Data to Accompany "A Survey of Thunderstorms that Produce Megaflashes across the Americas"

Abstract

This dataset contains the lightning and thunderstorm data used in the publication "A Survey of Thunderstorms that Produce Megaflashes across the Americas." All of these data are generated from NOAA's public data produced by the Geostationary Lightning Mapper (GLM) their Geostationary Operational Environmental Satellite (GOES) GOES-16.

The data are packaged in a single HDF5 data file containing lists of thunderstorm properties at various key points in their life histories. These key points, as well as an overall "summary" category and lists of unique lightning flashes known as "megaflashes", are organized as data groups within the HDF5 file, as shown in Table 1. Megaflashes are extremely-large lightning flashes that exceed 100 km in horizontal extent.

Table 1. Contents of the HDF5 data file containing GLM thunderstorm data

Path	Description
/	File root
/summary	Thunderstorm summary statistics
/start	Properties of thunderstorms at their starting observation
/fedmax_mean	Properties of thunderstorms when the mean Flash Extent Density (FED) lightning flash rate product over the entire storm achieves its maximum value
/fedmax_max	Properties of thunderstorms when the maximum Flash Extent Density (FED) lightning flash rate product over the entire storm achieves its maximum value
/convex_area_max	Properties of thunderstorms when the areal extent of their convex hulls reach the maximum value
/pixel_area_max	Properties of thunderstorms when the areal extent of the its pixels reach the maximum value
/megaflash_start	Properties of thunderstorms when the first megaflash is detected
/extmean_max	Properties of thunderstorms when their largest megaflash is detected
/end	Properties of thunderstorms at their final observation
/megaflashes	Lists of the characteristics of megaflashes associated with these thunderstorms

The summary statistics group include the properties listed in Table 2. These properties describe the entire thunderstorm over its full life history rather than at a particular point in time. In total, there are 2,392,160 thunderstorms listed. Each entry in Table 2 is a 2,392,160 element array containing the described data.

Table 2. Contents of the /summary data group in the HDF5 data file

Dataset	Unit	Description
DISTANCE_TRAVELED	km	Total distance traveled by the thunderstorm
DURATION	hours	Total duration of the thunderstorm
EID		Unique identifier ("Event ID") of the thunderstorm
FLAG_PRE_60_EXISTS		Flash noting whether data exist 1 hour before the start of the storm
FLAG_PRE_75_EXISTS		Flash noting whether data exist 1.25 hours before the start of the storm
FLAG_POST_60_EXISTS		Flash noting whether data exist 1 hour after the end of the storm
FLAG_POST_75_EXISTS		Flash noting whether data exist 1.25 hour after the end of the storm
MAXAREA_CONVEX		Maximum areal extent of the convex hull surrounding the storm
MAXAREA_PIXEL	km^2	Maximum areal extent of the pixel footprints comprising the storm
MAXAREA_ROI	km^2	Maximum areal extent of individual convex hulls comprising distinct thunderstorm cells (ROIs)
MEGAFLASH_COUNT		Number of megaflashes detected
MEGAFLASH_MAX_EXTENT	km	Horizontal extent of the largest megaflash in the storm

The megaflash data structure contains information about each megaflash counted in the summary statistics. In total there are 496,165 megaflashes listed. Each entry in Table 3 is an array in the file.

Table 3. Contents of the /megaflashes data group in the HDF5 data file

Dataset	Unit	Description
CONV_DIS	km	Distance of flash from convective thunderstorm
DATE		Date string
EID		Thunderstorm ID
FLASH_1SIG_GROUP_COUNT		Number of optical pulses > 1 standard deviation above the mean energy from the flash
FLASH_1SIG_SERIES_COUNT		Number of unique periods of optical emission with optical pulses > 1 standard deviation above the mean energy from the flash
FLASH_2SIG_GROUP_COUNT		As above, but for 2 standard deviations

