FloodScan Data Users Guide

FloodScan Version V05R01
Document Revision R02

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

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

26 August 2022

Table of Contents

1 Overview

1.1 Description of dataset

FloodScan's flood extent depiction products provide daily algorithmic delineation of temporarily flooded and unflooded areas from satellite remote sensing observations. FloodScan's core products give a simple indication of the flood state—whether an area of land (a grid cell) was flooded or not on a certain date. The products do not differentiate if there was flooding on all or part of the day or if all or part of a grid cell was flooded. FloodScan makes daily flooded/unflooded determinations over an entire region using the latest available observations as well as prior data to arrive at a best-estimate flood map. Regions covered in the current version include Africa, South America, and North America south of 55°N latitude. The FloodScan data record extends from early 1998 to the present with daily updates from a near real-time processing system.

FloodScan downscales coarse resolution (~22-km) passive microwave satellite data to depict flooding at ~90-m resolution. Microwave measurements are most sensitive to changes in water cover when land conditions are otherwise stable. As a result, FloodScan's primary products are best at depicting large scale, inland river flooding when landscapes are unfrozen. Flooding in smaller floodplains and within 5–10 km of coastlines is usually not depicted unless it is part of a larger flood event. FloodScan offers secondary products that may depict flooding in these areas but with higher false positive rates. This document includes guidance for selecting a product set appropriate for your application.

1.2 Primary and secondary products

FloodScan has two *primary* algorithm modes with distinct product types:

- *Standard flood extent depiction* (SFED): SFED is designed to prioritize low false positive rates for large scale flooding with algorithmic consistency over long time scales and large regions. SFED processing includes dynamic 2- to 3-day weighted averaging along with other spatiotemporal methods that minimize false positives and noise. As a result, *the SFED algorithm mode makes relatively conservative estimates of maximum flood extent, flood frequency, and flood duration.*
- *Maximum daily flood extent depiction* (MFED): MFED is designed for disaster response applications that prioritize (1) timely updates and (2) maximum flood extent depiction. MFED processing is identical to SFED's except that it skips SFED's false positive filtering and multi-day noise averaging steps. As a result, *the MFED algorithm mode is more likely to depict the maximum flood extent and do it sooner than SFED but with an increased risk of large-scale false positives*.

Users of the SFED and MFED products should be aware of the trades-offs between the two so they can make appropriate use of the data in their applications. In general, we recommend that users interested in singular, extreme events use MFED, while other users, particularly those interested in flood occurrence over time, use SFED.

FloodScan also has two *secondary* products that are variations on the primary products:

- *No detection threshold SFED* (NDT-SFED): Like SFED but produced without flooded fraction thresholding.

- *No detection threshold MFED* (NDT-MFED): Like MFED but produced without flooded fraction thresholding.

FloodScan processing produces an intermediate flooded fraction product at the passive microwave satellite data scale (~22-km). The algorithm downscales flooded fraction to make its flood depiction products (e.g., 90-m scale). However, the algorithm applies a *minimum detectable flooded fraction* (MDFF) threshold prior to downscaling to create the SFED or MFED products. This MDFF step filters out low level flooded fraction noise as potential false positives but has the unintended effect of suppressing small-scale floods or flooded areas on the margins of larger floods. NDT-SFED and NDT-MFED give users the option to work with flood maps without MDFF filtering[1]. However, *the NDT products should only be used with the understanding that they are more likely to have false positives than their respective primary data products.* For example, large-scale NDT false positives frequently occur due to signals from soil moisture variation, land cover change, or meteorological conditions.

1.3 Derivative products

Derivative products include (1) composite maps combining primary and secondary FloodScan products over one or more days with static masks (e.g., persistent open water) and (2) analytics such as number of days flooded or flood depth. Derivative data products may be based on either SFED or MFED. Section 2 describes the current derivative product set. New derivative products are added based on user interests.

1.4 Version numbering

The FloodScan product version number (e.g., V05R01, indicated on the title page) corresponds to the major version (e.g., V05) and minor revision (R01) of the algorithm associated with the dataset.

