Author Responses to Reviewer Comments: EMMA-Tracker v1.0: A lifecycle-based algorithm for identifying and tracking mesoscale convective systems in observations and climate models

Manuscript: egusphere-2026-357

David Kneidinger, Armin Schaffer, and Douglas Maraun

May 26, 2026

We would like to thank the reviewers for their time and constructive feedback provided on our manuscript. Their insightful comments helped us identify areas where the methodology required more transparency and where the physical reasoning needed further clarification. We believe that addressing these points has significantly strengthened the manuscript and improved the overall quality of the study. Below, we provide a point-by-point response to each comment.

Reviewer comments are shown in **red boxes**, our responses are in **blue**, and changes to the manuscript are highlighted in ***green***.

Most important comments

- We made more clear which types of MCSs the algorithm targets.
- Acknowledged the coexistence of atmospheric phenomena such as fronts and MCS and made clear why we chose to postprocess the raw tracking output.
- Clarified better why we chose to go with precipitation as the main tracking variable instead of e.g. IR/OLR.
- Addressing additional limitations in the discussion section of the manuscript.

Reviewer 1

Major Comments

Line 11: I understand what coherent means in the abstract after reading the full manuscript, but it was unclear to me on first read. I might suggest including specifics on what the authors use to assess MCS 'coherence' here.

We adapted the abstract and also the rest of the manuscript to define MCS as follows:

> Following the initial tracking, EMMA applies post-processing filters to retain only distinct MCSs—systems that maintain a clear shape and steady movement, remaining structurally distinct from any co-occurring synoptic triggers.

Line 23: I think that the authors' statement about MCS simulation being an important benchmark for climate model skill is not well-justified here. Please include additional literature (i.e., Chen et al. 2020) that better explains why this is the case.

We agree with the reviewer and went for a smoother transition to better link from the Derecho Events to the process based climate model evaluation can improve the readability. This transition also eliminates the benchmark for the climate model skill sentence and instead highlights the importance of process-based model evaluation to understand the capabilities of models in simulating MCSs.

> Given the devastating impacts of such extreme events, understanding how the characteristics of MCSs will evolve in a warming climate is critical. Climate models are our primary tools for projecting these changes, yet accurately simulating the complex spatiotemporal organization of MCSs remains a well-documented grand challenge (Feng et al., 2021b; Lin et al., 2022). Consequently, there is a crucial need for more sophisticated, process-based model evaluation.

Lines 49-60: The association of some extreme precipitation events with MCSs and fronts is not inherently contradictory. While I appreciate the framework the authors are arguing here, I would urge them to also incorporate the context of recent literature on co-occurring precipitation phenomena, such as Tsai et al. 2025. Their MCS definition also does not include the description in Schumacher and Rasmussen (2020) that describes externally-driven MCSs that are sustained by lifting associated with large-scale forcing of ascent, such as that along a frontal boundary. I agree with the overarching point argued that an evaluation dataset making a clear distinction between these two systems could be useful in evaluating certain aspects of model performance, but given that the central argument of the manuscript rests on a delineation between two phenomena that traditionally can be co-located, I would recommend the overall argument of this section be strengthened and would be aided by additional literature.

We sincerely thank the reviewer for this excellent point and for directing us to (Tsai et al., 2025). We completely agree that the association of extreme precipitation with both MCSs and fronts is not inherently contradictory. As Tsai et al. (2025) demonstrate, extreme precipitation often results from compound events where fronts and MCSs co-occur. Furthermore, as the reviewer correctly points out via Schumacher and Rasmussen (2020), many convective systems are strongly forced by large-scale frontal ascent, even if their mature mesoscale structure is ultimately internally driven (e.g., by cold pools). We agree that our original framing, which implied these phenomena are always strictly independent in nature, lacked this necessary nuance.

However, our strict separation of these systems is a deliberate and crucial methodological choice based on structural distinction, rather than meteorological forcing. We wish to clarify that the EMMA-Tracker does not blindly exclude frontally-forced convection. If a system forms near a front but organizes into a structurally distinct, coherently propagating mesoscale entity, EMMA successfully retains it.

What EMMA actively filters out are precipitation fields that lack distinct mesoscale organization and instead merge into broad, amorphous, synoptic-scale baroclinic zones. This strict structural filtering is absolutely critical for our primary objective: process-oriented evaluation of climate models.

When conducting process-oriented analyses—such as creating storm-centered composites or evaluating environmental conditions—it is essential that the datasets represent actual mesoscale convective structures. If an evaluation dataset conflates structurally distinct MCSs with massive baroclinic rainbands, the resulting reference and model composites become dominated by smeared, synoptic-scale frontal features, masking the actual mesoscale physics.

This risks generating severe structural ”false positives” regarding a climate model's convective skill. A regional climate model generally simulates synoptic-scale baroclinic boundaries well because the large-scale forcing is resolved by the model grid. If we allow these broad frontal systems into the tracking database, a model might produce composites that perfectly match the reference data simply by resolving a synoptic front, even if its parameterization of mesoscale convective physics is completely flawed. By demanding that systems remain structurally distinct and coherently propagating, we ensure models are genuinely evaluated on their ability to simulate organized mesoscale convection.

We have significantly revised the Introduction and Abstract to explicitly acknowledge the meteorological reality of co-occurring phenomena (incorporating Tsai et al. (2025) and Schumacher and Rasmussen (2020)), while clearly reframing our strict structural separation as a methodological necessity for robust process-oriented model evaluation. We changed the paragraph Line 47-62 accordingly:

> Developing a robust benchmark for Europe is difficult because synoptic-scale atmospheric fronts associated with mid-latitude cyclones frequently occur in the region (Catto et al., 2014; Hénin et al., 2019; Schaffer et al., 2024). Past literature attributes European precipitation extremes to different phenomena depending on the timescale. While some studies suggest mesoscale convective systems (MCSs) primarily drive hourly precipitation extremes during the warm season (Da Silva and Haerter, 2023), other research shows cold fronts also trigger a large fraction of these hourly events (Schaffer et al., 2024). Furthermore, when analyzing longer timescales (e.g., 6-hourly), atmospheric fronts directly cause the vast majority of mid-latitude precipitation extremes (Catto and Pfahl, 2013). As recent global quantifications demonstrate, these different attributions are not contradictory, because extreme precipitation can result from events where fronts and MCSs co-occur (Tsai et al., 2025), and large-scale frontal ascent frequently forces organized convection (Schumacher and Rasmussen, 2020).

