This will provide a more consistent regional comparison than isolated regional studies in the past.

Although IMERG is advantageous over IR, tracking MCSs in IMERG is challenging due to connected systems (Mapes and Houze, 1993; Virts and Houze, 2015) and differences in PMW sensor resolutions (You et al., 2020). Therefore, one of our goals is to overcome these challenges and track MCSs in IMERG using the FiT tracking algorithm. The other goal is to study the regional variability of MCSs' properties, such as rain contribution, frequency, area, precipitation intensity, lifetime, and propagation velocity. This article broadly compares the MCS properties over land vs. ocean and onshore vs. offshore regions. Nevertheless, regional differences exist due to location-specific atmospheric processes such as the dominating diurnal cycle over the maritime continent, easterly waves over West Africa, and high surface latent heat flux over the Amazon basin. We present regional comparisons as the frequency of low vs. high MCS property value categories (e.g., short-lived vs. long-lived MCSs, slow-moving vs. fast-moving MCSs), which complements the recent IMERG-based MCSs tracking studies (Feng et al., 2021; Hayden et al., 2021).

2 Data and Methods

2.1 IMERG Version 06B

IMERG has three runs: Early, Late, and Final, available at various latencies targeted for different applications. This paper uses the IMERG Final run, a research-quality product, with precipitation data available every 30 minutes and at 0.1° spatial resolution. IMERG combines the precipitation retrievals from PMW sensors onboard the virtual constellation of low earth orbit (LEO) satellites and retrievals from IR sensors onboard geostationary satellites (Huffman et al. 2019a, 2020). At the locations with PMW observations, IMERG uses only PMW-derived precipitation, considered higher quality than IR-derived precipitation. In the absence of PMW observations, IMERG V06B uses a Kalman Filter-based time morphing algorithm to compute the weighted average of IR precipitation, backward-advected PMW precipitation from future observations, and forward-advected PMW precipitation from past observations. Rajagopal et al. (2021) reported that at locations without PMW observations, IMERG V06B sometimes had spurious rain rates, mostly < 1 mm $hr^{-1,}$ and also underestimated heavy precipitation rates. SHARPEN, a new averaging method for IMERG, will significantly reduce these issues in

IMERG V07 (Tan et al., 2020). In Section 2.4, we discuss using a 1 mm h^{-1} precipitation threshold to mitigate the problem of spurious precipitation.

The PMW sensors on board the LEO satellites observe at multiple frequencies and spatial resolutions. You et al. (2020) lists the nominal precipitation resolution for various PMW sensors, which ranges from 10 km for Sounder for Probing Vertical Profiles of Humidity (SAPHIR) to 59 km for Special Sensor Microwave Imager/Sounder (SSMIS) sensors. IMERG maps the different PMW sensors' precipitation estimates from their nominal resolutions to 0.1° x 0.1° global grid using the nearest-neighbor interpolation.

The IMERG final run has global rainfall data for approximately 20 years, from June 2000 to the near present, with a latency of ~3.5 months. Many PMW sensor satellites are added or decommissioned from the virtual constellation during this period. The merging of various PMW sensor retrievals to produce a global precipitation product presents some challenges to tracking the precipitation systems, discussed in Section 2.3.

2.2 Tracking algorithm

We use the FiT algorithm to identify and track precipitation systems in the IMERG precipitation field. The algorithm tries to mimic object identification performed by a human via subjective analysis. Earlier versions of the FiT algorithm have been used to track precipitation systems in the TRMM MultiSatellite Precipitation Analyses (TMPA; Huffman et al., 2007) 3B42 precipitation field, a predecessor to IMERG (Skok et al., 2009, 2010, 2013; White et al., 2017).

The FiT and other tracking methods involve two major steps: (i) identifying objects at each time step and (ii) relating objects across multiple time steps as the same time-evolving precipitation system. In conventional tracking methods, objects are typically defined as contiguous pixels below or above a single threshold. In contrast, the FiT algorithm uses a technique called "cascading thresholds", with multiple thresholds to identify objects. This object identification method is similar to the Detect And Spread (DAS) technique used in many recent tracking algorithms (Boer and Ramanathan, 1997; Fiolleau and Roca, 2013; del Moral et al., 2018; Feng et al., 2021). A DAS technique typically has two thresholds – a first threshold detects the inner convective core, which spreads to the surrounding region defined by a lower second threshold. With the FiT, more than two thresholds can be used and are explained in detail by Skok et al. (2013).

