Description of the MCS Tracking Database using Global MergedIR and GridRad NEXRAD Radar Dataset over CONUS

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

1. Introduction

The mesoscale convective system (MCS) database over Contiguous United States (CONUS) is a long-term (2004-2017) high-resolution (4 km, 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) (Feng et al., 2018) algorithm described in detail in Feng et al. (2019) to the NASA Global MergedIR infrared brightness temperature (Janowiak et al., 2001) and the GridRad mosaic 3D NEXRAD radar dataset (Bowman & Homeyer, 2017).

2. Purpose

The purpose of the dataset is to provide lifecycle evolution of individual MCS events, along with a suite of important MCS characteristics including lifetime, precipitation feature characteristics, convective and stratiform separation, echo-top heights, propagation speeds, and more. The long-term high-resolution nature of the dataset provides MCS climatology over all seasons covering large regions of the CONUS east of the Rocky Mountains, which can be used to understand relationships between atmospheric environments and MCS characteristics, impacts of MCS on hydroclimate and severe weather events, provides context to ARM SGP observations, 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 T_b data, and subsequently identifies MCSs using radar defined convective features and precipitation features (PFs). Convective features are identified using the storm labeling in three dimensions (SL3D) classification (Starzec et al., 2017). Precipitation features are contiguous areas with rain rate > 1 mm h^{-1}. An MCS is defined as a large CCS (area > 6×10^4 km^2) containing a PF with major axis length > 100 km, a convective feature containing radar reflectivity > 45 dBZ at any vertical level, and the PF persists for at least 6 hours.

4. MCS Data Description

a) Hourly Pixel-level Data

The MCS database using GridRad mosaic 3D radar dataset is available from 2004-2017 January-December. Surface precipitation is also provided in the database, which is obtained from the Stage IV multi-sensor precipitation dataset produced by the 12 River Forecast Centers in the continental U.S. (Lin, 2011).

Sub-directory: /mcstracking_linkpf

The MCS tracking pixel-level data are produced on the native grid (~4 km resolution) of the input dataset: NASA global geostationary satellite infrared brightness temperature (T_b) data. The pixel-level data contains full field of T_b, precipitation (interpolated from Stage IV precipitation data), radar reflectivity at 2 km MSL (if not available, use 3 km reflectivity instead) (interpolated from the native GridRad, which has 2 km horizontal resolution), convective-stratiform classification using the SL3D algorithm (Starzec et al., 2017), echo-top height (with various reflectivity thresholds), deep convective cloud object identification, and MCS tracking identification. The data file name follows this format "mcstrack_yyyymmdd_hhmm.nc".

Within each sub-directory of pixel-level data, denoted by the start and end date of a continuous 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, containing most if not all of the precipitation in the vicinity of the MCS. Pcptracknumber is defined using precipitation > 1 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 statistics 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.

Missing and Incorrect Data

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. For a small number of analysis times, no volume scans area available from the NEXRAD network and the radar variables and Stage IV precipitation data will appear as all blank (missing), lists of the dates/times are provided in /missingfile_statistics/missingfilelistyyyy.txt. A small number of NEXRAD volume files have no valid meteorological echoes in them even though the radars were operating, lists of dates/times are provided in /missingfile_statistics/invalidfilelistyyyy.txt. There are also periods with bad volume scans from individual NEXRAD radars, which were removed in GridRad, a list of the dates/times is provided on GridRad's website (link). More details about the GridRad radar dataset can be found here: http://gridrad.org/data.html. Stage IV precipitation also occasionally contain erroneous values as they are operational data products. A manual quality control was applied to the Stage IV precipitation data to exclude analysis times with sudden very large precipitation values (> 40 mm h^{-1}) that cover unrealistically large area (> 300×4^2 km^2). A list of dates/times with Stage IV precipitation data excluded is provided in /missingfile_statistics/unrealistical_stage_IV_filelist.txt. Users of precipitation data are advised to be cautious on the quality of individual events.

b) Hourly Statistics Data

Sub-directory: /stats_linkpf

The statistics data contains many variables summarizing the key features of MCSs. Each continuous tracking period correspond to one statistics file. The data file name follows this format "robust_mcs_tracks_yyyymmdd_yyyymmdd.nc".

