areas and flags from water areas area only dilated to other water areas.
2. Secondary precipitation test: The algorithm retests the areas with dilated primary flags for scattering conditions with a slightly lower threshold: $SI > SI_0(d_c,x)$-1. Areas passing this test are also flagged as rain. Desert and arid tests are not reapplied based on the assumption that proximity to rain-detected areas provides a stronger indication that rain is the source of the scattering signature.

Figure 12 illustrates the primary and final AMSR-E flags on 7 September 2003. Figure 13 and Figure 14 show the overall results of the flag dilation step for year 2003. Dilation increases the incidence of rain flagging everywhere that flags occur. Areas along the Congo River and in the Barotse Plain that are inconsistent with their surroundings for the primary flag are notably more consistent for the final flag. Some areas with additional flags are likely to be false-positives but our assessment is that the benefit—reduction of rain-affected data entering the AFED algorithm—offsets the moderate loss of potentially valid data.

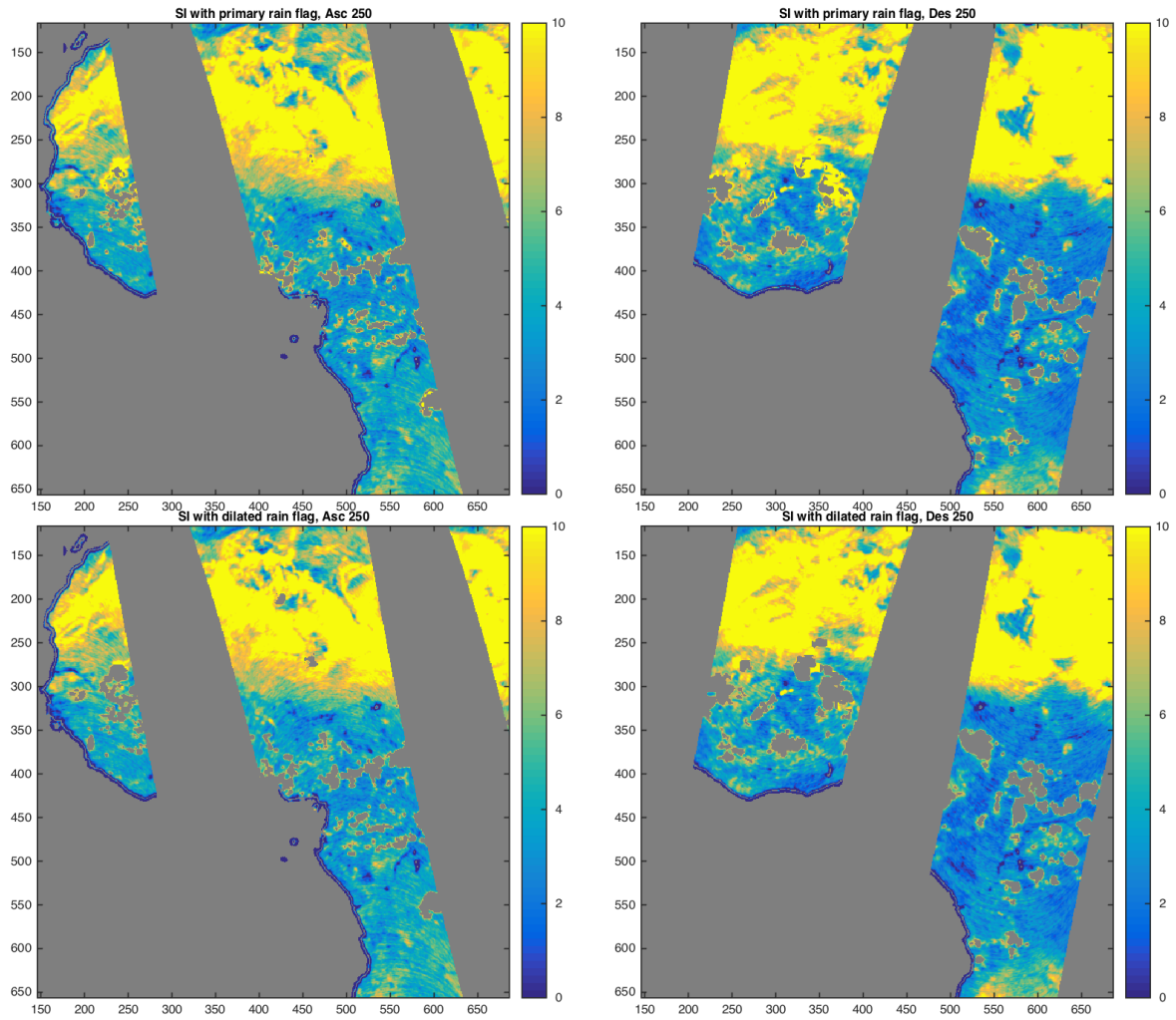

Figure 12: Example of preliminary precipitation flag (top row) and final flag after dilation and retesting (bottom) for AMSR-E ascending (left) and descending (right) passes on 7 September 2003. Flagged areas are gray. (Missing data and ocean areas are also gray; the Gulf of Guinea is in the lower left.) The background image is the scattering index, SI. Axes indicate grid cell row and column number.

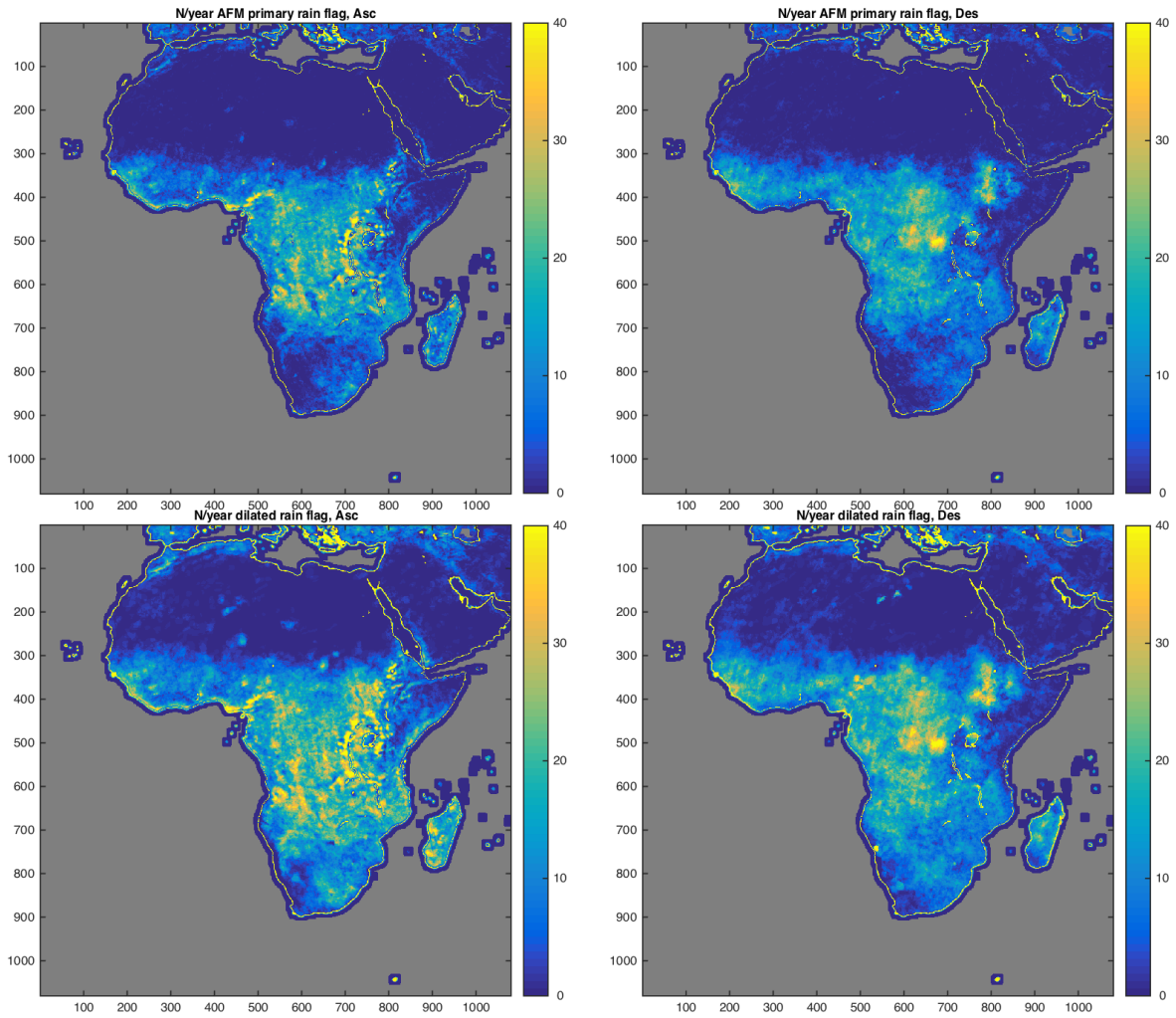

Figure 13: Number of AFED algorithm AMSR-E primary (top) and final (bottom) rain flags per year 2003. Left: ascending passes. Right: Descending.

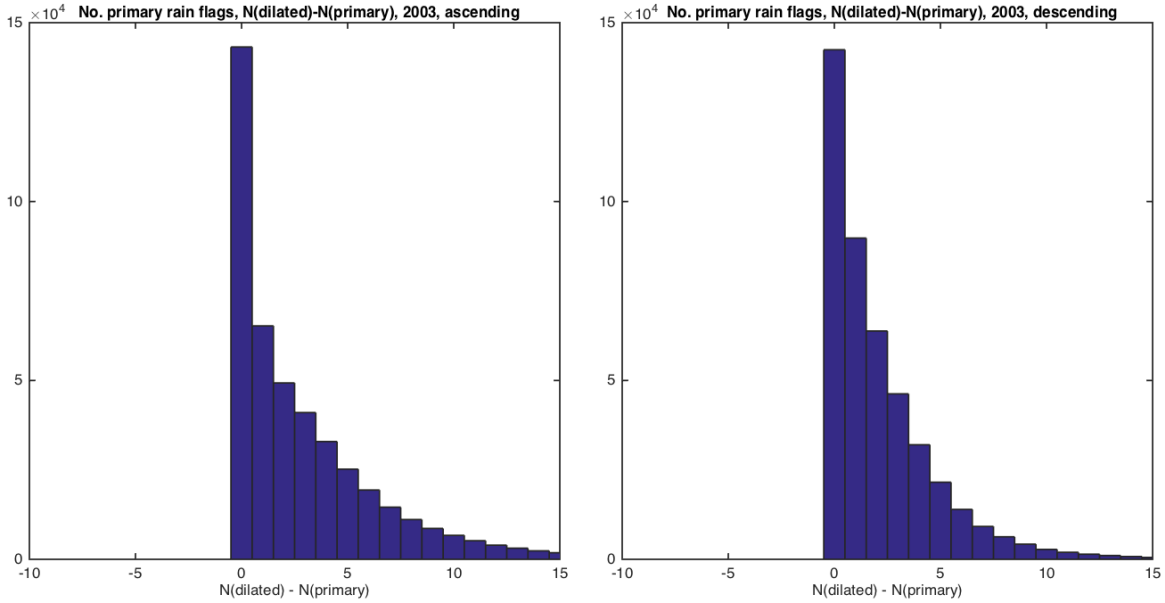

Figure 14: Difference in number of final (dilated and retest) and primary flags per year 2003, excluding areas where persistent water cover fraction is greater than 0.2. Left: Ascending pass data. Right: Descending pass.

3.8 Compute flooded fraction

Methods for the retrieval of flood area rely on the high sensitivity of microwave brightness temperature to water cover even in the presence of vegetation (e.g., Sippel et al., 1994; Galantowicz, 2001; Galantowicz, 2002; Prigent et al. 2001a, 2007; Papa et al., 2010; Schroeder et al., 2010). The AFED algorithm takes advantage of the spatial stability in the brightness temperature (TB) time series data provided by our footprint matching and earth-gridding processes, which help isolate temporal changes in emission even in spatially heterogeneous areas. The algorithm is formulated in terms of polarization ratio index (PRI), which minimizes thermal effects relative to the strong polarization signal caused by surface water. PRI is calculated as:

$$PRI_f = \frac{\left(T_{Bfv} - T_{Bfh}\right)}{\left(T_{Bfv} + T_{Bfh}\right)} \tag{9}$$

where T_{Bfv} and T_{Bfh} are vertical, *v*, and horizontal, *h*, polarization data at microwave frequency *f* (e.g., 19 or 37 GHz).

Empirical and theoretical results (see section 4.7) show that the relationship between PRI and water fraction can be approximately but accurately linearized using the following transform:

$$Q_f = sqrt(PRI_f + a) + bPRI_f \tag{10}$$

where a=0.11 and b=-0.58.

The algorithm uses a multi-end-member approach that solves the following linear mixing model equation at each frequency for flooded fraction, f_{flood}:

$$Q_{obs} = Q_{pow}f_{pow} + Q_{dry}f_{dry} + Q_{flood}f_{flood} \tag{11}$$

where Q_{obs} is observed data calculated from eqs. 9 and 10, f_{pow} is fraction of persistent open water, Q_{pow} is the empirically derived expected value for Q of persistent open water, f_{dry} is the fraction of dry land areas (neither open water nor flooded), Q_{dry} is the empirically derived expected value of Q for dry land, and Q_{flood} is the expected effective Q of flood water. f_{dry} is unknown but can be eliminated by applying:

$$f_{dry} = 1 - f_{pow} - f_{flood}. \tag{12}$$

Solving (11) for flooded fraction, f_{flood}:

$$f_{flood} = \frac{Q_{obs} - Q_{pow} f_{pow} - Q_{dry}\left(1 - f_{pow}\right)}{Q_{flood} - Q_{dry}} \tag{13}$$

The terms in eq. 13 are described below and in the following sections:

- f_{pow}: Calculation of persistent open water cover fraction (f_{pow}) is described in section 4.3. This term is used for excluding known open water from the f_{flood} retrieval. Total water cover fraction, f_w, may be expressed as $f_w = f_{flood} + f_{pow}$.
- Q_{pow}: We empirically derived Q_{pow} for Africa using f_{pow} data and historical average Q values observed in the Lake Victoria region (section 4.7). Q_{pow} has no seasonal dependence but ascending and descending microwave sensor passes are treated separately for data from sun-synchronous satellites to account for observed systematic differences.
- Q_{flood}: Q_{flood} is defined to be equivalent to Q_{pow} in the current algorithm but could be adjusted in future algorithms to account for the effect of vegetation cover in the flood plain. Vegetation can effectively reduce Q_{flood} relative to Q_{pow} due to the attenuation by canopies over floodwaters but it is difficult to estimate the magnitude of this effect because of the heterogeneity of vegetation cover, density, and height relative to flood depth.
- Q_{dry}: Computation of daily Q_{dry} values is described below.

