https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

CoCoMET v1.0: A Unified Open-Source Toolkit for Atmospheric Object Tracking and Analysis

Travis Hahn[1,*], Hershel Weiner[2,*], Calvin Brooks[3], Jie Xi Li[4], Siddhant Gupta[5], and Dié Wang[6]

[1]Department of Statistics, The Pennsylvania State University, University Park, PA 16802
[2]Department of Physics and Astronomy, University of Hawaii at Manoa, Honolulu, HI 96822
[3]Physics, Applied Physics, and Astronomy Department, Rensselaer Polytechnic Institute, Troy, NY 12180
[4]Applied Mathematics & Statistics, Stony Brook University, Stony Brook, NY 11794
[5]Environmental Sciences Division, Argonne National Laboratory, Lemont, IL 60439
[6]Environmental and Climate Sciences Department, Brookhaven National Laboratory, Upton, NY 11937
[*]These authors contributed equally to this work.

Correspondence: Dié Wang (diewang@bnl.gov)

Abstract. Advances in performance and analysis capabilities have accelerated the development of object tracking algorithms for atmospheric research. This has resulted in a growing number of studies using Lagrangian tracking techniques to analyze the evolution of atmospheric phenomena and the underlying processes. However, the increasing complexity and variety of tracking algorithms present a steep learning curve for new users and make it difficult for existing users to compare algorithm performance.

We introduce CoCoMET (Community Cloud Model Evaluation Toolkit), an open-source toolkit that addresses these issues. CoCoMET simplifies the process of running multiple tracking algorithms simultaneously and analyzing objects in both model and observational datasets by specifying parameters in a single configuration file. It standardizes input data from different sources into a consistent format and unifies the tracking output across algorithms. CoCoMET enhances the functionality of existing tracking methods by calculating additional properties such as cell growth and dissipation rates, perimeter, surface area, convexity, and irregularity. In addition, CoCoMET includes a novel method for identifying mergers and splits in 2D and 3D tracks and supports the integration of external Eulerian/stationary datasets for process studies. Its potential utility is demonstrated through examples of model intercomparison, model evaluation against observations, and comparisons between tracking algorithms. Designed for open-source environments, CoCoMET will continue to expand with future releases, incorporating more input data types and tracking algorithms.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

1 Introduction

For decades, object-tracking algorithms based on a Lagrangian framework have been used to identify and study meteorological phenomena. These algorithms enable users to link objects along their trajectories, allowing for detailed analysis of their evolution over time, which accelerates the process-level understanding of tracked systems (e.g., Leese et al., 1971; Dixon and Wiener, 1993; Johnson et al., 1998; Machado et al., 1998; Fiolleau and Roca, 2013). In recent years, the atmospheric sciences community has widely embraced these tracking algorithms, thanks to the growing availability of open-source tools (Table 1). The types of meteorological phenomena they are capable of tracking are getting broader as well, including individual convective clouds, mesoscale convective systems (MCSs), tropical and extratropical cyclones, atmospheric rivers, and equatorial waves (e.g., Prein et al., 2023).

Tracking algorithms often share similarities in certain aspects but differ significantly in others, as they might be originally designed to track different types of phenomena. These differences are particularly evident in the trackers' input data requirements, thresholds, and internal modules, which handle essential tasks such as object segmentation, linking objects across time steps, identifying splits and mergers, and enabling three-dimensional tracking (e.g., Prein et al., 2024; Feng et al., 2024). The performance of these algorithms can be optimized by applying them to datasets tailored to their specific modules, functions, and tuning parameters. For instance, the Tracking and Object-Based Analysis of Clouds (*tobac*) algorithm (Sokolowsky et al., 2024) has shown robust performance in tracking individual or isolated convective clouds (e.g., Oue et al., 2022; Gupta et al., 2024). In contrast, the Python FLEXible object TRacKeR (PyFLEXTRKR; Feng et al., 2023) was developed for tracking MCSs (Feng et al., 2021; Hayden et al., 2021; Cui et al., 2024; Robledo et al., 2024). Meanwhile, more versatile algorithms, such as the Multi-Object Analysis of Atmospheric Phenomena (MOAAP) algorithm (Prein et al., 2021), are designed to track a broad range of features, including MCSs, atmospheric rivers, and synoptic troughs.

Variations between algorithms can pose significant challenges for users trying to select the most suitable tracker for a specific environmental region or scientific question. Identifying the optimal tracker often demands a deep understanding of each algorithm's nuances, strengths, and limitations. While users may rely on suggestions from previous studies, it is more pedagogical to compare the results produced by different trackers and make an informed decision (e.g., Prein et al., 2024; Feng et al., 2024). However, this process is often prohibitively challenging and requires collaborative actions due to the significant computational resources needed and effort required to understand, install, and operate multiple trackers—particularly as new trackers continue to emerge. Unfortunately, familiarity with one tracker rarely translates to ease of use with another. Given that no single tracker can be perfect for all applications, an ensemble tracking approach would offer a more robust solution to mitigate discrepancies that may arise in downstream analyses since it helps account for the inequalities between different trackers.

Another significant challenge is ensuring consistency across data pre-processing, tuning parameters, tracking thresholds, and the calculation of tracked properties. Any mismatches in these aspects may introduce non-physical uncertainties, making direct comparisons of tracking results problematic. To address this, there is a clear need for an open-source toolkit that can

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

simultaneously run multiple trackers, unify thresholds, and standardize the calculation of properties associated with tracked features.

The use of object tracking for model evaluation is gaining popularity, both for model intercomparisons and for comparisons between models and observational data (e.g., Prein et al., 2024; Hahn et al.). However, applying trackers to simulations from different models poses several challenges, including inconsistencies in model output variables and the calculation procedures for observable quantities. For example, the Weather Research and Forecasting (WRF) model (Skamarock et al., 2019) and the Regional Atmospheric Modeling System (RAMS) model (Cotton et al., 2003; van den Heever et al., 2023) differ in key aspects such as variable naming conventions, output hydrometeor classifications, and the methods used to compute essential properties like radar reflectivity or precipitation rate. Similarly, discrepancies between models and observations often arise due to differences in available quantities, temporal and spatial resolution, or spatial coverage. To address these issues, a preprocessing step is crucial to standardize and reformat input data structures before running trackers on different inputs, ensuring compatibility and consistency.

In this study, we develop an open-source Python package, CoCoMET (Community Cloud Model Evaluation Toolkit), to streamline the pre-processing of model and observational data as inputs, enable simultaneous execution of multiple trackers, and standardize the analysis of tracking outputs. One of the key highlights of this package is its simplicity, allowing users to perform all necessary tasks by editing a single configuration text file. This package also includes a newly developed function for detecting 2D and 3D merging and splitting events (Hahn et al.), which can be seamlessly integrated with various tracking algorithms. Additionally, CoCoMET offers an important feature: the ability to link the system life cycle characteristics (tracking results) to their surrounding environmental conditions (e.g., sounding data), which is particularly useful and recommended for the studies of aerosol-environment-cloud interactions (e.g., Veals et al., 2022; Wang et al., 2024)

2 Overall Structure

CoCoMET operates by internally managing all data transformations and necessary variable calculations to prepare and execute various trackers, with users having the option to tune the corresponding parameters. It also computes key characteristics of tracked cells in a parallelized environment to enhance computational efficiency. The framework consists of four main components: input pre-processing, tracker implementation, output unification, and analysis functions (Figure 1), while also providing the flexibility to incorporate additional modules and extend functionality within each component. Note that here, 'feature' refers to a tracked object in each frame or time step, while 'cell' refers to all 'features' linked across frames, forming a lifecycle.

2.1 Input Pre-processing

CoCoMET can accommodate both observational datasets and model simulations as inputs for tracking features. Observational datasets supported by the current version of CoCoMET include brightness temperature from the Geostationary Operational Environmental Satellites (GOES), radar reflectivity from the Next Generation Weather Radar (NEXRAD) system, and any other gridded radar datasets. The required data format for the gridded radar datasets is explained in Appendix A. Additionally,

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

the package provides a function to handle input data from multiple adjacent NEXRAD radars to track organized convective clouds that cover a large domain. The NEXRAD radar data gridding is performed using the Py-ART open-source package (Collis and Helmus, 2013). CoCoMET is designed with flexibility to integrate other gridded observational data streams, such as the Stage IV precipitation dataset (Lin and Mitchell, 2005) and the European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA5; Hersbach et al., 2020), in the future as needed.

For model simulations, users can run built-in functions to precompute commonly tracked variables such as brightness temperature, precipitation rate, and radar reflectivity, if these variables are not directly available within the standard model output. Currently, CoCoMET supports three model outputs: WRF, RAMS, and the non-hydrostatic mesoscale atmospheric model (MesoNH; Lac et al., 2018). The package is designed to facilitate the integration of additional models in the future, such as the Simple Cloud-Resolving Energy Exascale Earth System (E3SM) Atmosphere Model (SCREAM; Donahue et al., 2024) and the ICOsahedral Non-hydrostatic model (ICON; partnership, 2024).

Given the variations in variable names, formats, dimensions, and sometimes even definitions across different models, each dataset undergoes a dedicated pre-processing step for standardization before being passed to the trackers. Users simply need to provide the model simulation data to CoCoMET and specify the field they want to track in a configuration file for the trackers. For commonly tracked variables such as vertical velocity, brightness temperature, precipitation rate, and radar reflectivity, standardized input names – ”wa”, “tb”, “pr”, and “dbz”, respectively - are assigned regardless of the type of numerical model. Users can also track any variable in the simulations by specifying its original name in the configuration file. No additional steps are required to prepare input files. We detail the standardization process for each model output in the following sections.

2.1.1 RAMS

RAMS is a highly versatile numerical model developed at Colorado State University for simulating and forecasting meteorological phenomena. It consists of three main components: an atmospheric model for executing the simulations, a data analysis package for processing initial meteorological data, and a post-processing tool for visualizing and analyzing model output (Cotton et al., 2003).

The raw output data for each RAMS simulation case consists of two files: a netCDF file containing the numerical data for simulated variables and a text file storing metadata for the simulation (e.g., variable dimensions, model grid spacing). A more detailed description of RAMS can be found at https://rams.atmos.colostate.edu/detailed.html. A separate function in CoCoMET (`rams_configure.py`) is implemented to parse the metadata file and extract essential information required for tracking, such as the simulation start time, variable dimensions, grid spacings, and map projection.

Radar reflectivity is computed following the methods outlined in the RAMS source code (https://github.com/RAMSmodel/RAMS/tree/main) which assumes the hydrometeors (rain, pristine ice, snow, aggregates, graupel, hail) are spherical and their number concentrations are represented by gamma distributions. Brightness temperature is computed from the outgoing longwave radiation using the Stefan Boltzmann law (Yang and Slingo, 2001), and the same method also applies to WRF and Meso-NH simulations.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

The precipitation rate can be calculated using different settings, determined by the user's input for `calculation_type` in the CoCoMET configuration file. If `calculation_type` is set to "surface time averaged precipitation rate" (default), the change in the sum of all surface accumulative hydrometeor mixing ratio rates (rain, pristine ice, snow, aggregates, graupel, hail, and drizzle) over consecutive model time steps is used for tracking. If set to "surface instantaneous precipitation rate", the sum of all surface hydrometeor rates simulated at each model time step is used for tracking. Finally, for "volumetric instantaneous precipitation rate", the sum of all 3D hydrometeor rates derived at each model time step is used for tracking.

2.1.2 WRF

The WRF model is a widely-used mesoscale numerical weather prediction system developed for both atmospheric research and operational forecasting (Skamarock et al., 2019). It includes two dynamical cores, a data assimilation system, and a software framework designed for parallel computing and system expansion. The model is applicable to a broad spectrum of meteorological studies, covering spatial scales from a few meters to thousands of kilometers. The output of WRF is in netCDF format, containing both numerical variable data and metadata.

In addition to the direct outputs of WRF, which may be selected for tracking (e.g., updraft velocity and rain mixing ratio), we also implement functions to calculate radar reflectivity (when it is not directly output by WRF) using the methodology from Koch et al. (2005). The calculation is based on the `wrf.dbz` function in the wrf-python open-source package (Ladwig, 2017). The hydrometeors involved in the calculation include rain, snow, and graupel. In the next version of CoCoMET, we plan to replace the current equations used for calculating radar reflectivity with a radar simulator (e.g., The Cloud-resolving model Radar SIMulator; CR-SIM; Oue et al., 2020). This upgrade will extend the reflectivity calculation's applicability to simulations using various microphysical schemes (different hydrometeor types), enhancing CoCoMET's versatility.

The precipitation rate from WRF at a given time step is calculated using the methodology used to describe "surface time averaged precipitation rate" in RAMS, and is provided as the input "pr" to CoCoMET. It is estimated from the difference in accumulated precipitation between consecutive time steps. The accumulated precipitation is the sum of RAINC (accumulated convective precipitation) and RAINNC (accumulated grid-scale precipitation).

2.1.3 Meso-NH

Meso-NH is a non-hydrostatic mesoscale atmospheric model developed by the French research community (Lac et al., 2018). The model supports simulations across a wide range of scales, from large eddy to synoptic, with advanced physical parameterizations for clouds and precipitation. Coupled with the SURFEX surface model, it represents surface-atmosphere interactions across different surface types. Meso-NH features grid nesting for multi-scale simulations, operates in 1D, 2D, or 3D, and it includes a chemistry module and a lightning module.

