Description of the Global MCS Tracking Database using NASA MergedIR and GPM IMERG Satellite Dataset

Prepared by Zhe Feng (zhe.feng@pnnl.gov)
Pacific Northwest National Laboratory

1. Introduction

The mesoscale convective system (MCS) database is a long-term high-resolution (10 km resolution, 1 hourly) storm system dataset that tracks individual MCS events. The MCS database is produced by applying an updated version of the FLEXible object TRaKeR (FLEXTRKR) algorithm (Feng et al., 2018) described in detail in Feng et al. (2021) to the NASA global MergedIR infrared brightness temperature (Janowiak et al., 2017) and GPM IMERG half hourly precipitation datasets (Huffman et al., 2019).

2. Purpose

The purpose of the dataset is to provide identification and lifecycle evolution of individual MCS events, along with a suite of important MCS characteristics including lifetime, cloud shield and precipitation feature characteristics, propagation speeds, and more. The long-term high-resolution nature of the dataset provides MCS climatology over all seasons for near global coverage where GPM MergedIR and GPM IMERG datasets were available. The database can be used to understand relationships between atmospheric environments and MCS characteristics, impacts of MCS on hydroclimate, and to evaluate and improve numerical weather prediction models and earth system models.

3. MCS Tracking Algorithm (FLEXTRKR)

The FLEXTRKR algorithm first identifies and tracks large cold cloud systems (CCSs) associated with deep convection using satellite infrared (T_b) data, and subsequently identifies MCSs using GPM IMERG precipitation feature (PF) characteristics. An MCS is defined as a large CCS ($T_b < 241$ K, area $> 4 \times 10^4$ km^2) containing a PF (contiguous area within a CCS with rain rate > 2 mm h^{-1}) with major axis length > 100 km, and PF area, PF mean rain rate, PF rain rate skewness, and heavy rain volume ratio exceeding certain lifetime dependent thresholds and persists for at least 5 hours. These PF-based criteria are derived based on the IMERG precipitation characteristics mapped onto the U.S. MCS tracking database (https://doi.org/10.5439/1571643) using satellite T_b and ground-based NEXRAD 3D mosaic radar dataset (Feng et al., 2019). More details of the MCS tracking algorithm can be found in Feng et al. (2021).

4. MCS Data Description

The global MCS tracking database is produced using GPM IMERG V06B dataset, which is available from 2000 to 2019. As of the writing of this documentation (April 2021), the global MCS database has been produced from June 2000 to November 2019, and preliminary data quality control and evaluation has been performed.

The globe is divided into three large regions (**Figure 1**) to conduct the MCS tracking on each region separately: Asia-Pacific (35°E – 180°E, 60°S – 60°N), Europe-North America (180°W – 50°E, 20°N – 60°N), and Africa-South America (180°W – 50°E, 60°S – 30°N). These three regions are large enough to contain most MCSs with minimal impact from those MCSs crossing from one region to another. The overlapping areas between the three regions further allow a "buffer zone" to reduce discontinuity if the regional data are to be stitched together to develop a global data. This stitching has only been performed on monthly mean or longer periods. The MCS tracking data has the same format for each region, therefore the description below applies to any specific region.

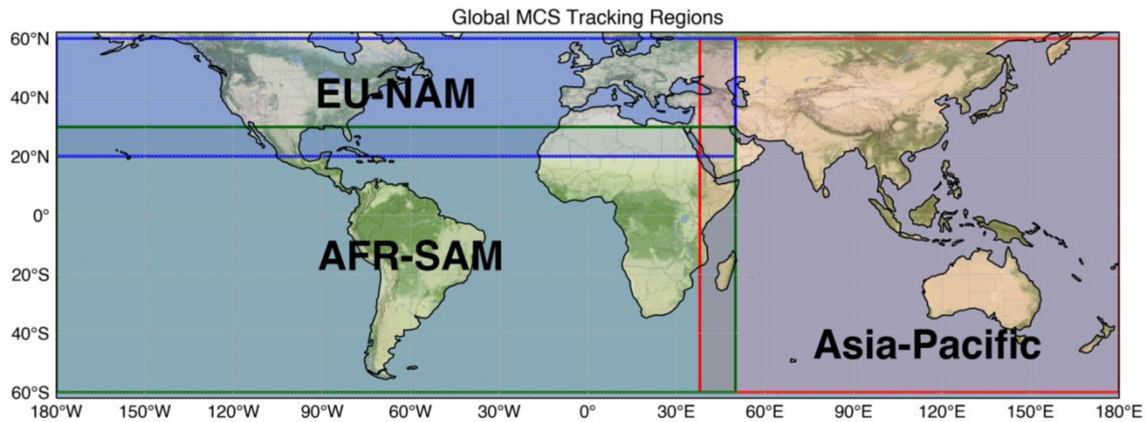

Figure 1. Three regions in the database to conduct MCS tracking.

a) Hourly pixel-level data

The MCS tracking pixel-level data are produced on the native grid (0.1°, ~10 km resolution) of the input dataset: NASA GPM IMERG precipitation data. The NASA global geostationary satellite infrared brightness temperature (T_b) data (0.04°, ~4 km resolution) is interpolated to match the IMERG data. The pixel-level data contains full field of T_b, precipitation, deep convective cloud object identification, and MCS tracking number. The data file name follows this format "mcstrack_yyyymmdd_hhmm.nc".