The next step in the FiT algorithm is to relate objects in consecutive time steps to the same precipitation system using the area overlap method. The term "system" here denotes a time-evolving entity representing a set of spatial objects that overlaps in consecutive time steps. Tracking a precipitation system becomes complex when it splits into many pieces or merges with another system. The FiT algorithm is capable of tracking a system after a split or merger, providing relatively accurate evolution and lifecycle information.

Compared to the older versions of the FiT algorithm that were used in previous studies, the new version used here includes two major improvements: (i) a new parameter called the "separation distance" and (ii) support for the periodic domain in the zonal direction.

After a precipitation system splits into multiple pieces, some can grow to large sizes and move farther away from the other pieces, particularly in regions such as the Intertropical Convergence Zone (ITCZ) and cold fronts. If these breakaway pieces are considered part of the same system, it results in an unrealistic estimation of the system's lifetime, area, accumulated rain volume, and frequency. This led to the development of a new user-defined parameter called "separation distance". If the distance between the centroid of a piece and the centroid of the system (all pieces) is greater than the separation distance, then the piece is treated as a new system. The second modification to the FiT allows for a periodic domain in the zonal direction. In the earlier versions of the FiT algorithm, the systems crossing the periodic domain boundary were terminated and tracked as a new system on the other side. In the newer version, we can successfully track the precipitation systems traversing the IMERG's domain boundaries at 180° E/180° W without artificial discontinuities.

2.3 Tracking challenges

Tracking MCSs in the IMERG precipitation field over the global tropics presents two significant challenges: (i) connected MCSs forming transient precipitation bands that span thousands of kilometers, and (ii) differences between PMW sensor resolution and retrievals. A precipitation band that spans thousands of kilometers (Fig. 1a) is not uncommon in the ITCZ region. Past studies over the Indo-Pacific warm pool region have observed superclusters similar to precipitation bands (Mapes and Houze, 1993; Virts and Houze, 2015). The precipitation bands in IMERG are transient with a lifetime < 3 hours, have distinct mesoscale structures, and lack coherent propagation. Since a precipitation system's lifetime increases with its size (Chen and

Houze, 1997; Machado et al., 1998), one would expect a precipitation band to live longer, yet it is brief.