Meso-NH outputs data in netCDF format, similar to WRF, containing both numerical variables and metadata. Radar reflectivity is calculated in the same way as WRF with plans to upgrade using CR-SIM. Unlike RAMS and WRF, Meso-NH provides output variables for surface precipitation rate, which can be directly used for tracking precipitation cores.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

2.2 Implemented Trackers

The current version of CoCoMET supports three tracking algorithms — *tobac* (v1.5.3), MOAAP (v1.1.1), and TAMS (v0.1.5). These trackers were prioritized based on their openly accessible source codes, demand from collaborators and users, and their familiarity to the CoCoMET developers. Nevertheless, CoCoMET is designed with a flexible, modular structure (Figure 2) to facilitate the incorporation of additional open-source, Python-based trackers in future releases based on user feedback and research needs. The independent integration of multiple trackers within CoCoMET ensures that the unique features of widely-used trackers are incorporated into the package. PyFLEXTRKR, ATRACKCS (An algorithm for TRACKing Convective Systems), and other open-source trackers listed in Table 1 are the next candidates for integration. The open-source nature of CoCoMET enables users to suggest updates whenever an incorporated tracker releases a new version. Future versions of CoCoMET will integrate any updated tracker versions to ensure the package remains current.

2.2.1 *tobac*

Tobac is a Python-based, open-source framework designed to identify, track, and analyze atmospheric features in both 2D and 3D datasets. The package is designed in a modular framework with three main steps. In the first step, regions satisfying progressively restrictive thresholds (e.g., 30 dBZ, 40 dBZ, 50 dBZ; Gupta et al., 2024) in a given variable field are identified as 'features'. Feature identification is followed by a 'segmentation' step wherein the area (for 2D datasets) or volume (for 3D datasets) associated with each feature is estimated at each time step. Finally, features identified across successive time steps are linked based on a search within a defined radius of their projected positions. More specific details associated with each step are provided by Heikenfeld et al. (2019) and Sokolowsky et al. (2024).

Recent updates in *tobac* v1.5.3 (Sokolowsky et al., 2024) introduced an increase in computational efficiency, tracking of 3D features, handling of feature splits and mergers, internal spectral filtering, and support for periodic boundary conditions, making it more robust for analyzing atmospheric data. *Tobac* has been widely used for tracking convective clouds, updrafts, precipitation systems, and other meteorological phenomena in both model simulations and observational datasets (e.g., Oue et al., 2022; Kukulies et al., 2023; Gupta et al., 2024). We ensure that unique elements of *tobac*, such as parallel computing and its ability to track any type of feature within a gridded field (even non-meteorological ones), are incorporated into CoCoMET. This tracker is integrated with all input data streams implemented in the current version of CoCoMET.

2.2.2 MOAAP

The MOAAP algorithm is designed to identify and track a wide range of atmospheric features within a unified framework. It processes both regional and global datasets to detect and analyze phenomena such as tropical and extratropical cyclones, mid-level cyclones, cut-off lows, anticyclones, atmospheric rivers, jet streams, fronts, MCSs, and tropical waves (Prein et al., 2023).

MOAAP begins by applying thresholding methods to classify grid points corresponding to specific atmospheric phenomena. It then generates label maps that delineate the spatial extent of each identified feature, with unique identification criteria tailored

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

to each type of feature (e.g., vorticity for cyclones, moisture flux for atmospheric rivers). Once features are identified, MOAAP employs a nearest-neighbor approach and spatial overlap techniques to track the movement of features across consecutive time steps. The linking process primarily relies on a simple connectedness principle, i.e. adjacency in space and time between detected features. Once features are initially detected, that information is fed to a two-pass binary connected-component labeling algorithm (Dillencourt et al., 1992) for linking. MOAAP accounts for features that merge, split, or dissipate during the tracking process, maintaining accuracy in complex scenarios. After tracking, MOAAP generates comprehensive outputs that include all detected features and their trajectories.

This tracker is currently integrated with only three types of model outputs in CoCoMET, as it requires a set of variables (see https://github.com/AndreasPrein/MOAAP/wiki) to track atmospheric phenomena that are often unavailable in a single observational dataset. In the next version of CoCoMET, we plan to incorporate multiple data sources (e.g., GOES plus Stage IV) to enable MOAAP analysis using observations.

2.2.3 TAMS

The Tracking Algorithm for Mesoscale Convective Systems (TAMS) is a Python-based tool which operates in a series of steps: identification, tracking, classification, and variable assignment. It is primarily used to analyze organized cloud systems in both observational and model data using Infrared brightness temperature (Tb) or cloud top temperature (Moon and Ocasio, 2024). TAMS can handle data from both structured and unstructured grids, which makes it adaptable for various types of input datasets.

By default, TAMS identifies cloud elements by applying the following threshold criteria: an edge Tb or cloud top temperature below 235 K (a free parameter in TAMS) within an area larger than 4,000 km^2 (a fixed parameter) and an embedded cold core with a Tb or cloud top temperature below 219 K (a free parameter) that covers an area larger than 10 km^2 (a fixed parameter) and is detected at least once during the lifetime of the MCS.

After identifying the cloud elements, TAMS uses backward linking to group cloud elements that show significant overlap with projected elements from the previous time steps, creating MCS tracks. TAMS classifies each tracked event into different categories based on the size, structure, and longevity of the tracked MCSs. The detailed explanation of each category can be found in Moon and Ocasio (2024). Note that this tracker is currently integrated with three types of model outputs and GOES observations in CoCoMET.

2.3 Output Unification

Trackers often share common outputs, such as feature IDs (identified at a specific point in time), cell IDs (the full spatial and temporal extent of a tracked system or a series of linked features across multiple time steps), cell location, start time, and end time. However, how other cell characteristics are defined, computed, formatted, and stored (e.g., csv, NetCDF) can vary significantly across trackers. Additionally, not all trackers have the capability to identify merging and splitting events. These inconsistencies present a significant challenge when analyzing and integrating results from different tracking algorithms.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

To overcome this, CoCoMET reformats universally available output variables (e.g., cell location, timing) from various trackers in a consistent manner. It then standardizes the calculation of key cell properties (e.g., cell height and area) through standalone functions. This approach ensures consistency across different trackers and facilitates robust cloud lifecycle analysis.

Moreover, CoCoMET goes beyond traditional metrics by incorporating additional properties like perimeter and irregularity, providing a more detailed and comprehensive characterization of tracked atmospheric systems. Calculating these additional parameters can expand the convection tracking literature by elucidating key properties such as updraft shape, width, and size, while also addressing research questions related to cloud evolution, entrainment, and mass flux (e.g., Heiblum et al., 2019; Chen et al., 2023a). CoCoMET incorporates a newly developed method for identifying merging and splitting events in both 2D and 3D tracking, addressing a key limitation in many existing trackers. These functions can be executed either after initial tracking has been completed or specified in the configuration file to be applied automatically during the tracking process.

2.3.1 2D Perimeter

The perimeter of a tracked 2D feature (in km) is calculated by identifying and summing the lengths of its edge segments. First, the segmentation mask is used to identify the feature's boundary by checking each grid point in the feature. A point is considered an edge if at least one of its four neighboring points lies outside the feature. The length of each edge segment is calculated using the projection coordinates, then all edge segments are summed to determine the total perimeter. This process is repeated for each feature at each time step. The feature perimeter function is used extensively in the newly developed methodology within CoCoMET to identify splits and mergers between objects tracked in two-dimensional space.

2.3.2 Feature Area

The 2D feature area is determined through segmentation, resulting in a masked array that identifies grid points occupied by the feature at each time step. The area, expressed in square kilometers, is calculated as the product of the number of grid points within the feature and the area of an individual grid box. For 3D tracking, the feature area at a given height level can be calculated by specifying the height (2 km by default) within the configuration file in the segmentation function. This variable is important in multiple aspects of atmospheric sciences. For example, the extent of cloud cover directly influences Earth's energy balance (e.g., Hartmann, 2016; Sokol and Hartmann, 2020). Additionally, the size and distribution of precipitation areas determine the volume and distribution of rainfall, affecting water resource availability (e.g., Yu et al., 2022; Lee et al., 2023).

2.3.3 Feature Surface Area

In 3D tracking, the surface area of each feature is calculated by checking every grid point in the feature segmentation. For each point, the algorithm looks at its six connected neighboring points in the cubic orientation. If any of these neighbors is not part of the feature, the current grid point is considered "exposed", and the area of the exposed side is calculated using the projection coordinates. This process is repeated for all grid points, and the surface areas are summed up. The final result is in

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

square kilometers and is stored for each feature in each time step for a tracked cell. The feature surface area function is used extensively in the newly developed methodology within CoCoMET to identify splits and mergers between objects tracked in three-dimensional space.

2.3.4 Feature Volume

The feature volume can be calculated when tracking cells using 3D data. It is determined by summing the volumes of all the grid boxes that define a feature according to the segmentation results. The final output is provided in cubic kilometers and is stored for each feature in each time step for a tracked cell. This variable has been frequently used to analyze cloud and precipitation characteristics. For instance, Prein et al. (2017) found that a warmer condition may increase the precipitation volume produced by MCSs by up to 80%.

2.3.5 Feature Irregularity

The irregularity or roughness of the edges of a 2D feature is quantified using convexity, a measure of how much a shape deviates from being convex. Convexity is defined as the ratio of the perimeter of the object's convex hull to the perimeter of the object itself. The convex hull is the smallest convex shape that fully encloses the object. A value close to 1 indicates a smooth, compact shape (Figure 3b), while lower values suggest jagged or complex boundaries (Figure 3a). This is useful in identifying cloud structures/morphology in meteorological data or model simulations and the potential influence by entrainment and mixing processes (e.g., Mellado, 2017; Chen et al., 2023b).

For a 3D feature, convexity is similarly used to evaluate the irregularity of an object's surface. Instead of using perimeter, the ratio is calculated using surface area—comparing the actual surface area of the object to the surface area of its convex hull. A value close to 1 indicates a compact, smooth structure, while a lower value suggests the object has indentations or rough, uneven surfaces. In both cases, convexity provides a quantitative measure of shape complexity of tracked clouds and/or system. Although such parameters have gained little popularity, they have the potential to facilitate studies, such as classifying different cloud fields (e.g., Lim and Daya Sagar, 2008) and determining whether an updraft is more thermal-like or plume-like (e.g., Morrison et al., 2020; Peters et al., 2020).

We also provide another way of quantifying the irregularity of a feature in 3D, which is called sphericity. Sphericity is a measure that compares the surface area of a perfect sphere with the same volume to the actual surface area of the feature being analyzed.

Mathematically, sphericity (ψ) is given by the equation:

$$\Psi = \frac{\pi^{\frac{1}{3}}(6V_p)^{\frac{2}{3}}}{A_p} \tag{1}$$

, where V is the volume of the feature, A is the surface area of the feature.

In Equation (1), the numerator represents the surface area of a sphere with the same volume as the feature, and the denominator is the actual surface area of the feature. By comparing these two quantities, sphericity provides a way to quantify how

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

close the feature is to a perfect sphere, with values approaching 1 for more spherical shapes and values much less than 1 for highly irregular shapes. The sphericity is calculated for each feature in each time step for a tracked cell.

2.3.6 Maximum Height of a Variable of Interest

The maximum height of a variable is defined as the 95th percentile of altitudes (in km above ground level) where the selected variable exceeds a specified threshold across the entire segmented feature area.

For 3D tracking, this calculation is straightforward on a 3D variable of interest, as height data is directly available within the segmented feature. For 2D tracking, if the variable of interest is 3D with a vertical dimension, the maximum height can still be determined. The process involves examining the vertical columns within the horizontally tracked feature area and identifying the highest altitudes where the variable exceeds the given threshold (an independent threshold from tracking and segmentation). For example, if tracking is performed based on updraft at 6 km altitude, and a threshold of 5 m/s is given for the calculation of the maximum height, the maximum height of the updraft corresponds to the highest altitude (95th percentile) on the same 2D area as the segmented feature where the updraft velocity surpasses 5 m/s. The strength of convective clouds or the depth of convective cloud cores is often represented using parameters such as the Echo Top Height (e.g., Wang et al., 2019, 2024) which can now be derived automatically using CoCoMET.

2.3.7 Cell Growth and Dissipation Rates

The cell growth and dissipation rates are determined by calculating the rate of change in a targeted cell property (e.g., area for 2D tracking, volume for 3D tracking). For each cell, this property is taken at each frame/time step where the cell exists. The difference in the cell property between consecutive frames is divided by the time interval between the frames to estimate the rate of change in cell property. The cell growth/dissipation rate for the last feature along an object's life cycle are left as NaN. Such estimates can help understand temporal changes in cloud updraft width and the role of entrainment to determine its influence on convective cloud depth (e.g., Varble et al., 2024).

2.3.8 Feature Propagation Velocity

The propagation speed of a tracked cell is determined by estimating its position change over consecutive time steps. For each feature, the velocity is calculated as the displacement of the cell between its current frame and its next frame divided by the time interval. The velocity is output as a unit vector including propagation speed along two (y, x) or three dimensions (z, y, x). The speed (m/s) is output as the magnitude of the original velocity vector. The velocity vector and speed for the last feature along an object's life cycle are left as NaN.

These quantities are particularly useful for examining the relative influence of local and synoptic-level forcings on cloud fields by comparing the propagation speed and direction of individual clouds against the background wind speed and direction (Corfidi, 2003; Zhang et al., 2021).

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

2.4 Identification of Mergers and Splits

Mergers and splits in cloud systems are key to understanding weather development and evolution (e.g., Westcott, 1984; Lu et al., 2022). Mergers can combine smaller cells into larger, more organized storm systems that may lead to increased rainfall or severe weather, while splits can fragment these systems, changing the distribution and intensity of precipitation. Grasping these processes is essential for enhancing weather forecasts and refining Earth system models by improving our understanding of atmospheric dynamics like stability, turbulence, and energy transfer. CoCoMET implements a novel technique for identifying merge and split events during the lifecycle of tracked systems. As not all tracking methods include this capability, integrating a standardized procedure within CoCoMET ensures consistency in detecting such events across different trackers.

