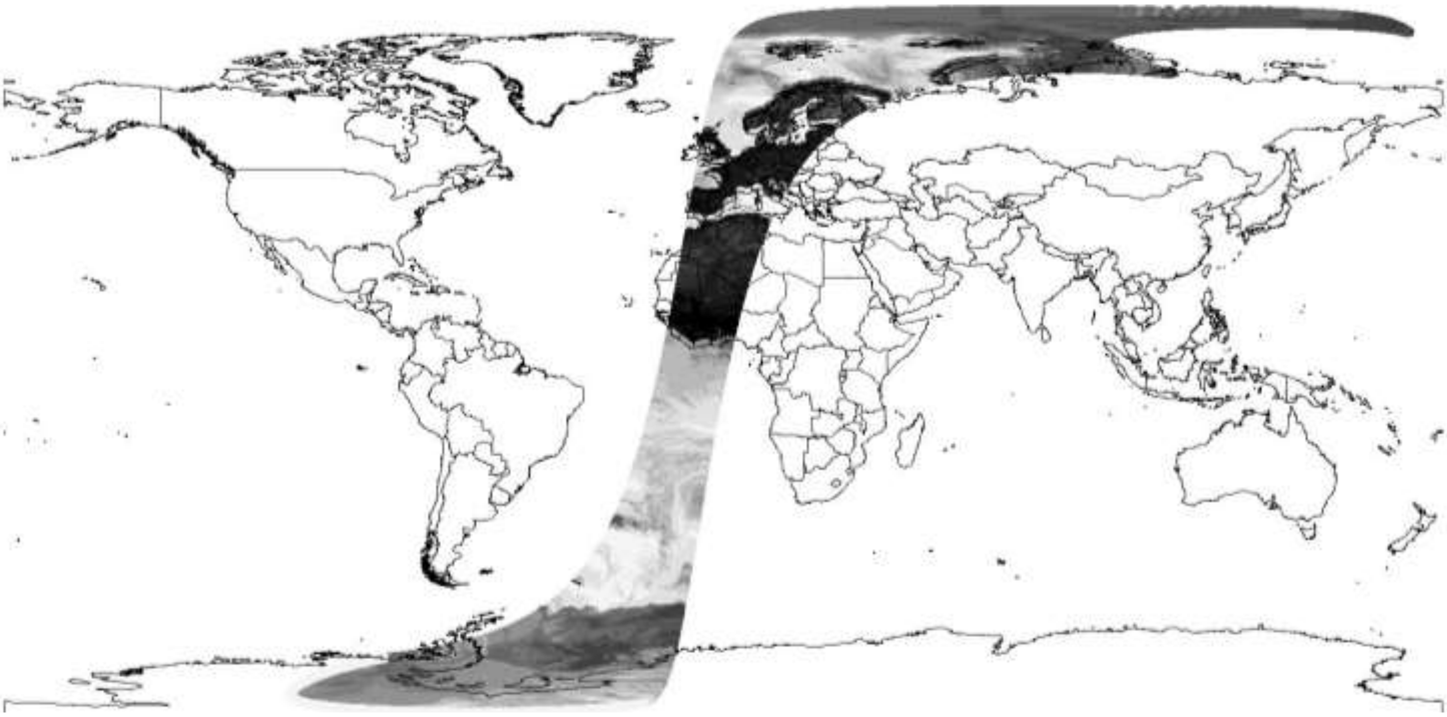
European Commission

JRC TECHNICAL REPORT



Global Flood Detection System: Data Product Specifications

Version 2015

Tom De Groeve, Robert G. Brakenridge, Stefano Paris

2015

Report EUR xxxxx xx

Joint Research Centre

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Tom De Groeve
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 680, 21027 Ispra (VA), Italy
E-mail: tom.de-groeve@jrc.ec.europa.eu
Tel.: +39 0332786340

JRC Science Hub
https://ec.europa.eu/jrc

JRC97421

EUR xxxxx xx

ISBN xxx-xx-xx-xxxxx-x (print)
ISBN xxx-xx-xx-xxxxx-x (PDF)