Figure 2 shows an example of the pixel-level MCS database over the Maritime Continent. IR brightness temperature and IMERG precipitation are both full fields from the original dataset. The color shadings behind large clusters of PFs show the MCS masks for individual MCSs. The same color shadings between different times denote the same MCS being tracked.

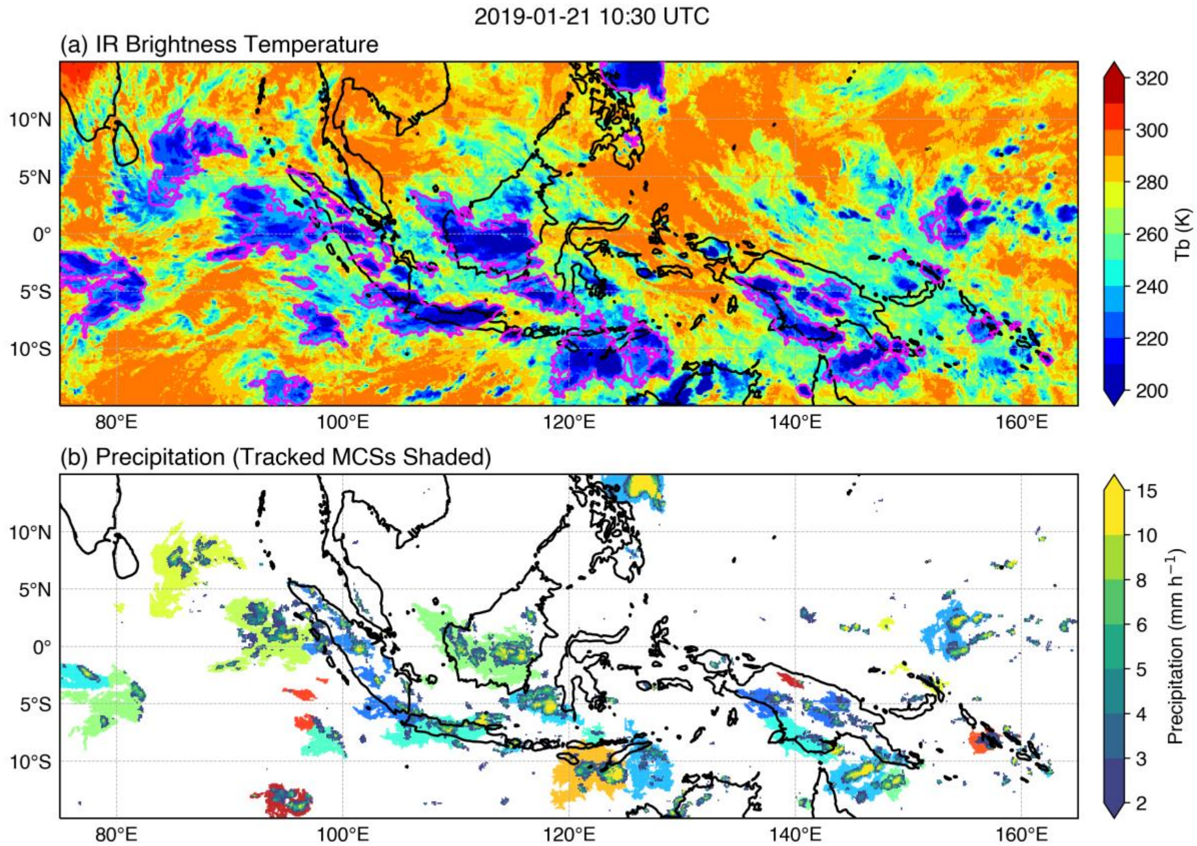

Figure 2. Example snapshot of pixel-level MCS database over the Maritime Continent during 21 January 2019. (a) IR brightness temperature, (b) GPM IMERG precipitation. The magenta contour in (a) and the color shadings behind large clusters of PFs in (b) denote MCS masks ("cloudtracknumber"). Weak rain rates < 2 mm h^{-1} in (b) are excluded for clarity. Animation for this day can be found in https://portal.nersc.gov/project/m1867/mcs_global/MCS_pixeldata_example_animation.mp4.

Table 1. MCS track pixel-level hourly data variables and descriptions.