3.8.1 *Initial flooded fraction*

The flooded fraction (FF) algorithm initially computes FF per each microwave band (19 and 37 GHz) and gridded data file independently; there are two gridded data files per day for sun-synchronous satellites and one to three for non-sun-synchronous, although each file does not have valid data at every grid point. This initial algorithm step includes three subcomponents: calculation of Q_{obs}; calculation of Q_{61E}, the expected value of the current 61-day time-centered median Q, by three methods; calculation of Q_{dry}; and calculation of initial FF values:

- The algorithm computes Q_{obs} from earth-leaving (EL) TBs using eqs. 9 and 10.
- The algorithm computes two point-wise expected 61-day Q medians, Q_{61E} and Q_{61Epw}, from two related quantities: (1) Q_{31R}, the current 31-day retrospective median value of Q spanning the period d-30 to d (the current day) and (2) Q_{61R}, the current 61-day retrospective median value of Q spanning d-60 to d. First, the algorithm adjusts Q_{31R} to approximate the time-centered median, Q_{31E}, (i.e., spanning d-15 to d+15), using the Q 31-day median time shift model (section 4.15):

  $$Q_{31E}(Q_{31R}(d)) = Q_{31R}(d) + b_{31}(d) \tag{14}$$

  Second, the algorithm adjusts Q_{61R} to estimate Q_{61E} using the Q 61-day median time shift model (section 4.14)

  $$Q_{61E}(Q_{61R}(d)) = a_{61}(d) Q_{61R}(d) + b_{61}(d). \tag{15}$$

The algorithm then computes Q_{61Epw} as the average of these two quantities:

$$Q_{61Epw}(d) = [Q_{31E}(Q_{31R}(d)) + Q_{61E}(Q_{61R}(d))]/2. \quad (16)$$

As of V04R01, the algorithm also computes the spatial expected 61-day Q median, Q_{61Esp}. The spatial algorithm uses data from an 11x11 grid point area centered on the target grid point but excluding the grid point's immediate neighbors. The median deviation, $dQ_{61} = Q_{61E,pw} - Q_{61C}$, is computed over all points for which Q_{61Epw} is greater than Q_{61C}, which prevents Q_{61Esp} from falling below Q_{61C} and triggering false positives. The median deviation, dQ_{61}, is an estimate of the grid point's expected Q_{61} departure from climatology. Hence, the spatial analysis yields $Q_{61Esp} = Q_{61C} + dQ_{61}$.

- The algorithm computes three daily Q_{dry} values per grid point—$Q_{dry,61}(d,x)$, $Q_{dry,pw}(d,x)$, $Q_{dry,sp}(d,x)$—using the $Q_{dc}(d_c,x)$ climatology and $Q_{61E}(d,x)$, $Q_{61Epw}(d,x)$ and $Q_{61Esp}(d,x)$, respectively. For grid points and days flagged as regular seasonal flooding (section 4.17), Q_{dry} is set to $Q_{dc}(d_c,x)$. For other areas, the algorithm computes Q_{dry} as the weighted average:

$$Q_{dry} = \frac{W_{Qdc}Q_{dc} + W_{Q61E}Q_{61E}}{W_{Qdc} + W_{Q61E}} \quad (17)$$

where W_{Qdc} is $1/\sigma_{Q61c}^2$ (section 4.13) and W_{Q61E} is $1/0.001^2$. The value for W_{Q61E} ($1/0.001^2$) was chosen to balance false positive reduction (which benefits from higher W_{Q61E} values) with flood detection (which benefits from lower W_{Q61E} values).

- The algorithm uses eq. 13 to compute three initial FF values— $f_{f,61}$, $f_{f,pw}$, and $f_{f,sp}$— based on $Q_{dry,61}$, $Q_{dry,pw}$ and $Q_{dry,sp}$, respectively.

All components in the initial FF algorithm use atmospherically corrected earth-leaving (EL) TBs at a common sensor-specific resolution. The algorithm skips flooded fraction computation at grid points with precipitation or frozen surface flags, failing data quality flags, or missing data. The algorithm computes FF regardless of Q value; the final FF algorithm includes steps to handle non-physical FF results.

3.8.2 *Flooded fraction false positive detection*

The FF false positive algorithm flags short duration, large-scale FF events (hereafter, "flashes") attributable to soil moisture or other meteorological phenomena distinguishable from typical flood temporal-spatial signatures. The algorithm uses only point-wise FF, $f_{f,pw}$, from the initial flooded fraction step. To distinguish between flashes and flooding, the algorithm searches for large areas of flooding not present in the most recent prior FF map. As of V05R00, the algorithm processes data from two sensors simultaneously when possible. The earliest observation time in the grid files determines the processing sequence. The algorithm retains the two prior FF maps and may use the older of the two when a grid file includes observation times that precede the corresponding times in the more recent prior FF map. The following logic flags initial FF pixels as false positives:

1. Initial detection
    i. Create a map of newly flooded pixels, where newly flooded is defined as a pixel with (a) current 19-GHz derived flooded fraction above 0.05 and (b) most recent previous 37-GHz flooded fraction below 0.05. The algorithm uses 19-GHz data for test (a) because 19-GHz is more sensitive to false positives and 37-GHz data

for test (b) because 37-GHz better represents true flood conditions. Both tests use courser-resolution coarse-res data (27, 36-, or 69-km).

ii. Pass a circular filter with a 10-pixel diameter over the map of newly flooded pixels and frozen surface condition flags. For each pixel, this returns the fraction of the 10-pixel diameter area around it that is newly flooded or frozen-flagged.

iii. Mark pixels whose newly flooded or frozen-flagged area fraction is greater than 75% as *possible* flashes.

2. Screen initial detections

i. Throw out detections for pixels that have not had a valid flooded fraction measurement (prior FF) within Δt_{flash} days, where

$$\Delta t_{flash} = 1.5 + 4.75\, f_{hop}$$

and f_{hop} is the flooded fraction historical outlier prevalence factor (section 4.19). Large jumps in flooded fraction over longer time periods may be consistent with true flooding.

ii. Throw out detections for pixels flagged as seasonal wetlands or flood plains (section 4.17) where flooded fraction jumps are likely to be true flooding.

iii. Throw out detections for pixels that have a current flooded fraction below a seasonally-varying noise threshold similar to that used in MDFF calculations (section 4.19). Relatively small changes in flooded fraction may indicate true flooding.

iv. Throw out detections for isolated pixels, i.e., lacking neighboring pixels also marked as detections.

3. Dilate

i. Expand remaining detections to include surrounding contiguous areas with (a) current flooded fraction above 0.025 or (b) current rain flagging if the last flash test result with non-frozen-flagged data was also either flash-flagged or rain-flagged. Part (b) of this test allows the dilation to spread from the initial detection area across rain-flagged areas to other areas likely to be a part of the same flash.

3.8.3 *Final flooded fraction*

The FF finishing step operates on initial FF results for each day and the two preceding days. The process uses only the finest resolution FF results, which come from 37-GHz band data footprint-matched to either 22D (AMSRX sensor type) or 50D (SSM/I) RFS (described in section 3.3). The finishing step includes five subcomponents: weighted averaging of all non-flagged FF results from the 3-day period; selection or merging of initial FF types; filling-in missing data; quality control; and application of the minimum detectable flooded fraction limit. As of V05R00, the algorithm combines data from two sensors when possible.

- The FF averaging algorithm computes the weighted mean of all unflagged 37-GHz band initial FF values from all satellite passes on the current day and the two preceding days. Over the three days there are up to six FF data points from sun-synchronous sensors, up to seven from TMI or GMI, and up to 13 total when data from two sensors can be combined. The weights depend on the FF historical outlier prevalence factor, f_{hop} (section 4.19) as follows:
  - Current day weight: $0.67 – 0.33\, f_{hop}$
  - 1-day prior weight: 0.33
  - 2-days prior weight: $0.33\, f_{hop}$

The algorithm tests for outliers in the group and reduces their weights by half. If there is no data available on the prior day, the last valid instantaneous FF value from earlier days is included in the average with weight reduced by half. If there is only one valid FF value in the 3-day period, it is used only if it falls on the current day. This step is repeated for all initial FF types.

- The final FF algorithm uses a 15-day record of prior final FF results and the current ensemble of initial FF types to select among and merge the initial FF estimates. For each grid point, the algorithm records whether or not the final FF exceeds MDFF each day (described in final step below). Transition from $f_{f,pw}$ to $f_{f,sp}$ is triggered at points where prior FF values exceed MDFF on at least 12 consecutive days out of the last 15. A contiguous dilation region is defined around each trigger point where the current $f_{f,sp}$ and $f_{f,61}$ both exceed MDFF. Within each region, the number of days the final prior FF exceeds MDFF, N_{FF}, at the region's trigger point is used to compute the merged FF:

$$f_{f,final} = \max\{[w_{sp}\, f_{f,sp} + (1 - w_{sp})\, f_{f,pw}],\, f_{f,pw}\} \quad (18)$$

  where $w_{sp} = (N_{ff} - 11)/3$ within 0-1 bounds. The maximum value function in eq. 18 guarantees that $f_{f,final}$ is never less than $f_{f,pw}$.
- The fill-in algorithm maintains copies of the latest two valid FF 3-day average values, f_{-1} and f_{-2}. When there are no valid FF inputs for a grid point on a given day, the fill-in algorithm uses these stored FF values and the effective time since f_{-1} to estimate current FF, f_0:

$$f_0 = f_{-1} - a \max(T_0 - T_{-1}, 0)/(f_{-1} - f_{floor})$$

  where the reduction factor, a, is 0.05 if $f_{-1} < (f_{-2}+0.015)$ and 0.2 otherwise and f_{floor} is f_{-2} if $f_{-1} > (f_{-2}+0.015)$ and zero otherwise. With this formulation, the larger the difference is between f_{-1} and f_{floor} the faster f_0 decreases during a data gap.
- The quality control algorithm enforces physical limits on FF: where FF is less than zero, the algorithm sets FF equal to zero; where FF is greater than the total land fraction, $(1 - f_{pow})$, the algorithm sets FF equal to $(1 - f_{pow})$.
- The minimum detectable flooded fraction limiting algorithm enforces a lower limit on FF to account for near-zero FF values due to noise and other factors unaccounted for in the algorithm: where FF is less than the minimum detectable flooded fraction (MDFF), the algorithm sets FF equal to zero. As of V04R01, an exception to this rule is made in coastal areas—specified in a static ancillary dataset—where MDFF is usually relatively high due to irregular day-to-day footprint sampling of the water body. If FF is below MDFF in a coastal area grid point that neighbors a point where FF exceeds MDFF, then instead of setting FF equal to zero at the point, the algorithm estimates final FF as the minimum of either (a) the FF already calculated for the point or (b) FF calculated by interpolation between zero and the point's greater-than-MDFF neighbor with the lowest FF value. This logic allows detected flooding in low-noise areas to override the usual false positive protections of MDFF in adjacent coastal areas without risking an extreme false positive FF value.

3.9 Perform downscaling to depict flood extent

The downscaling algorithm uses the flooded fraction threshold (f_{f0}) database (section 4.21) to derive 3-arcsecond flood maps from microwave flooded fraction algorithm estimates, f_{flood}. The

algorithm interpolates f_{flood} to each AFED 3-arcsecond grid point, (i,j), using two-dimensional cubic spline interpolation; the algorithm initially indicates flooding at grid points meeting the following inequality:

$$f_{flood}(i,j) \geq f_{f0}(i,j). \quad (19)$$

After the initial downscaling step, the algorithm computes the resulting flooded fraction by *upscaling* the results, i.e., computing flooded fraction from the footprint-weighted average of the 3-arcsecond flood depiction. The difference between the microwave-derived and upscaled FF is computed as the *residual* flooded fraction—a value greater than or less than zero representing the amount that the flood depiction under or over represents the flooded fraction map inputs. The algorithm then repeats the downscaling/upscaling steps multiple times, each time decreasing the residual magnitude and improving the flood map. Figure 15 illustrates the process. At least three and up to six total downscaling-upscaling iterations are performed followed by a convergence test on the residuals; a final downscaling step is performed after the maximum residual magnitude in a tile falls below a threshold level (Figure 16). To speed convergence and reduce the possibility of oscillatory behavior, the computation of FF per each iteration applies a damping term where successive prior residuals switch signs. The resulting binary flooded/not-flooded map is the primary AFED algorithm product.