ISSN xxxx-xxxx (print)
ISSN xxxx-xxxx (online)

doi:xx.xxxx/xxxxx

Luxembourg: Publications Office of the European Union, 2015

Printed in Italy

Abstract
The Global Flood Detection System (GFDS) is an operational flood monitoring system developed and run at the Joint Research Centre of the European Commission in collaboration with the Dartmouth Flood Observatory (Colorado University). Since 2006, GFDS calculates water surface metrics from brightness temperatures recorded at 36.5 GHz in the H-polarization. Full time series of four sensors are used (TRMM, GPM, AMSR-E and AMSR2), and are converted in a range of raster data products showing flood signal and magnitude, as well as time series (virtual hydrographs) for around 10000 observation sites. This technical note describes the specifications of the GFDS data products as a guide for scientists and practitioners to facilitate the exploitation of this unique data source in new research and applications.

Introduction

Floods are difficult to monitor on a global scale, because they are determined by local factors such as precipitation, slope of the terrain, drainage of the river, protection devices in place, etc. Each river must be monitored at different places along its course. Although some flood disasters occur annually, most happen unexpectedly. All rivers must therefore be monitored, and along their full course. The number of rivers in the world is hard to determine, but even the Digital Chart of the World (Danko 1992), a database at a scale of 1:1 million which shows only major rivers, has close to 1 million records, with a total length of 10 million km. Unlike for earthquakes where few measuring stations suffice to monitor the globe (the United States Geological survey global Seismographic network has less than 150 stations outside America), an in situ global flood monitoring system would need a dense network of gauging stations along all rivers. However, such stations are expensive (the United States Stream gauging Network costs US$89 million per year; USGS 1998), which makes this hardly feasible on a global scale (De Groeve, 2010).

In situ measurements can be replaced by remote sensing measurements, from airplanes or satellites. The change in surface water extent can be extracted from aerial or satellite imagery. While the use of sensors in the visible or infrared portion of the spectrum is limited due to cloud cover, the microwave portion of the spectrum can penetrate clouds. However, for most remote sensing solutions, the revisit frequency (i.e. the time between two measurements in the same place) is too low for monitoring purposes, or the spatial coverage is limited.

For satellite based imagery, the revisit time depends on the orbit and the image size, and several sensors have been launched since the late nineties having a daily revisit time and global coverage, and provide microwave data in near-real time free of charge. These are the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) instrument on board of the NASA EOS Aqua satellite (2002 to 2011), the Microwave Imager on board of the Tropical Rainfall Monitoring Mission (TRMM) (1997-2015), the Advanced Microwave Scanning Radiometer 2 (2013-ongoing) and the GPM Microwave Imager on board of the Global Precipitation Mission (2014-ongoing).

The Global Flood Detection System (GFDS) is an operational flood monitoring system developed and run at the Joint Research Centre of the European Commission in collaboration with the Dartmouth Flood Observatory (Colorado University). Brakenridge et al. (2005) demonstrated that AMSR-E can measure river discharge changes in various climatic conditions. Using AMSR-E data, De Groeve et al. (2007) developed a method for detecting major floods on a global basis in a systematic, timely and impartial way appropriate for humanitarian response.

The GFDS data and products have been used in many scientific studies, and several operational flood monitoring applications are using the live data.

This technical note describes the specifications of the GFDS data products as a guide for scientists and practitioners to facilitate the exploitation of this unique data source in new research and applications.

GFDS water surface metrics

The Global Flood Detection System (GFDS) calculates water surface metrics from brightness temperatures recorded at 36.5 GHz in the H-polarization. If the physical temperature remains constant, changes in brightness temperature will be related to changes in water surface extent in the pixel.

$$T_b = (1-w)T_{b,land} + wT_{b,water} \qquad (1)$$
$$w = \text{water portion of the pixel}$$