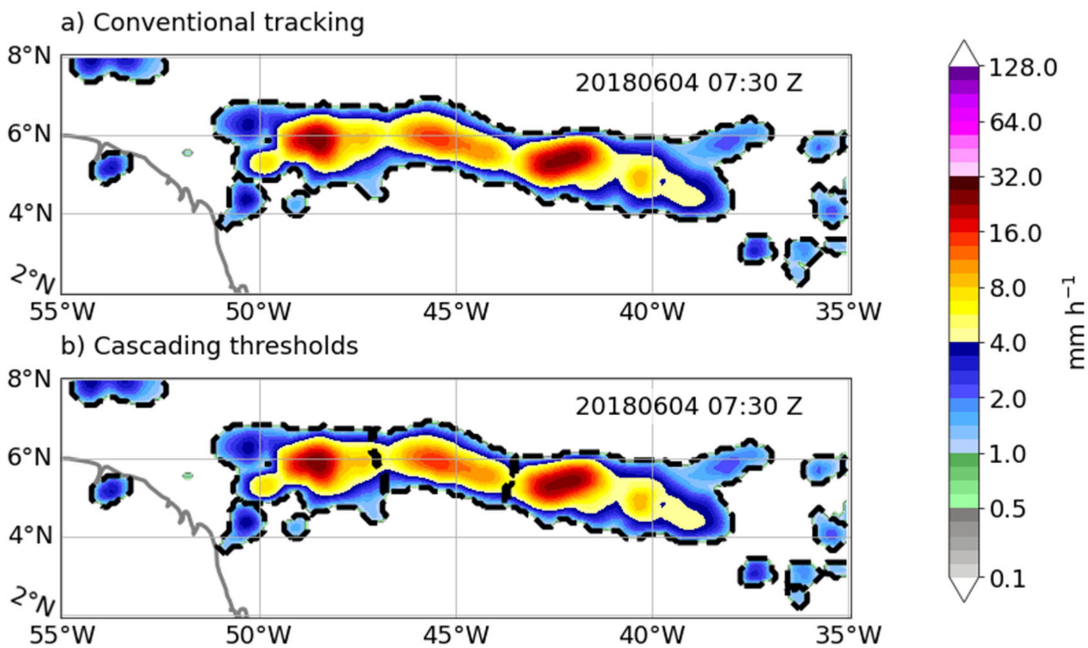

Fig. 1. A transient precipitation band in IMERG on 04 June 2018 at 07:30 UTC. The band spans approximately 15° (~1660 km) of longitudinal width and has three or more mesoscale structures. The dashed contour represents tracked object boundary using a) single threshold of 1 mm h^{-1} and b) cascading thresholds (1, 3, 9, 27)mm h^{-1}.

A conventional tracking method using a single threshold of 1 mm h^{-1} will identify contiguous pixels of a precipitation band as a short-lived precipitation system. In such a method, we lose the tracking information of MCSs that form the band briefly and break up later. The FiT algorithm's "cascading thresholds" technique is beneficial for identifying multiple mesoscale objects within a contiguous precipitation band and tracking each separately (Fig. 1b).

The other challenge to tracking MCSs is the differences between PMW sensors' resolution and retrieval. You et al. (2020) lists the PMW sensors' surface precipitation resolution, which varies from ~10 km (for the SAPHIR PMW sensor) to ~59 km (for the SSMIS PMW sensor). Fig. 2a and 2d present an MCS precipitation field observed from the SAPHIR and SSMIS sensors in a 30-minute interval. The precipitation retrieval from SAPHIR is very noisy compared to SSMIS retrieval. To mitigate the noise and the resolution differences, we apply the uniform smoothing filter (moving average with the square kernel) of width 0.5°, which is approximately the SSMIS resolution. Despite smoothing, the MCS has a multicellular structure in SAPHIR retrievals but is absent in SSMIS (Figs. 2b and 2e). When applying the fixed physical thresholds (1, 3, 9, 27) mm h^{-1}, the FiT identifies a single mesoscale size object in the SSMIS precipitation field. In contrast, it identifies multiple convective cell-sized objects in the SAPHIR precipitation field (Figs. 2b and 2e). This is because of the resolution and precipitation intensity differences between these PMW sensors.