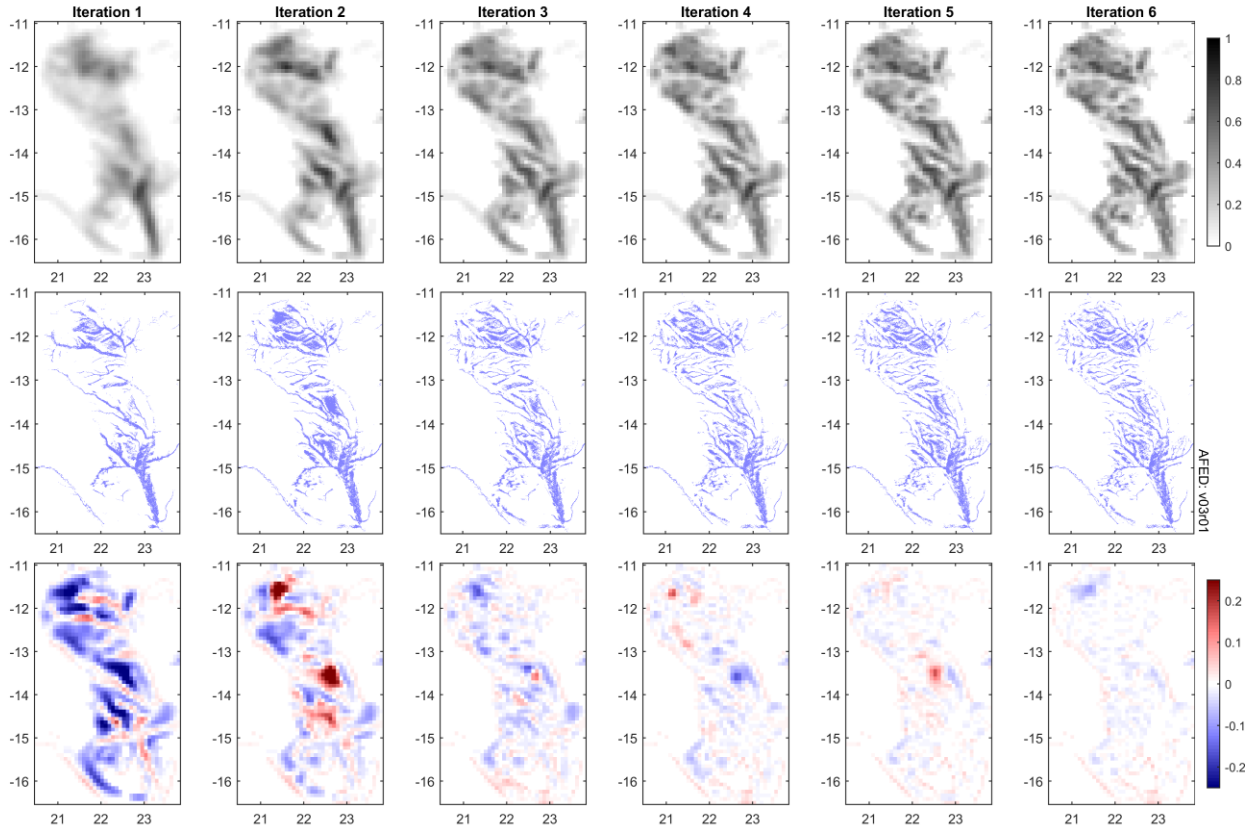

Figure 15: The iterative downscaling process, Barotse Flood Plain, 2007 March 4. Top row: Flooded fraction inputs for each iteration (1-6); iteration 1 represents observed flooded fraction. Middle: AFED following downscaling of FF inputs for the indicated iteration. Bottom: Residual flooded fraction expressed as the upscaled AFED per each iteration minus observed FF.

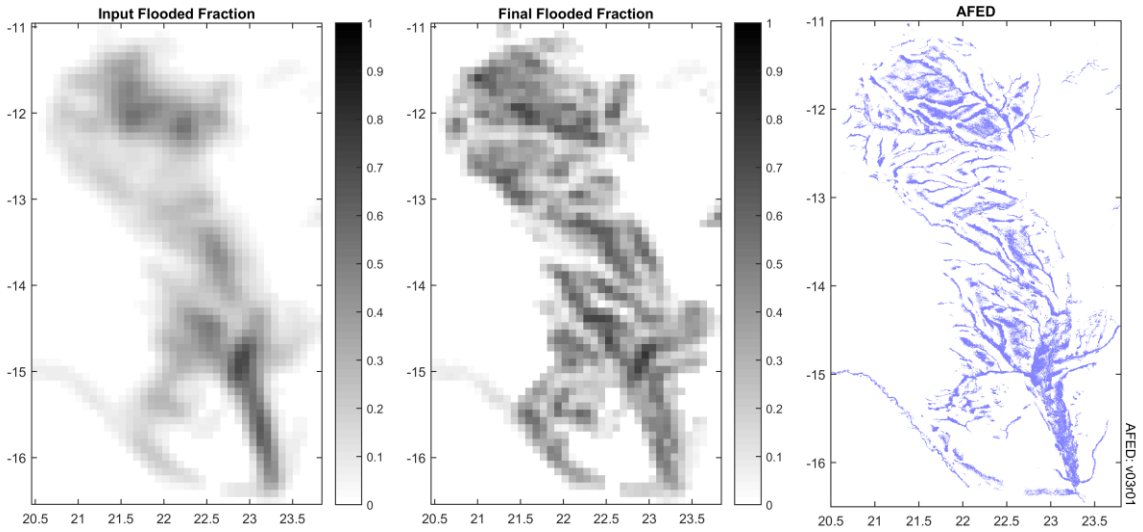

Figure 16: Final step in the downscaling process illustrated in Figure 15. Left: Observed flooded fraction. Center: Final flooded fraction including adjustments for the iteration 6 residual FF. Right: Final AFED results.

Because f_{f0} is pre-computed at each grid point, the downscaling step does not require computationally expensive high-resolution operations at retrieval time, enabling relatively rapid conversion of f_{flood} to a flood map.

The process for computing the flooded fraction and filling in missing data in the flooded fraction algorithm and the interpolation step in the downscaling algorithm mean that some of the source data for a flood extent depiction typically comes from a time outside the 24-hour AFED product UTC day. As described in section 3.8, a daily flooded fraction dataset may include 5-arcminute grid points with data from the current day, the next day, prior days, or a combination. Furthermore, when the AFED downscaling algorithm interpolates 5-arcminute flooded fraction data to the 3-arcsecond grid, the interpolated value may be based on data from the current day or a combination days depending on the time periods represented by neighboring 5-arcminute grid points. The effective time quality control metadata product (section 2.4.2) may be used to assess the degree to which the timing of the flooded fraction product varies from day-to-day.

4 Static ancillary data generation

4.1 Footprint matching coefficient sets

The footprint matching process uses weighting coefficient sets specific to each sensor, channel, and reference footprint shape to which the sensor data is to be matched.

To derive weighting coefficient sets, we used the methodology described in Galantowicz (2001) and Galantowicz (2004). We chose the reference shapes per each sensor to be a circularized version of the lowest-resolution channel of that sensor used by the flooded fraction algorithm. We specified coefficients for composite brightness temperature (CTB) samples that the footprint matching algorithm produces within the along-track, along-scan geometry of the satellite swath at a higher density than the original sensor samples. The regridding algorithm interpolates the CTB samples to the earth grid and the higher sampling density minimizes spatial sampling distortions in the interpolation process.

GMI operates and TMI operated with alternating fore- and aft-viewing geometries and TMI operated at two orbital altitudes. We computed separate footprint matching coefficient sets for each distinct operational condition set.

4.2 Land-water mask (LWM)

The AFED downscaling algorithm uses a 3-arcsecond resolution open water mask to identify grid points where water is most likely to persist regardless of inundation state; the algorithm treats these points as persistent water. The microwave flooded fraction algorithm uses the mask to account for the footprint-weighted persistent open water fraction (f_{pow}) in flooded fraction (f_f) retrievals.

We derived the LWM at 3-arcsecond (3″) resolution from a combination of three datasets (Table 16):

1. the Landsat-derived 1″ global water mask (*datamask*) from Hansen et al. (2013);
2. the 1″ Shuttle Radar Topography Mapping Mission (SRTM) Water Body Data (SRTMSWBD; Farr et al., 2007); and
3. the MOD44W 250-m binary mask (Carroll et al., 2009).

First, we combined the Hansen et al. *datamask* on a 1″ cell-by-cell "or" basis with SRTMSWBD. If either dataset identified a cell as water then the AFED land-water mask flags it as water. Second, we calculated the water cover fraction on 3″ cells from a weighted average of 4x4 blocks of the combined binary mask. The 4x4 weights account for the fact that the 1″ combined mask has center-aligned cells and the 3″ AFED grid has edge-aligned cells. Finally, we create the binary AFED land-water mask by flagging 3″ cells with at least 50% water cover fraction as water. The algorithm is capable of filling coverage gaps using MOD44W; no such gaps were found in the land area covered by AFED.

Table 16: Inputs for land-water mask computation

Input	Description
Global Forest Change 2000-2013 Data Mask	1-arcsecond land-water binary mask, center-aligned geographic grid (Hansen et al., 2013). Based on cloud free, growing season observations made from year 2000 to 2012. Data and license and attribution statement available at http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.0.html
Shuttle Radar Topography Mission Water Body Data (SRTMSWBD)	1-arcsecond land-water binary mask, center-aligned geographic grid (Farr et al., 2007). Based on satellite radar observations made in February 2000. Data available on-line from: https://lpdaac.usgs.gov/products/measures_products_table
MODIS Water Mask (MOD44W)	Complete global map of surface water at ~250 m spatial resolution on a geographic grid (Carroll et al., 2009). Data available on-line from: http://glcf.umd.edu/data/watermask/

We use the term "persistent" instead of "permanent" here to emphasize the transient nature of inland water bodies. The datasets used to define the land-water mask are the best available but represent only the conditions observed during certain periods; they do not represent annual or longer-term variations in water body size. The Hansen et al. *datamask* and SRTMSWBD datasets represent water detected at different times with different methodologies, each of which is sensitive to water in different marginal conditions and each of which is known to misidentify some land and water areas. Therefore, a combined dataset (using Boolean "or") has a higher likelihood of correctly identifying water than either dataset alone but it also has a higher

likelihood of incorrectly identifying non-water areas as water. For example, one of the input datasets may include data taken during a significant flood event or systematically misclassify certain land cover type as water (e.g., *datamask* flags lava fields in northern Ethiopia as water).

We used the following procedure to identify and resolve significant discrepancies between *datamask* and SRTMSWBD:

1. Detect areas where one of the input datasets significantly over-represents persistent water cover relative to the other:
    a. Divide the datasets into 2.5-arcminute areas (i.e., 150x150 1-arcsecond cells) in a grid covering Africa.
    b. For each dataset (*datamask* and SRTMSWBD), sum the number of 1-arcsecond cells in each 2.5-arcminute area where the dataset indicates water but the other does not.
    c. Identify 2.5-arcminute areas of interest, defined as areas where the sum in step 1b for one of the datasets is greater than ½ the total number of 1-arcsecond cells in the 2.5-arcminute area.
2. Determine the proximate cause of the discrepancy:
    a. Lava Fields: a group of 4 or more neighboring cells identified in step 1 for which *datamask* indicates more water than SRTMSWBD and Google Earth imagery indicates lava fields are present.
    b. Tidal Flooding: any cells identified in step 1 in a tidal area near the continental coast (within 20 km).
    c. Variable Reservoir Area: a group of identified cells associated with a reservoir identifiable in Google Earth imagery.
    d. Inland Flooding: a group of 4 (2x2 arrangement) or more neighboring cells identified in step 1 and not falling into categories a-c.
    e. Minor discrepancy: a single cell not falling into categories a-d.
3. Determine what change, if any, in the dataset combination algorithm is needed for the affected area.
    a. Lava Fields: do not use the dataset in which lava is misidentified as water (*datamask*) anywhere within the basin in which the lava was detected.
    b. Tidal Flooding: No change—continue to use both datasets (i.e., Boolean "or" combination). This approach means that the AFED LWM represents the highest tide conditions represented by either dataset.
    c. Variable Reservoir Area: No change—continue to use both datasets (i.e., Boolean "or" combination). This approach means that the AFED LWM represents the highest reservoir water level represented by either dataset.
    d. Inland Flooding: do not use the dataset with flooding anywhere within the basin in which inland flooding was detected. We used basin boundary data from Hoogeveen (2009).
    e. Minor Discrepancy: No change—continue to use both datasets (i.e., Boolean "or" combination).

Table 17 lists three example areas affected by dataset inconsistencies and the corrective actions taken using this methodology. Figure 17 shows areas where discrepancies were found and how they were resolved. (Areas with tidal flooding are not shown.) Note that three islands

fall into the "inland flooding" category because they are not represented at all in the *datamask* (i.e., *datamask* effectively represents these areas as ocean).

Table 17: Land-water mask input dataset inconsistency corrections

Affected area	Inconsistency	Corrective Action
Lava fields near L. Afrera and L. Karum, Ethiopia	Hansen et al. *datamask* incorrectly identifies some lava flow areas as water	LWM uses only SRTMSWBD data in this area.
Wetlands area north of Chibuto, Mozambique	SRTMSWBD indicates significantly more water than Hansen et al. *datamask* does. There was flooding in this area in Feb. 2000 during the SRTM observations.	LWM uses only Hansen et al. *datamask* in this area.
Makgadikgadi Pans, Botswana	SRTMSWBD indicates significantly more water than Hansen et al. *datamask* does. There was flooding in this area in Feb. 2000 during the SRTM observations.	LWM uses only Hansen et al. *datamask* in this area.