However, in spite of the great radiation dissimilarity of water and land cover, the raw brightness temperature observations cannot be used to reliably detect changes in surface water area. This is because brightness temperature (T_b) measures are influenced by other factors such as physical temperature, differences in emissivity and atmospheric moisture. While the relative contribution of these factors cannot be measured, the authors assume them to be constant over a larger area. As shown in equation 2, the ratio between two nearby pixel values is a function of *w* alone. Therefore, by comparing a "wet pixel" received over a river channel of a potential inundation location ($w > 0$) with a "dry pixel" without water cover ($w = 0$) the mentioned noise factors can be reduced. The brightness temperature values of the measurement/wet signal are divided by the calibration/dry observations, referred to as *M/C* ratio or signal *s*.

$$T_{b,measurement} = (1-w)T_{b,land} + wT_{b,water} = T_{measurement}\left((1-w)\varepsilon_{land} + w\varepsilon_{water}\right) \qquad (2)$$
$$T_{b,calibration} = T_{b,land} = T_{calibration}\varepsilon_{land}$$

If for nearby pixels we assume

$$\varepsilon_{land,measurement} \approx \varepsilon_{land,calibration} \approx \varepsilon_{land}, T_{measurement} \approx T_{calibration} \qquad (3)$$

then

$$s = \frac{M}{C} = \frac{T_{b,measurement}}{T_{b,calibration}} = \frac{T_{measurement}\left((1-w)\varepsilon_{land} + w\varepsilon_{water}\right)}{T_{calibration}\varepsilon_{land}} \approx 1 - w + w\frac{\varepsilon_{water}}{\varepsilon_{land}} = f(w) \qquad (4)$$

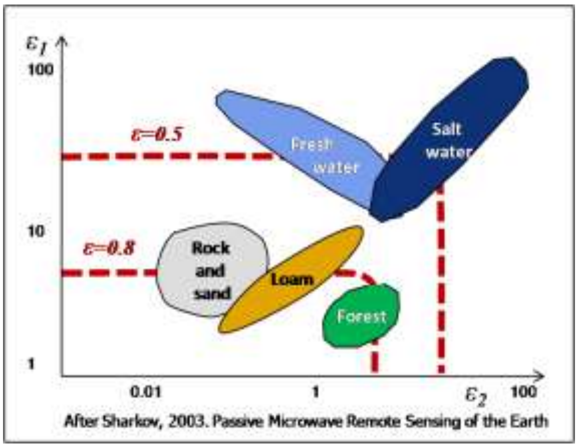


Figure 1. Schematic representation of microwave signatures of water and land.

In order to distinguish between areas with permanent water (e.g. lakes or wide rivers) and areas with flood waters, we look at change in flood signal over time. Based on a time series of 7 years (going back to June 2002 when the satellite was launched), anomalies are automatically detected using a method described in De Groeve *et al.* (2006). Since lower *M/C* signals generally accounts for increased water coverage, extreme events, or major floods, should represent negative anomalies in the time series of a given site. In order to detect anomalies, they first determined the reference value for normal flow, which varies for each site based on the local emissivity properties and river geometry. This reference value was calculated as the average *M/C* value for the site since the launch of the satellite. They then set flood level thresholds based on the statistics of the time series. Flood magnitude was defined as the number of standard deviations (*sd*) from the mean (*avg*):