2.4.1 Mergers and Splits in 2D

Mergers and splits occur when individual cells either combine into a single cell or break apart into multiple cells. The identification of mergers and splits follows the methodology of Hahn et al., with modifications to improve adaptability to users' needs (Figure 4).

First, we Identify candidate pairs. Features that share a common border (referred to as "touching" features) exceeding a user-defined percentage of their perimeter are flagged as potential merging/splitting features. The default value is set to 20%. Next, we define the search region for feature area calculation. Each feature is approximated as a circle with a radius $r_{feature}$, which is calculated by adding up all "edge" grid cells (those which are not completely surrounded by segmented grid points) to define a circumference and divide this by $2 \times pi$ to get the radius in "grid points." The radius is then expanded to rsearch using a user-defined weighting to define a search region for edge pixels for each feature. By default, the search region is set to 110% of the radius of each feature, making $r_{search} = 1.1 \times r_{feature}$. A square is created with the same center as the circle and with side length $2 \times r_{search}$. Within this square, each grid point is assigned a score based on its proximity to the feature core (point of maximum value) and its intensity relative to a background threshold. The calculation of the score (S) follows the formula below:

$$S_{i,j} = \omega_1 \frac{R_{i,j} - V}{R_{\max} - V} - \omega_2 \frac{\|(i,j) - (i_0, j_0)\|_2}{\sqrt{2} r_{\text{search}}}, \tag{2}$$

where R represents the variable field, R_{max} is the variable value at the feature core, i_0, j_0 is the location of the feature core, V is a new background threshold for the tracked variable, set by the user (20 by default for radar reflectivity input), and ω_1 and ω_2 present adjustable weights that allow users to fine-tune the relative influence of intensity versus distance in determining mergers or splits. The default values for ω_1 and ω_2 are both 1.

If the score for a given feature exceeds the threshold, which is by default 0, it is included as part of a new mask. After applying this process to both features in the pair, the resulting masks are overlaid, and the overlap area is calculated as a percentage of either feature's area mask. Pairs that exceed a certain percentage area overlap threshold (50% by default) move to the final step and are referred to as potential merging or splitting pairs.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

For these potential merging or splitting pairs, we look forward in time over a user-specified number of time steps (2 by default). If one cell persists and the other does not, this is classified as a merge event. However, if both features exist (or both do not exist) in the subsequent time steps, no merging event is confirmed. In cases where neither feature exists anymore, meaning both features are no longer tracked, they may have moved closer to each other but dissipated without merging, or they could have merged into other cells along the way.

At the same time, for these potential merging or splitting pairs, we look backward in time over a user-specified number of time steps (2 by default). If one feature exists consistently across the time steps while the other does not, this is confirmed as a splitting event. However, if both features exist (or both do not exist) in these time steps, no splitting event is recorded.

Unlike the original approach in Hahn et al., which focuses on radar reflectivity only, this version in CoCoMET allows tracking based on any user-selected variable. For variables with inverse behavior (e.g., Tb, where lower values reflect stronger convection), the calculation is adjusted by inverting the variable field before the calculations. Examples of merging and splitting events are shown separately in Figures 5 and 6, respectively.

The output of this function is stored in two Pandas DataFrames (two tables), which record the time of the event (merge or split) and the cell IDs involved. In the merger DataFrame, each row contains a list of the two frames during which the merger occurred, a tuple of the two cell IDs of the parent cells involved, and the cell ID of the merged cell. In the split DataFrame, each row contains a list of the two frames during which the split occurred, the cell ID of the split cell, and a tuple containing the cell IDs of the two child cells.

Overall, this approach provides a flexible and robust way to track structural changes in 2D convective cells while allowing users to customize parameters for different atmospheric variables.

2.4.2 Mergers and Splits in 3D

We apply a similar approach to identifying mergers and splits in 3D as described for 2D data, with adjustments to account for the third dimension. This functionality rarely exists in other trackers.

To speed up the process, before identifying mergers and splits, the dataset is filtered to focus on cells in the time steps immediately before feature disappearance or immediately after feature appearance. These are the only two instances when a merge or split can occur, so only the cells identified at these time steps are analyzed. This filtering step also pre-classifies all cells as being involved in either a merge or a split, which will be used in the final step of the identification process.

Instead of perimeter, we use the surface area of 3D objects to identify touching features at a given time step. In other words, for each feature, we evaluate how much of its surface area is shared with an adjacent feature to determine the potential for a merge/split. Rather than modeling the feature area as a circle in 2D, we use a sphere with an adjustable radius, extended in three dimensions. Similarly to the 2D process, we generate a cube with the same center as the sphere and side length $2 \times r_{search}$. Within the search cube, each grid point is assigned a score calculated using a formula similar to Equation (2) below:

$$S_{i,j,k} = \omega_1 \frac{R_{i,j,k} - V}{R_{\max} - V} - \omega_2 \frac{\|(i,j,k) - (i_0,j_0,k_0)\|_2}{\sqrt{3} r_{\text{search}}}, \tag{3}$$

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

where k is the third dimension. If the score of a grid point exceeds a user-defined threshold, it is considered part of the new mask. The overlap volume between two masks is then calculated to assess whether a merge or split has occurred. If the overlap volume exceeds a user-defined percentage of either mask's original volume, the event is confirmed as either a merge or split, based on the classification from the initial step.

2.5 Linkage to Other Data Sets

2.5.1 Linkage to Eulerian Data Sets

A key enhancement in the CoCoMET package is the integration of Lagrangian and Eulerian datasets for studying clouds and other atmospheric systems. While cell tracking provides a Lagrangian perspective on the lifecycle of a tracked property, not all atmospheric measurements are collected in a manner that is suitable for tracking. Many observations, such as those from the Atmospheric Radiation Measurement (ARM) user facility (Mather and Voyles, 2013), are collected at fixed locations with one dimension (time) or two dimensions (time, height). Examples include but not limited to vertical pointing cloud radars and lidars, aerosol measurements at the surface or along a tethered balloon system, and precipitation measurements (disdrometers). To fully leverage these diverse datasets, it is essential to link cloud lifecycle stages with time-height observations, enabling a more comprehensive analysis of cloud structure and evolution (e.g., Gupta et al., 2024; Wang et al., 2024).

To achieve this, we developed a function (`extract_arm_product`) that extracts Eulerian measurements at each time step of the tracked cells. This function aligns the tracking frames with their corresponding timestamps and matches them to the nearest available Eulerian data time. It also calculates the time difference (`time_delta`) between the tracked cell's timestamp and the Eulerian data time, where positive values indicate Eulerian data recorded after the tracked frame and negative values indicate earlier measurements. Additionally, the function identifies the closest feature and cell to a specified ground-based measurement site or location, providing their respective IDs along with the distance between the Eulerian measurement site and the nearest tracked feature in kilometers. Such capabilities can easily be extrapolated to include distance from observations collected using mobile platforms like research aircrafts or Uncrewed Aerial Vehicles (UAVs).

2.5.2 Linkage to the Environmental Conditions

To study aerosol-cloud-environment interactions and related processes, it is essential to link cloud properties with their surrounding environmental conditions. One key dataset we incorporate is the ARM INTERPSONDE data (Fairless et al., 2021), from which we derive convective indices such as Convective Available Potential Energy (CAPE), Convective Inhibition (CIN), and low-level wind shear (0-5 km), following the methodology of Wang et al. (2020).

These indices can be calculated using parcel theory under different assumptions, considering both irreversible pseudo-adiabatic and reversible moist adiabatic ascents. In the pseudo-adiabatic process, we assume an undiluted parcel ascent while neglecting hydrometeor loading. Users can also choose to include ice-phase processes, which introduce additional buoyancy above the melting level due to latent heat release during freezing. The function also allows users to specify initial parcels, with choices among the most unstable parcel, surface parcel, and mixed layer parcel. The surface-based parcel is defined as the

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

parcel at the lowest sounding data level; the most unstable parcel is defined as the parcel that has the greatest virtual temperature in the lowest levels above surface (700 mb as default and can be changed by users); the mixing-layer parcel is defined as the parcel with properties of the mean of the user-defined boundary layer (500 m as default).

Such capabilities can easily be extended to include environmental parameters derived from other datasets, such as ERA5 reanalysis, which will enrich the analysis and fill the spatial coverage gap between the tracking results and single point measurements.

3 Come to Practice

3.1 Configuration File

One of the key features of CoCoMET is its ability to seamlessly run multiple trackers or process multiple input datasets simultaneously with a single line of code or editing a single configuration file (referred to as CONFIG). Users only need to specify the necessary parameters in the CONFIG. This file defines essential settings for CoCoMET, including selected trackers, input data directories, and parameters for subsequent cell analysis.

Additionally, users have the flexibility to customize which cell properties will be included in the output. If certain cell properties are not specified in the CONFIG, users have the flexibility to compute them after running CoCoMET by calling the corresponding functions. This allows for a more streamlined workflow, enabling users to focus on essential parameters during the initial run while retaining the option to derive additional properties as needed. This approach enhances adaptability, ensuring that users can tailor their analysis without rerunning the entire tracking process.

Each tracker's inherent parameters are fully configurable, allowing users to fine-tune their settings based on specific research needs. To enhance computational efficiency, CoCoMET supports parallel processing, enabling faster data processing when multiple cores are available.

The CONFIG can be stored in either a .yaml format or as a Python dictionary object. Example `boilerplate.yml` files are provided in the CoCoMET repository and also in Weiner et al. (2025). A detailed breakdown of each parameter and its function within the CONFIG is summarized in Table 2 and Figure 2.

3.2 Output Formatting

Another key feature of CoCoMET is its standardized output structure, which ensures that outputs from all trackers can be uniformly converted for additional analysis and intercomparison. The output from any given tracker is a Python object—an instance of a class containing data and methods—with three components: feature identification (returned as a GeoDataFrame), linking (returned as a GeoDataFrame), and segmentation (returned as an Xarray Dataset). Each row in a GeoDataFrame corresponds to a single detected feature, maintaining a consistent format across different tracking methods. Additionally, a single Python object of class `Analysis_Object` is returned to the user allowing for the post-hoc calculation of additional analysis

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

variables if desired. This uniform structure enhances the ease of subsequent analyses, allowing users to compare results across different tracking approaches efficiently.

3.3 Examples using CoCoMET

3.3.1 Model Intercomparison

CoCoMET facilitates model intercomparison studies by enabling the comparison of different numerical model outputs for the same event. An example of this is illustrated in Figure 7, where we compare tracked 3D updrafts from outputs of WRF and RAMS for a case that occurred on June 19, 2013 over the Houston region. The thresholds for defining an updraft are set at 3 m/s, 5 m/s, and 10 m/s, and we use *tobac* for the tracking process. The temporal resolution of the model outputs is 5 minutes, with simulations running for a duration of 4 hours from 1600 - 2000 local time. The details of this case and the corresponding model setups are presented in Marinescu et al. (2021).

We present histograms of eight selected updraft characteristics in Figure 7. The maximum volume, maximum updraft velocity, and maximum updraft height represent the maximum values of these variables captured during the cell's lifetime. The remaining variables are feature-based, meaning that the values for each individual feature within the cell are plotted. In general, both models exhibit similar statistics for most variables although small differences appear for some variables. While the scientific reasons behind these differences are beyond the scope of this study, our primary goal is to demonstrate the utility of CoCoMET in facilitating this type of comparison. The CONFIG is available in Weiner et al. (2025). After running this CONFIG, users will obtain outputs from *tobac* for three inputs simultaneously without additional steps.

3.3.2 Model Evaluation

For model evaluation, conventional methods typically compare bulk properties of clouds or other atmospheric variables; however, there is an increasing trend toward comparing properties at the individual cloud level, particularly through the analysis of cell life cycles. This approach not only provides a more detailed assessment of clouds but also proves invaluable when evaluating regional and Earth system models, especially in relation to the diurnal cycle of cloud initiation and evolution on a global scale.

We have utilized CoCoMET to evaluate multi-case ensemble simulations of sea breeze convection days over the Houston region, as demonstrated in Hahn et al.. This example highlights the effectiveness of CoCoMET in facilitating such comparisons.

In Figure 8, we provide another example of evaluating RAMS simulations used in Figure 7 with NEXRAD observations for the same case. *Tobac* is employed for tracking, with tracking performed on radar reflectivity at 2 km height. The thresholds for defining tracked features are set at 30, 40, 50 dBZ. The horizontal grid spacings of the NEXRAD data and the RAMS simulations are both 1 km. The temporal resolution of both datasets are similar, around 5 minutes. In Figure 8, both RAMS and NEXRAD show peak cell area and reflectivity during the mature stage of convection (around lifecycle bin 4), as expected (e.g., Gupta et al., 2024). The cell area growth rate transitions from positive in the earlier stage to negative in the later stage, which

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

is consistent between observations and simulations for this particular case. These results give us confidence in the tracking method and the implementation in CoCoMET.

This example is achieved by specifying parameters in a single CONFIG (available in Weiner et al. (2025)), which highlights its streamlined workflow and allows for rapid analysis. This makes it a powerful tool for both single-case and multi-case evaluations with observational data, identifying key differences and similarities between the two, ultimately accelerating model evaluation and development. This time-efficient processing also ensures that comparisons can be made over extended periods, which is particularly valuable for long-term studies, such as those involving seasonal or annual model performance.

3.3.3 Tracker Intercomparison

Differences in tracked cell properties are expected when using different trackers due to the varying designs and underlying algorithms of these methods. Whether such differences significantly affect the conclusions drawn in studies related to the tracked systems is worth exploring further. One approach to quantify these differences is to track the same system using multiple trackers and analyze how variations in tracking results influence subsequent findings. This process can help improve the robustness of scientific conclusions.

CoCoMET is specifically designed to simplify this task by allowing users to configure and execute multiple trackers simultaneously through specifications in the configuration file. The toolkit returns results from all selected trackers in a standardized format, making it easier to compare and interpret differences.