> Line 73-74: It is true that infrared brightness temperature is not a standard model output, but outgoing longwave radiation, which can easily be converted, is, and as the authors note, most previous MCS trackers use brightness temperature only (i.e., Feng et al. 2025). Additionally, the co-location of precipitation in brightness temperature in MCS tracking has been shown to eliminate the cited issues with the incorporation of non-precipitating cirrus (i.e. Feng et al. 2021). I am a little bit confused about the choice to use precipitation only here. Could further justification be provided?

We selected precipitation as our primary tracking variable primarily due to its widespread availability at hourly temporal resolution. While it is true that OLR is a common model output variable it is often only available at 3h or 6h resolution (as in CMIP5 driven EURO-CORDEX or CORDEX FPS CONV and also in most reanalysis products such as CERRA, COSMO REA6, ...).

We now mentioned in the manuscript that methods such as in PyFLEXTRKR eliminate the non precipitating cirrus clouds by using precipitation in addition to brightness temperature. To demonstrate that our tracked systems align with established infrared thresholds, we analyzed the brightness temperatures of our track masks.

For that we took the minimum (a) and mean (b) brightness temperature (NASA Global Merged IR V1 infrared brightness temperature) of our MCS masks and compared with Tb threshold common in MCS trackers (pyFLEXTRKR: cold core < 225 K, cloud shield < 241 K).

Results (Figure 1) show that the median of our tracks is below these commonly used thresholds for almost all regions, except for the main storm track. The major storm track is the region where we hardly have any events, partly because the postprocessing filter successfully removes remaining tracked systems in that regions. Regions where the commonly used Tb thresholds are not met for the median are in reddish colors and hashed. We changed the manuscript so that it reflects the comment about the improvements with incorporating precipitation in addition to infrared brightness temperature and also explicitly mentioned the absence of OLR in 1h temporal resolution.

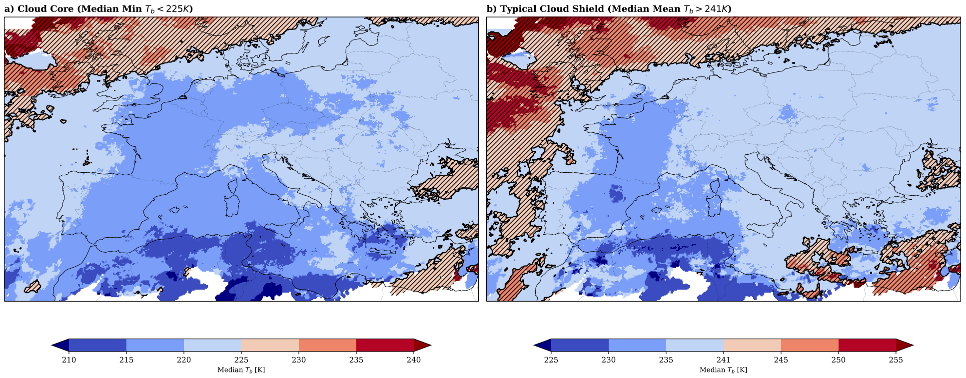

Figure 1: Median minimum brightness temperature (a) and median mean brightness temperature (b). Hashed red areas mark regions where often used thresholds for MCS detection (225K for cold core, and 241K for cloud shield) are not met. These hashed parts are in regions where hardly any MCS tracks are present in EMMA.

The pioneering climatology of Morel and Senesi (2002) identified systems by their

> large, cold cloud shields. Yet trackers relying solely on cloud-top temperatures often conflate active convection with non-precipitating cirrus anvils (Fiolleau and Roca, 2013). More recent objective tracking algorithms, such as the Tracking Algorithm for Mesoscale Convective Systems (TAMS; Núñez Ocasio and Moon (2024)) and the Python FLEXible object TRacKeR (PyFLEXTRKR; Feng et al. (2023)), have significantly improved detection by incorporating precipitation data alongside infrared brightness temperature (Feng et al., 2025, 2021a). However they still rely on satellite-derived brightness temperature thresholds as a primary identification variable. This creates a fundamental barrier for process-based evaluation because infrared brightness temperature and outgoing longwave radiation are not standard, hourly output variables in most regional climate model ensembles. Consequently, the reliance on variables unavailable in model output prevents a consistent, pan-European assessment of how MCSs are represented in currently available climate model ensembles.

Line 110-111: This argument could be strengthened by referencing the high-resolution radar dataset used in the supplemental figure evaluation directly in the text.

We added a direct citation of the INCA dataset into the manuscript.

Line 143 - 144: Interpolating temperature and specific humidity variables from 25 km to a higher resolution has the potential to introduce biases in the final data product for the lifted index, especially given that these fields are not always smoothly varying. Could the authors please explain/justify this choice further? and Line 426-428: Stronger justification is needed for why this dataset is useful for process-based evolution of models like EURO-CORDEX given that it has been stated above that it does not have the appropriate hourly resolution for direct comparison.

We address both of these comments together since we think that they are closely related:

It's important to note here that the lifted index is not the primary tracking variable used for the masks and overlapping criteria. It serves as a "check up variable": is this object's mask in an area of convective instability or not? Hence the tracking is not very sensitive to potential interpolation biases. The reason for interpolating is that the algorithm takes both variables (precipitation and LI) on the same grid. Additionally the lifted index is not used for small scale feature identification. And in the postprocessing it's a mean over the full lifetime of the track.

The concern about the temporal resolution of model data, such as EURO-CORDEX, primarily applies to the pressure-level variables used for the Lifted Index (LI) rather than the main tracking variable (precipitation). The EMMA-Tracker identifies and links storm candidates based on hourly precipitation fields, which are standard outputs in these climate model ensembles. The LI is only used as a secondary "convective flag" that a system must meet at least once during its mature phase, and it does not dictate the hourly movement or structural tracking of the system. Because atmospheric instability is a relatively smooth field, interpolating 6-hourly pressure data to an hourly frequency is sufficient to confirm a system's convective nature without significantly altering the results. This allows researchers to use the high-resolution precipitation data to evaluate the physical processes and organization of the storms while maintaining a consistent and reliable detection of their convective environment.

We did a sensitivity tracking run using a 6-hourly lifted index interpolated to hourly temporal resolution. Figure 2 (a) and (b) show the 10-year climatology from the original hourly LI (a) and the interpolated 6-hourly LI (b). Panel (c) shows the absolute bias in terms of MCS frequency, and (d) shows the relative bias.

The climatology hardly changes at all. Only the relative bias over the storm track is above 30%, but this is the area where hardly any events are present.