$$m = \frac{s - avg(s)}{sd(s)}$$

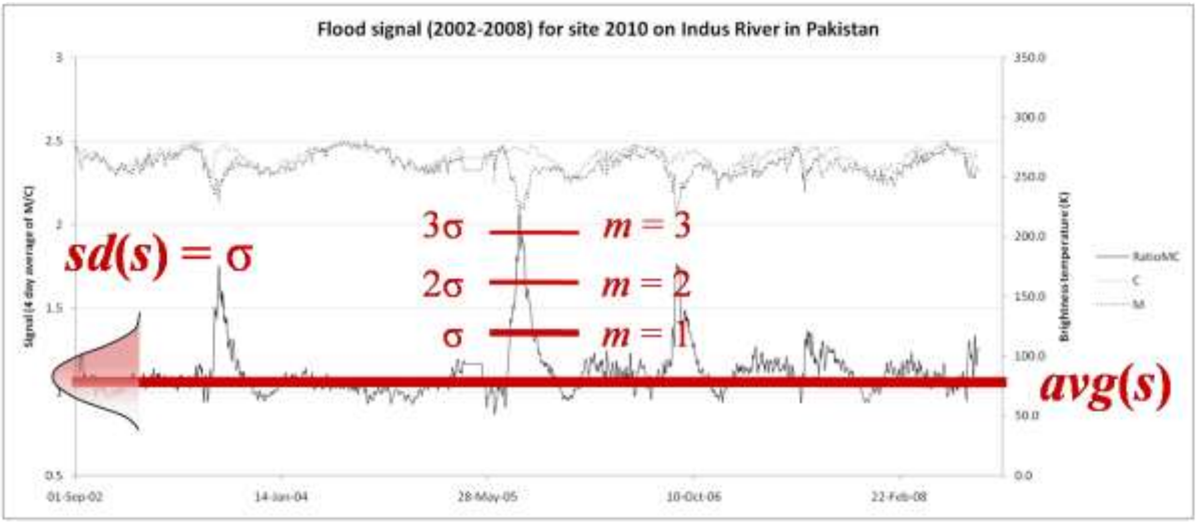


Figure 2. Example time series of M, C and signal (RatioMC) values. The flood magnitude is a statistical measure for the anomaly of the signal.

GFDS satellite-born sensors

GFDS calculates water surface metrics from brightness temperatures recorded at 36.5 GHz, both ascending and descending swaths. Full time series of four sensors constitute the input data (Table 1).

Table 1. GFDS sensors and characteristics

Sensor	Name	Satellite	Characteristics	Comments
AMSR-E	Advanced Microwave Scanning Radiometer - EOS	NASA's Earth Observing System (EOS) Aqua Satellite	2002-2011 Polar orbit, full geographic coverage 36.5 GHz (V, H)	AMSR-E antenna stopped spinning at 07:26GMT Oct 4, 2011
TRMM-TMI	TRMM Microwave Imager	Tropical Rainfall Measuring Mission	1997-2015 37.0GHz (V,H) 40S to 40N	Operations stopped on 15/04/2015
AMSR2	Advanced Microwave Scanning Radiometer 2	GCOM-W	2013-ongoing Polar orbit, full geographic coverage 36.5 GHz (V, H)	Integrated under ERCC1/MIC7
GPM-GMI	GPM Microwave Imager	Global Precipitation Mission	2015-ongoing 37 GHz (V, H) 65S to 65N	Integrated under ERCC2/MIC8

GFDS has a single sensor product ("flooddetection") and a multi-sensor product ("floodmerge"). Satellite observations are processed as soon as they are available at JRC. Lag times vary for different satellites from around 3 hours (AMSR2) to around 24 hours (GPM). GFDS has processed all data for TRMM, AMSR-E, AMSR2 and GPM, covering a time period from December 1997 to now. Observations are converted in raster products on a daily basis and with global coverage, effectively providing water surface metrics with daily frequency for any location in the world.