We illustrate this process with an example in Figure 9, where we track brightness temperature from CONUS404 WRF simulations (Rasmussen et al., 2023) using two thresholds, 219 K and 235 K, with three trackers: TAMS, MOAAP, and *tobac*. This case occurred on June 19, 2013, and the trackers were applied over a 24-hour period. The CONUS404 simulations are performed using WRF version 3.9.1.1, with horizontal grid-spacing of 4 km and temporal resolution of one hour. Figure 9 shows the initiation location and timing (colors) of tracked cells for all trackers.

Overall, all trackers identify cells in the same general region, but *tobac* detects more cells compared to TAMS and MOAAP. This difference arises because TAMS and MOAAP are specifically designed to track large, organized systems, whereas *tobac* is more focused on identifying isolated convective cells. Compared to TAMS, MOAAP applies additional thresholds internally which limits the number of cases tracked. Despite the difference in the number of tracked cells, the initiation timings are well captured by all trackers, demonstrating consistency in some aspects across the methods. The CONFIG file for running CoCoMET is available in Weiner et al. (2025).

4 Future Development Plan

To enhance the functionality and applicability of CoCoMET, we propose the following key areas for future development:

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

4.1 Expanding Tracking Capabilities

We will integrate additional, existing tracking algorithms such as PyFLXTRKR and TempestExtremes, to facilitate intercomparison among additional trackers and the use of model-ensemble approach to represent uncertainties in the subsequent analyses. We are exploring the development of a machine learning-based tracking algorithm to enhance cloud detection and tracking efficiency, accuracy, and adaptability.

Another key focus is the development of tracking methods that incorporate multiple atmospheric variables to provide a more comprehensive view of cloud and precipitation systems. Furthermore, we aim to implement multi-feature tracking (e.g., updrafts and precipitation) to better understand interactions between different atmospheric processes. On top of these, we plan to develop tracking capabilities for unstructured grid inputs, such as those used in the Model for Prediction Across Scales (MPAS) model (Heinzeller et al., 2016).

4.2 Increasing Input Data Support

We will extend compatibility with additional ARM scanning radar datasets and radars from other agencies to improve cloud and precipitation tracking. Additionally, we will integrate other global and regional observational precipitation datasets, such as Stage VI, to enable broader applications and better suit for global model evaluation. To accelerate the tracking process and improve user accessibility, we will develop streamlined methods for downloading and preprocessing various input datasets.

We also plan to incorporate additional model outputs, including those from ICON and SCREAM to further enhance CoCoMET's capability for facilitating model intercomparison studies (e.g., Marinescu et al., 2021).

4.3 Enhancing Linkages to External Datasets

Understanding the large-scale regime of tracked events/days is crucial for studying cloud-environment interactions and the related studies. To achieve this, we will incorporate synoptic weather regime classification products from ARM (Wang et al., 2022; 2024), based on self-organizing maps, to better link tracked cells to synoptic conditions. Additionally, we will link tracking results with low-orbit satellite data (such as the Cloud, Aerosol and Radiation Explorer EarthCare; Wehr et al., 2023) to assign lifecycle stages to clouds sampled at the time of the satellite overpasses at locations of interest. We will also integrate ERA5, AIRS (Atmospheric InfraRed Sounder; Tian et al., 2019), or other thermodynamic datasets to extract environmental conditions for tracked clouds, further supporting studies on cloud-environment interactions.

4.4 Visualization Enhancements

We will improve visualization tools for tracking outputs, including interactive maps and 3D representations. Additionally, we will develop user-friendly interfaces for exploring and analyzing tracking results with linked environmental data.

These future developments will significantly enhance CoCoMET's capabilities, making it more versatile for cloud, precipitation, and atmospheric studies.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

5 Conclusions

CoCoMET is a Python-based open-source package designed to facilitate the evaluation of cloud properties in both observational data and model simulations. With the growing interest in understanding cloud lifecycle characteristics and improving model representation of convection, CoCoMET addresses the challenge of efficiently conducting standardized, object-based comparisons across different datasets and trackers. The toolkit integrates multiple cell-tracking methods and supports data from various models (e.g., WRF, RAMS, MesoNH) and observations (e.g., radar, satellite).

CoCoMET provides a unified framework for defining and calculating fundamental cell characteristics, such as cell strength, size, height, and lifespan, ensuring consistency across datasets and trackers. The package's modular design allows users to easily configure and analyze single-platform or multi-platform datasets using a single configuration file. It provides developers with the flexibility to add new functions and incorporate additional input data streams in the future. The design of CoCoMET emphasizes computational efficiency, making it possible to perform long-term analyses or ensemble-based studies on large datasets, which are often limited by computational expense.

Finally, this manuscript highlights several case studies, including a model intercomparison using simulations of convective cells over Houston, an evaluation of WRF simulations against ground-based scanning radar observations, and a comparison between different trackers using CONUS404 simulations. These examples demonstrate CoCoMET's ability to streamline workflow, quantify differences, and identify key patterns of tracked cell properties. The package accelerates model evaluation and development by providing a robust, scalable, and time-efficient solution for object-based analysis.

Code availability.

The source code for the CoCoMET v1 package is available at https://github.com/ASCENT-BNL/CoCoMET and https://doi.org/10.5281/zenodo.15090741 (Hahn et al., 2025).

The configuration files for running CoCoMET for results in Section 3 are available at https://doi.org/10.5281/zenodo.15048051 (Weiner et al., 2025).

Data availability.

WRF simulation data used in Figures 5 and 6 can be downloaded from https://app.globus.org/file-manager?origin_id=0c079436-56af-11ed-b805-855d8beae885&origin_path=%2F&two_pane=false (Prein et al., 2022; Ramos-Valle et al., 2023).

Model outputs used in Figures 7 and 8 can be accessed in the U.K. CEDA JASMIN supercomputer following processes listed in this document: http://acpcinitiative.org/Docs/Instructions_Jasmin_Workspace_171011.pdf (Marinescu et al., 2021).

CONUS404 data used in Figure 9 can be downloaded from https://www.usgs.gov/data/conus404-four-kilometer-long-term-regional-hydroclimate-reanalysis-over-conterminous-united (Rasmussen et al., 2023).

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

Appendix A: CoCoMET Gridded Radar Data Requirements

To use gridded radar data as input in CoCoMET, the data must meet the following requirements (an example is available in the GitHub repository under `/examples/example_radar_standardized.py`):

A1 Data Format

- The data must be an xarray DataArray named `"reflectivity"`.

A2 Required Dimensions (in the following order)

- **time**: A `numpy.datetime64` list of radar scan times.
- **z**: A list of altitudes in meters above the radar, with attributes:
  - `standard_name`: `"altitude"`
  - `units`: `"m"`
- **y**: A list of y indices (e.g., `[0, 1, 2, ..., 500]`).
- **x**: A list of x indices (e.g., `[0, 1, 2, ..., 500]`).

A3 Required Coordinates (in no particular order)

- **proj_y** (follows `y` dimension): A list of y distances in meters from the radar (e.g., `[-25000, ..., 25000]`).
- **proj_x** (follows `x` dimension): A list of x distances in meters from the radar (e.g., `[-25000, ..., 25000]`).
- **south_north** (follows `y` dimension): A list of y indices, identical to the `y` dimension.
- **west_east** (follows `x` dimension): A list of x indices, identical to the `x` dimension.
- **model_level_number** (follows `z` dimension): A list of z indices (e.g., `[0, 1, 2, ..., 40]`).
- **altitude** (follows `z` dimension): A list of altitudes in meters above the radar, identical to the `z` dimension (attributes are not necessary).
- **lat** (follows `y, x` dimensions): An array of latitudes with attributes:
  - `standard_name`: `"latitude"`
  - `units`: `"degree_N"`
- **lon** (follows `y, x` dimensions): An array of longitudes with attributes:
  - `standard_name`: `"longitude"`
  - `units`: `"degree_E"`

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

A4 Required DataArray Attributes

- `long_name`: `"Reflectivity"`
- `units`: `"dBZ"`
- `standard_name`: `"equivalent_reflectivity_factor"`

Author contributions. TH: Coding, conceptualization, validation, writing; HW: Coding, conceptualization, validation, writing; CB: Coding, writing; JXL: Coding, writing; SG: Writing; DW: Coding, conceptualization, funding acquisition, supervision, writing.

Competing interests. The contact author has declared that none of the authors have any conflicts of interest.

Acknowledgements. This project was supported by the U.S. Department of Energy (DOE) Early Career Research Program, Atmospheric System Research (ASR) program, and the Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). This paper has been authored by employees of Brookhaven Science Associates, LLC, under Contract DE-SC0012704 with the U.S. Department of Energy (DOE). SG is supported by Argonne National Laboratory under U.S. DOE contract DE-AC02-06CH11357 and the ARM User Facility, funded by the Office of Biological and Environmental Research in the U.S DOE Office of Science.

We would like to acknowledge Dr. Aryeh Drager from Brookhaven National Lab for his help with the implementation of RAMS outputs and the RAMS developer team at Colorado State University for providing RAMS simulations. We also acknowledge the deep convection model intercomparison project (MIP) of the Aerosol, Cloud, Precipitation and Climate (ACPC) initiative for providing model simulations used in Figure 7 of the manuscript.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

References

Alvaro, R.-C., Vanessa, R., M., R. A. A., J, H. J., Hernandez, K. S., Sebastián, G.-R., and F., M. J.: Algorithm for Tracking Convective Systems (ATRACKCS), https://doi.org/10.5281/zenodo.7025990, 2022.

Borque, P., Kollias, P., and Giangrande, S.: First Observations of Tracking Clouds Using Scanning ARM Cloud Radars, Journal of Applied Meteorology and Climatology, 53, 2732 – 2746, https://doi.org/10.1175/JAMC-D-13-0182.1, 2014.

Chen, J., Hagos, S., Feng, Z., Fast, J. D., and Xiao, H.: The Role of Cloud–Cloud Interactions in the Life Cycle of Shallow Cumulus Clouds, Journal of the Atmospheric Sciences, 80, 671 – 686, https://doi.org/10.1175/JAS-D-22-0004.1, 2023a.

Chen, J., Hagos, S., Xiao, H., Fast, J., Lu, C., Varble, A., Feng, Z., and Sun, J.: The Effects of Shallow Cumulus Cloud Shape on Interactions Among Clouds and Mixing With Near-Cloud Environments, Geophysical Research Letters, 50, e2023GL106 334, https://doi.org/https://doi.org/10.1029/2023GL106334, e2023GL106334 2023GL106334, 2023b.

Corfidi, S. F.: Cold Pools and MCS Propagation: Forecasting the Motion of Downwind-Developing MCSs, Weather and Forecasting, 18, 997 – 1017, https://doi.org/10.1175/1520-0434(2003)018<0997:CPAMPF>2.0.CO;2, 2003.

Cotton, W. R., Pielke Sr., R. A., Walko, R. L., Liston, G. E., Tremback, C. J., Jiang, H., McAnelly, R. L., Harrington, J. Y., Nicholls, M. E., Carrio, G. G., and McFadden, J. P.: RAMS 2001: Current status and future directions, Meteorology and Atmospheric Physics, 82, 5–29, https://doi.org/10.1007/s00703-001-0584-9, 2003.

Crook, J., Klein, C., Folwell, S., Taylor, C. M., Parker, D. J., Stratton, R., and Stein, T.: Assessment of the Representation of West African Storm Lifecycles in Convection-Permitting Simulations, Earth and Space Science, 6, 818–835, https://doi.org/https://doi.org/10.1029/2018EA000491, 2019.

Cui, W., Galarneau Jr., T. J., and Hoogewind, K. A.: Changes in Mesoscale Convective System Precipitation Structures in Response to a Warming Climate, Journal of Geophysical Research: Atmospheres, 129, e2023JD039 920, https://doi.org/https://doi.org/10.1029/2023JD039920, e2023JD039920 2023JD039920, 2024.

Dillencourt, M. B., Samet, H., and Tamminen, M.: A general approach to connected-component labeling for arbitrary image representations, J. ACM, 39, 253–280, https://doi.org/10.1145/128749.128750, 1992.

Dixon, M. and Wiener, G.: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-based Methodology, Journal of Atmospheric and Oceanic Technology, 10, 785 – 797, https://doi.org/10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2, 1993.

Donahue, A. S., Caldwell, P. M., Bertagna, L., Beydoun, H., Bogenschutz, P. A., Bradley, A. M., Clevenger, T. C., Foucar, J., Golaz, C., Guba, O., Hannah, W., Hillman, B. R., Johnson, J. N., Keen, N., Lin, W., Singh, B., Sreepathi, S., Taylor, M. A., Tian, J., Terai, C. R., Ullrich, P. A., Yuan, X., and Zhang, Y.: To Exascale and Beyond—The Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM), a Performance Portable Global Atmosphere Model for Cloud-Resolving Scales, Journal of Advances in Modeling Earth Systems, 16, e2024MS004 314, https://doi.org/https://doi.org/10.1029/2024MS004314, e2024MS004314 2024MS004314, 2024.

Fairless, T., Jensen, M., Zhou, A., and Giangrande, S. E.: Interpolated Sounding and Gridded Sounding Value-Added Products, Tech. rep., Pacific Northwest National Lab. (PNNL), Richland, WA (United States), https://doi.org/10.2172/1248938, 2021.