Variable Name	Dimension	Description
time	time	Epoch time
basetime	time	Epoch time
julian_day	time	Julian Day Number (used in IDL)
lon	lon	Longitude
lat	lat	Latitude
longitude	lat, lon	Longitude of each grid
latitude	lat, lon	Latitude of each grid
numclouds	time	Number of CCS identified in this file
tb	time, lat, lon	Brightness temperature (full regridded MergedIR data)
precipitation	time, lat, lon	Precipitation (full original IMERG data)
cloudnumber	time, lat, lon	Number associated with a CCS at a given pixel
cloudtracknumber	time, lat, lon	MCS track number mask assigned to CCS
pcptracknumber	time, lat, lon	MCS track number mask assigned to precipitation feature (PF). This track number is the same with "cloudtracknumber" except only applied to precipitation > 1 mm/h.

Explanation of each variable in the pixel-level data is provided in **Table 1.** Within each sub-directory of pixel-level data, denoted by the start and end date of a continuous 1 year tracking period (e.g. 20140101_20141231), the MCS track number ("cloudtracknumber", and "pcptracknumber") monotonically increases from 1 to X (the maximum number of MCSs tracked during a period). For example, cloudtracknumber = 1 within multiple pixel-level data files denotes the first MCS tracked during this period. The MCS track number in "cloudtracknumber" and "pcptracknumber" is the same, but they differ in areal coverage. Cloudtracknumber is defined using IR temperature < 241 K, so it has a larger coverage, usually containing most if not all of the precipitation in the vicinity of the MCS. Pcptracknumber is defined using precipitation > 2 mm h^{-1}, so it has a smaller coverage, only containing the major precipitation features.

Note that the MCS track numbers all start from 1 for each individual tracking periods, i.e., they are not unique across different tracking periods. For example, there will be MCS track number = 1 in both tracking period 20140101_20141231 and 20150101_20151231. Because each tracking period are tied to one track file (described below), which contains time and location of each MCSs, they can be distinguished across different tracking periods if one desired to do so. Occasionally, if there is not a single CCS identified in the entire domain at a given time, or there are significant areas of missing T_b data, there will not be an MCS tracking pixel-level data for that time.

Location of hourly pixel-level data on NERSC HPSS:
/home/f/feng045/GPM/mcs_region/region/mcstracking_ccs4_4h/
Replace region with any of the three specific region names shown in **Figure 1**: asia, spac, nam.

b) Hourly track statistics data

The track data contains many variables summarizing the key features (i.e., statistics) of MCSs. Each continuous tracking period correspond to one track file. The data file name follows this format "robust_mcs_tracks_extc_yyyymmdd_yyyymmdd.nc" (e.g., robust_mcs_tracks_extc_20140101_20141231.nc).

The track data describes the timing, centroid location ("meanlat", "meanlon"), and evolution of each MCS during the tracking period. The variables are organized by [tracks, times]. The "tracks" dimension is the number of MCSs in the tracking period, and the tracks match the track number in the pixel-level data. For example, tracks = 0 is the first MCS in the pixel-level data (cloudtracknumber = 1, or pcptracknumber = 1), and tracks = 1 is the second MCS (cloudtracknumber = 2, or pcptracknumber = 2), …, etc. The "times" dimension is the relative time since an MCS is detected. For example, times = 0 is the convective initiation time (defined by the first detection of a CCS from satellite IR data) of an MCS. A variable named "base_time" (Epoch time in seconds since 1970-01-01T00:00:00) with dimension [tracks, times] designates the time for each MCS track at each time. Combining "base_time", which locates the time of an MCS, and the track index, which corresponds to the "cloudtracknumber" mask in the pixel-level data described above, one can link the MCS track data with the exact pixel-level location of the MCS at a given time.

For MCS characteristics, there are generally two types of variables, one derived from IR data, which has two dimensions [tracks, times]; and another type derived from precipitation data, which has three dimensions [tracks, times, nmaxpf].

Explanation of each variable in the MCS track statistics data is provided in **Table 2**.

Table 2. MCS track statistics hourly data variables and descriptions.