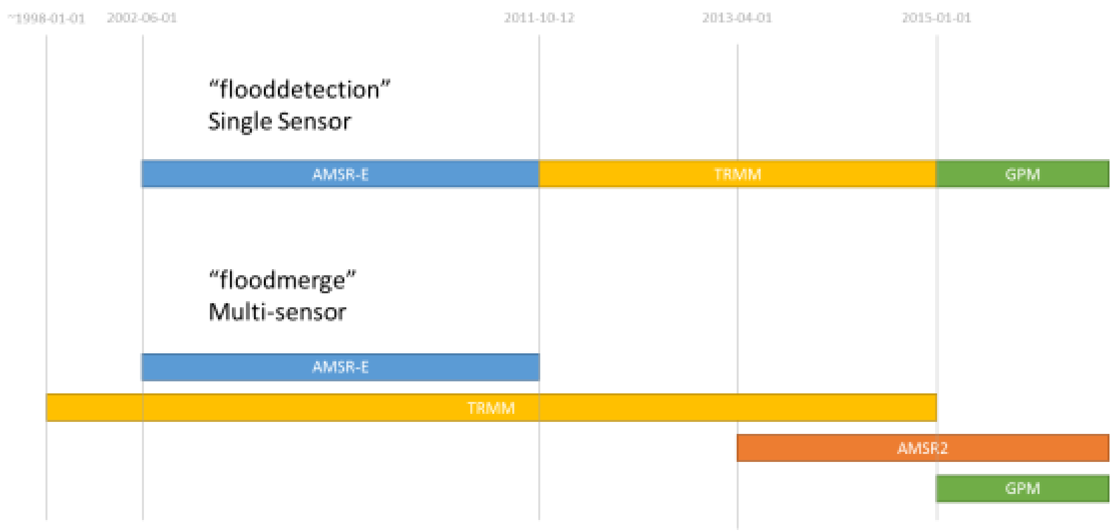


Figure 3. Timeline of GFDS data sources.

GFDS Raster Data Products

Satellite observations are processed as soon as they are available at JRC. Lag times vary for different satellites from around 3 hours (AMSR2) to around 24 hours (GPM). GFDS has processed all data for TRMM, AMSR-E, AMSR2 and GPM, covering a time period from December 1997 to now.

Observations are converted in raster products on a daily basis and with global coverage, effectively providing water surface metrics with daily frequency for any location in the world.

Daily Datasets (single sensor)

For each swath, the GFDS statistics are calculated for each measurement pixel *M*. The statistics and source data (measurement and calibration temperatures) are stored in separate raster files at the pixel location of the measurement pixel *M*. In the database, the values are tagged as "DAILY".

Table 2. Daily Datasets (single sensor)

	Description	Folder Name	Unit	Data type	Scale factor	Aggregation method
T_M	Brightness temperature of measurement pixel	SourceTiffs	K	Int32	100	Last
T_C	Brightness temperature of calibration pixel	CalibrationTiffs	K	Int32	100	Last
P	Relative position of calibration pixel	PositionTiffs	–	Int32	1	Last
s	Flood signal, $s = T_M/T_C$	SignalTiffs	–	Int32	1000000	Average
m	Flood magnitude, or number of standard deviations removed from the mean	MagTiffs	sd	Int32	1000	Average

The position value *P* indicates the relative position of the calibration pixel *C* to the measurement pixel *M*. The calibration pixel *C* is chosen as the 95 percentile of the pixels in a grid of 9by 9 pixels centered on the measurement pixel *M*. (It is not the hottest pixel to exclude outliers due to measurement error.) The position numbers of calibration pixels are listed in the figure below.

1	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	18
19	20	21	22	23	24	25	26	27
28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45
46	47	48	49	50	51	52	53	54
55	56	57	58	59	60	61	62	63
64	65	66	67	68	69	70	71	72
73	74	75	76	77	78	79	80	81

Figure 4. Position numbers of calibration pixels around the measurement pixel (41).

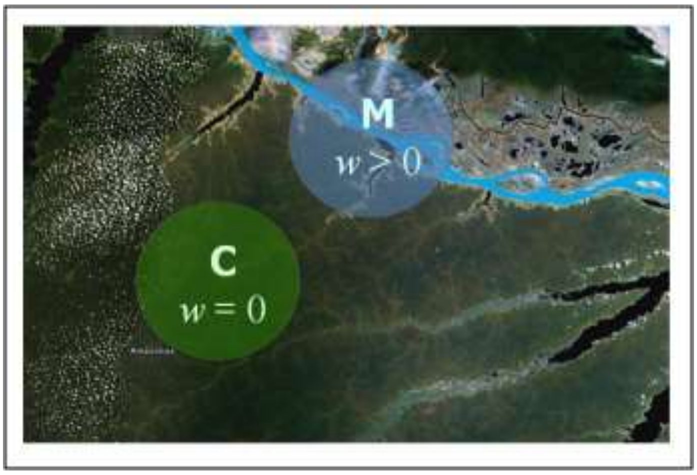


Figure 5. Example of a measurement M and calibration C pixel.

All values are calculated in the swath geometry, and then projected in a global grid of 4000 by 2000 pixels. When multiple samples for one pixel are available in one day, an aggregate value is calculated as follows: for *M*, *C* and *P* values that last sample value; for *s* and *m* values, the average of all samples of the day.