1.5 Terms and conditions

Purchasers of FloodScan data products agree to abide by the AER Services and Data Sales eCommerce Terms and Conditions.

2 Product details and usage guidance

The FloodScan product suite includes the primary and secondary data products, which are generated automatically by the FloodScan processing system, derivative data products, which are generally generated on demand, and static masks. The following tables provide details on the current product set and guidance for usage.

[1] NDT product processing retains flooded fraction thresholding near coastlines for quality control purposes.

Table 1: Primary and secondary data products and static masks

Short name	Full name	Data values	Usage guidance
SFED	Standard flood extent depiction	1: Flooded land 0: Unflooded land or masked	Best used for automatic detection and mapping of large, inland floods. SFED has the lowest false positive rate among the FloodScan products, but users should have some tolerance for large-scale false positives.
MFED	Maximum daily flood extent depiction		Best used only when and where the occurrence of a major flood event can be independently verified (e.g., from news reports, stream gauges, models, or other remote sensing methods). Expect high false positive rates in the absence of independent verification of flooding.
NDT-SFED	No detection threshold SFED		Use with SFED to delineate the possibility of either (1) smaller floods not represented in SFED or (2) flooding on the margins of larger floods. Expect many additional false positives outside of SFED flood extents.
NDT-MFED	No detection threshold MFED		Use with MFED to delineate the possibility of either (1) smaller floods not represented in MFED or (2) flooding on the margins of larger floods. Expect many additional false positives outside of MFED flood extents.
LWMASK	Static land-water mask	1: Persistent open water 0: Land	Use to mask areas FloodScan considers to be persistent open water. FloodScan does not make a flooded/unflooded determination for persistent open water areas.
WWMASK	Static woody wetlands mask	1: Woody wetlands 0: Other land or water	US only: Use to mask areas FloodScan considers to be woody wetlands. FloodScan does not make a flooded/unflooded determination for woody wetlands.[2]

[2] The FloodScan DEPTH product infers flooding in woody wetlands when flooding is indicated in adjacent areas.

Table 2: Derivative data products based on either SFED or MFED as primary product

Short name	Full name	Data values	Usage guidance
ACC	Accumulated flood extent	Over specified time period: 0: Unflooded land 1: Any flooded land (per selected primary product) 2: Any additional secondary product flooded land 3: LWMASK 4: WWMASK	• See usage notes for SFED, MFED, NDT-SFED, and NDT-MFED (Table 1). • Use to combine primary/secondary data products and masks into a single file. • Use to map maximum flood extent over a defined time period (i.e., known duration of a flood event).
NDAYS	Number of days flooded	0-*N* days per specified time period	• Use as a proxy indicator of flood severity. • Use to indicate flood detection confidence. Grid cells with NDAYS less than 2 or 3 days are more likely to be false positives.
DEPTH	Flood water depth	0-*D* meters maximum water depth over specified time period	• Use as a proxy indicator of flood severity. • Use to indicate flood detection confidence. Grid cells with lower flood depth are more likely to be false positives.
AREA	Area aggregate fractional flooding	0-1 fraction of unmasked area with flooding aggregated at specified spatial resolution (daily only)	• Use for regional/historical questions, e.g.: ○ locate areas most affected by flooding ○ compute flood return frequency ○ find examples of past floods

3 Product data files

FloodScan product data files contain primary, secondary, and/or derivative data reformatted and mosaiced to a user-specified area of interest. The files may include data from a single day or compiled over a user-specified range of contiguous dates. Product files are delivered in ZIP archives along with order information, copyright notice, and this users guide. FloodScan's internal raw data files are described in the Appendix.

3.1 Product data file formats

Users may select either GeoTiff (http://trac.osgeo.org/geotiff/) or self-documenting NetCDF-4 (https://www.unidata.ucar.edu/software/netcdf/) data formats. GeoTiff files store data with either one-bit (all raw data types), one-byte unsigned integer (ACC and NDAYS), or single-precision (DEPTH) encoding. NetCDF-4 files store data with either one-byte unsigned integer (all raw data types, ACC, and NDAYS) or single-precision (DEPTH) encoding.