Variable Name	Dimension	Description
length	tracks	Duration of each tracked MCS defined by CCS. This is the longest lifetime of MCS, potentially including times before and after precipitation.
mcs_length	tracks	Duration of each tracked MCS when the MCS CCS area exceeds 40,000 km^2.
mcs_type	tracks	Type of MCS (crudely defined as semi-circular MCS: 1; squall-line: 2) based on the shape of CCS during MCS stage.
starttrackresult	tracks	First time step status of each MCS (2 = cloud is the larger fragment of a merger; 10 = start of a new track; 13 = cloud is the smaller fragment of a split)
endtrackresult	tracks	Last time step status of each MCS (0 = cloud from previous file dissipated; 12 = cloud is the smaller fragment of a merger)
base_time	tracks, times	Epoch time of each MCS in a track
julian_day	tracks, times	Julian Day Number of each MCS (used in IDL)
datetimestring	tracks, times	Date-Time of each MCS (year-mo-dy_hr:mn:sc)
mcs_status	tracks, times	Flag denoting MCS stage defined by CCS area (-1: before cloud is tracked; 0: cloud is not MCS; 1: MCS; 2: Squall-line)
meanlat	tracks, times	Centroid latitude of MCS defined using CCS area
meanlon	tracks, times	Centroid longitude of MCS defined using CCS area
core_area	tracks, times	Area of cold cloud core
ccs_area	tracks, times	Area of cold cloud shield (CCS)
mintb	tracks, times	Minimum brightness temperature of CCS
meantb	tracks, times	Mean brightness temperature of entire cloud
ccs_meantb	tracks, times	Mean brightness temperature of CCS
core_meantb	tracks, times	Mean brightness temperature of cold cloud core
speed	tracks, times	Propagation speed of CCS (calculated using centroid location difference)
direction	tracks, times	Propagating direction of CCS (calculated using centroid location difference)
uspeed	tracks, times	Zonal propagation speed of CCS (calculated using centroid location difference)
vspeed	tracks, times	Meridional propagation speed of CCS (calculated using centroid location difference)

trackresult	tracks, times	Track status of MCS at each time (-1 = default (not MCS); 0 = CCS from previous file dissipated; 1 = CCS track from previous file continues; 2 = CCS is the larger fragment of a merger; 3 = CCS is the larger fragment of a split; 10 = start of a new track; 12 = CCS is the smaller fragment of a merger; 13 = CCS is the smaller fragment of a split)
cloudnumber	tracks, times	Cloud number in the corresponding cloudid file
majoraxislength	tracks, times	Major axis length of CCS
eccentricity	tracks, times	Eccentricity of CCS
mergecloudnumber	tracks, times, nmaxmerge	Cloudnumber of small merger at a given time
splitcloudnumber	tracks, times, nmaxmerge	Cloudnumber of small split at given time
pf_length	tracks	Duration of the MCS when a precipitation feature (PF) is detected, which can be considered as the active period of the convection.
pf_mcsstatus	tracks, times	Flag to denote MCS stage defined by PF major axis length > 100 km (0: not MCS; 1: MCS)
pf_heavyrainratio	tracks, times	Volumetric heavy rain (rain rate > 10 mm/h) ratio, defined as heavy rain volume divided by total rain volume
total_rain	tracks, times	Total precipitation under CCS (all rainfall included). No cosine latitude weighting is applied.
total_heavyrain	tracks, times	Total heavy precipitation (rain rate > 10 mm/h) under CCS. No cosine latitude weighting is applied.
rainrate_heavyrain	tracks, times	Mean heavy precipitation rate (rain rate > 10 mm/h) under CCS. No cosine latitude weighting is applied.
npf	tracks, times	Number of PF under the CCS
pf_landfrac	tracks, times	Fraction of PF over land
pf_area	tracks, times, nmaxpf	PF area under CCS
pf_lon	tracks, times, nmaxpf	Centroid longitude of each PF
pf_lat	tracks, times, nmaxpf	Centroid latitude of each PF
pf_rainrate	tracks, times, nmaxpf	Mean precipitation rate of each PF
pf_maxrainrate	tracks, times, nmaxpf	Maximum precipitation rate of each PF
pf_accumrain	tracks, times, nmaxpf	Area total precipitation rate of each PF
pf_accumrainheavy	tracks, times, nmaxpf	Area total heavy precipitation (rain rate > 10 mm/h) of each PF
pf_skewness	tracks, times, nmaxpf	PF pixel-level rain rate skewness

pf_majoraxislength	tracks, times, nmaxpf	PF major axis length
pf_aspecratio	tracks, times, nmaxpf	PF aspect ratio (major axis length / minor axis length)
pf_rr8mm_npix	tracks, times, nmaxpf	Number of pixels with rain rate > 8 mm/h
pf_rr10mm_npix	tracks, times, nmaxpf	Number of pixels with rain rate > 10 mm/h
pf_nuniqpix	tracks	No longer used
location_idx	tracks, nmaxpix	No longer used
pixel_duration	tracks, nmaxpix	No longer used
pixel_pcp	tracks, nmaxpix	No longer used
movement_r	tracks, times	PF movement along angle movement_theta relative to lag
movement_r_meters_per_second	tracks, times	PF propagation speed [m/s]
movement_theta	tracks, times	PF propagation direction
movement_storm_x	tracks, times	PF zonal propagation distance in [km/hour]
movement_storm_y	tracks, times	PF meridional propagation distance in [km/hour]
movement_time_lag	tracks, times	Time Lag between consecutive advection estimates

Important notes about propagation speed and direction: Because the propagation speeds from IR data are simply calculated based on the centroid locations, the CCS propagation speed is prone to large fluctuations in the shapes of the clouds and hence requires cautions for using in accurate estimates of propagation speed. Further, the propagation direction reflects the movement of the upper-level clouds which could differ from the convection/precipitation at the lower-level (e.g. when lower-level and upper-level wind directions differ), caution is advised when using these values in a quantitative manor.