The statistics data describes the timing, centroid location, and evolution of each MCSs. The variables are organized by [tracks, times]. The "tracks" dimension is the number of MCSs in the tracking period, and the tracks matches 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 initiation time (defined by 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" in the pixel-level data described above, one can link the MCS statistics data with the exact pixel-level location of the MCS at a given time.

There are 3 variables related to the lifetime of the MCS: length, mcs_length, pf_length. "length" is the entire duration of the tracked CCS, potentially including times before and after precipitation. "mcs_length" is the duration when MCS cloud shield area exceeds 60,000 km^2. "pf_length" is the duration of the MCS when a precipitation feature is detected, which can be considered as the "active" period of the convection. It is recommended to use "pf_length" as the lifetime of MCS.

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 radar data, which has three dimensions [tracks, times, nmaxpf].

The variables derived from IR data contains the time, centroid location ("meanlat", "meanlon"), cloud shield area ("core_area", "ccs_area"), MCS status ("mcs_status"), merging/splitting information ("trackresult", "starttrackresult", "endtrackresult") and some additional parameters such as shapes of cloud ("majoraxislength", "eccentricity"), 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 not recommended for use 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 have opposite sign), caution is advised when using these values in a quantitative manor.

Within each CCS, the variables derived from radar data contains up to 10 of the largest PFs, and up to 10 of the largest convective features. The "nmaxpf"/ "nmaxcore" dimensions are organized by size, e.g., nmaxpf = 0 contains the largest PF (similarly, nmaxcore = 0 contains the largest convective feature) at any given time associated with an MCS. The PF variables contain MCS status (pf_mcsstatus) as defined by radar, PF centroid locations (pf_lon, pf_lat), various PF characteristics (pf_area, pf_rainrate, pf_skewness, pf_majoraxislength, pf_aspectratio, pf_rr8mm_npix, pf_rr10mm_npix). Convective and stratiform feature characteristics include: convective/stratiform precipitation area (pf_ccarea, pf_sfarea), convective feature centroid locations (pf_corelon, pf_corelat), convective feature morphology (pf_corearea, pf_coremajoraxislength, pf_coreaspectratio, pf_coreeccentricity, pf_coreorientation), convective feature mean echo-top heights (pf_ccdbz10, pf_ccdbz20, pf_ccdbz30, pf_ccdbz40), convective feature max echo-top heights (pf_coremaxdbz10, pf_coremaxdbz20, pf_coremaxdbz30, pf_coremaxdbz40), area exceeding some reflectivity threshold at 2 km MSL (pf_dbz40area, pf_dbz45area, pf_dbz50area).

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").

While convective and stratiform rainfall variables (pf_ccrate, pf_sfrate, pf_ccvolrate, pf_sfvolrate) are available in the dataset, it is not recommended using for quantitative purpose. This is because of the mismatch between hourly snapshot of GridRad radar reflectivity data (near the top of the hour at 00 min, which is used for separating convective/stratiform precipitation area,) and the hourly accumulation of precipitation from StageIV data. Due to the propagating nature of MCSs, hourly accumulated precipitation field cannot be accurately separated into convective/stratiform with a single snapshot of reflectivity data near the top of the hour.

c) Monthly Statistics Data

Sub-directory: /monthly_mean

The monthly data contain several sets of monthly accumulated or averaged MCS data to facilitate climate-oriented applications. All monthly data are calculated on the native 4 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. (2019).

Monthly MCS precipitation data ("mcs_rainmap_yyyymm.nc") includes: total and MCS precipitation amount ("precipitation", "mcs_precipitation"), number of MCS precipitation hours ("mcs_precipitation_count"). The MCS fractional to total precipitation can be easily calculated by dividing the MCS precipitation to the total precipitation.