3.2 Product data file spatial coverage

File latitude and longitude bounds are user selectable up to 100 square degrees total area.

3.3 Product data file temporal coverage

File date or date range are user selectable up to 30 contiguous days.

3.4 Product data file internal organization

File includes one N_{rows} x $N_{columns}$ pixel raster data array. Each pixel represents a 3-arcsecond latitude x 3-arcsecond longitude area. An arcsecond is $1/60^{th}$ degree.

3.5 Product data file naming conventions

Template for single-day product file names:

aer_<product>_<resolution>_<YYYYMMDD>_<version>.<ext>

- <product>: one of *sfed*, *mfed*, *ndt_sfed*, *ndt_mfed*, *lwmask*, *wwmask*, *sfed_acc*, *mfed_acc*, *sfed_ndays*, *mfed_ndays*, *sfed_depth*, *mfed_depth*, *sfed_area*, *mfed_area*.
- <resolution>: spatial resolution in arcseconds, e.g., *3s* (3 arcseconds), *300s*, etc.
- <YYYYMMDD>: 4-digit year, 2-digit month, and 2-digit day of month. MM and DD include a leading zero for numbers less than 10.
- <version>: FloodScan product version number.
- <ext>: .tif (GeoTiff) or .nc (netCDF).

Examples:

aer_sfed_3s_20210215_v05r01.tif
aer_ndt_mfed_3s_20210215_v05r01.nc
aer_sfed_area_300s_20210215_v05r01.tif

Template for multi-day product file names:

aer_<product>_<resolution>_<YYYYMMDD>-<YYYYMMDD>_<version>.<ext>

- <product>: one of *sfed_acc*, *mfed_acc*, *sfed_ndays*, *mfed_ndays*, *sfed_depth*, *mfed_depth*.
- <YYYYMMDD>-<YYYYMMDD>: *start date - end date*.

Examples:

aer_sfed_acc_3s_20210215-20210224_v05r01.tif
aer_mfed_depth_3s_20210215-20210224_v05r01.tif

4 Dataset spatial representation

4.1 Horizontal datum

World Geodetic System 1984 (WGS84, EPSG:4326).

4.2 Vertical datum

Not applicable.

4.3 Coordinate system

Data values represent 3-arcsecond (~90-m at the equator) grid cells on a geographic coordinate system grid. Grid cells are edge aligned (Figure 1).

4.4 Spatial coverage

Spatial coverage includes Africa, South America, and North America south of 55°N latitude (Figures 2 to 4).

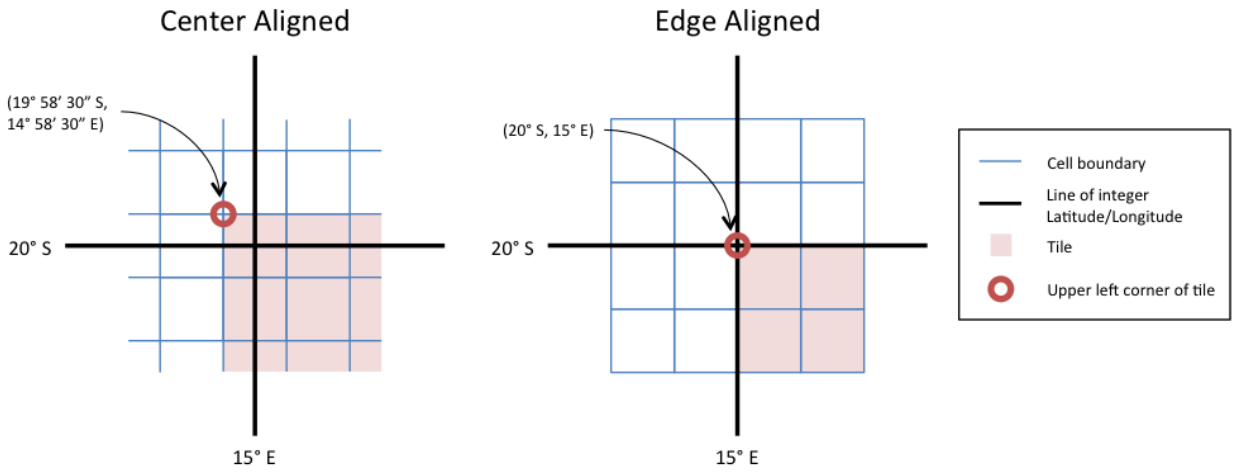

Figure 1: Definition of center-aligned and edge-aligned grid cells. FloodScan grid cells are edge-aligned.