Feng, X., Swann, A. L. S., Breshears, D. D., Baldwin, E., Cheng, H., Derbridge, J. J., Fei, C., Lien, A. M., López-Hoffman, L., McCarl, B., McLaughlin, D. M., and Soto, J.: Distance decay and directional diffusion of ecoclimate teleconnections driven by regional-scale tree die-off, Environmental Research Letters, 18, 114 013, https://doi.org/10.1088/1748-9326/acff0d, 2023.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

Feng, Z., Leung, L. R., Liu, N., Wang, J., Houze Jr, R. A., Li, J., Hardin, J. C., Chen, D., and Guo, J.: A Global High-Resolution Mesoscale Convective System Database Using Satellite-Derived Cloud Tops, Surface Precipitation, and Tracking, Journal of Geophysical Research: Atmospheres, 126, e2020JD034 202, https://doi.org/https://doi.org/10.1029/2020JD034202, e2020JD034202 2020JD034202, 2021.

Feng, Z., Prein, A. F., Kukulies, J., Fiolleau, T., Jones, W. K., Maybee, B., Moon, Z., Ocasio, K. M. N., Dong, W., Molina, M. J., Albright, M. G., Feng, R., Song, J., Song, F., Leung, L. R., Varble, A., Klein, C., and Roca, R.: Mesoscale Convective Systems tracking Method Intercomparison (MCSMIP): Application to DYAMOND Global km-scale Simulations, https://doi.org/10.22541/essoar.172405876.67413040/v1, 2024.

Fiolleau, T. and Roca, R.: Composite life cycle of tropical mesoscale convective systems from geostationary and low Earth orbit satellite observations: method and sampling considerations, Quarterly Journal of the Royal Meteorological Society, 139, 941–953, https://doi.org/https://doi.org/10.1002/qj.2174, 2013.

Gupta, S., Wang, D., Giangrande, S. E., Biscaro, T. S., and Jensen, M. P.: Lifecycle of updrafts and mass flux in isolated deep convection over the Amazon rainforest: insights from cell tracking, Atmospheric Chemistry and Physics, 24, 4487–4510, https://doi.org/10.5194/acp-24-4487-2024, 2024.

Hahn, T., Wang, D., Chen, J., and Jensen, M.: Evaluating Sea Breezes and Associated Convective Cloud Evolution in the Model Gray Zone, in review.

Hahn, T., Weiner, H., Brooks, C., Li, J. X., Gupta, S., and WANG, D.: CoCoMET: Community Cloud Model Evaluation Toolkit v1.0, Zenodo, https://doi.org/https://doi.org/10.5281/zenodo.15090741, 2025.

Hartmann, D. L.: Tropical anvil clouds and climate sensitivity, Proceedings of the National Academy of Sciences, 113, 8897–8899, https://doi.org/10.1073/pnas.1610455113, 2016.

Hayden, L., Liu, C., and Liu, N.: Properties of Mesoscale Convective Systems Throughout Their Lifetimes Using IMERG, GPM, WWLLN, and a Simplified Tracking Algorithm, Journal of Geophysical Research: Atmospheres, 126, e2021JD035 264, https://doi.org/https://doi.org/10.1029/2021JD035264, e2021JD035264 2021JD035264, 2021.

Heiblum, R. H., Pinto, L., Altaratz, O., Dagan, G., and Koren, I.: Core and margin in warm convective clouds – Part 1: Core types and evolution during a cloud's lifetime, Atmospheric Chemistry and Physics, 19, 10 717–10 738, https://doi.org/10.5194/acp-19-10717-2019, 2019.

Heikenfeld, M., Marinescu, P. J., Christensen, M., Watson-Parris, D., Senf, F., van den Heever, S. C., and Stier, P.: tobac 1.2: towards a flexible framework for tracking and analysis of clouds in diverse datasets, Geoscientific Model Development, 12, 4551–4570, https://doi.org/10.5194/gmd-12-4551-2019, 2019.

Heinzeller, D., Duda, M. G., and Kunstmann, H.: Towards convection-resolving, global atmospheric simulations with the Model for Prediction Across Scales (MPAS) v3.1: an extreme scaling experiment, Geoscientific Model Development, 9, 77–110, https://doi.org/10.5194/gmd-9-77-2016, 2016.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

Huang, X., Hu, C., Huang, X., Chu, Y., Tseng, Y.-h., Zhang, G. J., and Lin, Y.: A long-term tropical mesoscale convective systems dataset based on a novel objective automatic tracking algorithm, Climate Dynamics, 51, 3145–3159, https://doi.org/10.1007/s00382-018-4071-0, 2018.

Johnson, J. T., MacKeen, P. L., Witt, A., Mitchell, E. D. W., Stumpf, G. J., Eilts, M. D., and Thomas, K. W.: The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm, Weather and Forecasting, 13, 263 – 276, https://doi.org/10.1175/1520-0434(1998)013<0263:TSCIAT>2.0.CO;2, 1998.

Koch, S. E., Ferrier, B. S., Stoelinga, M. T., Szoke, E. J., Weiss, S. J., and Kain, J. S.: THE USE OF SIMULATED RADAR REFLECTIVITY FIELDS IN THE DIAGNOSIS OF MESOSCALE PHENOMENA FROM HIGH-RESOLUTION WRF MODEL FORECASTS, https://api.semanticscholar.org/CorpusID:56388139, 2005.

Kukulies, J., Lai, H.-W., Curio, J., Feng, Z., Lin, C., Li, P., Ou, T., Sugimoto, S., and Chen, D.: Mesoscale convective systems in the third pole region: Characteristics, mechanisms and impact on precipitation, Frontiers in Earth Science, 11, https://doi.org/10.3389/feart.2023.1143380, 2023.

Lac, C., Chaboureau, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., Leriche, M., Barthe, C., Aouizerats, B., Augros, C., Aumond, P., Auguste, F., Bechtold, P., Berthet, S., Bielli, S., Bosseur, F., Caumont, O., Cohard, J.-M., Colin, J., Couvreux, F., Cuxart, J., Delautier, G., Dauhut, T., Ducrocq, V., Filippi, J.-B., Gazen, D., Geoffroy, O., Gheusi, F., Honnert, R., Lafore, J.-P., Lebeaupin Brossier, C., Libois, Q., Lunet, T., Mari, C., Maric, T., Mascart, P., Mogé, M., Molinié, G., Nuissier, O., Pantillon, F., Peyrillé, P., Pergaud, J., Perraud, E., Pianezze, J., Redelsperger, J.-L., Ricard, D., Richard, E., Riette, S., Rodier, Q., Schoetter, R., Seyfried, L., Stein, J., Suhre, K., Taufour, M., Thouron, O., Turner, S., Verrelle, A., Vié, B., Visentin, F., Vionnet, V., and Wautelet, P.: Overview of the Meso-NH model version 5.4 and its applications, Geoscientific Model Development, 11, 1929–1969, https://doi.org/10.5194/gmd-11-1929-2018, 2018.

Ladwig, W.: wrf-python (Version x.x.x) [Software], https://doi.org/10.5065/D6W094P1, 2017.

Lee, J., Byun, J., Baik, J., Jun, C., and Kim, H.-J.: Estimation of raindrop size distribution and rain rate with infrared surveillance camera in dark conditions, Atmospheric Measurement Techniques, 16, 707–725, https://doi.org/10.5194/amt-16-707-2023, 2023.

Leese, J. A., Novak, C. S., and Clark, B. B.: An Automated Technique for Obtaining Cloud Motion from Geosynchronous Satellite Data Using Cross Correlation, Journal of Applied Meteorology and Climatology, 10, 118 – 132, https://doi.org/10.1175/1520-0450(1971)010<0118:AATFOC>2.0.CO;2, 1971.

Lim, S. L. and Daya Sagar, B. S.: Cloud field segmentation via multiscale convexity analysis, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/https://doi.org/10.1029/2007JD009369, 2008.

Lin, Y. and Mitchell, K.: The NCEP stage II/IV hourly precipitation analyses: Development and applications, 2005.

Lu, J., Qie, X., Xiao, X., Jiang, R., Mansell, E. R., Fierro, A. O., Liu, D., Chen, Z., Yuan, S., Sun, M., Yu, H., Zhang, Y., Wang, D., and Yair, Y.: Effects of Convective Mergers on the Evolution of Microphysical and Electrical Activity in a Severe Squall Line Simulated by WRF Coupled With Explicit Electrification Scheme, Journal of Geophysical Research: Atmospheres, 127, e2021JD036 398, https://doi.org/https://doi.org/10.1029/2021JD036398, e2021JD036398 2021JD036398, 2022.

Machado, L. A. T., Rossow, W. B., Guedes, R. L., and Walker, A. W.: Life cycle variations of mesoscale convective systems over the Americas, Monthly Weather Review, 126, 1630–1654, https://doi.org/10.1175/1520-0493(1998)126%3C1630%3ALCVOMC%3E2.0.CO;2, 1998.

Marinescu, P. J., van den Heever, S. C., Heikenfeld, M., Barrett, A. I., Barthlott, C., Hoose, C., Fan, J., Fridlind, A. M., Matsui, T., Miltenberger, A. K., Stier, P., Vie, B., White, B. A., and Zhang, Y.: Impacts of Varying Concentrations of Cloud Condensation

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

Nuclei on Deep Convective Cloud Updrafts—A Multimodel Assessment, Journal of the Atmospheric Sciences, 78, 1147 – 1172, https://doi.org/10.1175/JAS-D-20-0200.1, 2021.

Mather, J. H. and Voyles, J. W.: The Arm Climate Research Facility: A Review of Structure and Capabilities, Bulletin of the American Meteorological Society, 94, 377 – 392, https://doi.org/10.1175/BAMS-D-11-00218.1, 2013.

Mellado, J. P.: Cloud-Top Entrainment in Stratocumulus Clouds, Annual Review of Fluid Mechanics, 49, 145–169, https://doi.org/https://doi.org/10.1146/annurev-fluid-010816-060231, 2017.

Moon, Z. and Ocasio, K. M. N.: knubez/TAMS: v0.1.5, https://doi.org/10.5281/zenodo.13273150, 2024.

Morrison, H., Peters, J. M., Varble, A. C., Hannah, W. M., and Giangrande, S. E.: Thermal Chains and Entrainment in Cumulus Updrafts. Part I: Theoretical Description, Journal of the Atmospheric Sciences, 77, 3637 – 3660, https://doi.org/10.1175/JAS-D-19-0243.1, 2020.

Oue, M., Tatarevic, A., Kollias, P., Wang, D., Yu, K., and Vogelmann, A. M.: The Cloud-resolving model Radar SIMulator (CR-SIM) Version 3.3: description and applications of a virtual observatory, Geoscientific Model Development, 13, 1975–1998, https://doi.org/10.5194/gmd-13-1975-2020, 2020.

Oue, M., Saleeby, S. M., Marinescu, P. J., Kollias, P., and van den Heever, S. C.: Optimizing Radar Scan Strategies for Observing Deep Convection Using Observing System Simulation Experiments, EGUsphere, 2022, 1–29, https://doi.org/10.5194/egusphere-2022-346, 2022.

partnership, I.: ICON release 2024.01, https://doi.org/10.35089/WDCC/IconRelease01, 2024.

Peters, J. M., Morrison, H., Varble, A. C., Hannah, W. M., and Giangrande, S. E.: Thermal Chains and Entrainment in Cumulus Updrafts. Part II: Analysis of Idealized Simulations, Journal of the Atmospheric Sciences, 77, 3661 – 3681, https://doi.org/10.1175/JAS-D-19-0244.1, 2020.

Pilon, R. and Domeisen, D. I. V.: cloudbandPy 1.0: an automated algorithm for the detection of tropical–extratropical cloud bands, Geoscientific Model Development, 17, 2247–2264, https://doi.org/10.5194/gmd-17-2247-2024, 2024.

Prein, A. F., Liu ChangHai, L. C., Ikeda, K., Trier, S. B., Rasmussen, R. M., Holland, G. J., and Clark, M. P.: Increased rainfall volume from future convective storms in the US., 7, 880–884, https://doi.org/10.1038/s41558-017-0007-7, 2017.

Prein, A. F., Rasmussen, R. M., Wang, D., and Giangrande, S. E.: Sensitivity of organized convective storms to model grid spacing in current and future climates, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379, 20190 546, https://doi.org/10.1098/rsta.2019.0546, 2021.

Prein, A. F., Ge, M., Valle, A. R., Wang, D., and Giangrande, S. E.: Towards a Unified Setup to Simulate Mid-Latitude and Tropical Mesoscale Convective Systems at Kilometer-Scales, Earth and Space Science, 9, e2022EA002 295, https://doi.org/https://doi.org/10.1029/2022EA002295, e2022EA002295 2022EA002295, 2022.

Prein, A. F., Mooney, P. A., and Done, J. M.: The Multi-Scale Interactions of Atmospheric Phenomenon in Mean and Extreme Precipitation, Earth's Future, 11, e2023EF003 534, https://doi.org/https://doi.org/10.1029/2023EF003534, e2023EF003534 2023EF003534, 2023.

Prein, A. F., Feng, Z., Fiolleau, T., Moon, Z. L., Núñez Ocasio, K. M., Kukulies, J., Roca, R., Varble, A. C., Rehbein, A., Liu, C., Ikeda, K., Mu, Y., and Rasmussen, R. M.: Km-Scale Simulations of Mesoscale Convective Systems Over South America—A Feature Tracker Intercomparison, Journal of Geophysical Research: Atmospheres, 129, e2023JD040 254, https://doi.org/https://doi.org/10.1029/2023JD040254, e2023JD040254 2023JD040254, 2024.