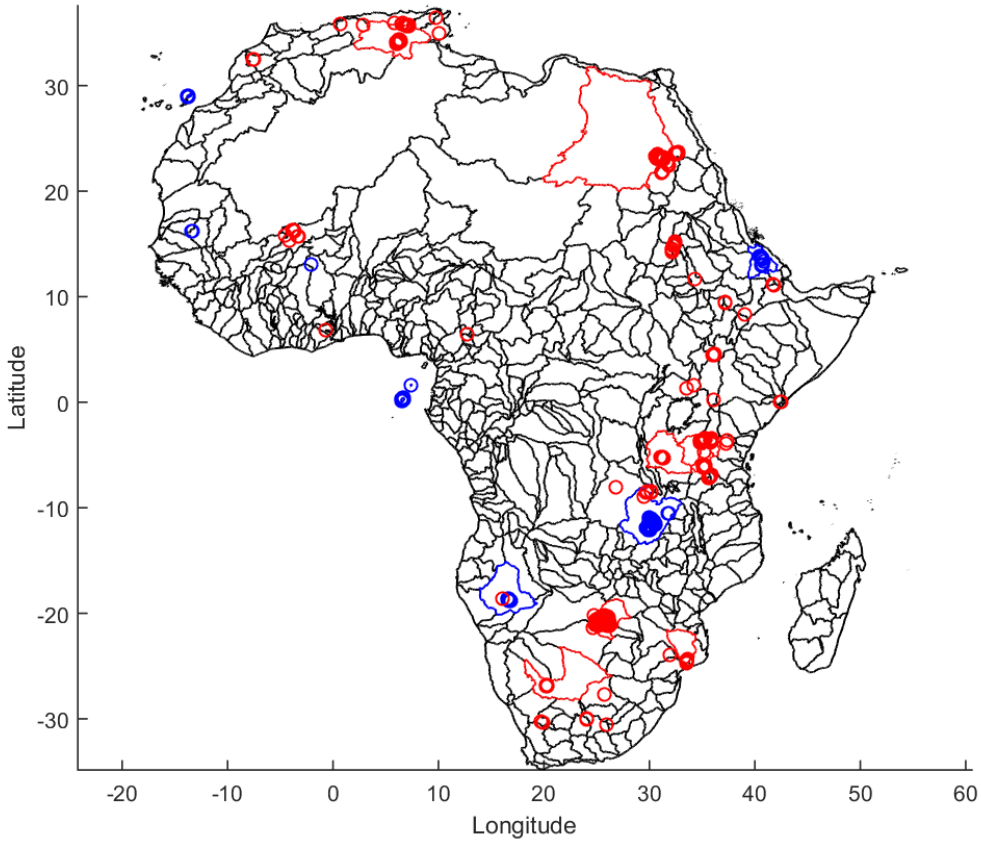

Figure 17: Red circles are areas where SRTMSWBD over-represents water. Blue circles are areas where Hansen *datamask* over-represents water. Black lines delineate basins in which SRTMSWBD are combined using Boolean "or.". Red lines delineate basins where SRTMSWBD was not used. Blue lines delineate basins where Hansen *datamask* was not used.

4.3 Persistent open water area fraction

The AFED flooded fraction algorithm uses footprint-weighted persistent open water fraction (f_{pow}) in flooded fraction retrievals.

A value for f_{pow} is calculated at each point on the 5′ AFED grid by the weighted sum of the binary 3-arcsecond land-water mask. The weights are determined from a circular Gaussian-like function of distance from the grid point; the weights are normalized to sum to one. The half-weight diameter of the circular Gaussian-like function is equal to the footprint resolution of the microwave sensor data with which the f_{pow} data are used in the flooded fraction algorithm. The circular Gaussian-like spatial weighting function at each resolution is the same as that used in the footprint matching process (section 3.3). Figure 21 shows f_{pow} for the 22D RFS.

4.4 Tree cover area fraction

The AFED flooded fraction algorithm uses footprint-weighted tree cover area fraction (f_{tree}) to derive the dry land Q end-member climatology.

A value for f_{tree} is calculated at each point on the 5′ AFED grid by the weighted sum of the Hansen et al. (2013) 1-arcsecond tree canopy cover for year 2000 (*treecover2000*) dataset. The weights are determined from a circular Gaussian-like function of distance from the grid point; the weights are normalized to sum to one. The half-weight diameter of the circular Gaussian-like function is equal to the footprint resolution of the microwave sensor data with which the f_{tree} data are used. The circular Gaussian-like spatial weighting function at each resolution is the same as that used in the footprint matching process (section 3.3).

4.5 Elevation model

As of V05R00, the AFED algorithm uses the MERIT 3-arcsecond digital elevation model, (DEM; Yamazaki et al., 2017). MERIT is based on data from the Shuttle Radar Topography Mission (SRTM; Farr et al., 2007) and improves upon the SRTM datasets in several respects, including correction for bias induced by tree cover, removal of radar processing artifacts (known as “striping”), and reduction in pixel-wise noise. The MERIT developers have also provided flow direction and accumulation area datasets based on MERIT DEM data. All three datasets are used in relative floodability calculations (section 4.21.1).

The AFED algorithm requires an edge-aligned, 3-arcsecond DEM to produce AFED data on edge-aligned tiles that promote ease of use. The MERIT DEM and associate flow direction and accumulation area datasets are distributed on 3-arcsecond, center-aligned grids. We shifted the MERIT data by 1.5-arcseconds in each direction to align it with the AFED grid.

4.6 Normalized difference vegetation index climatology

The microwave flooded fraction algorithm uses seasonal vegetation indices to tabulate $Q_{dc}(d_c,x)$: the Q end-member for dry land for each climatological day-of-year, d_c, and 5-arcminute grid point, x.

We use normalized difference vegetation index (NDVI) values from the 2000-2010 Long Term Average Vegetation Index dataset provided by the Vegetation Index and Phenology Laboratory at the University of Arizona (VIP Lab., 2011). The dataset provides 15-day average NDVI on a 0.05° latitude x 0.05° longitude grid.

A value for NDVI is calculated at each point on the 5′ AFED grid by a weighted sum. The weights are determined from a circular Gaussian-like function of distance from the grid point;

the weights are normalized to sum to one. The half-weight diameter of the circular Gaussian-like function is equal to the footprint resolution of the microwave sensor data with which the $Q_{dc}(d_c,x)$ data are used. The circular Gaussian-like spatial weighting function at each resolution is the same as that used in the footprint matching process (section 3.3).

The final step interpolates the 15-day average NDVI values to each day-of-year to create a 365-day NDVI climatology at each point on the 5′ AFED grid.

4.7 CSU SSM/I calibration alignment parameters

The microwave flooded fraction algorithm uses the CSU SSM/I calibration alignment parameters to modify *Q* values computed from CSU SSM/I source data such that they align with *Q* values computed from RSS SSM/I source data.

The CSU calibration alignment model is a 2-parameter linear equation:

$$Q_{RSS}(Q_{CSU}) = p_1 Q_{CSU} + p_2.$$

We calculated the parameters (p_1,p_2) per sensor band (19 and 37 GHz) and satellite pass (ascending and descending) using linear least-squares fitting trained to daily coincident Q_{RSS} and Q_{CSU} SSM/I data from December 2011 on days with at least 10,000 valid points. The final (p_1,p_2) values, given in Table 18, were computed as the mean over all valid days.

Table 18: CSU SSM/I calibration alignment parameters

Satellite pass	Microwave band			
	19 GHz		37 GHz	
	p_1	p_2	p_1	p_2
Ascending	1.0133	-0.0059	1.0050	0.0001
Descending	1.0091	-0.0039	1.0028	0.0011

4.8 AMSR2-to-AMSR-E calibration alignment parameters

The microwave flooded fraction algorithm uses the AMSR2-to-AMSR-E calibration alignment parameters to modify *Q* values computed from AMSR2 source data such that they can be used with historical data compiled from combined AMSR-E and AMSR2 data. The AFED algorithm uses historical AMSR-E and AMSR2 data for computing seasonal averages and other statistics (described in later sections). AMSR-E and AMSR2 share a common design and orbit but *Q* differences may still arise from onboard calibration and sensor data processing system differences.

The AMSR2-to-AMSR-E calibration alignment model is a 2-parameter linear equation:

$$Q_{AMSR\text{-}E}(Q_{AMSR2}) = A_{xAMSR2} Q_{AMSR2} + B_{xAMSR2}.$$

We calculated the parameters (A_{xAMSR2},B_{xAMSR2}) per sensor band (19 and 37 GHz) and satellite pass (ascending and descending) using robust linear least-squares fitting trained to separately-computed AMSR-E (2003-2010) and AMSR2 (2013-2016) 61-day median climatologies. We compute daily (A_{xAMSR2},B_{xAMSR2}) values from locations with less than 0.0005 year-to-year 61-day median *Q* variation (section 4.13). Final (A_{xAMSR2},B_{xAMSR2}) values, given in Table 19, were computed as the mean over all valid training results.

Table 19: AMSR2 to AMSR-E calibration alignment parameters

Satellite pass	Microwave band			
	19 GHz		37 GHz	
	A_{xAMSR2}	B_{xAMSR2}	A_{xAMSR2}	B_{xAMSR2}
Ascending	1.0554	-0.0188	1.0466	-0.0154
Descending	1.0603	-0.0215	1.0438	-0.0141

4.9 GMI-to-AMSR-E calibration alignment parameters

The microwave flooded fraction algorithm uses the GMI-to-AMSR-E calibration alignment parameters to modify *Q* values computed from GMI source data such that they can be used with historical data compiled from AMSR-E and AMSR2. The AFED algorithm uses historical AMSR-E and AMSR2 data for computing seasonal averages and other statistics (described in later sections). The GMI design is similar to AMSR-E and AMSR2 but differences—including that GMI's earth-incidence angle (52.8°) differs from that of AMSR-E and AMSR2 (55°)—are sufficient to warrant a correction. Furthermore, because GMI is non-sun-synchronous, there is no systematic time-of-day distinction between ascending and descending pass data. As a result, the AFED algorithm uses the average of ascending and descending pass AMSRX statistics for GMI processing.

The GMI-to-AMSR-E calibration alignment model is a 2-parameter linear equation:

$$Q_{AMSR\text{-}E}(Q_{GMI}) = A_{xGMI}Q_{GMI} + B_{xGMI}.$$

We calculated the parameters (A_{xGMI},B_{xGMI}) per sensor band (19 and 37 GHz) using robust linear least-squares fitting trained to coincident Q_{GMI} (single time sample data) and Q_{AMSR2} (daily average data) from 2015/1/1 to 2016/12/31. AMSR2-to-AMSR-E cross-calibration correction was applied to all AMSR2 data prior to computing daily averages. Only GMI datasets with at least 10,000 valid AMSR2 overlap points were included in the training. The final (A_{xGMI},B_{xGMI}) values, given in Table 20, were computed as the mean over all valid training results.

Table 20: GMI to AMSR-E calibration alignment parameters

Satellite pass	Microwave band			
	19 GHz		37 GHz	
	A_{xGMI}	B_{xGMI}	A_{xGMI}	B_{xGMI}
Ascending & descending	1.0963	-0.0324	1.0835	-0.0276

4.10 TMI-to-AMSR-E calibration alignment parameters

The microwave flooded fraction algorithm uses the TMI-to-AMSR-E calibration alignment parameters to modify *Q* values computed from TMI source data such that they can be used with historical data compiled from AMSR-E and AMSR2. The AFED algorithm uses historical AMSR-E and AMSR2 data for computing seasonal averages and other statistics (described in later sections). The TMI design is similar to AMSR-E and AMSR2 but differences—including that TMI's earth-incidence angles (52.8° and 53.4° at 350-km and 402-km altitude, respectively) differ from that of AMSR-E and AMSR2 (55°)—are sufficient to warrant a correction. Furthermore, because TMI is non-sun-synchronous, there is no systematic time-of-day distinction between ascending and descending pass data. As a result, the AFED algorithm uses the average of ascending and descending pass AMSRX statistics for TMI processing.

The TMI-to-AMSR-E calibration alignment model is a 2-parameter linear equation:

$$Q_{AMSR\text{-}E}(Q_{TMI}) = A_{xTMI}Q_{TMI} + B_{xTMI}.$$

We calculated the parameters (A_{xTMI},B_{xTMI}) per sensor band (19 and 37 GHz) for the two TMI orbital altitudes using methods developed for AMSR2 and GMI cross-calibration to AMSR-E.

- **TMI 402-km altitude**: TMI operations at 402-km altitude overlaps with the AMSR-E operational period. To derive (A_{xTMI},B_{xTMI}) we used using robust linear least-squares fitting trained to coincident Q_{TMI} (single time sample data) and $Q_{AMSR\text{-}E}$ (daily average data) from 2003/1/1 to 2010/12/31. Only TMI datasets with at least 10,000 valid AMSR-E overlap points were included in the training. The final (A_{xTMI},B_{xTMI}) values, given in Table 21, were computed as the mean over all valid training results.
- **TMI 350-km altitude**: TMI operations at 350-km altitude do not overlap with the AMSR-E, AMSR2, or GMI operational periods. To derive (A_{xTMI},B_{xTMI}) we used robust linear least-squares fitting trained to separately-computed AMSR-E (2003-2010) and TMI (1998-2000) 61-day median climatologies. We compute daily (A_{xTMI},B_{xTMI}) values from locations with less than 0.0005 year-to-year 61-day median Q variation (section 4.13). The final (A_{xTMI},B_{xTMI}) values, given in Table 21, were computed as the mean over all valid training results.

Table 21: TMI to AMSR-E calibration alignment parameters

TMI satellite altitude	Microwave band			
	19 GHz		37 GHz	
	A_{xTMI}	B_{xTMI}	A_{xTMI}	B_{xTMI}
350 km	1.1132	-0.0371	1.1120	-0.0368
402 km	1.0938	-0.0309	1.0787	-0.0258

4.11 Persistent open water *Q* end-member

The microwave flooded fraction algorithm uses the persistent open water Q end-member, Q_{pow}, in eq. 13 to predict the value of Q for persistent open water areas. The algorithm uses different Q_{pow} values for each microwave frequency and satellite pass direction. Q_{pow} has no seasonal dependence but ascending and descending microwave sensor passes are treated separately to account for observed systematic differences.

We empirically derived Q_{pow} for Africa using f_{pow} data and average Q values observed in the Lake Victoria region over the years 2003-2010 for AMSRX and 1994-2001 for SSM/I.

Figure 18 illustrates the first calibration step. We accumulated all valid (non-rain- and non-frozen-flagged) atmospherically-corrected microwave data by climatological day-of-year, d_c. We removed points with f_{pow} between 0.4 and 0.95 because the rain flagging algorithm is not meant to distinguish between the similar signatures of partial land cover and rain in this situation. We then used a robust fitting algorithm to calculate the best-fit line; $Q_{pow}(d_c)$ is the value of the best-fit line where water fraction is 1. (This process also yields Q_{dry} values at the point where the water fraction is 0; however, this value is only used in the Q_{pow} derivation process; it is not used in the flooded fraction algorithm itself.)