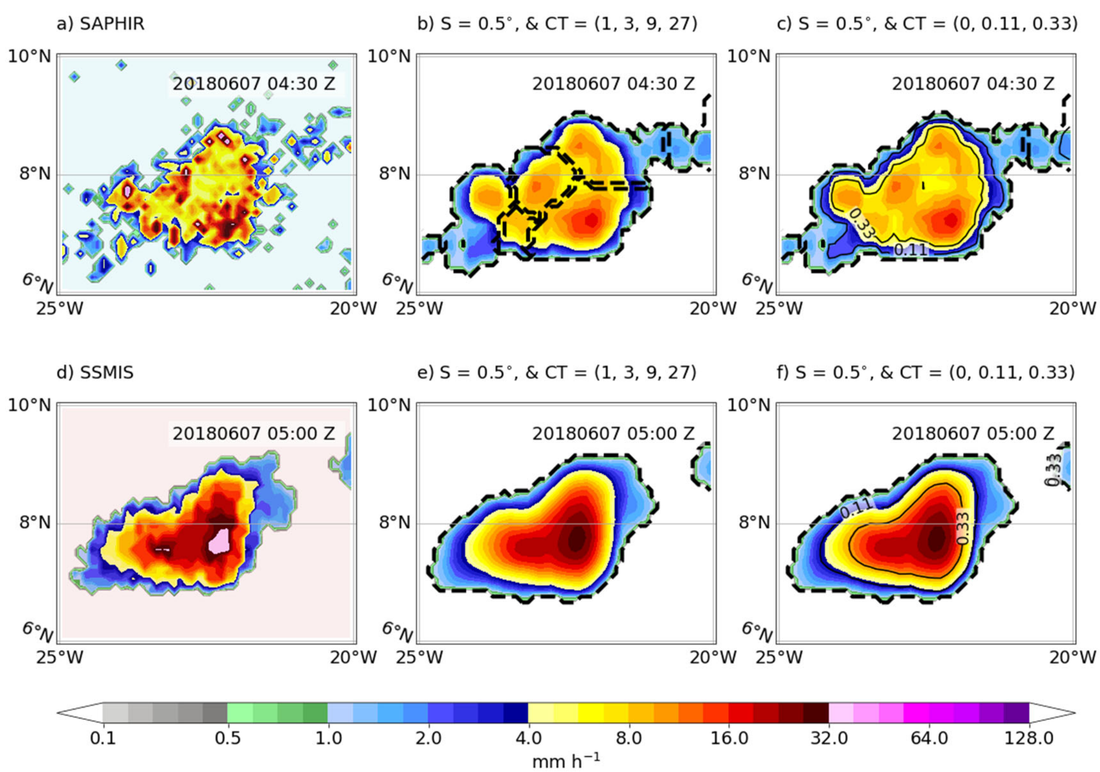

Fig. 2. PMW observations of an MCS on 07 June 2018 during subsequent half-hour periods starting at 04:30 UTC and 05:00 UTC. a) precipitation estimates from the SAPHIR PMW sensor at ~10 km resolution, b) Apply smoothing with filter width, S = 0.5°, and the physical cascading thresholds, CT = (1, 3, 9, 27) mm h^{-1};thick dashed contours represent objects c) Apply the normalized cascading thresholds, CT = (0, 0.11, 0.33); thick-dashed contours represent objects; thin solid contours are normalized threshold levels d) precipitation estimates from the SSMIS PMW sensor at 59 km resolution, e) same as panel b but for SSMIS precipitation, and f) same as panel c but for the SSMIS precipitation.

The cascading thresholds help identify multiple mesoscale objects in a synoptic-scale precipitation band (Fig. 1b). But, it also breaks an MCS into numerous convective cells (Fig. 2b).

This undesirable behavior exists even in other DAS-type tracking methods (Huang et al., 2018). To overcome this issue, we normalize the IMERG precipitation field by the maximum precipitation rate within each contiguous blob. The normalization transforms the precipitation field to values between 0 and 1. The motivation to use normalization comes from the understanding that an MCS's precipitation field is multicellular, with a large precipitation gradient around heavy precipitation cores. This structure is the same despite the resolution and retrieval differences between various PMW sensors.

The cascading thresholds for the normalized precipitation field are chosen as a series of decimal values between 0 and 1. For example, the cascading thresholds (0, 0.11, 0.33) imply that the normalized precipitation areas with values ≥ 0.33 will represent heavy rain areas, and the values ≥ 0.11 and <0.33 will be relatively moderate rain areas. The areas with values > 0 and < 0.11 will represent relatively low rain areas. For the MCS in Figs. 2c and 2f, the normalized cascading thresholds identify only a single mesoscale object in the SAPHIR and SSMIS precipitation fields.

The normalized cascading thresholds can be interpreted as variable or dynamic thresholds that change with a PMW sensor or location. In Fig. 2c, the precipitation blob has a maximum precipitation rate of ~18 mm h^{-1}, so the normalized thresholds of (0, 0.11, 0.33) translate to physical thresholds of 18 mm h^{-1} * (0, 0.11,0.33), which are approximately (0, 2, 6) mm h^{-1}. For the same MCS in the next time step (Fig. 2f), the maximum precipitation rate is 30 mm h^{-1}, so the normalized thresholds approximately translate to (0, 3.3, 10) mm h^{-1}. The normalized thresholds act as dynamically changing physical thresholds that help identify the mesoscale objects in different PMW sensors' precipitation fields.

2.4 Tracking procedure

Before running the FiT tracking, we perform three custom pre-processing steps on the IMERG precipitation field. The values for parameters in pre-processing steps are chosen based on literature and our understanding of the IMERG product.