5 Dataset temporal representation

5.1 Reporting frequency

Raw data reporting frequency is daily.

5.2 Reporting time

The nominal reporting time is 0000 UTC (Coordinate Universal Time) of the FloodScan product date indicated in the file name.

5.3 Time period represented by a report

The nominal time period represented by a report is the 24-hour period following the reporting time. The FloodScan algorithm attempts to use only satellite observations from the nominal time period and the two days before. When no observations are available during this time, the algorithm uses data from the most recent prior day.

5.4 Temporal coverage

Table 3 indicates historical coverage start dates for each region.

Table 3: Historical coverage start dates

Region	Subregion	Coverage starting
Africa	All	1998-01-12
North America	South of 38°N latitude	1998-01-12
	South of 55°N latitude	2002-07-02
	North of 55°N latitude	Not available
South America	North of 38°N latitude	1998-01-12
	All	2002-07-02

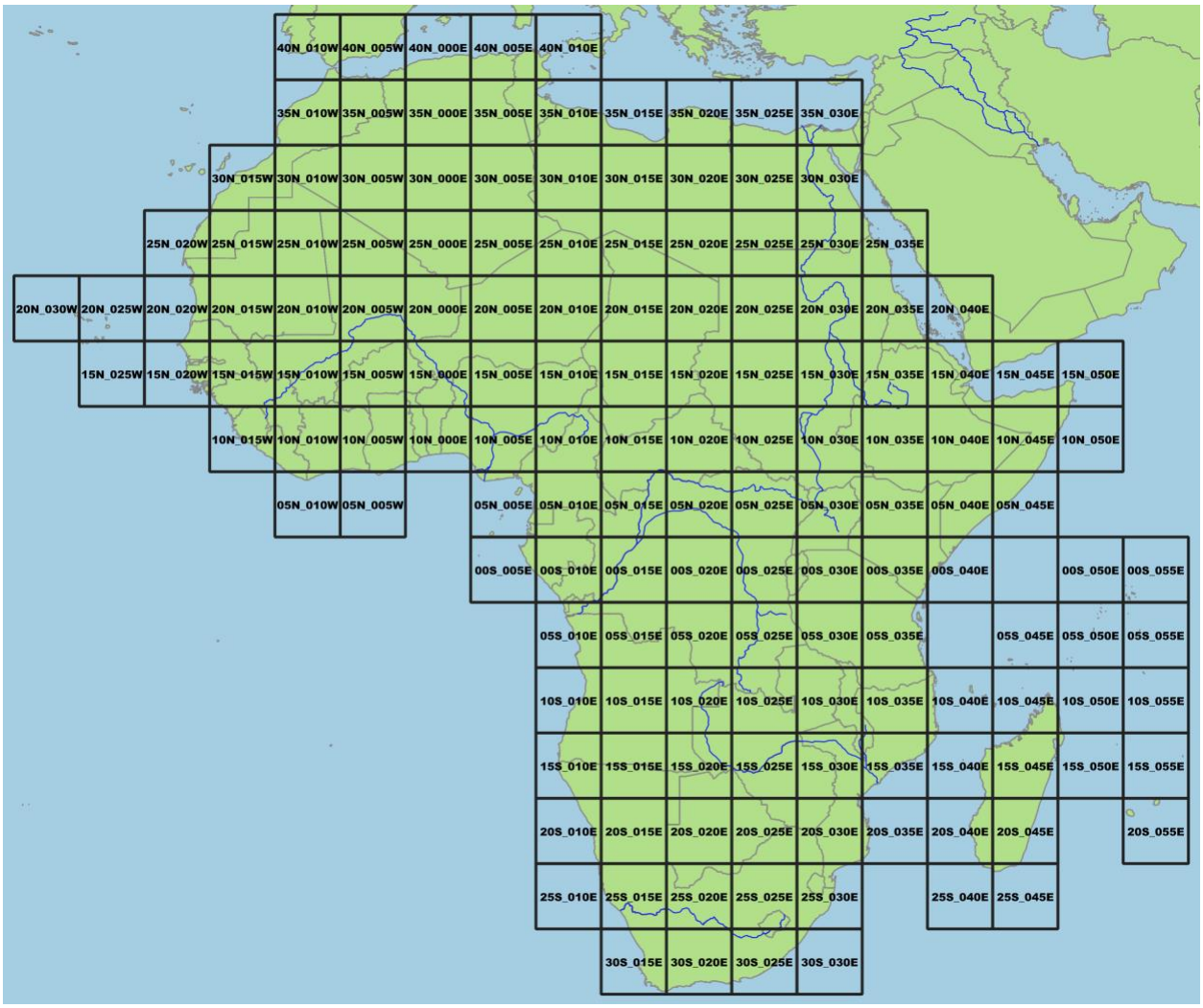

Figure 2: FloodScan coverage and 5° raw data tile layout – Africa region.

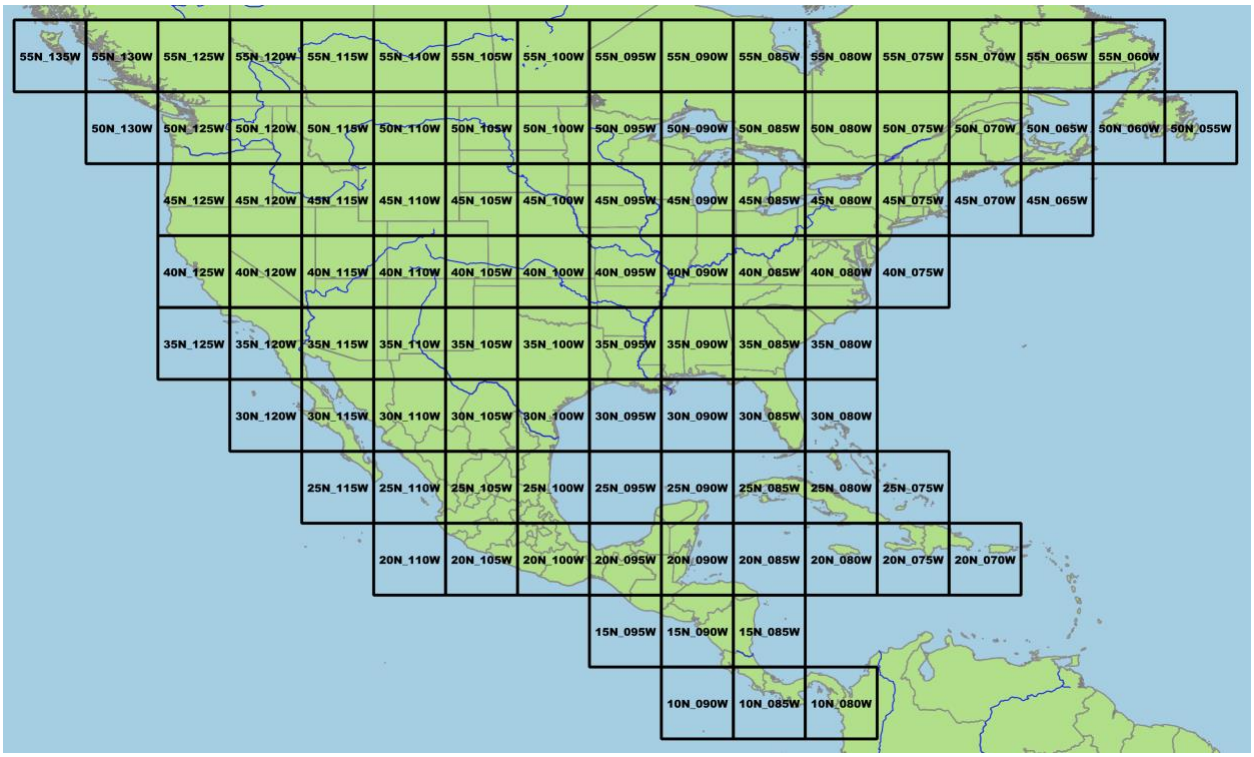

Figure 3: FloodScan coverage and 5° raw data tile layout – North America region.