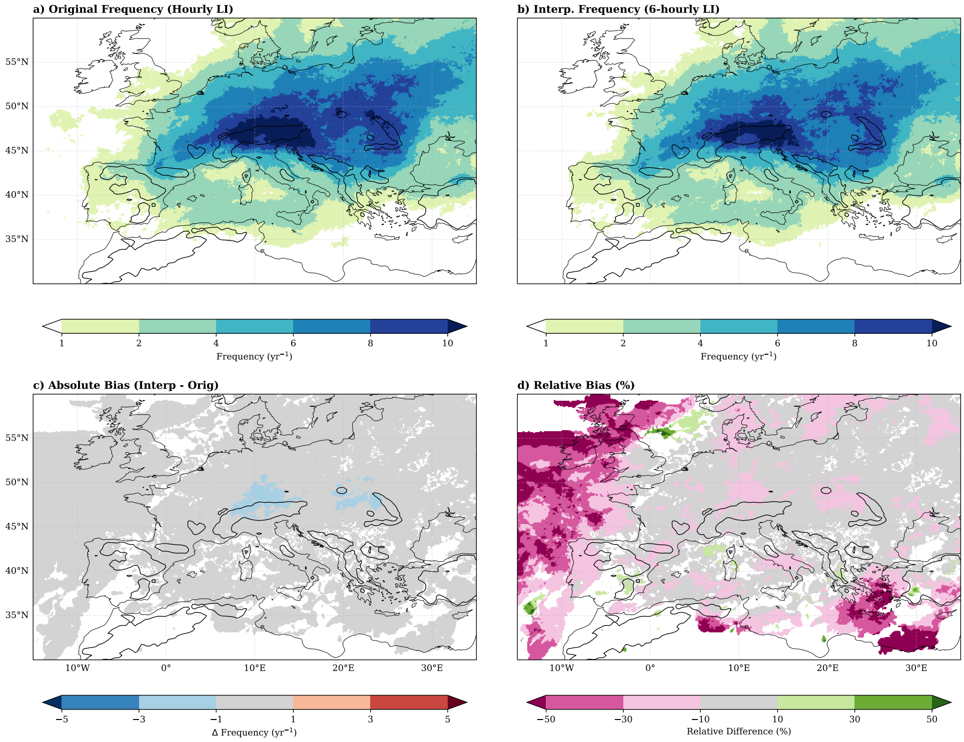

Figure 2: Comparison of tracking results from a sensitivity run (2005-2014) using hourly ERA5 derived LI (a) and every 6 hour ERA5 LI interpolated to hourly LI (b). (c) shows the absolute bias and (d) the relative bias. Results show that the climatology does not significantly change, with only a relative bias above 30% over the storm track, where hardly any events are present.

Line 148-149: Are the authors concerned that this smoothing would reduce the reliability of tracking on fine-scale features? Could they please explain this choice further, especially given that small-scale grid noise is not widely considered a problem in IMERG?

We agree that 'noise' was misleading; IMERG is a high-quality dataset. We apply the mild, one-grid-box smoothing as a spatial low-pass filter to address high-frequency spatial patchiness. Without this step, applying our thresholds could artificially fracture cohesive mesoscale systems into disconnected pixels. Since our algorithm specifically targets large systems (> 3500 km^2), this smoothing is essential to reliably track the unified mesoscale structure rather than individual fine-scale cells. We have revised the manuscript to clarify this physical reasoning.

> Prior to the first step, a mild spatial smoothing is applied to the IMERG precipitation field using a Gaussian filter with a standard deviation of one grid box (Fig. 1a). This acts as a spatial low-pass filter to address grid-scale patchiness, ensuring the threshold-based connected-components algorithm identifies large convective cores as unified structures rather than artificially fracturing them.

Line 155: I understand the rationale for using a threshold-based algorithm here to build thresholds that work for datasets of different resolutions, but I am not sure why having a threshold that adapts to a systematic bias of a model dataset is a good thing if the goal is to use this dataset for a process-oriented evaluation of models. Could this choice please be explained/justified further?

Using a relative threshold (99th percentile of wet hours) is a deliberate design choice to ensure the tracker identifies physically equivalent features across datasets, regardless of their absolute intensity biases. While in some data products certain features are of different intensity, they might still be the same physical feature. Setting an absolute threshold could result in over or under detection or even no detection. This is done because the main aim of this Tracker is to be used for Model Evaluation (EMMA: Evolution-based MCS Model Assessment-Tracker). As an example: Maybe a given GCM has too little humidity, but otherwise can provide the environmental conditions. Then the RCM may pick that up and organize an MCS, but because of an overall humidity bias it would just produce less intense rain. But still it would be the same physical phenomenon. And these events can then be used for process based analysis.

> For this study, we define the heavy precipitation threshold as the domain-wide 99th percentile of all wet hours ($> 0.1 \ mm\, h^{-1}$). This relative approach is a deliberate design choice to ensure the tracker identifies physically equivalent systems across different datasets, regardless of absolute intensity biases. By defining convective cores in the context of each datasets own climatology, we can isolate the evaluation of mesoscale organization from simple magnitude biases. This threshold targets the convective regime, as the 99th percentile of station-based hourly precipitation is almost entirely convective and exhibits the physical scaling behavior characteristic of convective storms (Haslinger et al., 2025; Berg et al., 2013). In the observational IMERG dataset, this 99th percentile corresponds to an absolute threshold of 6.8 mm h^{-1}.

Line 225: Figure A1b could also be interpreted as, rather than the centroid shifting erratically the whole time, two distinct systems, and an erratic jump in detection between the two. What is the rationale for implementing this track straightness criteria rather than just implementing a criteria that would identify these two different components of the frontal system as two unique areas of convection?

The erratic track jumps shown in Figure A1b are a "detection artifact" that occurs when the tracker switches between different rainfall peaks (and the resulting precipitation weighted center) within a single large frontal system. Rather than being a problem, this behavior is a useful signal that helps the algorithm identify and remove synoptic-scale fronts/ cyclones that do not have the smooth, steady movement of the targeted distinct MCSs. While these fronts may contain some convective parts, the algorithm is designed to remove the entire system to keep the final dataset focused on mesoscale processes. This is a deliberate choice to provide a clean reference for model evaluation, even though it may occasionally remove systems like squall lines if they are physically merged within a larger front, as noted in the limitations in Section 4.3.