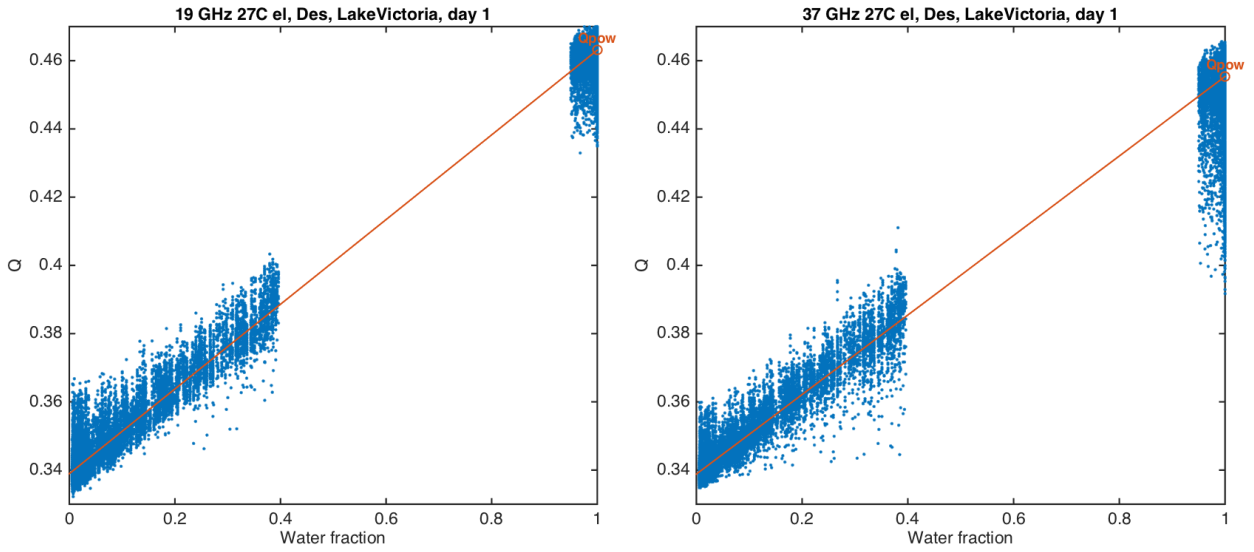

Figure 18: 19 (left) and 37-GHz AMSR-E Q_{pow} calibration data and best-fit line for 1 January, descending pass. Data were accumulated over the years 2003-2010.

Figure 19 shows 19 GHz Q_{pow} values derived for each day-of-year. We chose to average Q_{pow} values over the year because the values do not appear to have a strong seasonal dependence relative to day-to-day variation. The resulting Q_{pow} values are given in Table 22.

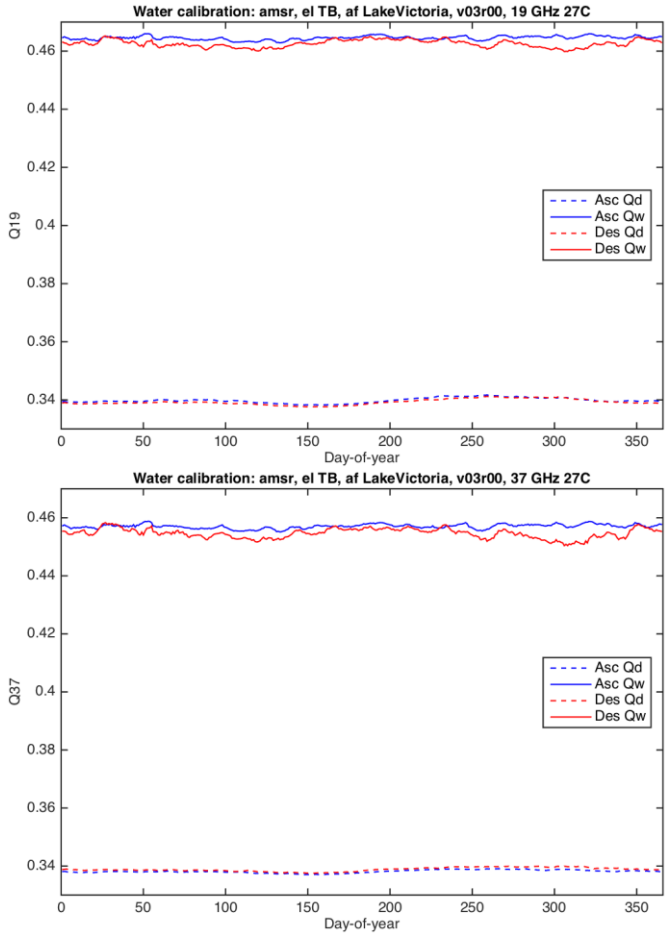

Figure 19: 19 GHz (top plot) and 37 GHz (bottom) AMSR-E Q_{pow} values for ascending and descending passes derived per day-of-year. Best-fit line Q_{dry} values are also shown.

Table 22: Q_{pow} values derived from AMSR-E and SSM/I data.

Satellite pass	Microwave band and sensor			
	19 GHz		37 GHz	
	AMSR-E	SSM/I	AMSR-E	SSM/I
Ascending	0.465	0.450	0.457	0.448
Descending	0.463	0.446	0.455	0.443

4.12 *Q* 31- and 61- day median climatologies

The microwave algorithm uses the *Q* 31- day and 61-day time-centered median climatologies ($Q_{61c}(d_c,x)$ and $Q_{31c}(d_c,x)$, respectively) in several contexts: (1) seasonal wetlands detection (section 4.17), (2) dry-land *Q* end-member climatology calculations (section 4.18), and (3) *Q* 61-day median model parameter calculations (section 4.14).

To calculate the climatologies, we first calculated and stored the 61- and 31-day running median *Q* (Q_{61} and Q_{31}, respectively) for each day and grid point in the historical record, adjusting for f_{pow} as $Q = (Q_{obs} - f_{pow}Q_{pow})/(1-f_{pow})$ and applying cross-calibration to AMSR2 data. We then computed Q_{61c} and Q_{31c} per climatological day-of-year, d_c, as the mean of the yearly daily values from the years 2003-2010 and 2013-2016 for AMSRX and 1992-2001 for SSM/I.

4.13 *Q* 61-day historical year-to-year variance

The microwave algorithm uses the *Q* 61-day year-to-year variance, $\sigma_{Q61c}^2(d_c,x)$, in: (1) seasonal wetlands detection (section 4.17), (2) dry-land *Q* end-member climatology calculations (section 4.18), and (3) daily dry-land end-member calculations (section 3.8).

The $\sigma_{Q61c}^2(d_c,x)$ computation used the 61-day running median, Q_{61}, described in section 4.12. We first computed the year-to-year Q_{61} variance separately for satellite ascending and descending pass data per day-of-year using the yearly values from the years 2003-2010 and 2013-2016 for AMSRX and 1992-2001 for SSM/I. We then averaged the ascending and descending pass variance values for an initial point-wise variance, σ_{pw}^2. We ran a 9x9 point spatial median filter over the point-wise variance each day to yield σ_{smooth}^2. We then computed the year-over-year trend in the yearly mean Q_{61} and from that the worst-case trend-induced Q_{61} bias at the end dates of the AMSRX and SSM/I time periods. We added the squared bias to the point-wise variance to account for, e.g., changing land cover conditions, yielding $\sigma_{pw+bias}^2$. We then flagged areas where the point-wise variance showed signs of being affected by annual flooding and set the final $\sigma_{Q61c}^2(d_c,x)$ to σ_{smooth}^2 in these and coastal areas and $\sigma_{pw+bias}^2$ elsewhere. We used the spatial median to minimize the effects of regular seasonal flooding on the year-to-year variance because the variance is intended to represent variability in dry-land conditions only; we found that the application of a 9x9 filter was effective in isolating dry-land conditions in the vicinity of seasonal floodplains.

4.14 *Q* 61-day median time shift model parameters

The flooded fraction algorithm uses the *Q* 61-day median time shift model parameters to compute the expected value of the current 61-day time-centered median *Q*, Q_{61E}, from the current 61-day retrospective median *Q*, Q_{61R}. A 61-day time-centered *Q* median on day *d* represents the median of *Q* values spanning the period from *d*-30 days to *d*+30 days.

The 61-day time-centered median model is a 2-parameter linear equation:

$$Q_{61E}(Q_{61R}) = a_{61}Q_{61R} + b_{61}. \tag{20}$$

We calculated the initial parameters (a_{61i}, b_{61i}) per climatological day-of-year, d_c, using a modified simple linear regression model trained to Q_{61} data from the years 2003-2010 and 2013-2016 for AMSRX and 1993-2001 for SSM/I. Ascending and descending pass data were combined in the calculation. The modification biases the slope parameter, a_{61i}, toward 1 in cases where Q_{61R} variance is small:

$$a_{61i} = \frac{w_0 \mathrm{cov}(Q_{61R},Q_{61R})/\mathrm{var}(Q_{61R})+1/\mathrm{var}(Q_{61R})}{w_0+1/\mathrm{var}(Q_{61R})} \tag{21}$$

$$b_{61i} = \bar{Q}_{61} - a_{61i}\bar{Q}_{61R} \tag{22}$$

where $w_0 = 1/10^{-4}$ and overbars indicate inter-year mean values.

As of AFED V04R01, we take three additional steps to compute a final (a_{61}, b_{61}) set. First, to account for extreme years when seasonal changes increase *Q* more quickly than average, we calculate the mean error residual, ER, of (a_{61i}, b_{61i}) from the four pass-year datasets with the largest positive 30-day *Q* change. We then spatially smooth (a_{61i}, b_{61i}) using a 7x7 grid point median filter and ER using a 5x5 filter (ER_{smooth}). Median spatial smoothing helps to minimize the undesirable effects of regular seasonal flooding by considering data over a larger area, only a

portion of which is usually affected by flooding. We then set the final a_{61} to the spatially smoothed a_{61i} and define the final $b_{61} = b_{61i} + 0.9\ ER_{smooth}$.

4.15 *Q* 31-day median time shift model parameter

The flooded fraction algorithm uses the *Q* 31-day median time shift model parameter to compute the expected value of the current 31-day time-centered median *Q*, Q_{31E}, from the current 31-day retrospective median *Q*, Q_{31R}. A 31-day time-centered *Q* median on day *d* represents the median of *Q* values spanning the period from *d*-15 days to *d*+15 days.

The 31-day time-centered median model is a 1-parameter linear equation:

$$Q_{31E}(Q_{31R}) = Q_{31R} + b_{31}. \quad (23)$$

We calculated the parameter b_{31} per climatological day-of-year, d_c, initially from the 31-day median climatology, Q_{31C}:

$$b_{31} = Q_{31C}\,(d) - Q_{31C}\,(d\text{-}15). \quad (24)$$

As of V04R01 we also spatially smooth the initial b_{31} using a 7x7 grid point median filter. Median spatial smoothing helps to minimize the undesirable effects of regular seasonal flooding on b_{31} by considering data over a larger area, only a portion of which is usually affected by flooding.

4.16 *Q* tabulated by NDVI and tree cover fraction

Look-up tables (LUTs) relating *Q* to seasonal NDVI and tree cover fraction, $Q_{dLUT}(NDVI,f_{tree})$, are used in the algorithms to detect seasonal wetlands signatures (section 4.17) and to calculate the dry-land end-member *Q* climatology (section 4.18).

We computed $Q_{dLUT}(NDVI,f_{tree})$ from a 20% subsample of *Q* data from the years 2003-2010 and 2013-2016 for AMSRX and 1992-2001 for SSM/I. Prior to subsampling, we excluded locations with the following land cover types unlikely to have *Q* correlations to seasonal NDVI and tree cover fraction: urban and built up, permanent wetlands, and snow and ice. To determine land cover type, we computed the footprint-weighted fractions of all 500-m MODIS Land Cover Type (MCD12Q1[7]) classes. We excluded locations where the sum of all excluded class fractions exceeded 5%. Using the footprint-averaged NDVI climatology (section 4.6) and f_{tree} data (section 4.4), we binned the *Q* data subsample in increments of 0.05 NDVI and 0.05 f_{tree} irrespective of location and calculated the 75th percentile *Q* values of these bins to populate the table. Figure 20 shows the AMSRX look-up table results.

[7] https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1

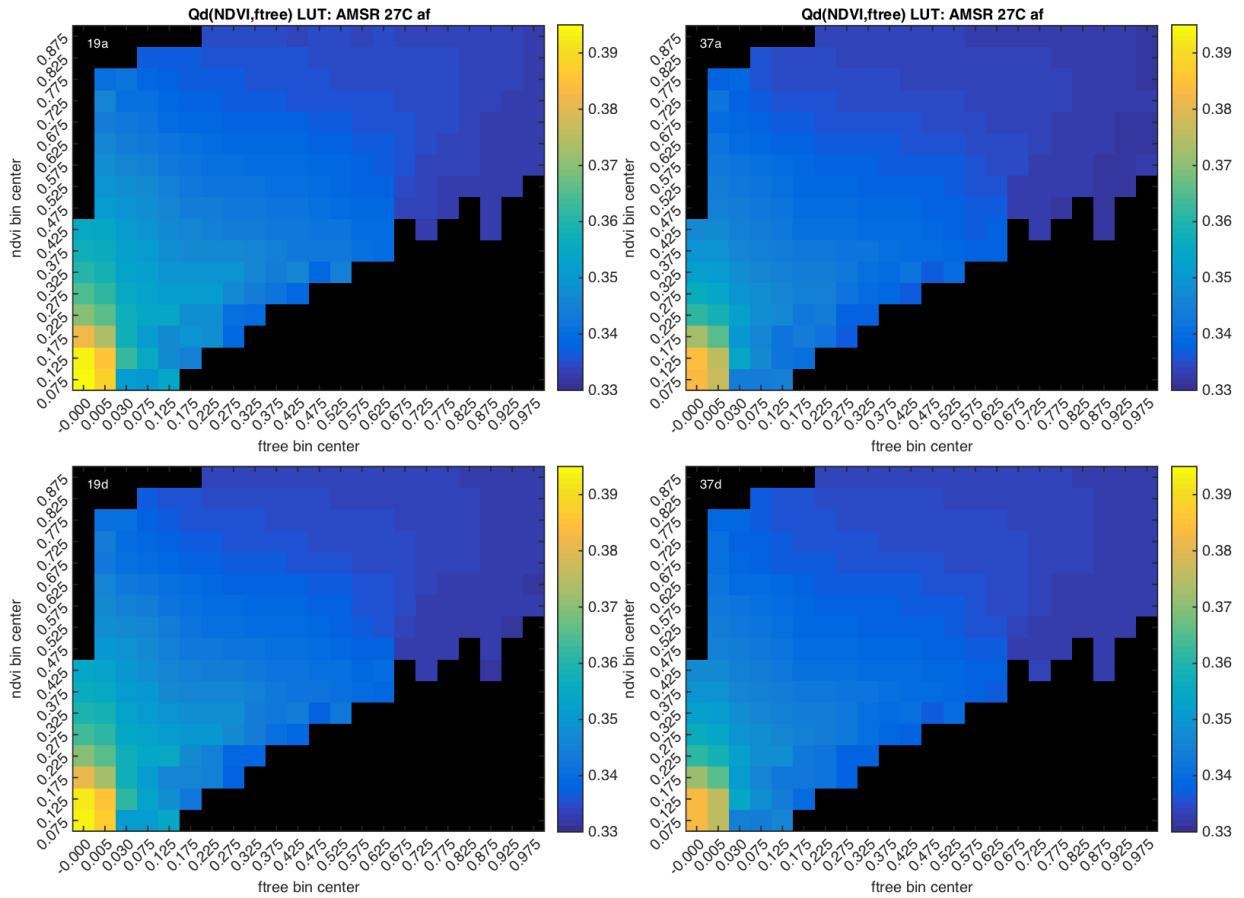

Figure 20: V04R01 $Q_{dLUT}(NDVI,f_{tree})$ tables for AMSRX 27C RFS 19 GHz (left) and 37 GHz (right) bands and ascending (top) and descending (bottom) satellite pass data.