Ramos-Valle, A. N., Prein, A. F., Ge, M., Wang, D., and Giangrande, S. E.: Grid Spacing Sensitivities of Simulated Mid-Latitude and Tropical Mesoscale Convective Systems in the Convective Gray Zone, Journal of Geophysical Research: Atmospheres, 128, e2022JD037 043, https://doi.org/https://doi.org/10.1029/2022JD037043, e2022JD037043 2022JD037043, 2023.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

Rasmussen, R. M., Chen, F., Liu, C., Ikeda, K., Prein, A., Kim, J., Schneider, T., Dai, A., Gochis, D., Dugger, A., Zhang, Y., Jaye, A., Dudhia, J., He, C., Harrold, M., Xue, L., Chen, S., Newman, A., Dougherty, E., Abolafia-Rosenzweig, R., Lybarger, N. D., Viger, R., Lesmes, D., Skalak, K., Brakebill, J., Cline, D., Dunne, K., Rasmussen, K., and Miguez-Macho, G.: CONUS404: The NCAR–USGS 4-km Long-Term Regional Hydroclimate Reanalysis over the CONUS, Bulletin of the American Meteorological Society, 104, E1382 – E1408, https://doi.org/10.1175/BAMS-D-21-0326.1, 2023.

Raut, B., Jackson, R., Picel, M., Collis, S., Bergemann, M., and Jakob, C.: An adaptive tracking algorithm for convection in simulated and remote sensing data, Journal of Applied Meteorology and Climatology, 60, 513–526, https://doi.org/10.1175/JAMC-D-20-0119.1, 2021.

Robledo, V., Henao, J. J., Mejía, J. F., Ramírez-Cardona, A., Hernandez, K. S., Gomez-Ríos, S., and Rendon, A. M.: Climatological Tracking and Lifecycle Characteristics of Mesoscale Convective Systems in Northwestern South America, Journal of Geophysical Research: Atmospheres, 129, e2024JD041 159, https://doi.org/https://doi.org/10.1029/2024JD041159, e2024JD041159 2024JD041159, 2024.

Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Liu, Z., Berner, J., Wang, W., Powers, G., Duda, G., Barker, D. M., and Huang, X.: A Description of the Advanced Research WRF Model Version 4, https://api.semanticscholar.org/CorpusID:196211930, 2019.

Sokol, A. B. and Hartmann, D. L.: Tropical Anvil Clouds: Radiative Driving Toward a Preferred State, Journal of Geophysical Research: Atmospheres, 125, e2020JD033 107, https://doi.org/https://doi.org/10.1029/2020JD033107, e2020JD033107 10.1029/2020JD033107, 2020.

Sokolowsky, G. A., Freeman, S. W., Jones, W. K., Kukulies, J., Senf, F., Marinescu, P. J., Heikenfeld, M., Brunner, K. N., Bruning, E. C., Collis, S. M., Jackson, R. C., Leung, G. R., Pfeifer, N., Raut, B. A., Saleeby, S. M., Stier, P., and van den Heever, S. C.: *tobac* v1.5: introducing fast 3D tracking, splits and mergers, and other enhancements for identifying and analysing meteorological phenomena, Geoscientific Model Development, 17, 5309–5330, https://doi.org/10.5194/gmd-17-5309-2024, 2024.

Tian, B., Fetzer, E. J., and Manning, E. M.: The Atmospheric Infrared Sounder Obs4MIPs Version 2 Data Set, Earth and Space Science, 6, 324–333, https://doi.org/https://doi.org/10.1029/2018EA000508, 2019.

Ullrich, P. A. and Zarzycki, C. M.: TempestExtremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids, Geoscientific Model Development, 10, 1069–1090, https://doi.org/10.5194/gmd-10-1069-2017, 2017.

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., and Reed, K. A.: TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in large datasets, Geoscientific Model Development, 14, 5023–5048, https://doi.org/10.5194/gmd-14-5023-2021, 2021.

van den Heever, S. C., Saleeby, S. M., Grant, L. D., Igel, A. L., and Freeman, S. W.: RAMS - the Regional Atmospheric Modeling System, https://doi.org/10.5281/zenodo.8327421, 2023.

Varble, A. C., Feng, Z., Marquis, J. N., Zhang, Z., Geiss, A., Hardin, J. C., and Jo, E.: Updraft Width Modulates Ambient Atmospheric Controls on Convective Cloud Depth, Journal of Geophysical Research: Atmospheres, 129, e2024JD041 769, https://doi.org/https://doi.org/10.1029/2024JD041769, e2024JD041769 2024JD041769, 2024.

Veals, P. G., Varble, A. C., Russell, J. O. H., Hardin, J. C., and Zipser, E. J.: Indications of a Decrease in the Depth of Deep Convective Cores with Increasing Aerosol Concentration during the CACTI Campaign, Journal of the Atmospheric Sciences, 79, 705 – 722, https://doi.org/10.1175/JAS-D-21-0119.1, 2022.

Wang, D., Giangrande, S. E., Schiro, K. A., Jensen, M. P., and Houze Jr., R. A.: The Characteristics of Tropical and Midlatitude Mesoscale Convective Systems as Revealed by Radar Wind Profilers, Journal of Geophysical Research: Atmospheres, 124, 4601–4619, https://doi.org/https://doi.org/10.1029/2018JD030087, 2019.

Wang, D., Jensen, M. P., D'Iorio, J. A., Jozef, G., Giangrande, S. E., Johnson, K. L., Luo, Z. J., Starzec, M., and Mullendore, G. L.: An Observational Comparison of Level of Neutral Buoyancy and Level of Maximum Detrainment in Tropical Deep Convective Clouds, Journal

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

of Geophysical Research: Atmospheres, 125, e2020JD032 637, https://doi.org/https://doi.org/10.1029/2020JD032637, e2020JD032637 10.1029/2020JD032637, 2020.

Wang, D., Kobrosly, R., Zhang, T., Subba, T., van den Heever, S., Gupta, S., and Jensen, M.: Causal Analysis of Aerosol Impacts on Isolated Deep Convection: Findings from TRACER, EGUsphere, 2024, 1–43, https://doi.org/10.5194/egusphere-2024-2436, 2024.

Wehr, T., Kubota, T., Tzeremes, G., Wallace, K., Nakatsuka, H., Ohno, Y., Koopman, R., Rusli, S., Kikuchi, M., Eisinger, M., Tanaka, T., Taga, M., Deghaye, P., Tomita, E., and Bernaerts, D.: The EarthCARE mission – science and system overview, Atmospheric Measurement Techniques, 16, 3581–3608, https://doi.org/10.5194/amt-16-3581-2023, 2023.

Weiner, H., Hahn, T., and WANG, D.: Configuration files for running CoCoMET - Examples, Zenodo, https://doi.org/10.5281/zenodo.15048050, 2025.

Westcott, N.: A Historical Perspective on Cloud Mergers, Bulletin of the American Meteorological Society, 65, 219 – 226, https://doi.org/10.1175/1520-0477(1984)065<0219:AHPOCM>2.0.CO;2, 1984.

Yang, G.-Y. and Slingo, J.: The Diurnal Cycle in the Tropics, Monthly Weather Review, 129, 784 – 801, https://doi.org/10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2, 2001.

Yu, Z., Wang, Y., Yu, H., and Duan, Y.: The Relationship Between the Inner-Core Size and the Rainfall Distribution in Landfalling Tropical Cyclones Over China, Geophysical Research Letters, 49, e2021GL097 576, https://doi.org/https://doi.org/10.1029/2021GL097576, e2021GL097576 2021GL097576, 2022.

Zhang, X., Shen, W., Zhuge, X., Yang, S., Chen, Y., Wang, Y., Chen, T., and Zhang, S.: Statistical Characteristics of Mesoscale Convective Systems Initiated over the Tibetan Plateau in Summer by Fengyun Satellite and Precipitation Estimates, Remote Sensing, 13, https://doi.org/10.3390/rs13091652, 2021.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

Table 1. List of algorithms commonly used to track convective clouds and precipitation

Tracking Algorithm	Codebase	Reference
tobac	github.com/tobac-project/tobac	Heikenfeld et al. (2019); Sokolowsky et al. (2024)
TAMS	github.com/knubez/TAMS	Moon and Ocasio (2024)
cloudbandPy	github.com/romainpilon/cloudbandPy	Pilon and Domeisen (2024)
MOAAP	github.com/AndreasPrein/MOAAP	Prein et al. (2023)
PyFLEXTRKR	github.com/FlexTRKR/PyFLEXTRKR	Feng et al. (2023)
TempestExtremes	github.com/ClimateGlobalChange/tempestextremes	Ullrich and Zarzycki (2017); Ullrich et al. (2021)
ATRACKCS*	github.com/alramirezca/ATRACKCS	Alvaro et al. (2022)
TINT*	github.com/openradar/TINT	Raut et al. (2021)
simpleTrack*	github.com/thmstein/simple-track	Crook et al. (2019)
KFyAO*	doi.org/10.1594/PANGAEA.877914	Huang et al. (2018)
CITA	N/A	Borque et al. (2014)
TOOCAN*	N/A	Fiolleau and Roca (2013)
ForTraCC*	N/A	Machado et al. (1998)
TITAN*	github.com/NCAR/lrose-titan	Dixon and Wiener (1993)

**Denotes lack of a public codebase or lack of code updated within 12 months.*

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

Table 2. CoCoMET Configuration File Specifications

Parameter	Type	Description	Notes
verbose	Boolean (bool)	Controls whether CoCoMET outputs text during execution.	True enables text output; False disables it.
parallel_processing	Boolean (bool)	Enables multi-core processing when available.	True enables processing; False disables it.
max_cores	Integer (int)	Specifies the maximum number of cores CoCoMET may use if `parallel_processing` is True.	Limited by system resources.
bounds	Array ([float, float, float, float])	Sets spatial bounds for inputs.	Optional; Format: [min_longitude, max_longitude, min_latitude, max_latitude].
path_to_data	String (str)	Path to input data files.	Supports glob-like patterns (e.g., "wrfout_d02*").
path_to_header	String (str)	Path to RAMS metadata .txt files.	RAMS only; Supports glob-like patterns (e.g., "RAMS_meta*.txt").
is_idealized	Boolean (bool)	Flag indicating if input data is from an idealized simulation.	Default is "False".
min_frame_index	Integer (int)	Minimum frame index to select a subset of input data.	Optional; 0-based, inclusive; Each frame corresponds to a single input file.
max_frame_index	Integer (int)	Maximum frame index to select a subset of input data.	Optional; 0-based, inclusive.
feature_tracking_var	String (str)	Variable used for feature tracking.	e.g., "dbz", "tb", "wa", "pr", or other variable names from the input data.
segmentation_var	String (str)	Variable used for segmentation.	Same options as `feature_tracking_var`.
calculation_type	String (str)	Specifies the type of precipitation calculation.	RAMS only; Options: "surface time averaged precipitation rate", "surface instantaneous precipitation rate", or "volumetric instantaneous precipitation rate".
gridding	Dictionary	Parameters for NEXRAD data gridding.	Optional; Uses Py-ART gridding functions.
Tracker	Dictionary	Specifies the tracking method (e.g., *tobac*).	Tracker-specific parameters need to be defined; see the CoCoMET user guide for details.
Analysis	Dictionary	Contains computed variables as keys and required parameters as values.	Post-processing variables and parameters must be specified; see the CoCoMET user guide for details.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

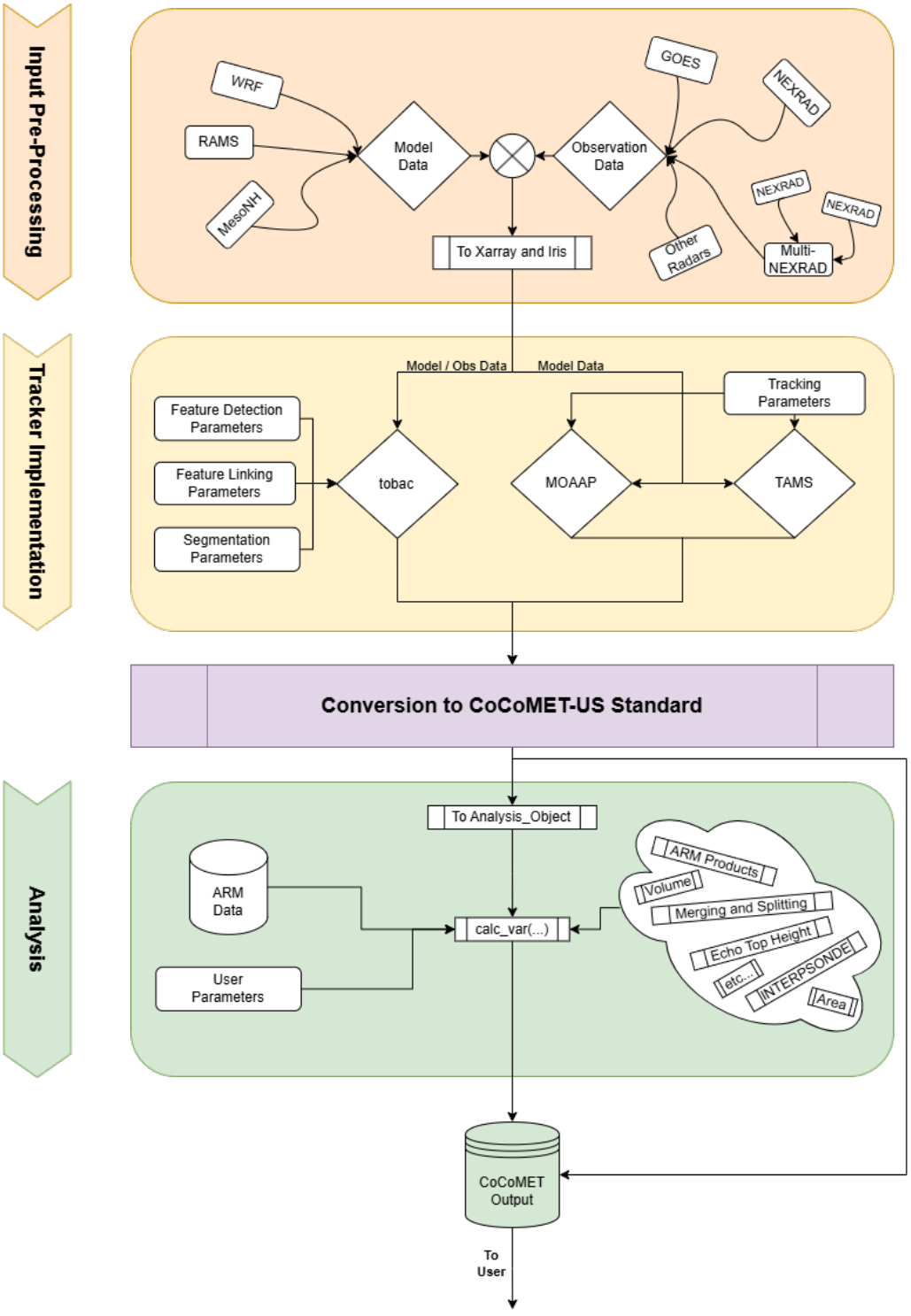

Figure 1. Flow chart highlighting the different components of the CoCoMET framework and the options available for user-based customization of parameters, input, and output fields.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