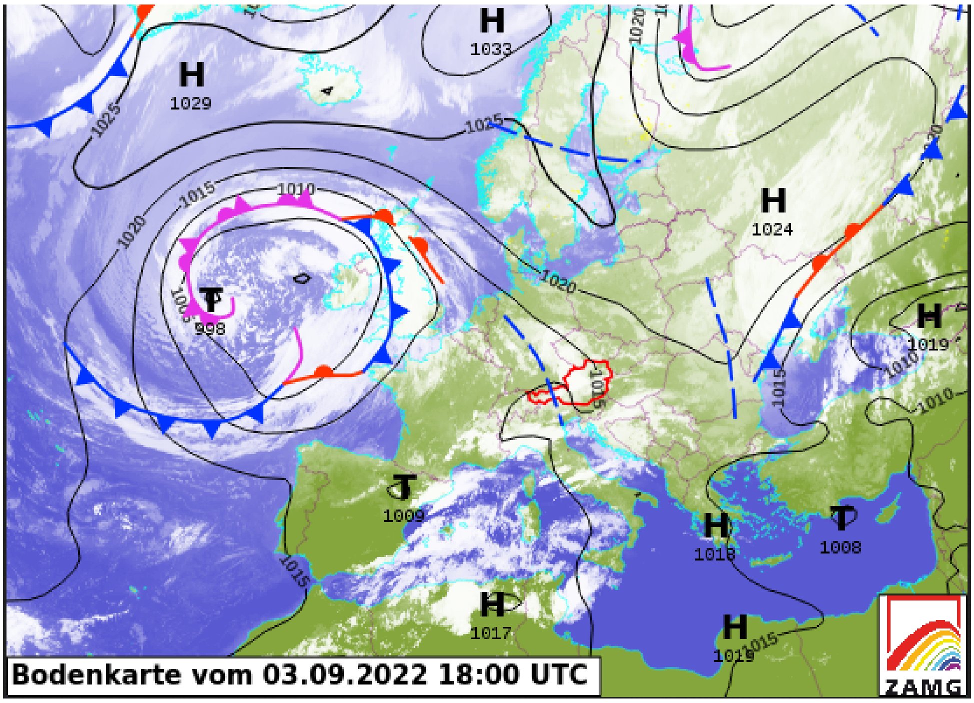

Figure 3: Weather map showing the synoptic situation corresponding to the track in Figure A1b. The erratic jumps in the track are due to the tracker switching between different rainfall peaks within a large frontal system. This behavior helps identify and remove synoptic-scale fronts that do not have the smooth, steady movement of distinct MCSs.

> Line 250: Similar to above, Figure A1d appears to show an MCS track that looks relatively reasonable if it were stopped a few time steps earlier, and I have seen valid MCSs masks that are more elongated in the horizontal. Could the authors please explain/justify further why the emphasis here is on eliminating this track in its entirety?

This is again the design choice of the tracker: Using this "detection artifact" to remove the full frontal systems. When looking at the weather map we can see that, while there have been convective lines further to the south in Spain (which was successfully tracked by the way) the system over UK/ Norway was a frontal system, where the algorithm detected the warm frontal rain and then jumped to the occlusion front and warm front in one timestep. Which could be used to remove the full system. See Figure 4 for the weather map corresponding to the track in Figure A1d.

> Line 259: I do not see a prominent maximum over the North Atlantic, Great Britain, and the North Sea in Figure 3a. Could the authors please clarify?

We agree with the reviewer that Figure 3a does not show a prominent maximum over the North Atlantic or Great Britain compared to the continent. We have corrected the text to state accurately that occurrence frequencies in this region are lower. To clarify how our methodology handles these systems, we have added a new panel which was requested by Reviewer 2 (Fig. 4e) showing the relative total rejected events resulting from the combination of all filters. This panel demonstrates that our physical filters successfully distinguish and remove these synoptic-scale systems, with the rejected events predominantly aligning with the North Atlantic storm track.

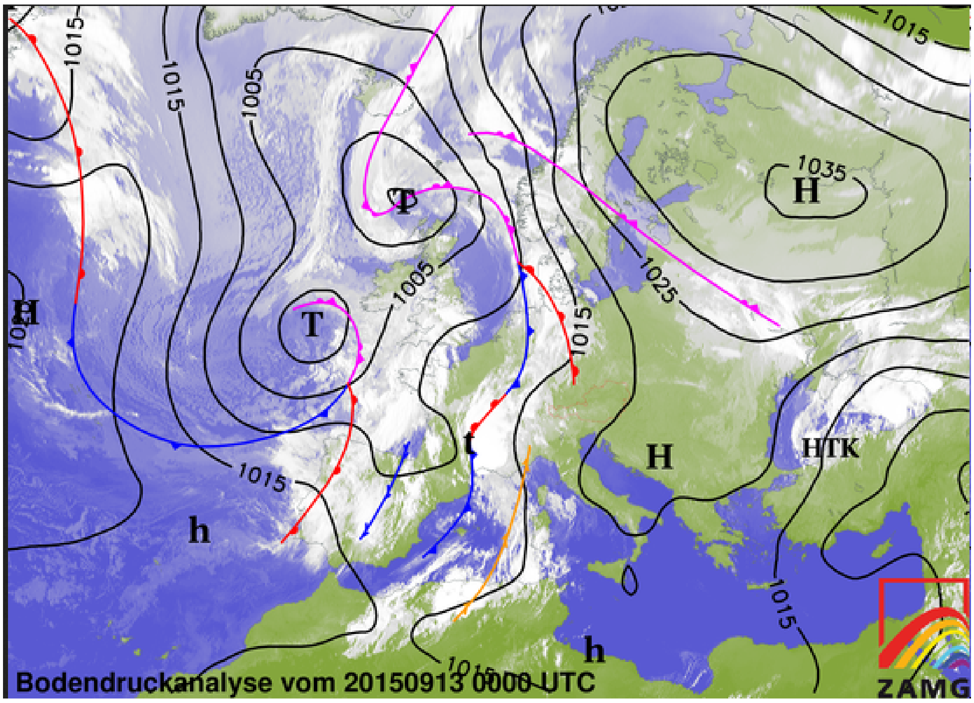

Figure 4: Weather map showing the synoptic situation corresponding to the track in Figure A1d. The track's erratic jumps between different frontal features indicate that it is part of a larger frontal system rather than a distinct MCS, justifying the removal of the entire track to maintain a clean reference for model evaluation.

We have updated the manuscript accordingly.