4.17 Detected seasonal wetlands climatology

The microwave flooded fraction algorithm uses seasonal wetlands flags in dry-land *Q* end-member climatology calculations (section 4.18) and daily dry-land *Q* end-member calculations (section 3.8). The purpose of the flags is to indicate where and when climatological data are affected by regular flooding and therefore unreliable as predictors of dry-land conditions. The seasonal wetlands detection algorithm is divided in three parts representing detection of longer-duration seasonal wetlands, shorter-duration seasonal floodplains, and irregularly flooded saltpans. All parts are applied across Africa without regard for geographic location.

The longer-duration wetlands detection algorithm flags areas whose data match one of several microwave time series signatures. The primary algorithm tests use the difference $H_{61} = Q_{61c}(d_c,x)–Q_{dLUT}(NDVI,f_{tree})$: seasonal wetlands are a possibility where H_{61} is much larger than zero over a long period of time. In addition to tests on the magnitude of H_{61} and the length of time that $H_{61}>0$, the algorithm tests several other factors: NDVI and correlation between NDVI and Q_{61c}; change in Q_{61c} value when $H_{61}>0$; and the 7-day variance relative to change in Q_{61c} when $H_{61}>0$. These additional factors are needed to rule out seasonal changes caused by phenomena other than wetlands. The algorithm tests seven rule-groups to detect wetlands. Multiple rule-groups are needed to account for variations in vegetation amount and type in and near the wetlands, length of the wetlands season, and other factors affecting temporal and spatial variability in different parts of Africa.

The algorithm to detect saltpans is a modified version of the wetlands detection algorithm intended to indicate irregularly flooded areas with low vegetation amounts. The primary tests use the maximum annual difference $Q_{61c}(d_c,x)–Q_{dLUT}(NDVI,f_{tree})$ and the ratio of the annual Q_{61c} range to the 7-day variance when $H_{61}<0$. Areas detected as saltpans do not have a regular seasonal flooding signature and as a result climatological data are treated as unreliable predictors of flooding through the year in these areas.

The algorithm to detect consistent shorter-duration seasonal flooding signatures uses $Q_{31c}(d_c,x)$. Where $Q_{31c}(d_c,x)–Q_{dLUT}(NDVI,f_{tree})>0$ for a period of days, we calculate $H_{31} = Q_{31c}(d_c,x) – \max(Q_{31c}(d_s,x), Q_{31c}(d_e,x))$, where d_s and d_e are the dates at the start and end of the period. The algorithm flags the period and location using rules based on: (a) duration of the period, (b) maximum H_{31} during the period, (c) difference between $Q_{31c}(d_s,x)$ and $Q_{31c}(d_e,x)$, (d) the range of $Q_{dc}(d_c,x)$ during the period, and (e) the maximum NDVI during the period. After the initial rules-based detection, the algorithm uses image processing techniques to dilate the daily flag to surrounding areas meeting more liberal detection thresholds.

The algorithm combines wetlands and seasonal flooding flags and flags neighboring areas as transitional. The number of days-per-year with the combined wetlands, seasonal flooding, and saltpan flags are shown in Figure 21.

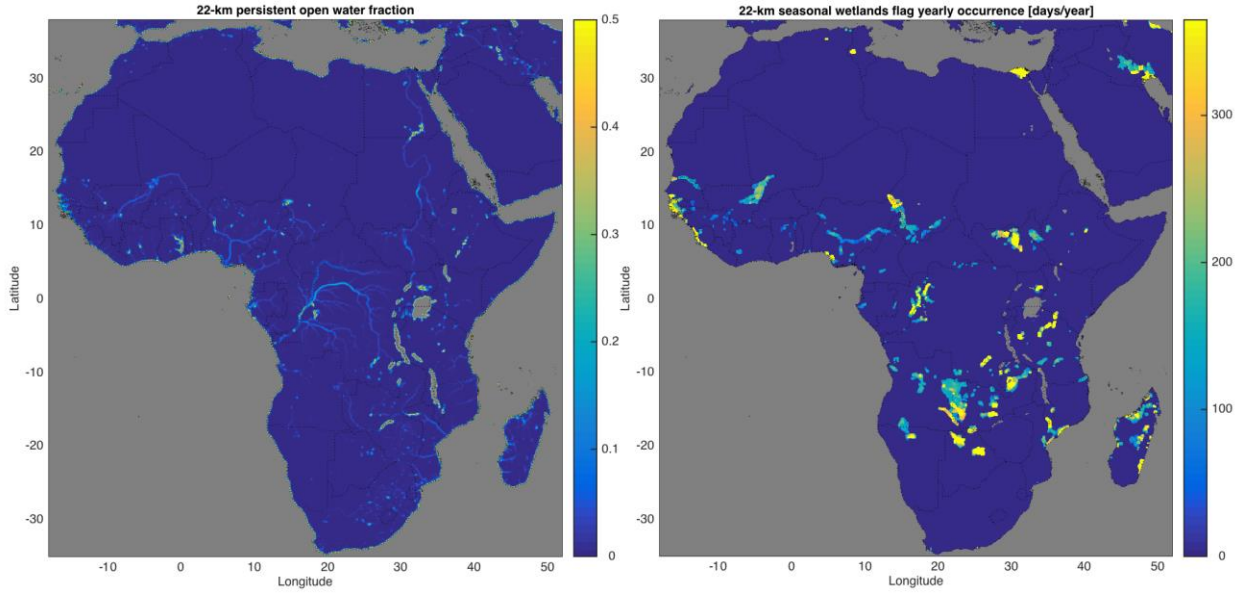

Figure 21: 22D RFS persistent open water fraction (left) and wetlands signature detection algorithm results (right) indicating number of days per year when wetlands signature was detected.

4.18 Dry-land *Q* end-member climatology

The microwave flooded fraction algorithm uses a tabulated dry land *Q* end-member climatology, $Q_{dc}(d_c,x)$, to help predict the current value of *Q* for dry land areas (neither open water nor flooded). The algorithm uses different $Q_{dc}(d_c,x)$ tables for each microwave frequency, satellite pass direction, resolution, and sensor type (AMSRX and SSM/I).

We empirically derived $Q_{dc}(d_c,x)$ for Africa using (a) *Q* 61-day median climatology, $Q_{61c}(d_c,x)$ (section 4.12), (b) footprint-averaged NDVI climatology (section 4.6), and (c) *Q* tabulated by NDVI and tree cover fraction, $Q_{dLUT}(NDVI,f_{tree})$ (section 4.16).

We derived $Q_{dc}(d_c,x)$ from the following steps:

1. *Tabulate initial Q_{dc} climatologies, $Q_{dci}(d_c,x)$*: The initial Q_{dc} is intended to adjust $Q_{dLUT}(NDVI,f_{tree})$ for local extremes not represented in the Africa-wide look-up table statistics. First, we found the local minimum and maximum $Q_{61c}(x)$ values (*minQ61c*, *maxQ61c*) and the local average *Q* values at minimum and maximum NDVI (*Q*(*minNDVI*), *Q*(*maxNDVI*)). Second, we flagged areas that (1) had anomalous *Q*-NDVI relationships (i.e., *Q* increasing with NDVI) and (2) were not also possible wetland areas (flagged separately as wetlands with anomalous *Q*-NDVI). To adjust $Q_{dLUT}(NDVI,f_{tree})$ for local extremes, we applied the following rules for points not flagged as anomalous:
   a) where NDVI(d_c,x) ≤ *minNDVI*, set $Q_{dci}(d_c,x)$ to *Q*(*minNDVI*).
   b) where *minNDVI*<NDVI(d_c,x)≤*minNDVI*+0.25, calculate $Q_{dci}(d_c,x)$ by interpolating in NDVI between *Q*(*minNDVI*) and Q_{dLUT}(minNDVI+0.25,f_{tree}).
   c) where NDVI>*minNDVI*+0.25, look up $Q_{dci}(d_c,x)$ using $Q_{dLUT}(NDVI,f_{tree})$.

   A final step adjusted $Q_{dci}(d_c,x)$ to be consistent with the minQ61c. For areas flagged as anomalous, *Q*(*NDVI*,*ftree*) was not used at all and instead $Q_{dci}(d_c,x)$ was computed by interpolating in NDVI between *Q*(*minNDVI*) and *Q*(*maxNDVI*).
2. *Tabulate final Q_{dry} climatologies, $Q_{dc}(d_c,x)$*. The algorithm computes the final $Q_{dc}(d_c,x)$ per climatological day-of-year and location as a weighted average of the initial $Q_{dci}(d_c,x)$ and $Q_{61c}(d_c,x)$:

$$Q_{dry}(d,x) = \frac{W_{Qdc}Q_{dci} + W_{Q61c}Q_{61c}}{W_{Qdc} + W_{Q61c}} \tag{25}$$

   where W_{Qdc} and W_{Q61c} are initially set to 1 then:
   a. for wetlands and $Q_{61c}(d_c,x) \le Q_{dLUT}(NDVI,f_{tree})$: $W_{Qdc} = 0$
   b. for non-wetlands and $Q_{61c}(d_c,x) \le Q_{dLUT}(NDVI,f_{tree})$: $W_{Qdc} = 1/(25 + 200(|Q_{61c}| - Q_0))$, where $|Q_{61c}|$ is the annual mean of $Q_{61c}(d_c,x)$ and $Q_0 = 0.32$
   c. for non-wetlands and $Q_{61c}(d_c,x) > Q_{dLUT}(NDVI,f_{tree})$: $W_{Qc} = 1/25$,
   d. for wetlands and $Q_{61c}(d_c,x) > Q_{dLUT}(NDVI,f_{tree})$ or for salt pans on any date: $Q_{dry}(d_c,x) = \min(Q_{61c}(d_c,x), Q_{dLUT}(NDVI,f_{tree}))$.

   Finally, for areas flagged as transitional Q_{dry} is recomputed as the average of the value from eq. 25 and $\min(Q_{61c}(d_c,x), Q_{dLUT}(NDVI,f_{tree}))$.

The use of climatological average NDVI here has two advantages: first, it reduces the chance that flooding itself may bias NDVI values derived during flood events, which would lead to biases in the flooded fraction retrieval; second, it eliminates the need for the currently-derived NDVI as an AFED algorithm input, improving the latency of the AFED algorithm for near real time operations.

4.19 Flooded fraction historical outlier prevalence factor

The false positive detection and final flood fraction algorithms use the flooded fraction historical outlier prevalence factor, f_{hop}, to indicate how likely outlier flooded fraction values. The factor takes values from zero for each grid point and day of year.

We calculate f_{hop} from the time series of initial flooded fraction values. As of V05R00, we use the AMSR-E initial FF time series at 22D resolution spanning the full years 2003 to 2010. We remove the 15-day running mean and set negative FF values to zero. We then compute the

time derivative at each point with valid data and compute the kurtosis of the FF time derivatives, k_{df}, for each day of year using data from all years in a 61-day window centered on the day. The f_{hop} factor is given as a function of k_{df} by:

$$f_{hop} = (\log10(k_{df}) - 0.75) / 0.75$$

with limits applied to enforce $0 < f_{hop} < 1$. Near coasts, $f_{hop} = 5f_{pow}$ to account for higher noise due to uncorrected geolocation error.

Figure 22 shows f_{hop} annual statistics. The factor is zero for much of Africa at least half the year and where it does reach one it does so only seasonally.