The three custom pre-processing steps are: i) smooth the IMERG precipitation field, ii) apply precipitation or no precipitation condition, and iii) normalize the precipitation field. In the first pre-processing step, we smooth the IMERG precipitation field with a uniform filter of 0.5° width (coarsest PMW resolution) to mitigate PMW resolution differences and reduce the noise (sharp changes). In the next step, 1 mm h^{-1} is used as a precipitation or no-precipitation condition to remove light precipitation that may be spurious. Removal of precipitation rates < 1 mm h^{-1} affects the precipitation system's area but has little effect on rain volume and other MCSs properties (Rajagopal et al., 2021). As a final pre-processing step, we identify the contiguous areas (blobs) in the precipitation field and normalize each blob by its maximum precipitation rate.

After pre-processing steps, the FiT tracking is run on smoothed and normalized precipitation field. The FiT is a generic algorithm to track objects in a given two- or three-dimensional data. Therefore, we need to determine the optimal value for FiT parameters such as cascading thresholds and separation distance that will track MCSs in the IMERG precipitation field. We performed sensitivity tests and visually analyzed tracking animations over a small region for a short period. The details of these tests are available as supporting material. These tests show that the normalized thresholds (0, 0.11, 0.33) and separation distance of 2° (~200 km) tracked MCSs reasonably well. An animation of FiT's tracking of IMERG precipitation systems over northern Brazil is provided as supporting material to showcase FiT's ability to handle splits and mergers, and the performance of optimal parameter values.

Smoothing reduces the precipitation rates, and normalization changes the precipitation field to values between 0 and 1. Therefore, we use the smoothed and normalized precipitation field to track the precipitation systems and unsmoothed and unnormalized precipitation overlaid on tracked data to compute precipitation systems' properties, such as rain volume and maximum rain rate. Dr. James Russell used the optimal tracking parameters (from sensitivity tests) to track precipitation systems over the global tropics from 2011 to 2020. The tracked objects in the precipitation field and the derived MCSs' properties are stored in a publicly available dataset called Tracked IMERG Mesoscale Precipitation System (TIMPS; Russell et al., 2022).

2.5 TIMPS dataset

On average, there are ~2.3 million tracked precipitation systems per year, but only ~160,000 systems satisfy our MCS criteria and are stored in the TIMPS dataset. A precipitation system is classified as an MCS if it meets all three criteria: i) lives for six hours or longer, ii) attains a maximum area ≥ 3000 km^2 during its lifetime, and iii) has an IMERG pixel with rain rate ≥ 10 mm h^{-1} at least once during its lifetime. These are subjective choices and may affect the quantitative results, but our conclusions will remain the same.

For each MCS, its properties such as area, volumetric rain rate (also referred to as rain volume), maximum rain rate (also referred to as precipitation intensity), weighted centroid, and propagation velocity, are computed at every timestep. Area, A (km^2), is the sum of all pixel areas inside an MCS. It is important to note that IMERG's pixel area changes with latitude, and it is taken into account in an MCS's area calculation. Volumetric rain rate, VRR (km^2 mm h^{-1}), is the sum of the products of pixel rain rate and pixel area. The maximum rain rate, MaxRR (mm h^{-1}), is defined as a maximum pixel rain rate within an MCS object boundary, computed every 30 minutes. Weighted centroid, expressed in latitude and longitude, is the rain rate weighted mean of latitudinal and longitudinal values of all pixels, respectively. The weighted centroid computation also accounts for the zonal boundary at 180°E. Propagation velocity, PV (ms^{-1}), is the ratio of the geodesic distance MCS's weighted centroid moved in consecutive timesteps and duration, i.e., 1800 seconds. Lifetime, L (h), is not stored in the file, but it is computed as half the number of timesteps since each timestep is 30 minutes.

MCSs' properties, such as area, volumetric rain rate, maximum rain rate, and propagation velocity, are available at each timestep and change as an MCS evolves. The property values at each timestep are referred to as instantaneous values and denoted with a subscript "inst". In addition, we also compute lifetime statistics such as a maximum or an average from instantaneous property values; for example, the instantaneous area at each timestep is denoted as A_{inst}, and the maximum area attained during an MCS's lifetime is denoted as A_{Lmax}. We use either instantaneous or lifetime statistics as we see fit to describe the probability and spatial distribution of MCSs' properties.