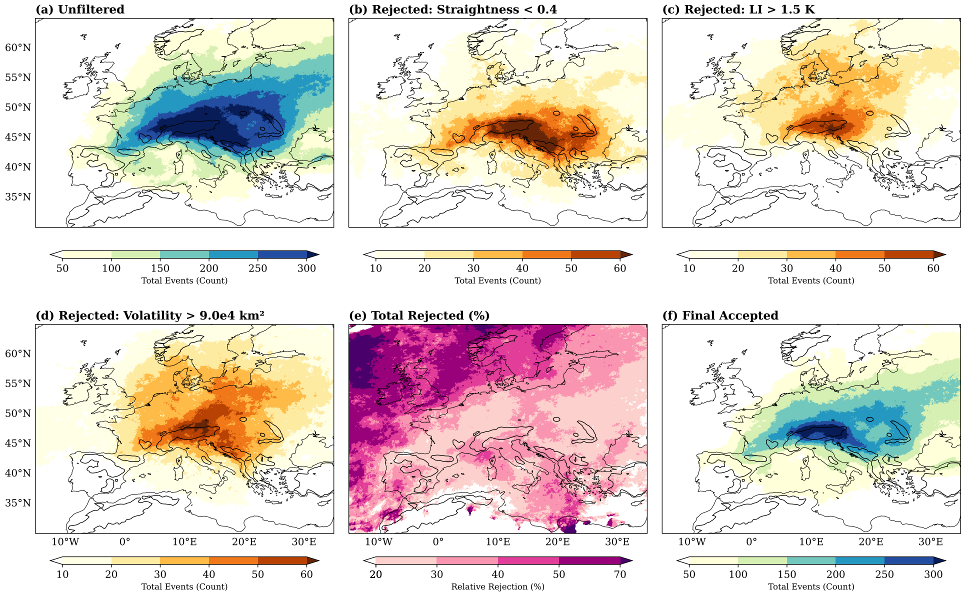

Figure 5: Impact of the physical filters on the spatial distribution of accepted and rejected systems. Panel (e) highlights the relative amount of total rejected events, which highlights that the combination of these physical filters predominantly rejects events along the North Atlantic storm track.

Line 278-280: A visual comparison with Morel and Senesi (2002) makes me wonder about the missing peak over the southern Mediterranean that is seen in that climatology; given that the tracking domain includes 30 degrees and northward. Could the authors please explain this discrepancy?

Both our climatology and the one from Morel and Senesi show a transition from continental towards more coastal/ marine in September (Note the different time spans: 1993-1997 vs 1998-2024). But it seems that there is very little heavy precipitation in that region at least according to IMERG. Maybe it's because Morel and Senesi have only looked at cloud top and note included precipitation (which more recent trackers do in addition to IR).

Over the coast of Tunisia and the Atlas Mountains in general it seems that the precipitation estimates from IMERG are rather low and our Climatology matches more closely to others which also tracked based on precipitation (DaSilva2023, Figure 3). We had a look at this by examining the amount the precipitation threshold of 6.82 mm is exceeded at a certain grid cell (Figure 6).

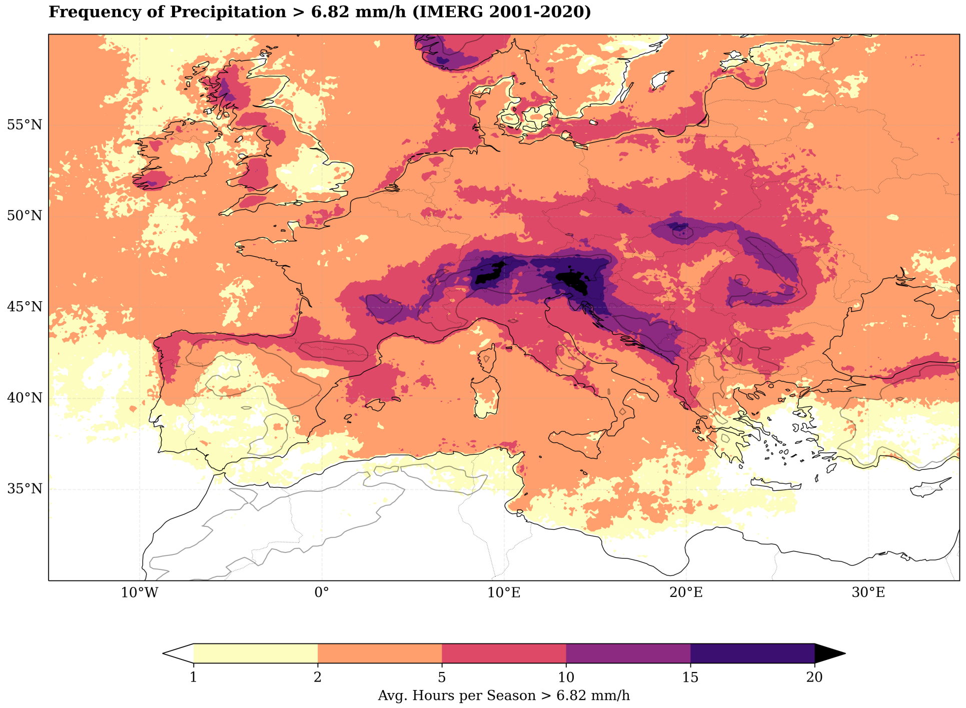

Figure 6: Frequency of grid cells exceeding the 99th percentile precipitation threshold (6.82 mm h^{-1}) from 1998-2024. The low frequency of exceedance in the southern Mediterranean region suggests that the discrepancy with Morel and Senesi (2002) may be due to differences in detection methods, with our precipitation-based approach capturing fewer events in this area compared to their cloud-top temperature method.

Figure 15: Why is masking only conducted in panel (b) and not (d)?

Thank you for pointing this out. That was a typo: it's masked in (b) and (d).

Line 302: I don't necessarily see a secondary maximum of rejected frontal systems over the Balkans and Croatia in Figure 4c (where is the primary maximum?); rather, I see one cohesive maximum on the eastern portion of the Alps. Could this description please be clarified?

We adapted this in the manuscript and changed the phrasing:

> A cohesive region of frequently rejected systems extends from the eastern Alps southward along the Croatian coast and into the Balkans (Fig. 7c), highlighting another important regional interaction. This area is known for significant thunderstorm activity, particularly in late summer and autumn (Galanaki et al., 2018). This activity is often linked to the interaction of synoptic systems with the Dinaric Alps and Adriatic regions (Mikuš et al., 2012).

Line 322: I don't think that the general pattern over eastern France in Figure 5a can be considered a maxima, given that there is a broad region showing an average of one MCS per year. Could this characterization be please explained/ clarified?

This was formulated inaccurately and we fully agree and have improved the phrasing in the manuscript:

> In May (Fig. 6 a), activity remains relatively widespread across the continent. While scattered systems emerge over eastern France (e.g., Vosges region), the northern Carpathians, and the Balkans, overall frequencies remain much lower than during peak summer.