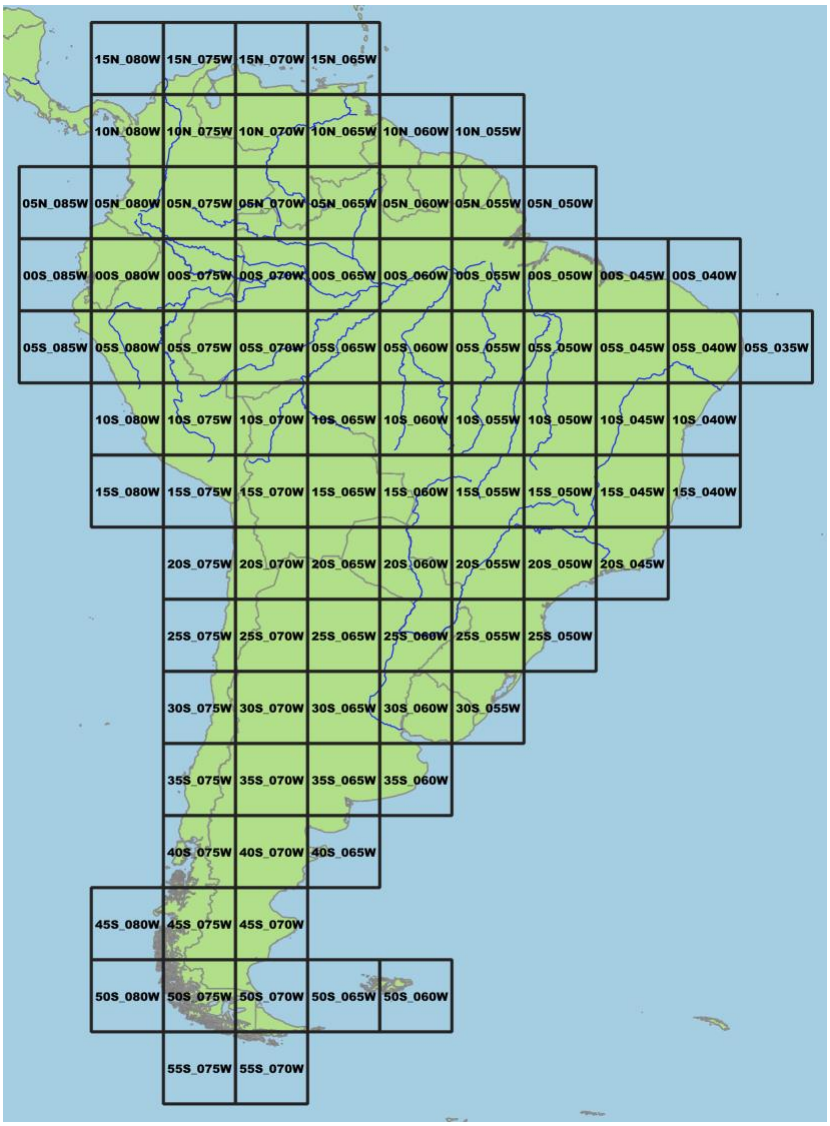

Figure 4: FloodScan coverage and 5° raw data tile layout – South America region.

6 Copyright and attribution

A copyright notice is included with all FloodScan product deliveries. If a FloodScan product is reprinted, copied, or otherwise used in full, copies must reflect the copyright notice actually included with the product. If a FloodScan product is reprinted, copied, or otherwise used in part, the following legend must appear on each page so used: “Includes copyrighted material of Atmospheric and Environmental Research, with its permission.”