3 Results

The following subsections discuss the regional variability of MCSs' rain contribution and properties in the global tropics. We ignore the MCSs near tropical cyclones (identified using International Best Track Archive for Climate Stewardship data; Knapp et al. 2010) and MCSs touching the north and south domain boundaries. Therefore, the results near the domain boundaries need to be interpreted accordingly.

Journal of Geophysical Research: Atmospheres

Supporting Information for

Tracking Mesoscale Convective Systems in IMERG and Regional Variability of their Properties in the Tropics

Manikandan Rajagopal[1], James Russell[1, 2*], Gregor Skok[3] and Edward Zipser[1]

[1]Department of Atmospheric Sciences, University of Utah, Salt Lake City, Utah.
[2]University Corporation for Atmospheric Research, Boulder, Colorado.
[3]Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana.

*Currently affiliated to University Corporation for Atmospheric Research, Boulder, Colorado.

Contents of this file

Additional Supporting Information (Files uploaded separately)

Captions for movies S1 to S2

Introduction

The following sections and figures describe the tests performed to determine the optimal parameter values for the Forward in Time (FiT) program to track MCSs reasonably in IMERG. The animations, uploaded separately from this document, showcase the FiT tracking with physical vs. normalized cascading thresholds. Tracking MCSs is more reasonable using normalized thresholds than direct physical thresholds, which sometimes result in premature termination.

S1 Optimal tracking parameters

The FiT tracking parameters need to be tuned to work with IMERG precipitation to track MCSs. Therefore, we performed extensive sensitivity tests and visual analysis of tracking animation to determine optimal values for tracking parameters such as cascading thresholds and separation distance. We tested four choices of normalized cascading thresholds: (0, 0.25, 0.50), (0, 0.11, 0.33), (0, 0.06, 0.25), and (0, 0.04, 0.20), and four choices of separation distance: 1°, 2°, 3°, and 4°. Each cascading threshold choice corresponds to a reduction factor of 2, 3, 4, and 5, respectively. For example, the thresholds (0, 0.25, 0.50) are obtained when the maximum normalized value of 1 is reduced by a factor of 2 to get 0.50. A further reduction by a factor of 2 is 0.25. This technique is similar to Skok et al. (2013) and White et al. (2017), who used fixed physical thresholds (40, 56, 80, 120) mm day^{-1} with a reduction factor of 1.5. Due to systems with low precipitation intensity, they required an additional lower threshold of 24 mm day^{-1} for the northeast Pacific ocean. We do not require such adjustments for different tropical regions or PMW sensors since the normalization will effectively handle it. For the other tracking parameter, separation distance, we decided on four options: 1°, 2°, 3°, and 4°, which are approximately 100, 200, 300, and 400 kilometers, representing mesoscale dimensions.

The four cascading thresholds choices and four separation distance choices give 16 possible combinations. We visually analyzed the tracking animation of each combination to determine the choices that were close to human identification and tracking. Since the visual examination is tedious and time-consuming, we performed this analysis only for 15 days (01 – 15 June 2018) over the tropical Atlantic Ocean (60° W to 10° W and 10° N to 5° S). We prioritized choices that tracked MCSs without premature termination and may miss out on a small cell in the periphery. Missing out on a small cell will slightly affect the system's area and rain volume, but other properties, such as lifetime and propagation velocity, are least affected. However, a premature termination will severely affect properties such as lifetime, propagation velocity, accumulated rain volume, and MCSs frequency. There was no perfect choice, but the normalized thresholds (0, 0.11, 0.33) and separation distance of 2° did relatively well compared to other combinations. Two animations of an MCS over the Amazon basin are provided as supporting materials, each with different tracking parameter values. One showcases the working of the FiT algorithm with optimal values for normalized thresholds (0, 0.11, 0.33) and for a separation distance of 2°. The other animation illustrates the issue of premature termination for fixed physical thresholds (1, 3, 9, 27) mm h^{-1} and a separation distance of 2°.