The variables derived from IMERG data contains up to 3 of the largest Precipitation Features (PFs) within each CCS. The "nmaxpf" dimension is organized by size, e.g., nmaxpf = 0 contains the largest PF at any given time associated with an MCS. The PF variables contain MCS status ("pf_mcsstatus") as defined by PF, fraction of PF over land ("pf_landfraction"), PF centroid locations ("pf_lon", "pf_lat"). Rainfall amount variables are: total precipitation ("total_rain", sum of all raining pixels in mm h^{-1}), total heavy precipitation ("total_heavyrain"), mean heavy rain rate (> 10 mm h^{-1}, "rainrate_heavyrain"). Volumetric rain rate can be calculated by multiplying "total_rain" with the area of a pixel (~10×10 km^2).

Various PF characteristics variables include: "pf_area", "pf_rainrate", "pf_accumrain", "pf_skewness", "pf_majoraxislength", "pf_aspectratio", "pf_maxrainrate", "pf_accumrainheavy". In addition, propagation of PFs are estimated using a more advanced 2-D cross-correlation map between two consecutive hours of the MCS PFs, as detailed in Feng et al. (2018). These PF propagation characteristics contains propagation speed ("movement_r", "movement_r_meters_per_second"), direction ("movement_theta"), and propagation distance ("movement_storm_x", "movement_storm_y").

Location of hourly track statistics data on NERSC HPSS:
/home/f/feng045/GPM/mcs_region/region/stats/robust.tar
Replace region with any of the three specific region names shown in **Figure 1**: asia, spac, nam.

c) Linking hourly track statistics data with pixel-level data

Here is an example usage case to link the hourly track statistics data with pixel-level data: where exactly is the MCS feature mask for track #2 at time = 2010-07-01 03:00 UTC? **Figure 5** shows an example on how to link the pixel-level MCS mask with the track statistics data. For a given hourly pixel-level file, the MCS mask values in the variable "cloudtracknumber" correspond to the "tracks" dimension in the MCS track statistics file (offset by 1 for 0-based indexing). **Figure 5a** shows two MCSs (track # 2, 3) coexist at 2010-07-01 03:00 UTC. These two MCSs correspond to the "tracks" indices (1, 2) and "times" indices (0, 1) in the track statistics file, because the values in the track statistics "base_time" variable (Epoch time) matches the pixel-level file time 2010-07-01 03:00 UTC.

In practice, since both sets of pixel-level and track statistics files contain Epoch time ("time" in pixel-level file, "base_time" in track statistics file), which contains the full date/time (e.g., 2010-07-01T03:00:00), one can simply search for the matching Epoch time between "base_time" in track statistics file and "time" in a specific pixel-level file to find all the indices of MCS tracks that exist in that hour.

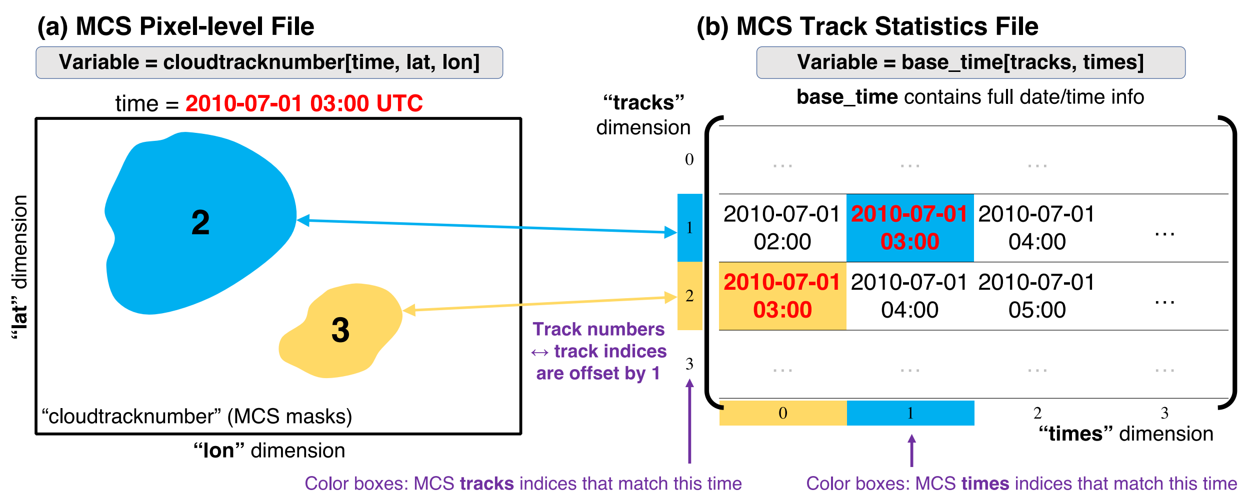

Figure 3. An example to link a track statistics variable to a pixel-level file. (a) MCS pixel-level file, showing variable "cloudtracknumber" containing MCS masks at 2010-07-01 03:00 UTC, (b) MCS track statistics file, showing variable "base_time" containing the Epoch time of each MCS. The MCS "cloudtracknumber" values correspond to the "tracks" indices (offset by 1 for 0-based indexing). For each matching MCS track, the matching time index can be located by matching the "base_time" value (Epoch time) with the pixel-level file "time" (Epoch time), shown in red colors (2010-07-01 03:00).