In publications, please use the following attributions:

- General attributions in text: "Atmospheric and Environmental Research (AER) FloodScan".
- Map legends: "AER FloodScan” with “flooding (date range)" abbreviated as needed.
- Source data citations: "Flood depictions: AER FloodScan v05r01".

7 Acknowledgments

FloodScan algorithm development was supported by the African Risk Capacity/World Food Programme (contracts QRSA-I01B-13 and QRSA-040-16) and the NASA Terrestrial Ecology Program (NNH10CC61C) and Terrestrial Hydrology Program (NNH13CH27C).

8 Document revision history

Revision	Date	Description
R00	2021/03/02	Initial Data Users Guide for FloodScan version v05r01.
R01	2021/03/26	Added AREA product details.
R02	2022/08/26	Corrected A.5 Internal raw data file naming convention.

Appendix A Internal raw data files

FloodScan raw data types include SFED, MFED, NDT-SFED, and NDT-MFED as well as the land-water mask (LWMASK) and woody wetlands mask (WWMASK). FloodScan's internal data raw data files may be useful for regional analytics or long-term flood monitoring. Contact AER to request access to this data.

A.1 Internal raw data file format

Data are stored as georeferenced raster imagery in GeoTIFF (http://trac.osgeo.org/geotiff/) files with one-bit encoding and *deflate* compression.

A.2 Internal raw data single file (tile) coverage

One file covers a 5° latitude x 5° longitude area (tile) described by the coordinates of the upper-left (UL) corner (Figures 2 to 4).

A.3 Internal raw data 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 area. An arcsecond is 1/60th degree.

A.4 Internal raw data file (tile) organization

There are 152, 117, and 92 5° latitude x 5° longitude tiles covering land areas in Africa, North America, and South America, respectively (Figures 2 to 4). Note that internally North America includes 18 tiles that extend coverage to from 55°N to 60°N latitude. This northern row of tiles goes beyond the standard FloodScan coverage and are provided for informational purposes only.

A.5 Internal raw data file naming convention

File naming template:

<prefix>_<product>_<resolution>_<UL latitude>_
<UL longitude>_<YYYYMMDD>_<version>.<ext>

- <prefix>: *aer* — except for Africa region where daily product files have no prefix.

- <product>: one of *sfed*, *mfed*, *ndt_sfed*, *ndt_mfed*, *lwmask*, or *wwmask* — except for Africa region where *afed* and *ndt_afed* are used in place of *sfed* and *ndt_sfed*.
- <resolution>: 3s (3 arcseconds).
- <UL latitude>: 2-digit upper-left tile latitude plus N or S character.
- <UL latitude>: 3-digit upper-left tile longitude plus E or W character.
- <YYYYMMDD>: 4-digit year, 2-digit month, and 2-digit day of month. MM and DD include a leading zero for numbers less than 10. *lwmask* and *wwmask* files do not include <YYYYMMDD>.
- <version>: for *sfed*, *mfed*, *ndt_sfed*, and *ndt_mfed*, <version> is the FloodScan product version number; for *lwmask* and *wwmask*, <version> is a dataset-specific version number always less than or equal to the FloodScan product version number.
- <ext>: .tif

Examples, all regions except Africa:

aer_sfed_3s_05N_010W_20070211_v05r01.tif
aer_ndt_sfed_3s_05N_010W_20070211_v05r01.tif
aer_mfed_3s_05N_010W_20070211_v05r01.tif
aer_ndt_mfed_3s_05N_010W_20070211_v05r01.tif

Examples, Africa region:

afed_3s_05N_010W_20070211_v05r01.tif
ndt_afed_3s_05N_010W_20070211_v05r01.tif
mfed_3s_05N_010W_20070211_v05r01.tif
ndt_mfed_3s_05N_010W_20070211_v05r01.tif

Examples, all regions:

aer_lwmask_3s_05N_010W_v01r01.tif
aer_wwmask_3s_05N_010W_v05r01.tif