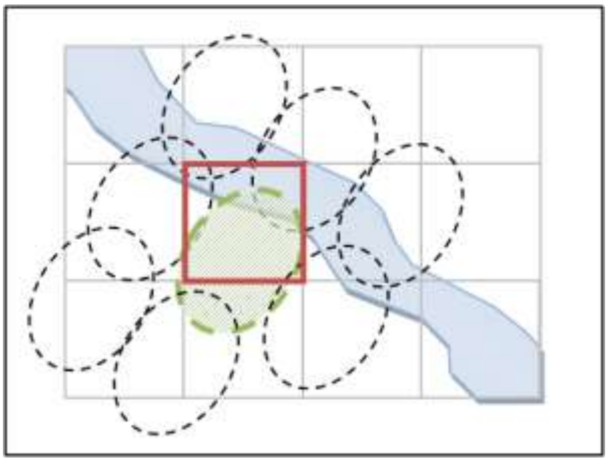


Figure 6. Schematic representation of the projection of swath pixels (ellipses) into a global grid (boxes).

4-day average Datasets (single sensor)

To handle missing data and provide smoothing, a running backwards-looking average of 4 days is produced in raster format. These files are stored in the same location as the daily files, but the folders are prefixed with “Avg”. In the database, the values are tagged with “4DAYS”.

Table 3. 4-day average Datasets (single sensor)

	Description	Folder Name	Unit	Data type	Scale factor	Aggregation method
T_M	Brightness temperature of last measurement pixel	AvgSourceTiffs	K	Int32	100	Last
T_C	Brightness temperature of last calibration pixel	AvgCalibrationTiffs	K	Int32	100	Last
P	Relative position of last calibration pixel	AvgPositionTiffs	–	Int32	1	Last
s	Average flood signal	AvgSignalTiffs	–	Int32	1000000	Average
m	Average flood magnitude	AvgMagTiffs	sd	Int32	1000	Average

The number of samples in four days may vary between 0 and 4, depending on the swath geometry. Typically, all pixels have at least one sample in 4 days. When multiple samples for one pixel are available in one day, an aggregate value is calculated as follows: for *M*, *C* and *P* values that last sample value; for *s* and *m* values, the average of all samples in 4 days.

Merged Daily Datasets (multiple sensors)

GFDS is using multiple sensors as input. The technique of ratioing *M* and *C* values from the same swatch (recorded synchronously) provides a robust surface water metric independent of sensor and time of day. (It is assumed that flood conditions remain stable in a period of 24h.) It is therefore meaningful to integrate data from multiple sensors in a single merged product, mainly to accommodate for missing data due to swath geometry and to reduce measurement noise through increased sampling. However, it is not meaningful to integrate individual brightness temperature measurements across satellites and from different times of day, as they are influenced strongly by the environmental temperature. As only one (*M, C, P*) triplet per day could be stored in a daily product, it is not calculated for the merged products.

For each sensor, the GFDS statistics are averaged and stored in separate raster files at the pixel location of the measurement pixel *M*. In the database, the values are tagged as “DAILY”.

Table 4. Merged Daily Datasets (multiple sensors)

	Description	Folder Name	Unit	Data type	Scale factor	Aggregation method
s	Flood signal, $s = T_M/T_C$	SignalTiffs	–	Int32	1000000	Average
m	Flood magnitude, or number of standard deviations removed from the mean	MagTiffs	sd	Int32	1000	Average

Merged 4-day average Datasets (multiple sensors)

Similarly to the single sensor datasets, also a 4-day average is calculated. This is the product containing most samples per pixel, taking from all available sensors over a 4 day period.These files are stored in the same location as the daily files, but the folders are prefixed with "Avg". In the database, the values are tagged with "4DAYS".

Table 5. Merged 4-day average datasets (multiple sensors)

	Description	Folder Name	Unit	Data type	Scale factor	Aggregation method
s	Average flood signal	AvgSignalTiffs	–	Int32	1000000	Average
m	Average flood magnitude	AvgMagTiffs	sd	Int32	1000	Average

Data access

The GFDS raster data products are available from http://www.gdacs.org/flooddetection/data. There are six folders available. The internal structure of the folders is described above.