d) Monthly statistics data

The monthly data contains monthly accumulated or average MCS data to facilitate climate-oriented applications. The three tracking regions have been combined to create the global monthly mean data. All monthly data are calculated on the native 10 km grid to retain the highest spatial resolution. These monthly files are used to produce many of the MCS climatology results reported in Feng et al. (2021).

Explanation of each variable in the pixel-level monthly data is provided in **Table 3**. Most variables are self-explanatory as they are monthly averages. Some variables are single values for each PF at a given hour (e.g. lifetime, PF area, max rain rate) that are mapped to a PF at that hour, then further averaged in time to retain the native grid resolution.

Table 3. MCS track pixel-level monthly data variables and descriptions

Variable Name	Dimension	Description
time	time	Epoch time
lon	lon	Longitude
lat	lat	Latitude
mcs_number_ccs	time, lat, lon	Number of MCS defined by cold cloud shield
mcs_number_pf	time, lat, lon	Number of MCS defined by precipitation feature
mcs_nhour_ccs	time	Number of CCS identified in this file
mcs_nhour_pf	time, lat, lon	Number of MCS hours defined by cold cloud shield
mcs_number_pf	time, lat, lon	Number of MCS hours defined by precipitation feature
mcs_nhour_speedmcs	time, lat, lon	Number of MCS hours for propagation speed (MCS status is met)
lifetime_mean	time, lat, lon	MCS averaged lifetime (defined by precipitation feature)
ccs_area_mean	time, lat, lon	MCS cold cloud shield averaged area
pf_area_mean	time, lat, lon	MCS precipitation feature averaged area
rainratemax_mean	time, lat, lon	MCS maximum rain rate temporal mean
initiation_ccs	time, lat, lon	MCS convective initiation hours defined by cold cloud shield
pf_speed_mcs	time, lat, lon	MCS averaged propagation speed
pf_uspeed_mcs	time, lat, lon	MCS averaged propagation speed (x-direction)
pf_vspeed_mcs	time, lat, lon	MCS averaged propagation speed (y-direction)
precipitation	time, lat, lon	Total precipitation amount
mcs_precipitation	time, lat, lon	MCS precipitation amount
mcs_precipitation_count	time, lat, lon	Number of hours MCS precipitation is recorded
ntimes_asia	time	Number of hours from ASIA region
ntimes_spac	time	Number of hours from SPAC region
ntimes_nam	time	Number of hours from NAM region

Location of monthly data on NERSC HPSS:
/global/project/projectdirs/m1867/zfeng/gpm/mcs_region/global/monthly/monthly_mean.tar

5. Data quality

The main data quality issue in the MCS database is missing satellite T_b data, as the IMERG precipitation data rarely have missing values within 60°S–60°N. Infrequent or continuous missing T_b data typically would not significantly affect MCS tracking but frequent missing data at fixed intervals (e.g. one missing frame every two hour) would impact MCS identification, as tracking of clouds would terminate and restart. Throughout the 20-year period (2000-2019), several generations of geostationary satellite fleets have been used to produce the global MergIR T_b dataset (for a complete list of satellites see https://docserver.gesdisc.eosdis.nasa.gov/public/project/GPM/CPC-4kmIR-Sats.pdf). During the early 2000s, certain regions have larger fractions of missing T_b data (**Figure 4**), such as the West Pacific (130°E - 180°E) and the Southeast Pacific (140°W - 80°W, 60°S - 5°S). The missing IR T_b data and their impacts on the MCS tracking results have been examined on a monthly timescale and results show that the impact in the West Pacific region during 2003-2005 is significant. Frequent missing IR T_b data at 4 hours of the day (03, 09, 15, 21 UTC) in that region during those 3 years results in a noticeable reduction of long-lived MCSs and a significant increase of short-lived (5-hour lifetime) ones. Caution is advised in using the MCS data over that region during 2003-2005. A more in-depth data quality check for individual years and regions can be found in this document: Global_MCS_data_quality_check_20210125.pdf. Users of the dataset are encouraged to read the data quality document to avoid known data issues.