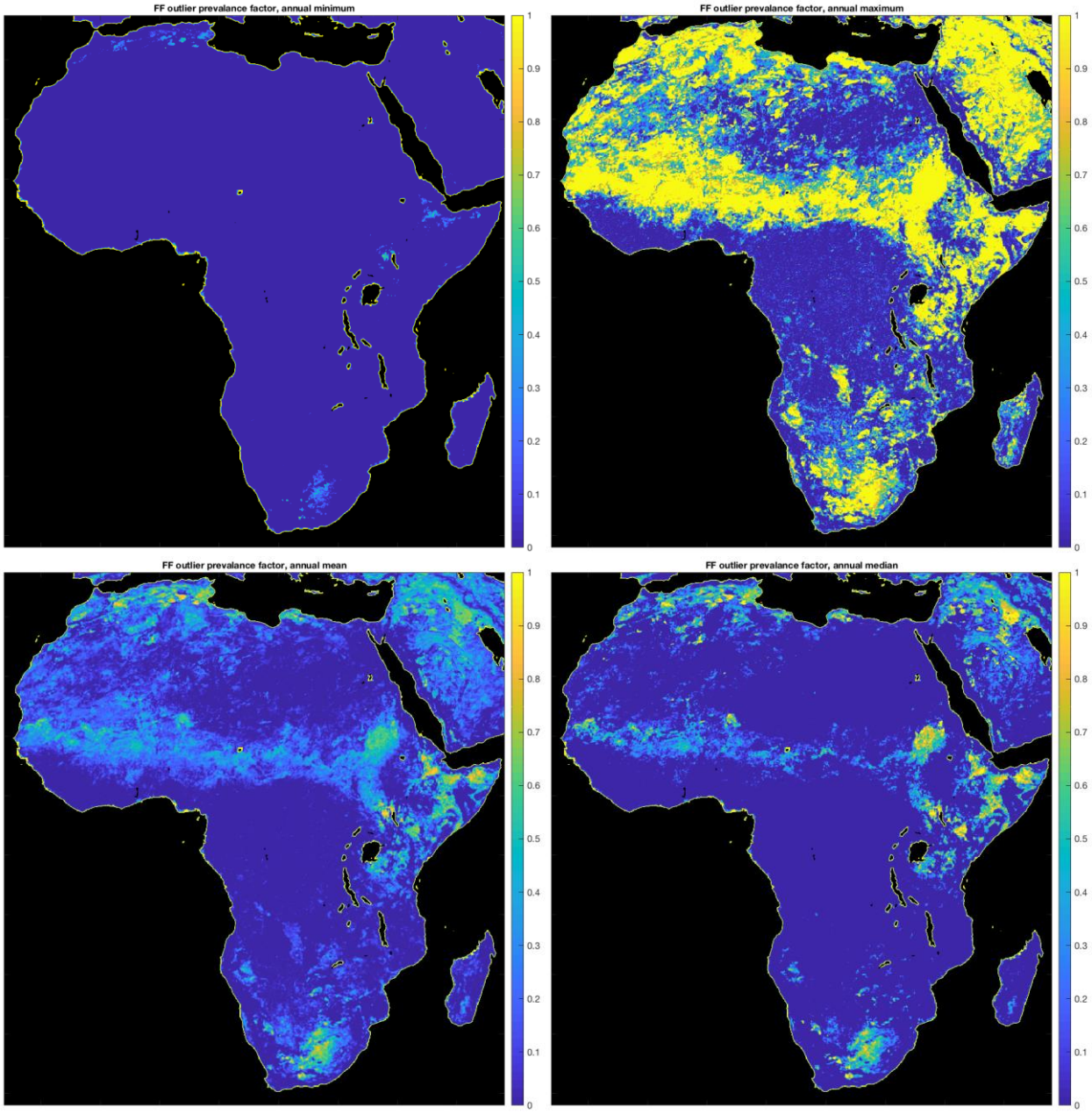

Figure 22: Minimum, maximum, mean, and median annual flooded fraction historical outlier prevalence factor.

4.20 Minimum detectable flooded fraction

The flooded fraction algorithm uses minimum detectable flooded fraction (MDFF) to determine when flooded fraction estimates are too low to be confidently distinguishable from zero. The MDFF derivation algorithm uses a modified version of flooded fraction algorithm, eq. 13, where (a) Q_{obs} is replaced by the $Q_{dc}(d_c,x)$ climatology (section 4.12) plus a local estimate of the magnitude of Q noise (σ_Q) in unflooded conditions and (b) Q_{dry} is set to the $Q_{dc}(d_c,x)$ climatology. The algorithm uses this method to compute $f_f(\sigma_Q)$ for each day-of-year.

The algorithm estimates σ_Q using a 7-day Q variance climatology representing typical Q variability in terrain in the absence of flooding: variance is calculated for each pass using Q data over 7-day windows with rain-, frozen-, and flash-flagged data removed; the mean variance for each day-of-year is calculated over all calibration years; the mean of ascending and descending passes is computed; a 3x3 mean filter is applied away from coastlines to reduce spatial noise; an 11x11 median spatial filter is applied away from coastlines to minimize the effect of flood plain variability; and 31-day averages are calculated to arrive at σ_Q. The algorithm defines points as near-coastline if $f_{pow}>0.01$ and there are at least three points with $f_{pow}>0.75$ within the 11x11 filter area. The algorithm does not apply spatial filtering to near-coastline points to avoid underestimating variance caused by day-to-day changes in true footprint-weighted persistent water fraction attributable to imperfections in composite footprint shape and location knowledge.

We calculate MDFF from $f_f(\sigma_Q)$ beginning with $MDFF_0 = 3f_f(\sigma_Q)$. In areas with $f_{pow}<0.49$, we cap $MDFF_0$ at a maximum value of 0.25. We compute final MDFF as the average of $MDFF_0$ and the annual maximum value of $MDFF_0$. With this approach, MDFF approximates the FF detectable above at least 3-sigma noise.

4.21 Flooded fraction threshold

The downscaling algorithm uses the flooded fraction threshold (f_{f0}) database to derive 3″ flood maps from microwave flooded fraction (f_f) estimates retrieved on circular Gaussian-like footprints. We derived the f_{f0} database from relative floodability.

4.21.1 *Relative floodability calculation*

Relative Floodability (RF) is a static database defined on the AFED 3-arcsecond grid. The inputs to the algorithm that derives RF are the MERIT DEM and its associated flow direction and accumulation area, the AFED persistent open water mask, and the Global Surface Water Annual Water Recurrence dataset (Pekel et al., 2016). The primary goal of RF is to relate all points on the AFED 3-arcsecond grid to a downhill point on a streamline under the assumption that the lower a point's elevation is relative to a floodable river channel the more likely it is to be flooded during a flood event. As a result, RF is primarily a function of a grid point's elevation above a channel as measured in the downhill direction. A secondary RF feature is the inclusion of historical water recurrence data to increase the flood likelihood prediction where the DEM-based method would otherwise underestimate it.

The steps for computing RF are as follows.

1. We define a two-tiered network of stream points on the grid to serve as the effective *starting* points for flooding (i.e., points most likely to flood during an event). We define large and small streams as grid points with accumulation area greater than 5000 and 100 km^2, respectively. This stream network is not dependent on the AFED persistent open

water mask and as a result the algorithm maps floodable river channels in dry areas based on topography alone.

2. We apply a one-dimensional median filter to the DEM flow-wise along the floodable river channels after applying a 5x5 minimum filter to the DEM to ensure the elevations along the stream align with the actual lowest local point. It is especially important to remove noise along channels because it is propagated through RF to all upstream points. The filter has length of 150 grid points, which we found to remove pixel level noise along channels without significantly sacrificing representation of the underlying terrain.
3. Using the results of the previous steps, we initially compute relative floodability for most points as the difference in elevation between each grid point and the closest channel (or ocean), with distance measured along the flow direction. For points that are part of the stream network, the initial RF is the difference between the grid point elevation prior to and after step 2. We compute initial RF separately for the large and small stream networks and then take a weighted average of the two results, with large stream RF weighted 0.75 and small stream RF weighted 0.25. We then apply a 21x21 point spatial filter over the initial RF that computes the best fit linear relationship between it and the DEM and then adjust RF to reduce ambiguities at drainage basin boundaries.
4. We override the initial RF for grid points that have non-zero Annual Water Recurrence values or are included in the persistent open water mask.
    a. Global Surface Water: The Global Surface Water dataset provides Annual Water Recurrence as observed by Landsat, with values ranging from water observed 100% of years to 0% of selected years (Pekel et al., 2016). Points with a Recurrence value greater than zero are assigned relative floodability proportionally, e.g. a point with a value of 90 has a higher relative floodability than a point with a value of 50. All points with a Recurrence value of 0 are then assigned relative floodability by methods b-d, and always have a lower relative floodability than points with a Recurrence value greater than 0.
    b. Persistent open water: Relative floodability is set to a null value. The AFED algorithm treats areas treats these grid cells as permanently 100% water covered.

4.21.2 *Flooded fraction threshold calculation*

We derived the flooded fraction threshold, f_{f0}, from RF and the microwave footprint weighting function. Flooded fraction threshold represents the minimum footprint-weighted flooded fraction at which the RF map indicates that a point must be flooded to produce a high-resolution flood depiction consistent with both RF and flooded fraction. That is, given a satellite f_{flood} estimate interpolated to grid point (i,j), $f_{flood}(i,j) \geq f_{f0}(i,j)$ indicates that grid point (i,j) is likely to be flooded. We derived f_{f0} by calculating the footprint-weighted fractional area of points with RF≥RF(i,j) for the footprint centered on point (i,j).

As described in section 3.3, the flooded fraction algorithm uses footprint matching to regularize microwave footprints to a circular Gaussian-like shape with half-peak diameter specific to each sensor. This allows us to neglect footprint orientation when calculating f_{f0}, which is a computationally intensive task due to the grid resolution and the need to evaluate a ~4,000-km^2 area around each grid point to capture the bulk of the footprint weight. The f_{f0} computation is performed using a general-purpose graphical processing unit (GPGPU) computation method.

The GPGPU method reduces (by a factor of ~1,400 compared to conventional computing in one test) the processing time required to calculate f_{f0} for every 3-arcsecond cell in the AFED grid.

5 References

Alsdorf, D. E., E. Rodrıguez, and D. P. Lettenmaier (2007), Measuring surface water from space, *Rev. Geophys.*, 45, RG2002, doi:10.1029/2006RG000197.

Backus, G., and F. Gilbert (1970), Uniqueness in the inversion of inaccurate gross earth data, *Phil. Trans. Roy. Soc. London*, A266:123-192.

Berg, W., M. Sapiano, J. Horsman, and C. Kummerow (2013), Improved Geolocation and Earth Incidence Angle Information for a Fundamental Climate Data Record of the SSM/I Sensors, *IEEE Transactions on Geoscience and Remote Sensing*, 51, 1504-1513.

Carroll, M., Townshend, J., DiMiceli, C., Noojipady, P., Sohlberg, R. 2009. A New Global Raster Water Mask at 250 Meter Resolution. *Int. J. Digital Earth*, 2, 4, 291-308, 2009.

Eluskiewicz, J., J. F. Galantowicz, T. Nehrkorn, R. Hoffman, A. Lipton, and J. L. Moncet (2005), *Algorithm Theoretical Basis Document (ATBD) for the Conical-Scanning Microwave Imager/Sounder (CMIS) Environmental Data Records, Volume 1: Overview, Part 3: Imagery EDR and Gridding*, P1130-TR-I-ATBD-OVERVIEW-IMG-V1.6-20050419, AER, Lexington, MA. Available at http://www.aer.com/scienceResearch/mwrs/pubs.html.

Farr, T.G., P.A Rosen, E. Caro, R. Crippen, R. Duren, S.Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth and Others, "The shuttle radar topography mission," *Rev. Geophys*., 45, 2, 2007. doi:10.1029/2005RG000183. Data available on-line from: https://lpdaac.usgs.gov/products/measures_products_table.

Ferraro, R. R., F. Weng, N. C. Grody, and A. Basist (1996), An eight-year (1987-1994) time series of rainfall, clouds, water vapor, snow cover, and sea ice derived from SSM/I measurements, *Bull. Am. Met. Soc.*, 77(5):891-905.

Galantowicz, J. F. (2001), *Algorithm Theoretical Basis Document (ATBD) for the Conical-Scanning Microwave Imager/Sounder (CMIS) Environmental Data Records, Volume 11: Vegetation/Surface Type EDR*, AER Document P757-TR-I-ATBD-VST-20010315, AER, Lexington, MA. Available at http://www.aer.com/scienceResearch/mwrs/pubs.html.

Galantowicz, J. F. (2002), High-resolution flood mapping from low-resolution passive microwave data, Proc. of IGARSS'02, Toronto, Canada, June 24-28, III:1499-1502.

Galantowicz, J. F. (2004), Metrics for the quality of footprint-matched passive microwave measurements, *Proc. of IGARSS'0*4, Vol. 6, pp. 3762-3766.

Galantowicz, J. F. and A. W. England (1991), *The Michigan Earth Grid: Description, Registration Method for SSM/I Data and Derivative Map Projections.* Technical Report 027396-2-T. Radiation Laboratory Dept. of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor.

Galantowicz, J. F., T. Nehrkorn, R. Hoffman, A. Lipton (2003), *Algorithm Theoretical Basis Document (ATBD) for the Conical-Scanning Microwave Imager/Sounder (CMIS) Environmental Data Records, Volume 1: Overview, Part 2: Footprint Matching*, P1130-TR-I-ATBD-OVERVIEW-FM-20030319, AER, Lexington, MA. Available at http://www.aer.com/scienceResearch/mwrs/pubs.html.

Galantowicz, J. F., J. Picton, and B. Root (2018), ARC Flood Extent Depiction Algorithm Performance Document: AFED Version V05R00, *AER document P1908-AFM-APD-V05R00-R00*, AER, Lexington, MA, USA.

Gallant, J.C., and A. Read (2009), Enhancing the SRTM Data for Australia, *Proc. Of Geomorphometry 2009,* August 31-Sepetember 2.

Gu, H. and A. W. England (2007) AMSR-E data resampling with near-circular synthesized footprint shape and noise/resolution tradeoff study, *IEEE Trans. Geosci. Rem. Sens.*, vol. 45, no. 10, pp.3193-3203.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." Science 342 (15 November): 850–53. Data available on-line from: http://earthenginepartners.appspot.com/science-2013-global-forest.

Hilburn, K. A., F. J. Wentz (2008), Intercalibrated Passive Microwave Rain Products from the Unified Microwave Ocean Retrieval Algorithm (UMORA). *J. Appl. Meteor. Climatol.*, 47, 778–794.

Kalnay, E., M. Kanamitsu, and W.E. Baker (1990), Global numerical weather prediction at the National Meteorological Center, *Bull. Amer. Meteor. Soc.*, *71*, 1410-1428.

Kanamitsu, M., Description of the NMC global data assimilation and forecast system, (1989) *Wea. and Forecasting*, *4*, 335-342.

Kummerow, Christian D., Wesley K. Berg, Matt R. P. Sapiano, Jennifer Horsman, Darren McKague, and NOAA CDR Program (2013): NOAA Fundamental Climate Data Record (FCDR) of Microwave Brightness Temperatures from the SSM/I and SSMIS Sensors, CSU Version 1.0. SSMI_FCDR. NOAA National Climatic Data Center. doi:10.7289/V5CC0XMJ, downloaded 2016.