Line 426-428: Stronger justification is needed for why this dataset is useful for process-based evolution of models like EURO-CORDEX given that it has been stated above that it does not have the appropriate hourly resolution for direct comparison. Line 401 - 402: I understand that MCSs are much less common over Europe in winter months, but if this tracking algorithm was specifically designed to filter out synoptically driven precipitation, why is it necessarily more unreliable under these conditions? Could further justification please be provided?

We thank the reviewer for pointing this out, as the original phrasing made it sound as though the algorithm itself physically fails in the winter. This is not the case.

The issue in winter is mainly a problem of "false positives" due to the very high number of fronts compared to MCSs. During the cold season, synoptic frontal systems are extremely common in Europe, while MCSs are very rare. Even in an algorithm with very high accuracy, the sheer volume of frontal events means that the small percentage of fronts that "sneak" past the filters would still likely outnumber the actual MCSs. This would result in a dataset dominated by frontal convergence.

We have revised the manuscript to clarify these points and provide a stronger justification for the seasonal limitations of the dataset. To ensure the final climatology remains a reliable benchmark for convective processes, we chose to limit the analysis to the warm season when MCSs are more frequent.

> Finally, this study is restricted to the warm season (May–September). In winter, European weather is so dominated by synoptic fronts that any attempt to track rare MCSs would result in a dataset overwhelmed by frontal false positives.

Technical Comments

Line 1 and 17: Mesoscale convective system (MCS) (and future text that is defined as an acronym, such as wet hour frequency (WHF)) should always be lowercase.

We corrected this in the manuscript.

Line 13: This statement is unclear. I suggest replacing 'exceeding' with 'accounting for' or another phrase to make it clear what 60% refers to here.

We improved the wording and changed the sentence to:

> Their contribution systematically increases with hourly precipitation intensity, accounting for over 60% of heavy precipitation (P99.9) across most of continental Europe and 80% over parts of the Mediterranean.

Line 63: It is not clear what a 'mixed dataset' refers to here.

This paragraph was adapted to make this more clear:

> Although fronts and MCSs coexist in nature, evaluating climate models requires researchers to cleanly separate their physical structures. If an observational reference dataset mixes these two systems, the evaluation process can become fundamentally flawed. Because Europe's precipitation climatology is heavily front-dominated, the environmental signal in any mixed dataset can become overwhelmingly dictated by large-scale baroclinic forcing. Even coarse-resolution RCMs represent these broad baroclinic dynamics reasonably well (Prein et al., 2015; Schaffer et al., 2025). However, accurately simulating the severe, short-duration hazards characteristic of organized convection, such as extreme sub-daily precipitation and associated wind gusts, strictly requires convection-permitting scales (Kendon et al., 2017). Consequently, if researchers generate storm-centered composites or analyze environmental conditions using a mixed dataset, the well-resolved synoptic features dominate the signal, making RCMs and CPMs appear structurally similar. This baroclinic dominance completely masks the intense, localized physics of organized convective storms, which

> often develop in entirely different environments such as the prefrontal warm sector (Pacey et al., 2023).

> Line 82: It is not clear what is meant by ‘genuine’ MCSs here. Internally driven / non-frontal associated MCSs?

We totally agree that this wording was not ideal and replaced it in the whole manuscript with ”distinct”. Which we defined now in both the abstract and the introduction as follows:

> ...systems that maintain a clear shape and steady movement, remaining structurally distinct from any co-occurring synoptic triggers.

> Line 120: Starting here and for the remainder of this section, the authors refer to IMERG where I believe they are intending to refer to ERA5.

Thank you, it should be ERA5.

> Line 355: The manuscript transitions from discussing MCS characteristics back to geospatial climatology here with no explanation, and left me as the reader confused.

We adapted the manuscript to avoid this jump and included a transition.

> Line 410 - 414: These sentences have multiple grammatical and punctuation errors that could be corrected for readability.
> Line 419: I think this should read ‘IMERG precipitation and ERA5 data.
> Line 435: I believe the sentence should end with ‘...for evaluating organized convection.

Corrected in the manuscript.

Reviewer 2

We would like to thank the reviewer for their time and constructive feedback. Their insightful comments helped us identify areas where the methodology required more transparency and where the physical reasoning needed further clarification. Below, we provide a point-by-point response to each comment.

Reviewer comments are shown in **red boxes**, our responses are in **blue**, and changes to the manuscript are highlighted in **green**.

Major Comments

Limitation of ERA5 as an instability proxy: One of my main concerns with the proposed method is that the instability proxy is derived from ERA5, which may not fully capture the atmospheric instability and convective environment, as it relies on convective parameterizations. However, I understand that the authors intend to present this tracking algorithm as a novel tool for evaluating high-resolution climate models. In the context of model evaluation, this approach is more consistent, since the simulated atmospheric instability should align with the model's simulated convective precipitation. With this in mind, using ERA5 as the only available observational proxy for instability is still reasonable, but its limitations should be explicitly acknowledged and discussed in Section 4.3.

We thank the reviewer for raising this important point. We completely agree that ERA5's reliance on parameterized convection means there is not a perfect, one-to-one spatiotemporal link between the reanalysis environment and the observed IMERG precipitation. We have updated Section 4.3 to explicitly acknowledge this limitation. To further clarify our rationale, we emphasize that the Lifted Index (LI) is strictly used in our algorithm as a qualitative, binary check to flag generally favorable convective environments, rather than as a quantitative predictor of precipitation intensity. Furthermore, the usage of ERA5-derived instability as an observational proxy is supported by Da Silva and Haerter (2023), supporting material, who found that using ERA5 CAPE yielded MCS climatologies highly consistent with those tracked using direct lightning observations. Supporting Material from Da Silva and Haerter (2023) can be found here: `https://agupubs.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1029%2F2023JD039045&file=2023JD039045-sup-0001-Supporting+Information+SI-S01.pdf` Finally, relying on these standard thermodynamic variables is a deliberate methodological choice to ensure the EMMA-Tracker can be applied consistently across climate model ensembles for evaluation.

> ERA5, which relies on parameterized convection and may not perfectly align spatiotemporally with the explicitly observed IMERG precipitation. We do not intend to use this proxy as a quantitative predictor of precipitation intensity, but rather as an effective qualitative threshold to identify environments broadly supportive of convection. Furthermore, this observational mismatch is naturally avoided when applying the algorithm to climate model data, since simulated precipitation is a direct consequence of the model's own representation of atmospheric instability.

We also added some related additional clarification to the method section, which we missed in the original manuscript preprint:

> Line 171+: Because precipitation is an accumulated variable and LI is an instantaneous snapshot, the algorithm evaluates the LI at the beginning of the precipitation accumulation period. This ensures the pre-storm environment is assessed before the convective event consumes the atmospheric instability.