Used for MCS Tracking

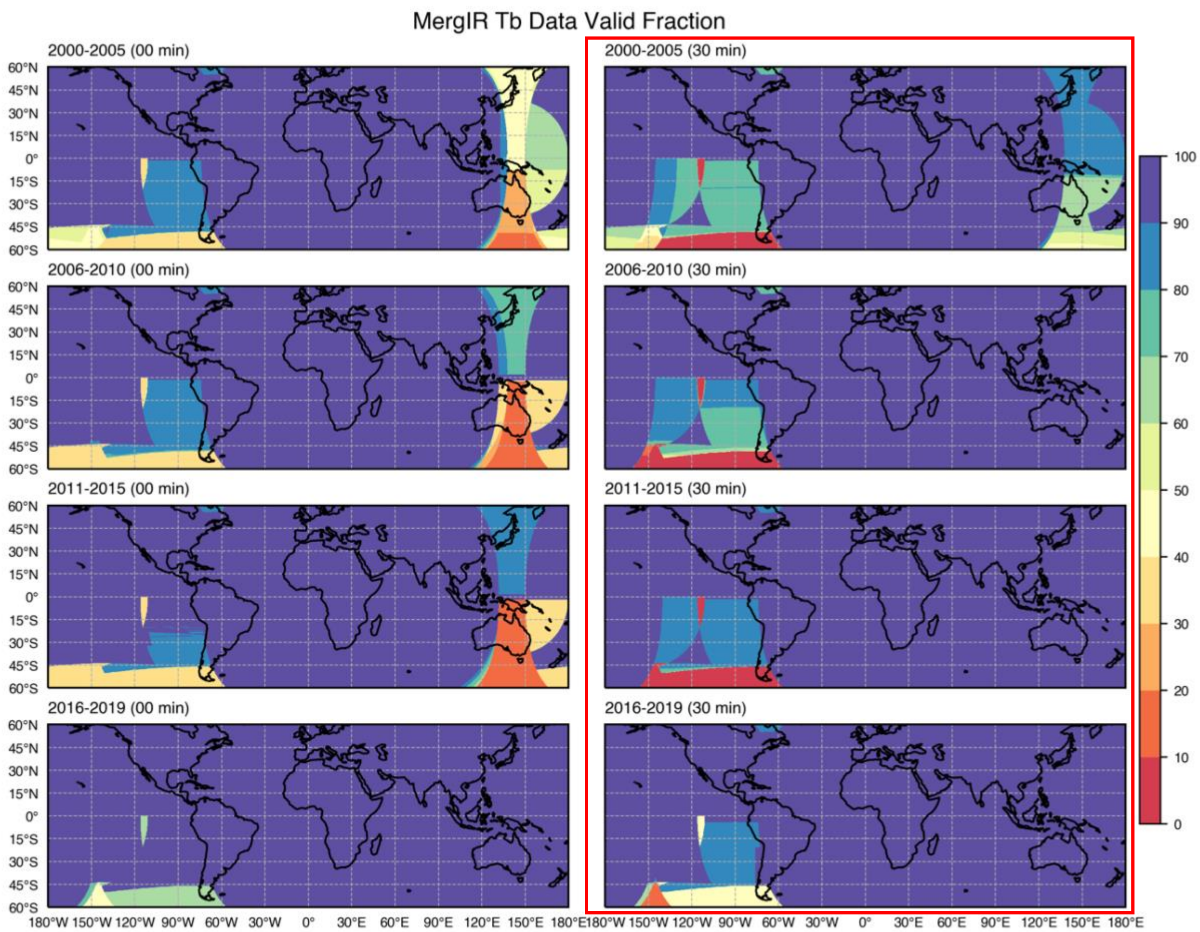

Figure 4. Fraction of valid MergIR Tb data for 00 min (left column) and 30 min (right column) past the hour for each 5-year period between 2000-2019. The **30 min data** are used for MCS tracking.

Additional data quality check is conducted by calculating global monthly mean MCS statistics such as the number of MCS, MCS precipitation amount and frequency and compare that against 20-year climatological mean for the same months. Significantly lower number of MCSs over a region that corresponds to large missing T_b data fraction indicates the missing T_b data adversely affect overall MCS identification, in addition to affecting proportion of short-lived vs. long-lived MCSs. **Figure 5** shows an example of notably lower total number of MCS over western Pacific during January – March 2006, likely caused by similar missing T_b data in that region between 2003-2005. The issue seems to reduce since April 2006. The January 2003 – March 2006 period only affects the western Pacific, while for June 2017 the global T_b data production were interrupted from June 11 to June 30, therefore no MCS data are produced for the entire globe during that period.