Miliaresis G, Delikaraoglou D. (2009) Effects of Percent Tree Canopy Density and DEM Misregistration on SRTM/NED Vegetation Height Estimates. *Remote Sensing*, 1(2):36-49.

Moncet, J.-L., P. Liang, J. F. Galantowicz, A. E. Lipton, G. Uymin, C. Prigent, and C. Grassotti (2011), Land surface microwave emissivities derived from AMSR-E and MODIS measurements with advanced quality control, *J. Geophys. Res.*, 116, D16104.

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. 1980. *NCEP FNL Operational Model Global Tropospheric Analyses, July 1976 to April 1997*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. http://dx.doi.org/10.5065/D6K07288. Accessed 02 June 2016.

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. 1997. *NCEP FNL Operational Model Global Tropospheric Analyses, April 1997 through June 2007*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. http://dx.doi.org/10.5065/D6FB50XD. Accessed 02 June 2016.

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. 2000, updated daily. *NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999*. Research Data Archive at the National

Center for Atmospheric Research, Computational and Information Systems Laboratory. http://dx.doi.org/10.5065/D6M043C6. Accessed 02 June 2016.

Papa, F., C. Prigent, C. Jimenez, F. Aires, and W. B. Rossow (2010), Interannual variability of surface water extent at global scale, 1993-004, in press, *J. Geophys. Res*

Pekel, J., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418-422. doi:10.1038/nature20584

Poe, G. A. (1990), Optimum interpolation of imaging microwave radiometer data, *IEEE Trans. Geosci. Rem. Sens.*, 28(5):800-810.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, 1986.

Prigent, C., E. Matthews, F. Aires, and W. B. Rossow (2001a), Remote sensing of global wetland dynamics with multiple satellite data sets, *Geophys. Res. Let.*, 28 , 4631-4634.

Prigent, C., F. Papa, F. Aires, W. B. Rossow, and E. Matthews (2007), Global inundation dynamics inferred from multiple satellite observations, 1993-2000, *J. Geophys. Res*., 112, D12107, doi:10.1029/2006JD007847.

Schroeder, R., M. A. Rawlins, K. C. McDonald, E. Podest, R. Zimmermann, and M. Kueppers (2010), Satellite microwave remote sensing of North Eurasian inundation dynamics: development of coarse-resolution products and comparison with high-resolution synthetic aperture radar data, *Environ. Res. Lett.*, 5, doi:10.1088/1748-9326/5/1/015003.

Sippel, S. J., S. K. Hamilton, J. M. Melack, and B. J. Choudhury (1994), Determination of inundation area in the Amazon River floodplain using the SMMR 37 GHz polarization difference, *Remote Sens. Environ.*, 48:70-76.

Stogryn, A. (1978), Estimates of brightness temperatures from scanning radiometer data, *IEEE Trans. Ant. and Prop.*, AP-26(5):720-726.

Van Leer, K. W., and J. F. Galantowicz (2011), Comparisons of flood affected area derived from MODIS and Landsat imagery. *Tenth Annual AMS Student Conference,* Seattle, WA, American Meteorological Society, S29.

VIP Lab. (2011), *Long Term Average Vegetation Index, Version 3*. Vegetation Index and Phenology Lab., The University of Arizona (VIP Lab.). Available online at http://vip.arizona.edu.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., Oloughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., Bates, P. D. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters, 44*(11), 5844-5853. doi:10.1002/2017gl072874

Zhu, Z., J. Bi, Y. Pan, S. Ganguly, A. A, L. Xu, A.Samanta, S. Piao, R. R. Nemani, R. B. Myneni (2012), Global Data Sets of Vegetation LAI3g and FPAR3g derived from GIMMS NDVI3g for the period 1981 to 2011, *Remote Sens.*, *5*(2), 927-948; doi:10.3390/rs5020927.

APPENDIX

A. Data use policies

1. SSM/I

SSM/I data is distributed by NOAA and are covered by the following use agreements:

- RSS-calibrated SSM/I: http://www1.ncdc.noaa.gov/pub/data/sds/cdr/CDRs/SSMI_S_Brightness_Temperatures_RSS/UseAgreement.pdf
- CSU-calibrated SSM/I: http://www1.ncdc.noaa.gov/pub/data/sds/cdr/CDRs/SSMI_S_Brightness_Temperatures_CSU/UseAgreement.pdf

Both use agreements include the following statement: "[T]he CDR data sets are non-proprietary, publicly available, and no restrictions are placed upon their use."

2. AMSR-E

AMSR-E data is distributed by NASA. The overarching NASA Data and Information Policy can be found at http://science.nasa.gov/earth-science/earth-science-data/data-information-policy/. There is no restriction on commercial use.

3. AMSR2

Section 6 of the Terms of Use for the GCOM-W1 Data Providing Service (http://gcom-w1.jaxa.jp/useagreement.html) states:

Ownership of Data etc

The copyrights of the standard products and other materials provided in the Service are the property of JAXA.

Please use them in compliance with the conditions stipulated in "Scope and Conditions for Use of the Contents of the Site" in the JAXA's site (http://global.jaxa.jp/policy.html).

Any user can utilize the data for commercial purposes without an agreement of intellectual property utilization authorization with JAXA, and without royalties.

For details, Please refer to "Implementation Guideline for GCOM-W1 Data Provision" in the JAXA's site.(http://gcom-w1.jaxa.jp/contents/GCOM-W1_data_prov_guideline_en.doc)

4. GMI and TMI

The TRMM and GPM Data Policy (https://pmm.nasa.gov/data-access) states:

TRMM and GPM data are freely available at all levels for which the particular sensor or sensor combination has been processed by GPM. For the GPM Core Observatory this is for Levels 0 through 3 products (as applicable).

The AFED model uses Level 1 TMI and GMI data from the GPM Core Observatory covered under this clause. The overarching NASA Data and Information Policy can be found at http://science.nasa.gov/earth-science/earth-science-data/data-information-policy/. There is no restriction on commercial use.

5. GFS and NCEP

GFS and NCEP data are distributed through the NOAA National Operational Model Archive and Distribution System (NOMADS, http://nomads.ncep.noaa.gov/). NOMADS is covered by the National Weather Service terms of data usage (http://www.weather.gov/disclaimer), which states that the information is in the public domain with no restriction for commercial use:

> *The information on National Weather Service (NWS) Web pages are in the public domain, unless specifically noted otherwise, and may be used without charge for any lawful purpose so long as you do not: 1) claim it is your own (e.g., by claiming copyright for NWS information -- see below), 2) use it in a manner that implies an endorsement or affiliation with NOAA/NWS, or 3) modify its content and then present it as official government material. You also cannot present information of your own in a way that makes it appear to be official government information.*

B. Near real time operations

This section provides information useful for monitoring near real time operations and planning future modes of operation, including: contacts and email alerts for NRT data providers; AMSR2 and GMI data latency statistics; AFED delivery timeliness; logic for triggering automatic recovery of the definitive AFM AFED product stream when expected sensor data are delayed; long term archive scheme; NRTPS failure alert email procedures; and an analysis of NRTPS-HPS flooded fraction differences attributable to NRTPS use of GFS data for atmospheric correction in place of NCEP reanalysis data.

Table 23: Contact and email alert details for NRT data providers

Data type	Data URL and contact	Email alerts
AMSR2	URL: https://gcom-w1.jaxa.jp/auth.html Contact: GCOM-W Data Providing Service Help Desk 2-1-1, Sengen, Tsukuba-city, Ibaraki 305-8505 Japan Business hours (JST): Monday - Friday, 10:00 - 12:15/13:00 - 17:45 TEL:+81-50-3362-6599 FAX:+81-29-859-5574	All registered users of GCOM-W1 Data Providing Service received email alerts. Use the URL to register.
GMI	URL: http://disc.gsfc.nasa.gov/ Contact: Name: GES DISC Help Desk E-mail: gsfc-help-disc@lists.nasa.gov Phone: 301-614-5224 Fax: 301-614-5228 Address: Goddard Earth Sciences Data and Information Services Center Attn: Help Desk Code 610.2 NASA Goddard Space Flight Center Greenbelt, MD 20771, USA	Register here for notifications from NASA Precipitation Processing System (PPS): https://registration.pps.eosdis.nasa.gov/registration/
GFS	URLs: 1) https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs 2) http://www.emc.ncep.noaa.gov/index.php?branch=GFS Contact: For NOAA NOMADS systems questions and problems: http://nomads.ncep.noaa.gov/	Register for GFS service change notices here: http://www.nws.noaa.gov/os/notif.htm Subscribed to NOMADS data service announcements here: https://www.lstsrv.ncep.noaa.gov/mailman/listinfo/ncep.list.nomads-ftpprd

Table 24: AMSR2 and GMI data latency statistics, 2016/12/1-2016/12/7

Sensor	Latency from time of final sample in granule [hours]			
	Minimum	Maximum	Mean	Outliers
AMSR2	9.2	12.7	10.7	None
GMI	3.9	13.5	5.4	6% > 6.5 hours

We tested AFED delivery timeliness during the V03R01 NRTPS test period from 2016/11/3 to 2016/12/12. Measuring delivery from 0000 UTC at the end of the product date, the production latency was 52-55 hours for AMSR2 and 55-61 hours for GMI. For consistency with the HPS, the V03R01 NRTPS used data observed up to two days after the end of the product date, which explains a large part of the delay. As of V05R00, the algorithm does not require data from days after the product date. With this change, the production latency could decrease by about 48 hours relative to V03R01, e.g., to about 4-7 hours for AMSR2 and 4-15 hours for GMI.

As of V05R00, the NRTPS delivers a single *definitive* AFM AFED product from a combined AMSR2 and GMI processing stream. NRTPS implements the following logic for populating the definitive AFED data stream from the combined AMSR2 and GMI stream:

- NRTPS will wait for AMSR2 and GMI completion for a maximum period to be defined by ARC (e.g., now 48 hours beyond the end of the product date, or about 17-20 hours beyond the maximum typical GMI latency noted above). This period is set in concert with the period designated for the NRTPS automatic recovery process (described in section 2.7).
- Within this period, NRTPS will deliver the definitive AFM products as soon as combined AMSR2 and GMI production is complete (meaning that the NRTPS has incorporated all or, after a shorter waiting period, nearly all swaths expected for use in the product).
- If combined AMSR2 and GMI processing has not completed, then the NRTPS automatic recovery feature will be triggered for whichever sensor is incomplete. Recovering processing for the combined streams should be complete about three hours after automatic recovery is trigger, at which time NRTPS will ship the AFM definitive product.

As a part of the definitive AFED product, NRTPS will continue to provide separate AMSR2 and GMI QC metadata for informational and diagnostic purposes.

NRTPS currently archives the AFED and metadata products from the definitive AFM feed on the FTP site where ARC accesses the data (ftp://54.175.148.68). On the NRTPS server, NRTPS implements the following storage clearing plan:

- Cleared after 70 days: AFED and metadata product files, flooded fraction intermediate and final process files, and gridded data files needed for 61-day retrospective median calculations (discussed in section 3.8).
- Cleared after 10 days: All other intermediate process files (e.g., downloaded sensor and atmospheric data files, converted sensor data files, footprint matched files).

NRTPS also archives all of the items now cleared after 70 days on the AWS S3 (Simple Storage Service). Storing some intermediate files—and particularly flooded fraction process results—in addition to AFED and metadata product files will aid diagnostics and troubleshooting. Items now cleared after 10 days are more space consuming and need not be archived because they are easily replicated if needed.

In addition to the product delivery emails described in section 2.7, NRTPS will also send email alerts to designated recipients when certain failures occur. We configured email alerts using the AWS CloudWatch service for the NRTPS, FTP, and GFS Provider server instances. CloudWatch is a monitoring service for AWS cloud resources accessed through the EC2 console. For example, CloudWatch will check system status every few minutes and send an alert email when system status checks fail. NRTPS routinely handles missing data and related processing delays through repeated attempts at executing each point in the processing stream. As a result, NRTPS will send internal email alerts only when unusual delays are encountered, e.g., when triggering processing with fewer than the total number of expected swaths or when 6-hourly GFS files are delayed. We will add more CloudWatch and NRTPS internal alerts over time as experience with the system exposes other events that would be of value to track.

The primary functional difference between the NRTPS and HPS is that the NRTPS uses GFS data for atmospheric correction while HPS uses NCEP reanalysis data. We tested the effect of this difference by comparing flooded fraction results from the two systems for the NRTPS test period, 2016/11/3 to 2016/12/12. The following table provides a compilation of the results for the AMSR2 and GMI data streams. Statistics are for the absolute value of the flooded fraction difference, $|\Delta FF|$, for coincident points on the same day. The statistics are divided in two groups,

one covering all points and one specific to the subset of points where flooded fraction was above zero in both the NRTPS and HPS results (FF>0).

Table 25: Comparison of NRTPS and HPS flooded fraction

Data type	Subset	|ΔFF|		% |ΔFF|>3σ (outliers)
		Mean	Standard deviation[1], σ	
AMSR2	All points	8.3E-6	5.4E-4	0.03%
	FF>0	3.8E-4	9.0E-4	0.58%
GMI	All points	3.8E-5	1.5E-3	0.10%
	FF>0	1.3E-3	6.3E-3	1.35%

[1]Root mean square error and standard deviation results were approximately equal

The vast majority of points at any time have zero flooded fraction because the MDFF limit resets small FF results to zero. As a result, statistics using all points reflect the degree to which NRTPS and HPS agree on the areas with and without any flooding; for both AMSR2 and GMI this agreement is extremely good. In areas where FF>0, differences are higher but still extremely small compared, for example, to MDFF, which is typically greater than 0.05. The occurrence of outliers indicates non-linear processes in the algorithm (such as the false positive detection process) that can amplify the effect of a small change (like that of the atmospheric correction) by, for example, changing a point from flooded to unflooded. Our overall assessment is that the effect of the GFS-NCEP difference is small with respect to the rate at which the NRTPS and HPS detect and mapping of large, long-lasting floods.