Table 6. Overview of data access links for 6 raster products.

Folder name	Content	Comments
ALL	Merged daily and 4-day average datasets 1997-current, updated every 3h	Best dataset for operational applications. Highest sampling rate, global coverage. May have artifacts due to multi-sensor integration. Some known spatial calibration issues exist among sensors (1/2 pixel systematic shift).
SINGLE	Single sensor daily and 4-day average datasets 1997-current, updated every 3h	Best dataset for scientific studies. Lower sampling rate, not always global coverage. No noise from multi-sensor integration. Some known spatial calibration issues exist among sensors (1/2 pixel systematic shift).
GPM	Data from Global Precipitation Mission Updated every 3h	
TRMM	Data from Tropical Rainfall Monitoring Mission No longer updated	
AMSR2	Data from AMSR2 Updated every 3h	
AMSR-E	Data from AMSR-E No longer updated	

GFDS Time Series Data Products

From the GFDS raster products, time series are constructed and stored in a database for effective dissemination through web services. This done for over 11000 locations in the world, selected by JRC, the Dartmouth Flood Observatory and partners. For each location (consisting of one or more pixels), the GFDS statistics are available as time series in a variety of formats: Excel, CSV, HTML and KML.

For sites constituting of more than one pixel, the values of signal and magnitude are averaged. This results in a more stable signal with less noise, in particular if the area is chosen as two or three pixels perpendicular to the river bed. For multi-pixel sites, position, source and calibration values are not provided, unless they refer to the same calibration pixel for all participating pixels.

For each site, the daily (DAILY) and 4-day running average (4DAYS) versions are available.

API location

Version 3

There are two version of the API: one for single sensor data and one for multi-sensor data. The data is derived from various sensors, as shown in the table below (for exact dates, see introduction).

	URL	Period	Sensors
Single sensor	http://www.gdacs.org/flooddetection/data.aspx	2002-now	01/06/2002: AMSR-E 12/10/2011: TRMM 01/01/2015: GPM
Merged product	http://www.gdacs.org/floodmerge/data.aspx	1997-now	1997: TRMM 2002: TRMM + AMSR-E 2011: TRMM 2013: TRMM + AMSR2 2015: AMSR2 + GPM

Version 4

	URL	Period	Sensors
Single and merged sensor	http://www.gdacs.org/flooddetection/data_v2.aspx or (identical) http://www.gdacs.org/floodmerge/data_v2.aspx	2002-now	Specified in API: - DFO - DFOMERGE - GPM

API Query parameters

The API allows to retrieve data for individual observation areas and for a specific time period. Other options specify the output data format and other parameters. The parameters are:

	API	Description	Default value	Required
areaid	All	The unique identifier of the observation area	All	Optional
siteid	All	The DFO site ID (for backwards compatibility). areaid has precedence.	All	Optional
type	All	The output format: • txt: text separated by semi-colon (;) • html: HTML table • rss: GeoRSS format (http://www.georss.org/) • kml: OGC KML (http://www.opengeospatial.org/standards/kml)	txt	Optional
from	All	Start date of extraction	Yesterday	Optional
to	All	End date of extraction	Today	Optional
datatype	All	Daily value or running average for smoothing and accounting for missing data: • DAILY: Daily value (with missing data) • 4DAYS: average of past 4 days (less missing data)	4DAYS	Optional
alertlevel	All	A magnitude based threshold. Magnitude is the number of standard deviation the current signal is above the mean of 2002-2008. Values exceeding the threshold are extracted. For point sites (one pixel): • RED: magnitude > 4 • ORANGE: magnitude > 2 • GREEN: all the rest For area sites (multiple pixels) • RED: more than 20% of pixels have magnitude > 4 • ORANGE: more than 20% of pixels have magnitude > 2 • GREEN: otherwise	RED	Optional
source	V4	The sensor data source • DFO: see single sensor product • DFOMERGE: see merged product • GPM: only GPM samples	n/a	Required

API Output

The API produces the following output in a variety of formats.