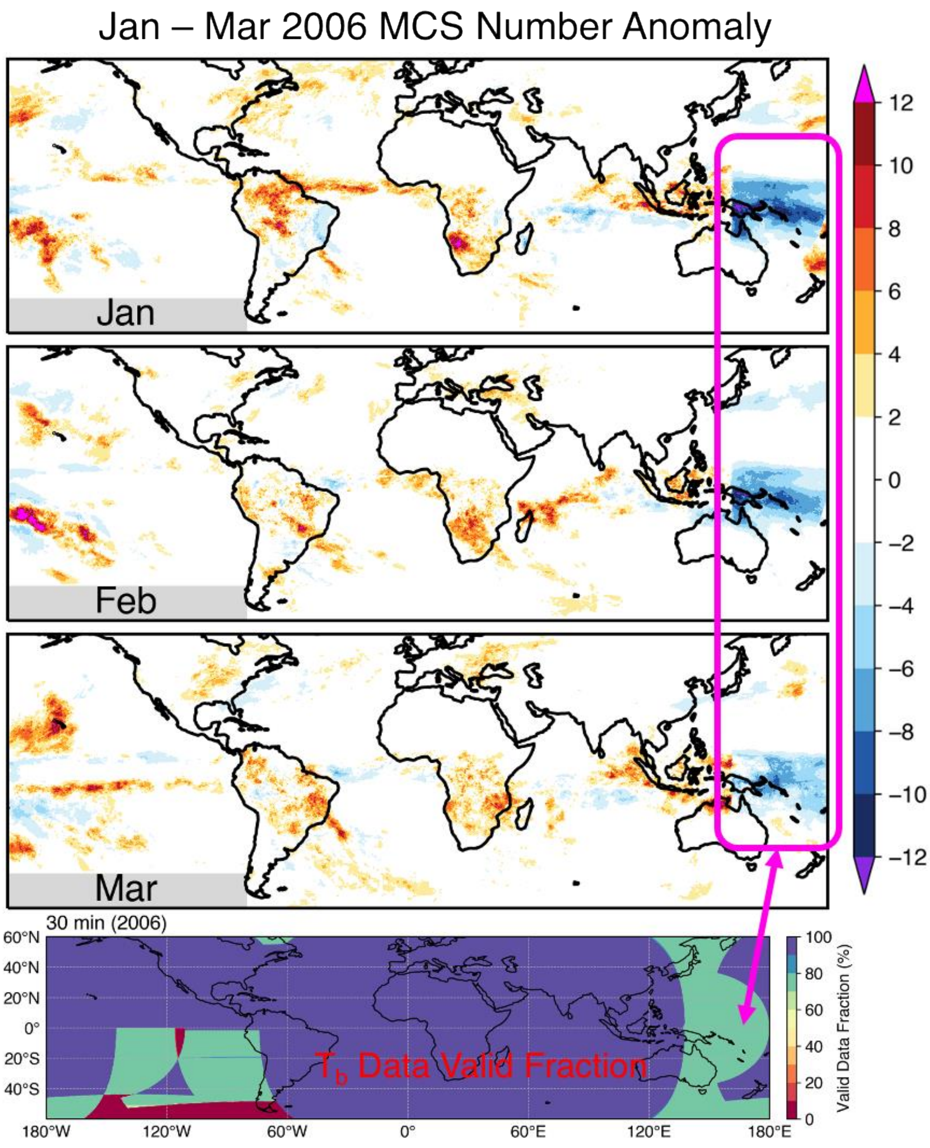

Figure 5. Example monthly MCS number anomaly against 20-year climatology for Jan – Mar 2006, and corresponding T_b data valid fraction map for 2006. Magenta box highlights the region with significantly lower MCS number that correspond to missing T_b data in that region.

6. Analysis codes

Post-processing Python codes are available on GitHub:
https://github.com/WACCEM/gpm_mcs_post

References

Feng, Z., Houze, R. A., Leung, L. R., Song, F., Hardin, J. C., Wang, J., et al. (2019). Spatiotemporal Characteristics and Large-scale Environments of Mesoscale Convective Systems East of the Rocky Mountains. *Journal of Climate, 32*(21), 7303-7328. https://doi.org/10.1175/JCLI-D-19-0137.1

Feng, Z., Leung, L. R., Houze, R. A., Hagos, S., Hardin, J., Yang, Q., et al. (2018). Structure and Evolution of Mesoscale Convective Systems: Sensitivity to Cloud Microphysics in Convection-Permitting Simulations Over the United States. *Journal of Advances in Modeling Earth Systems, 10*(7), 1470-1494. https://doi.org/10.1029/2018MS001305

Feng, Z., Leung, L. R., Liu, N., Wang, J., Houze Jr, R. A., Li, J., et al. (2021). A Global High-resolution Mesoscale Convective System Database using Satellite-derived Cloud Tops, Surface Precipitation, and Tracking. *Journal of Geophysical Research - Atmosphere*. https://doi.org/10.1029/2020JD034202

Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2019). *GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06*. Retrieved from: https://doi.org/10.5067/GPM/IMERG/3B-HH/06

Janowiak, J., Joyce, B., & Xie, P. (2017). *NCEP/CPC L3 Half Hourly 4km Global (60S - 60N) Merged IR V1*. Retrieved from: https://doi.org/10.5067/P4HZB9N27EKU