Tracking on precipitation only: Likewise, I think the authors should add a brief discussion of the limitation of tracking precipitation fields only while also using spatial overlap as the method to link systems between timesteps. Feng et al. (2024) and others have used precipitation together with brightness temperature (or outgoing longwave radiation from model output) data to overcome the limitation of potential discontinuous precipitation within a MCS. Why did the authors decide to not follow this approach? Do you have any concerns that the full lifecycle of MCSs may not be captured with this approach?

We thank the reviewer for raising this important point. We completely agree that precipitation-only tracking has limitations regarding the full lifecycle of an upper-level cloud shield, and that hybrid trackers like PyFLEXTRKR (Feng et al., 2023) beautifully address this using OLR/brightness temperatures.

The primary reason we did not adopt an OLR-based approach is data availability: hourly OLR is frequently missing in many regional climate model ensembles such as the CMIP5 driven EURO-CORDEX, CORDEX-FPS CONV runs and also in most reanalysis products such as CERRA, COSMO REA6. Furthermore, for process-oriented studies in the mid-latitudes, a precipitation-only approach offers a distinct advantage. As Da Silva and Haerter (2023) note, large synoptic frontal rainbands over Europe often exhibit low brightness temperatures indistinguishable from MCSs. By tracking precipitation morphology directly, we avoid conflating frontal cloud shields with convective organization.

Finally, we mitigate the issue of discontinuous precipitation by using a low initial threshold (1 mm h^{-1}), spatial smoothing, and the Moseley et al. (2013) recovery strategy.

> Tracking MCSs using only precipitation can artificially split systems and prematurely terminate their lifecycles, because precipitation beneath a continuous convective cloud shield is often patchy and spatially discontinuous (Fiolleau and Roca, 2013; Feng et al., 2021a, 2023). Recent hybrid trackers overcome this by incorporating infrared brightness temperatures or simulated outgoing longwave radiation (OLR) to follow the continuous upper-level cloud canopy (Feng et al., 2021a). However, this approach requires hourly OLR, which is rarely available in standard regional climate model output due to data storage constraints. Despite this limitation, precipitation has a high socioeconomically relevance for assessing convective impacts. To mitigate track splitting within discontinuous precipitation, our method uses a low initial threshold (1 mm h^{-1}), applies spatial smoothing to connect fragmented cores, and utilizes a temporal recovery strategy to bridge short gaps (Moseley et al., 2013).

Choice of thresholds: The effectiveness analysis of each of the added criteria is great and helps a lot to better understand how the chosen criteria eliminate misclassified MCSs in what regions. However, I am wondering if the thresholds for the three criteria that are tailored to exclude frontal systems were determined through trial and error and a subjective analysis of what looks good or can the numbers be connected to a more physical reasoning? I share the other reviewer's concern about global applicability and wonder what is the method to adjust these thresholds to make them suitable for another region?

The threshold selection was guided by a combination of physical reasoning and empirical sensitivity testing. Initially, we manually labeled precipitation masks into distinct classes (MCS, ambiguous, non-MCS) to identify structural differences. While we have not yet developed a universal automated method to transfer these specific numerical thresholds globally, the underlying physical concepts—such as the requirement for atmospheric instability—remain universally applicable. We note that the necessity of regional threshold tuning is well-recognized in the literature. For instance, Feng et al. (2021a) explicitly state that precipitation thresholds can and should be adapted to avoid missing storms in specific regions. For application in other regions, researchers using EMMA could generate regional sensitivity curves (similar to those provided in our supplement) to identify the optimal threshold at the 'knee' of the curve, visually verifying that the rejected systems align with local non-convective phenomena.

Choice of criteria to exclude fronts: The authors use track straightness, area volatility and environmental instability as additional criteria to distinguish MCSs from frontal systems. I am wondering why the geometric structure is not considered as this seems like a natural and easy-to-implement criterion that may be more straightforward than the track straightness or area volatility.

While geometric structure is an intuitive metric, it presents significant practical challenges for automated tracking. Frontal precipitation does not necessarily rain out uniformly along the entire frontal boundary for its entire lifetime, meaning the geometric footprint of a front can be highly diverse and fragmented. Furthermore, both MCSs (e.g., squall lines) and frontal rainbands can exhibit highly elongated geometries, making aspect ratio an unreliable differentiator. We found that utilizing dynamic tracking 'artifacts' (such as area volatility) provided a much more robust and straightforward separation.

Related to the comment above, why were no criteria commonly used for front tracking, such as those in Berry et al. (2011) (Berry, G., Reeder, M. J., & Jakob, C., 2011. A global climatology of atmospheric fronts. Geophysical Research Letters, 38(4)), applied? If frontal systems were explicitly identified and included in the dataset, it would allow a more rigorous assessment of the relative importance of MCSs vs. fronts in climate models' representation of precipitation. Another alternative would have been to combine the MCS tracking with the dataset presented in Fig. 4 by Schaffer et al. (2024). Can you clarify why the chosen criteria were judged to be the more robust solution than explicitly tracking and excluding fronts?

We opted against explicitly tracking and excluding fronts primarily for practical and methodological robustness. Implementing a parallel front-tracking algorithm (such as Berry et al., 2011) would require additional input variables (e.g., wind fields, temperature gradients) that are not always available at hourly resolutions in regional climate model outputs. Moreover, applying multiple complex tracking algorithms across large climate model ensembles substantially increases computational demands and data storage requirements.

Furthermore, combining two separate tracking algorithms introduces compounded uncertainties into the analysis. Distinguishing whether a resulting bias originates from the climate model's underlying physics, the MCS tracker, or the front tracker becomes exceedingly difficult. By relying on a single, self-contained precipitation tracker, we avoid these compounding errors and provide a cleaner, more interpretable reference for process-based model evaluation.

Detailed Comments

Title: First, I suggest slightly revising the title to better reflect the scope of the study. In its current form, it does not fully convey that an observation-based MCS climatology and a corresponding dataset are already presented. At the same time, the applicability to climate model data is prospective rather than demonstrated. Clarifying this distinction in the title would improve its accuracy and better align it with the actual contributions of the paper. In addition, the authors could also consider replacing "lifecycle-based" (which I believe all tracking algorithms are in some way) with what makes this tracker unique: the integration of instability data.

We now aimed for a rather descriptive and short title that should reflect the content of the paper better.

> EMMA-Tracker v1.0: A mesoscale convective system tracker and 27-year European observational climatology

Python: It does not matter where, but I would certainly recommend clarifying in the manuscript that this is a python-based tracking algorithm.