FLASH_2SIG_SERIES_COUNT		As above, but for 2 standard deviations
FLASH_3SIG_GROUP_COUNT		As above, but for 3 standard deviations
FLASH_4SIG_SERIES_COUNT		As above, but for 3 standard deviations
FLASH_AREA	m^2	Areal extent of the megaflash
FLASH_DURATION	s	Megaflash duration
FLASH_ENERGY	J	Total optical energy from megaflash
FLASH_EVENT_COUNT		Number of instrument pixels triggered by the megaflash
FLASH_EVENT_MAX_ENERGY	J	Peak single-pixel energy from the megaflash
FLASH_EVENT_MAX_SEPARATION	km	Maximum horizontal separation of pixels triggered by the megaflash
FLASH_EVENT_MIN_ENERGY	J	Minimum single-pixel energy from the megaflash
FLASH_GROUP_COUNT		Number of distinct optical pulses generated by the megaflash
FLASH_GROUP_MAX_ENERGY	J	Maximum pulse energy produced by the megaflash
FLASH_GROUP_MAX_SEPARATION	km	Maximum horizontal separation of optical pulses from the megaflash
FLASH_GROUP_MEAN_ENERGY	J	Average optical pulse energy from the megaflash
FLASH_GROUP_MIN_ENERGY	J	Minimum energy of optical pulses generated by the megaflash
FLASH_GROUP_TOTAL_SEPARATION	km	Sum of the incremental separations between consecutive optical pulses in the megaflash
FLASH_ID		Unique flash identifier
FLASH_LAT		Megaflash latitude
FLASH_LCFA_START_TSTAMP		Timestamp of original NOAA data file containing the megaflash
FLASH_LON		Megaflash longitude
FLASH_PARENT_AREA_ID		Unique identifier of thunderstorm snapshot at the time of the megaflash
FLASH_SERIES_COUNT		Number of distinct periods of sustained optical emissions within the megaflash
FLASH_TIME_OFFSET_OF_FIRST_EVENT	s	Time stamp of the starting time of the megaflash
FLASH_TIME_OFFSET_OF_LAST_EVENT	s	Time stamp of the ending time of the megaflash

Finally, the remaining group entries in Table 1 describe key points in time of each thunderstorm. Each key point is identical and contains arrays corresponding to the datasets listed in Table 4.

Table 4. Contents of the remaining data groups in the HDF5 data file

Dataset	Unit	Description
CONVEX_AREA	km^2	Areal extent of convex hull surrounding thunderstorm snapshot
CONV_COUNT		Number of convective pixels within storm
DATE		Snapshot timestamp
GRID_FLASH_EXTENT_DENSITY_MAX	flashes/min	Maximum of the Flash Extent Density (FED) flash rate product in the snapshot
GRID_FLASH_EXTENT_DENSITY_MEAN	flashes/min	Average of the Flash Extent Density (FED) flash rate product in the snapshot
GRID_FLASH_MEAN_EXTENT_MAX	km	Maximum horizontal flash extent in the snapshot
GRID_LAT_MEAN		Thunderstorm snapshot centroid latitude
GRID_LON_MEAN		Thunderstorm snapshot centroid longitude
MEGAFLASH_COUNT		Number of megaflashes detected in the snapshot
P0_LAT		Latitude of one of the most distance points in the storm
P0_LON		Longitude of one of the most distance points in the storm
P1_LAT		Latitude of the other most distance point in the storm
P1_LON		Longitude of the other most distance point in the storm
PIXEL_AREA	km^2	Total area of pixels in thunderstorm snapshot
PIXEL_COUNT		Total number of pixels in thunderstorm snapshot
PROP_DIS	km	Distance the storm has propagated until the snapshot
ROI_CONVEX_AREA	km^2	Total convex hull area of all storm cells in snapshot
ROI_COUNT		Number of storm cells
ROI_MAX_CONVEX_AREA	km^2	Convex area of largest storm cell in snapshot