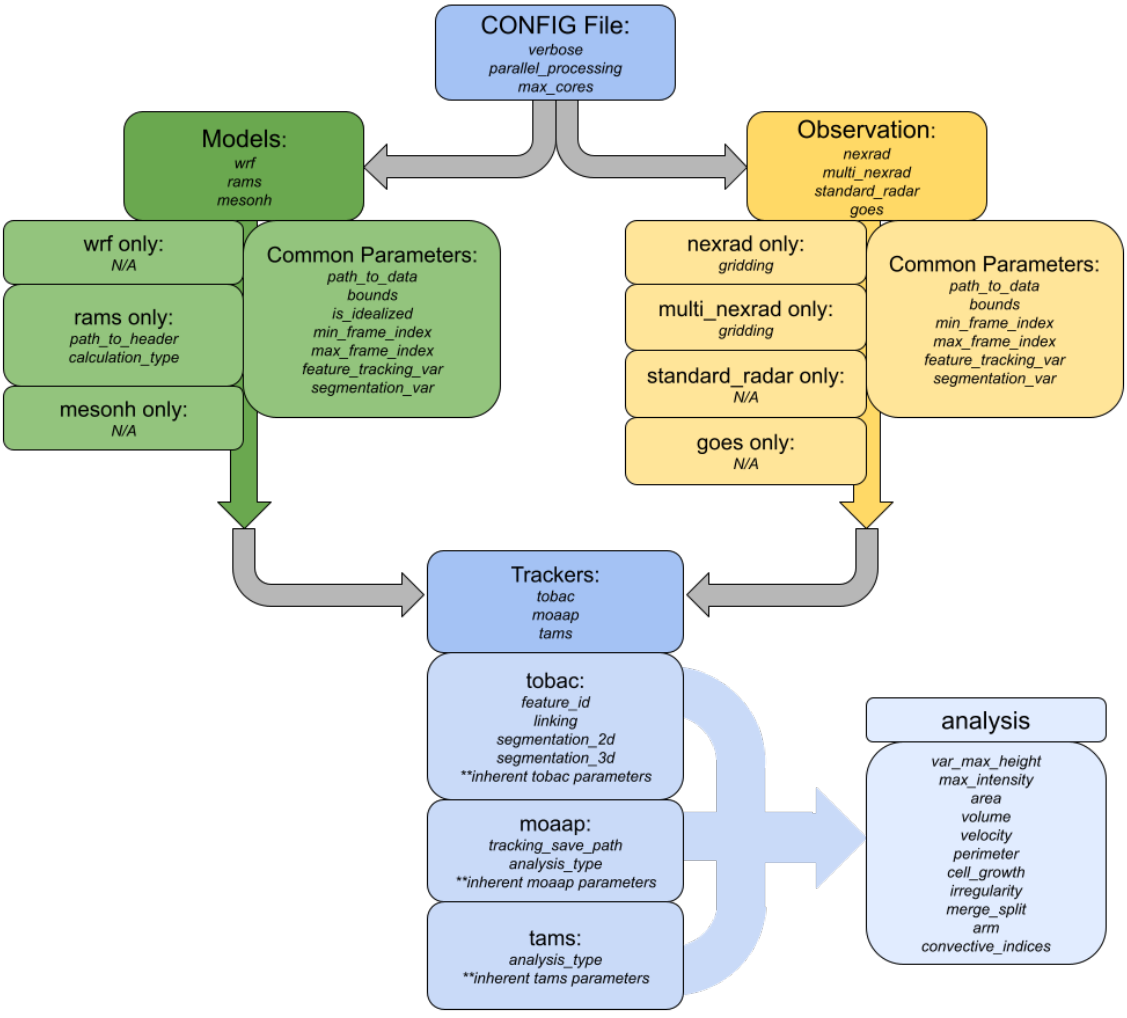

Figure 2. CoCoMET workflow highlighting the different input parameters for the configuration file, input data, and tracker options.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

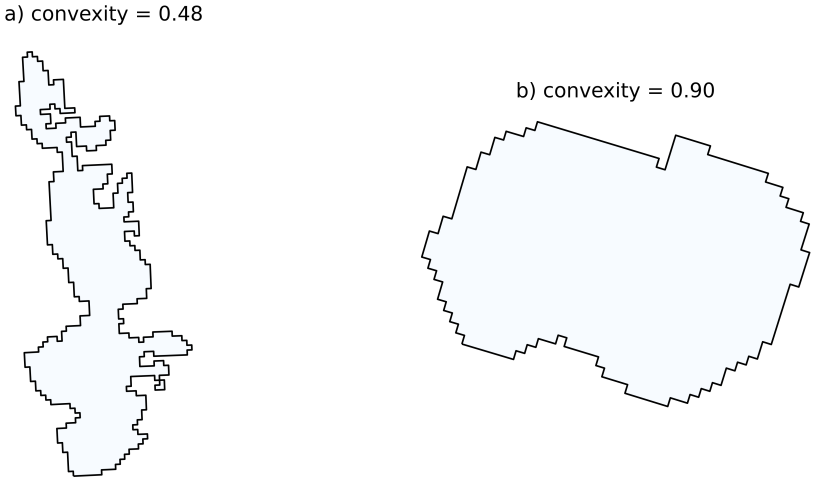

Figure 3. Illustration of convexities calculated for two tracked features.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

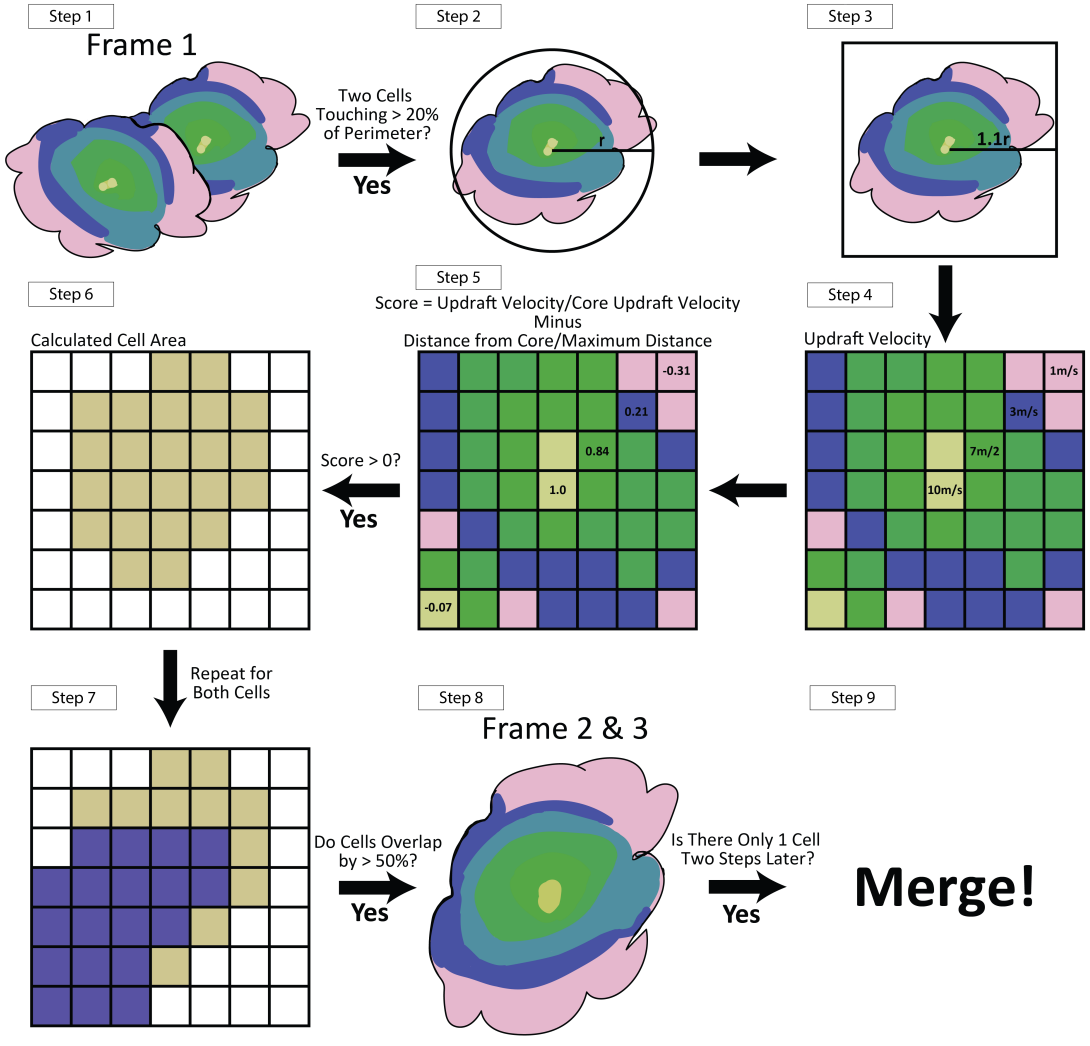

Figure 4. Illustration of steps for identifying mergers.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

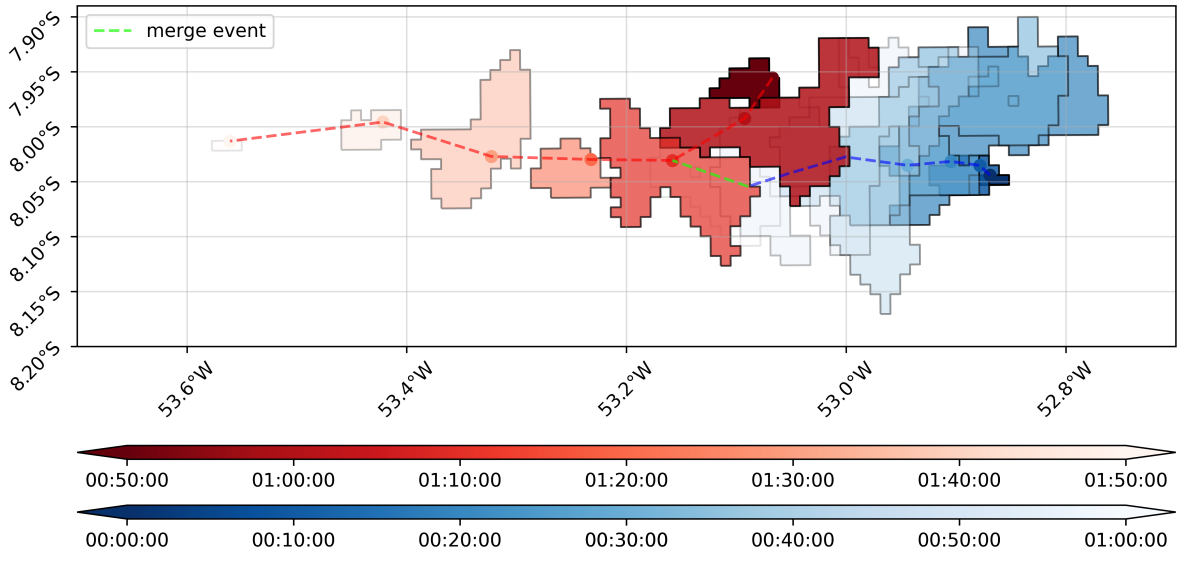

Figure 5. Example of a merging event detected by our new method using tracking outputs from *tobac*. The inputs for *tobac* are WRF simulations (Prein et al., 2022) of 2-km radar reflectivity at 1 km grid spacing for a deep convective case that occurred on April 1, 2014, in the Amazon. The tracking thresholds are 30, 40, and 50 dBZ. The cell in blue merges into the red one as both cells propagate toward the west.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