Monthly MCS statistics data ("mcs_statsmap_yyyymm.nc") includes: number of MCS passing each grid point ("mcs_number_ccs", "mcs_number_pf"), number of hours MCS passing each grid point ("mcs_nhour_ccs", "mcs_nhour_pf"), MCS average lifetime ("lifetime_mean"), MCS mean area ("ccs_area_mean", "pf_area_mean"), MCS mean convective echo-top height ("cc_dbz20_mean", "cc_dbz45_mean"), MCS mean convective core size ("core_area_mean", "core_majoraxislength_mean"), number of MCS initiation and genesis ("initiation_ccs", "initiation_pf", "genesis_ccs", "genesis_pf"), MCS mean propagation speed ("pf_speed_mcs", "pf_uspeed_mcs", "pf_vspeed_mcs").

Monthly Hovmöller precipitation data ("mcs_rainhov_*region*_yyyymm.nc") includes: total precipitation, MCS precipitation and MCS precipitation area fraction averaged in a specific longitudinal band for each region, designated in the file name by *region*. For example, "mcs_rainhov_sgp_201105.nc" contains Hovmöller precipitation data averaged within the Southern Great Plains region; and "mcs_rainhov_201105.nc" (without the *region* name) contains average precipitation within majority of the CONUS land region (31°-48°N).

5. V2 Updates

After the initial release of the MCS database in December 2019, several bug fixes in the data processing and tracking codes were implemented. In addition, MCS tracking data for the year 2017 has been added to the database. The continuous tracking period has been changed to a full year starting from 1 January and ending on 31 December of each year. The MCS database has been updated to V2 (February 2020).

Long-term MCS statistics in V2 are consistent with those reported in Feng et al. (2019). The averaged number MCSs has increased slightly, and their fractional contribution to total precipitation has increased as well. Detail visualizations of the V2 updated MCS statistics are provided in the presentation "2020.02.17_GridRad_MCS_statistics_v2_release.pdf".

6. Source codes

Sub-directory: /source_code/

Python source codes that produce monthly MCS statistics data are provided.

Code Name	Function
define_mcs_track_region.py	Defines which 4 climate region each MCS track belongs to
calc_gridrad_mcs_monthly_precipmap_single.py	Calculate MCS monthly precipitation map
calc_gridrad_mcs_monthly_preciphov_single.py	Calculates MCS monthly precipitation Hovmöller diagram over 4 different climate regions: SGP, NGP, Southeast, Northeast
map_mcs_statsspeed_bymonth.py	calculates monthly mean MCS statistics and maps them on the native pixel-level grid
plot_gridrad_mcs_monthly_rainhov.py	plots monthly MCS Hovmöller precipitation over the Central U.S.

7. Examples

Figure 1 shows an example snapshot of the pixel-level MCS database over the domain. IR brightness temperature and radar reflectivity are both full fields from the original dataset. The color shadings behind large clusters of PFs show the CCS mask for individual MCSs.