Field name	Description	Value lists	Only multi-pixel areas
AreasDataId	Unique data point ID		
AreaId	Area ID		
Country	Country name		
AlertLevel	Anomaly level: GREEN, ORANGE or RED (see table above for explanation)	GREEN ORANGE RED	
Description	Area name or description		
TypePointArea	Type of area: single pixel or multiple pixels	P: Single pixel A: Multiple pixel	
PointsInJsonFormat	List of coordinates of pixels in json format		
PointsNumber	Number of pixels		
BoundingBoxLonMin	Minimum longitude of bounding box	[-180,180]	
BoundingBoxLonMax	Maximum longitude of bounding box	[-180,180]	
BoundingBoxLatMin	Minimum latitude of bounding box	[-90,90]	
BoundingBoxLatMax	Maximum latitude of bounding box	[-90,90]	
Population	Population near location (deprecated)		
River	River near location (deprecated)		
Cities	Cities near location (deprecated)		
Slope	Maximum slope (derived from 1km resolution DEM) (deprecated)		
Xml	List of critical features (deprecated)		
Dams	List of critical features (deprecated)		
Agriculture	Area of agricultural land (deprecated)		
Urban	Area of urban land (deprecated)		
Comments	Comments		
DataType	Daily sample or running average over 4 days	DAILY 4DAYS	
RecordDate	Date in format D/M/YYYY 12:00:00		
RecordDateInteger	Data in format YYYYDDMM		
SignalAvg	For single-pixel areas: signal value For multi-pixel areas: average signal over all pixels		
SignalSd	Standard deviation of signal values		Yes

SignalMin	Minimum signal value of all pixels in the area		Yes
MagnitudeAvg	For single-pixel areas: magnitude value For multi-pixel areas: average magnitude value over all pixels		
MagnitudeSd	Standard deviation of magnitude values		Yes
MagnitudeMax	Maximum magnitude value of all pixels in the area		Yes
MagGreatEqual2Count	Number of pixels with magnitude > 2		Yes
MagGreatEqual4Count	Number of pixels with magnitude > 4		Yes
PixelIndexesMagGE2	Index of pixels exceeding magnitude 2		Yes
PixelIndexesMagGE4	Index of pixels exceeding magnitude 4		Yes
MValue	Brightness temperature of measurement pixel (last value of the day)		
CValue	Brightness temperature of calibration pixel (last value of the day)		
PValue	Position of calibration pixel (last value of the day).	[1-81]	

Note that the signal is not always equal to the ratio of M/C. The signal is the average of all samples of the day, while the M, C, P values are those of the last sample of the day.

References

Brakenridge, G.R., Nghiem, S.V., Anderson, E. and Chien, S. (2005). Space-based measurement of river runoff. Eos, Transactions American Geophysical Union 86: doi: 10.1029/2005EO190001. issn: 0096-3941.

De Groeve, T., Z. Kugler, G. R. Brakenridge, 2007. Near Real Time Flood Alerting for the Global Distaser Alert and Coordination System. Proceedings of the 4th International ISCRAM Conference (B. Van de Walle, P. Burghardt and C. Nieuwenhuis, eds.) Delft, the Netherlands, May 2007, pp.33-40.

De Groeve, T., 2010. Flood monitoring and mapping using passive microwave remote sensing in Namibia. Geomatics, Natural Hazards and Risk Vol. 1, No. 1, March 2010, 19–35. doi:10.1080/19475701003648085.

European Commission
EUR xxxxx XX – Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: Global Flood Detection System: Data Product Specifications

Author(s): Tom De Groeve, Robert G. Brakenridge, Stefano Paris

Luxembourg: Publications Office of the European Union

2015 – 16 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN xxxx-xxxx (print), ISSN xxxx-xxxx (online)

ISBN xxx-xx-xx-xxxxx-x (print)
ISBN xxx-xx-xx-xxxxx-x (PDF)

doi:xx.xxxx/xxxxx

JRC Mission

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new methods, tools and standards, and sharing its know-how with the Member States, the scientific community and international partners.

Serving society
Stimulating innovation
Supporting legislation

doi:xx.xxxx/xxxxx

ISBN xxx-xx-xx-xxxxx-

Publications Office