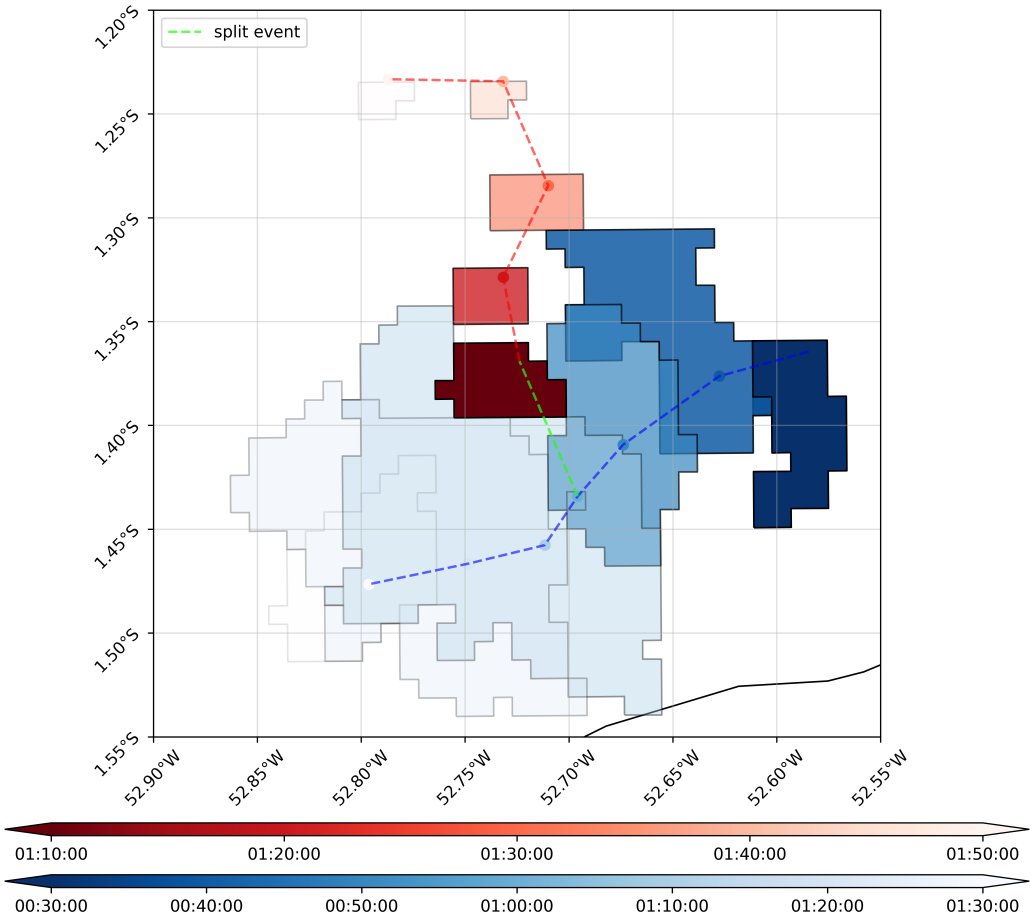

Figure 6. Example of a splitting event detected by our new method using tracking outputs from *tobac*. The inputs for *tobac* are WRF simulations (Prein et al., 2022) of 2-km radar reflectivity at 1 km grid spacing for a deep convective case that occurred on April 1, 2014, in the Amazon. The tracking thresholds are 30, 40, and 50 dBZ. The cell in blue splits into the blue and red cells.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

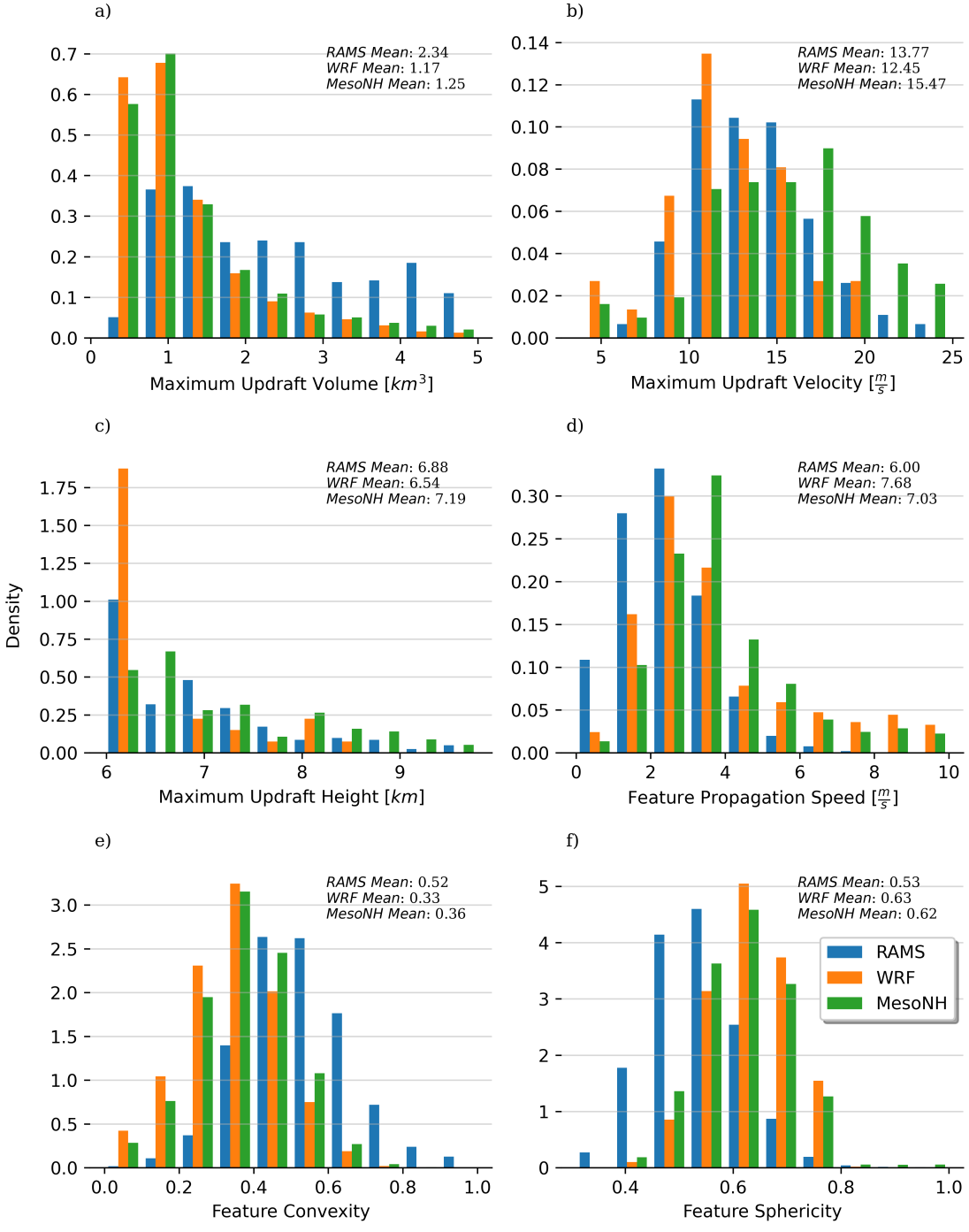

Figure 7. Histograms of tracked updraft properties using *tobac* based on simulations from WRF, MesoHN, and RAMS for cases occurred on June 19, 2013 over the Houston region. b) and c) are plotted only for updraft cells that extend higher than 6 km to highlight those deeper cores.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

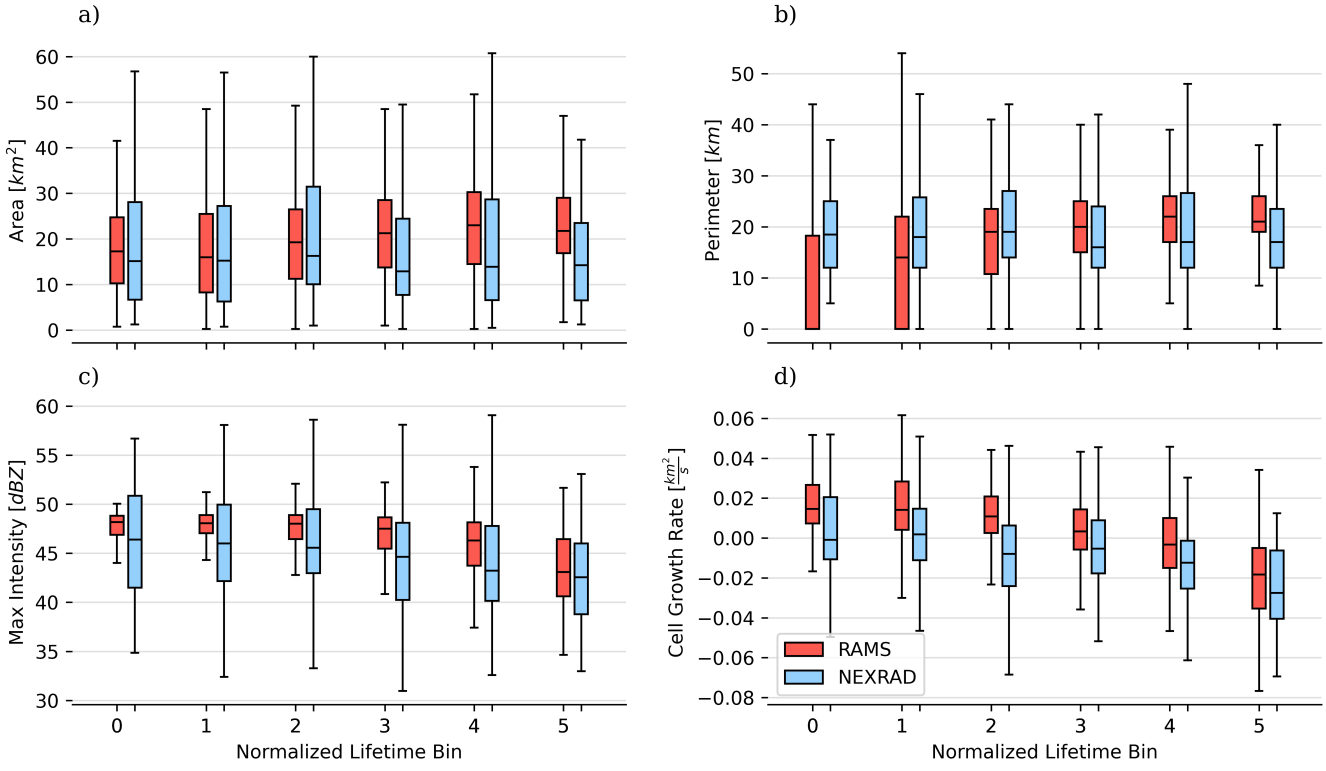

Figure 8. Box-whisker plot of tracked cell properties (both observed and simulated) as a function of the normalized cell lifetime bin. Only cells that last longer than 40 minutes and do not experience merging during their lifetime are included. 0 represents the first identification of the cell, and 5 indicates the termination of the tracked cell.

https://doi.org/10.5194/egusphere-2025-1328
Preprint. Discussion started: 10 April 2025

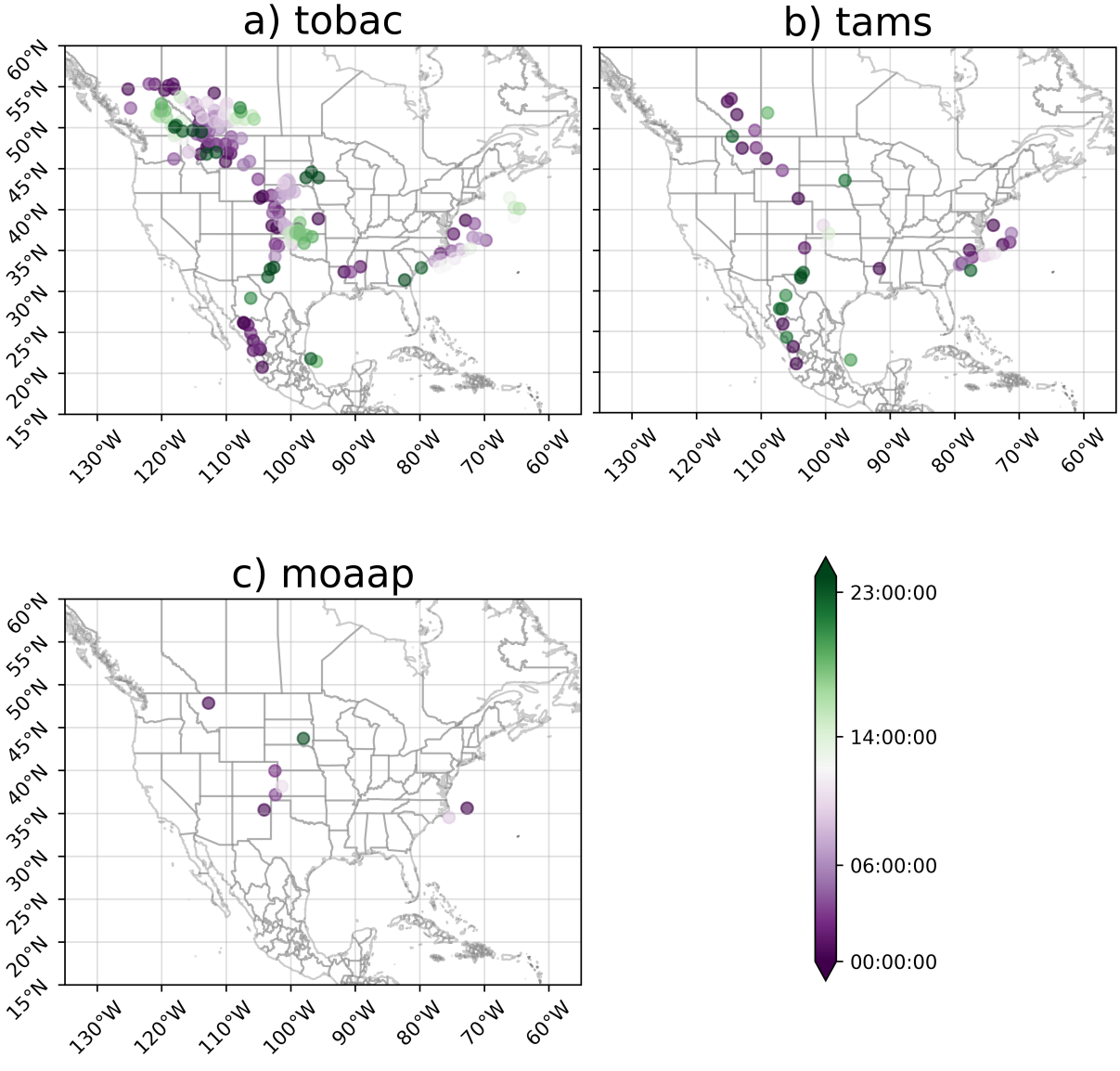

Figure 9. Initiation location and times (colors) for cells identified on June 19, 2013 in the CONUS404 WRF simulations using (a) *tobac*, (b) TAMS, and (c) MOAAP.