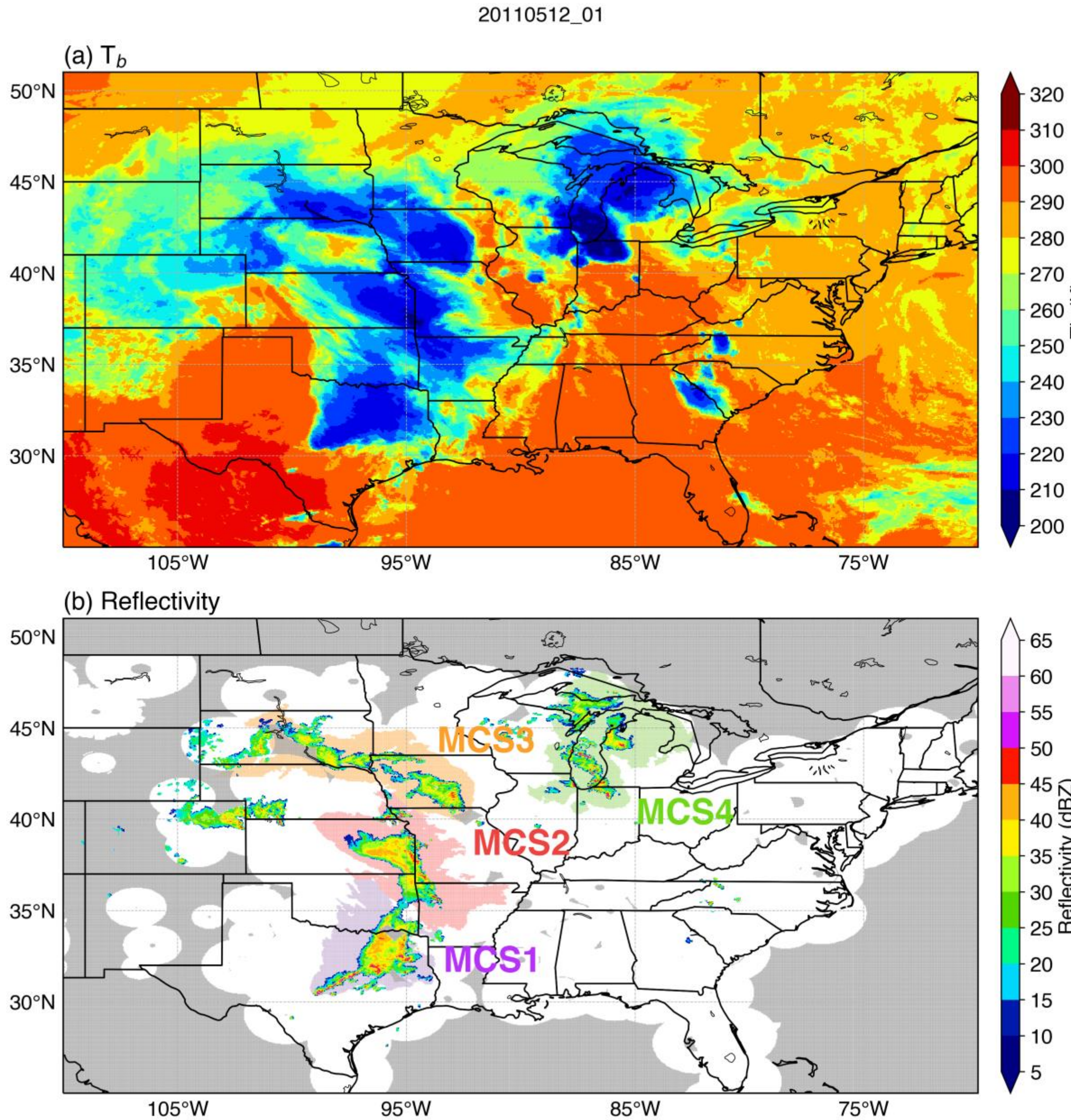

Figure 1. Example snapshot of pixel-level MCS database at 01 UTC on 12 May 2011. (a) IR brightness temperature, (b) GridRad radar reflectivity at 2 km MSL. The color shadings in (b) behind large clusters of PFs denote MCS masks ("cloudtracknumber"). The gray shadings in (b) are regions with no radar coverage at 2 km MSL.

References

Bowman, K. P., & Homeyer, C. R. (2017). *GridRad - Three-Dimensional Gridded NEXRAD WSR-88D Radar Data*. Retrieved from: https://doi.org/10.5065/D6NK3CR7

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://journals.ametsoc.org/doi/abs/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://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001305

Janowiak, J. E., Joyce, R. J., & Yarosh, Y. (2001). A Real–Time Global Half–Hourly Pixel–Resolution Infrared Dataset and Its Applications. *Bulletin of the American Meteorological Society, 82*(2), 205-217. http://journals.ametsoc.org/doi/abs/10.1175/1520-0477%282001%29082%3C0205%3AARTGHH%3E2.3.CO%3B2

Lin, Y. (2011). *GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Version 1.0*.

Starzec, M., Homeyer, C. R., & Mullendore, G. L. (2017). Storm Labeling in Three Dimensions (SL3D): A Volumetric Radar Echo and Dual-Polarization Updraft Classification Algorithm. *Monthly Weather Review, 145*(3), 1127-1145. http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-16-0089.1