The visual analysis of tracking animation is still a subjective test. Therefore, we performed sensitivity tests over the global tropics (30° N to 30° S) and tracked the precipitation systems for 24 hours on 100 random days between 2001 and 2020. The sensitivity test is run only for seven combinations of parameters since the tests are computationally expensive. We varied the cascading threshold in the first sensitivity test but used the same separation distance of 2° (Fig. S1a). Similarly, we varied the separation distance in the second sensitivity test but used the same cascading thresholds (0, 0.11, 0.33) (Fig. S1b). We assess the choices for tracking parameters by comparing the precipitation object or system's longitudinal width (W), as shown in Fig. S1. The longitudinal width of an object or a system is the longitudinal difference between the westernmost and the easternmost grid cell. If a precipitation system has multiple pieces after splitting, then the longitudinal width is computed for the entire group.

Fig. S1a shows that the conventional tracking method identifies precipitation bands of ~10° longitudinal width (W) at least once every 30 minutes in the global tropics. We imitate conventional tracking by running the FiT algorithm with a single fixed physical threshold of 1 mm h^{-1} and a separation distance of 2°. The use of cascading thresholds does not completely eliminate the identification of precipitation bands but reduces their occurrence from one per half-hour to one per day and increases the number of mesoscale objects (Fig. S1a). Amongst various choices of "cascading thresholds", the values (0, 0.11, 0.33) have a lower frequency of precipitation bands ($W \geq 10°$) by two orders of magnitude than conventional tracking.

When comparing the system's longitudinal width for different choices of separation distance (Fig. S1b), the frequency of precipitation bands ($W \geq 10°$) increased with separation distance because an object that broke off will continue as part of the parent system for large separation distances. Though separation distances of 1° and 2° have a similar precipitation band frequency, the visual analysis of tracking animation show that the separation distance of 2° did better than the other choices. Hence, we chose the cascading thresholds (0, 0.11, 0.33) and separation distance of 2° as optimal tracking parameters.

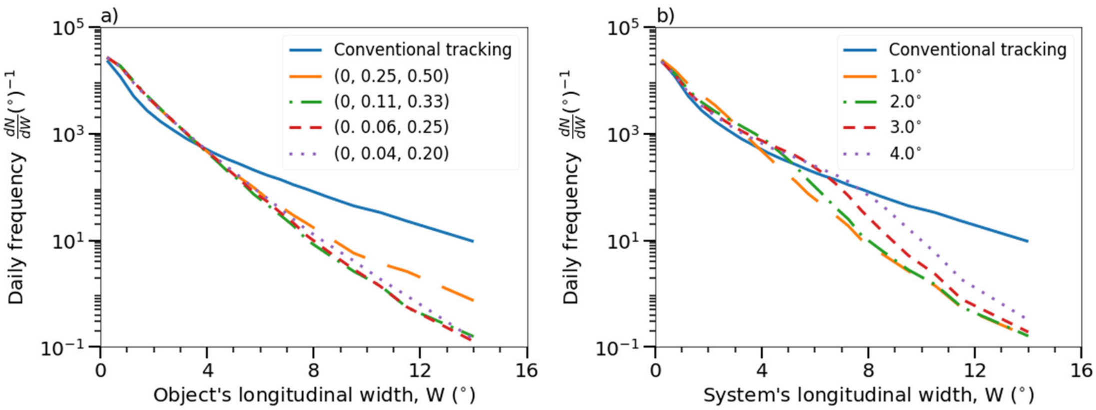

Figure S1. *Sensitivity tests for two major tracking parameters a) normalized cascading thresholds and b) separation distance. The identified object's size and the system's size are expressed as longitudinal width, which is the difference between the easternmost and westernmost pixel longitude values.*

Caption for movie S1

The animation shows the IMERG precipitation and FiT tracking with physical thresholds (1, 3, 9, 27) mm h^{-1} over northern Brazil from 20180604 03:00 20180605 21:00 UTC. One of the MCSs in the region develops as a small system and grows into a large MCS while going through multiple mergers and splits. Different colored contours represent different FiT objects. The change in contour color for the same blob through animation would imply the termination of a system's tracking and the start of a new system.

Caption for movie S2

Same as movie S1, but with normalized thresholds (0, 0.11, 0.33). The normalized thresholds track the MCS reasonably without premature termination.