Good Point. We mentioned it directly in the bottom of the Introduction.

Data documentation: There are some additional variables in the published netcdf files such as "active_track_touches_boundary" and "active_track_id" etc that I could not find in the documentation. Please add a description of those variables as well.

Done.

l. 55-58: This sentence seems confusing because there is some overlap between meso-alpha and synoptic scales and it also seems like the authors actually include systems larger than 200 km as well. It makes sense to distinguish between the spatial scales of organized convection and frontal systems, but it would be helpful to be more precise to which variables and processes these scales apply (primarily motion or also precipitation).

We reframed the whole paragraph to align also better with all the other review comments. The updated paragraph is in the answer to the next question.

l. 63-66: I am not fully convinced by this line of reasoning. If convection is better represented but still misidentified as frontal systems, would this not simply appear as an increase in frontal systems? I suggest reformulating this argument to more clearly articulate the value of explicitly distinguishing between frontal systems and MCSs in the context of model evaluation.

We thank the reviewer for pushing us to clarify this point. The reviewer is correct that in terms of pure occurrence statistics, misclassification might simply look like a shift in category counts. However, the true value of explicitly distinguishing these systems lies in process-oriented model evaluation (such as storm-centered composites or environmental analysis).

Because Europe's precipitation climatology is heavily front-dominated, the environmental signal of any mixed dataset will be overwhelmingly dictated by large-scale baroclinic forcing. If researchers evaluate an ensemble of Regional Climate Models (RCMs) and Convection-Permitting Models (CPMs) against a mixed reference dataset, the resulting composites will look broadly similar across all models because they all capture large-scale baroclinic zones reasonably well.

However, genuine mesoscale convective organization often occurs in entirely different environments, such as the prefrontal warm sector hundreds of kilometers away from the primary baroclinic zone. If fronts are not explicitly removed from the evaluation dataset, their dominant large-scale signal completely masks the localized, instability-driven convective processes. Therefore, cleanly separating MCSs from frontal systems is an absolute prerequisite for isolating and evaluating the specific mesoscale processes where CPMs are expected to add value. We have reformulated this argument in the manuscript to clearly articulate this physical reasoning.

> Although fronts and MCSs coexist in nature, evaluating climate models requires researchers to cleanly separate their physical structures. If an observational reference dataset mixes these two systems, the evaluation process can become fundamentally flawed. Because Europe's precipitation climatology is heavily front-dominated, the environmental signal in any mixed dataset can become overwhelmingly dictated by large-scale baroclinic forcing. Even coarse-resolution RCMs represent these broad baroclinic dynamics reasonably well (Prein et al., 2015; Schaffer et al., 2025).
>
> However, accurately simulating the severe, short-duration hazards characteristic, such as the severe wind gusts of convective lines usually require convection-permitting scales (Kendon et al., 2017).
>
> Consequently, if researchers generate storm-centered composites or analyze environmental conditions using a mixed dataset, the well-resolved synoptic features dominate the signal, making RCMs and CPMs appear structurally similar.
>
> This baroclinic dominance completely masks the intense, localized physics of organized convective storms, which often develop in entirely different environments such as the prefrontal warm sector (Pacey et al., 2023).

l. 72: . . . together with satellite-retrieved surface precipitation

We changed this to explicitly acknowledge the advantage of pyFLEXTRKR in using precipitation to avoid the problem with non precipitating cirrus clouds.

l. 78: "which purposefully includes convectively active frontal systems" - Do you mean here that the work of Da Silva and Haerter (2023) includes frontal systems but distinguishes them from MCSs or they just label large organized precipitation as the relevant system, no matter the underlying physical processes? That could be useful to know because if it is the latter, your dataset could be used together with their dataset.

This is from DaSilva2023 manuscript: *Unlike previous studies, our MCS detection algorithm has the specificity of using cloud to ground lightning flashes to discriminate between convective and nonconvective precipitation features. It enables our algorithm to include large frontal thundery rainbands and warm cloud/low-topped convective clusters in the definition of MCSs, which may be missed by methods combining cloud top brightness temperatures with thresholds on the precipitation field. Whether these "ambiguous" systems should be classified as MCS is debatable (Feng et al., 2021) and beyond the scope of this study.*

They label large enough organized precipitation features that have at least one lightning strike as MCSs. Their data unfortunately is not publicly available. Depending on the studies in mind a combination of datasets (also together with frontal datasets) can be useful. As mentioned above for the specific case of climate model evaluation and process based analysis in climate models, using a dataset that relies on lightning data can not be used for fair reference comparison.

l. 81: Write out the abbreviation here as well (it can only be found in the abstract as of now)

Done.

l. 111: Figure S3 is referenced before Figure S1 and S2 - maybe consider changing their order?

We have adjusted the order of the supplementary figures accordingly.

l. 123: IMERG geopotential → This seems like an error or typo

Correct, we changed this to ERA5.

l. 164: I suggest replacing the description of this methodological step "System merging" with something like "Object identification" or more commonly used "Segmentation". Otherwise, it sounds like this is already about the merging and splitting of systems.

Thank you for pointing this out. We addressed this and went for "System Segmentation".

l. 191: It is not entirely clear how the overlap criterion is applied in the identification of merging and splitting events. Does the overlap required to define these events also need to exceed the 10% area threshold? In addition, how are more complex situations handled? For example, if multiple non-contiguous cells at time step t are in close proximity but only some of them overlap with a single cell at time step t+1, how is this classified? Conversely, if several cells exist at time step t+1 but only one overlaps with the cell at the previous time step, how is this treated? Clarifying these cases would help improve the transparency of the tracking methodology.

1. **The 10% Threshold:** To ensure consistency and avoid links based on accidental single-pixel contacts, the 10% area overlap threshold is applied to all tracking operations, including merging and splitting events. A candidate only participates in a merge or split if it meets this minimum overlap requirement. We added this explicitly in the manuscript.

2. **Proximity vs. Overlap:** Because the EMMA-Tracker is fundamentally based on spatial overlap, cells in close proximity that do not physically touch are treated as independent systems. They are not merged unless they overlap by at least 10% (or are captured by the rescue step using displacement vectors and then make the overlap criteria).

3. **Complex Splitting:** In cases where multiple candidates exist at time t+1 but only one overlaps with the parent system from time t, the event is classified as a simple continuation rather than a split. The non-overlapping candidates are initialized as new, independent tracks.

> In all three cases, a link is only established if the precipitation areas overlap by at least 10%.

l. 209: Since the dataset is published alongside this paper, I suggest clarifying that the mask files with the same dimensions of the regridded input data are also available. This is important since they usually are more useful for more advanced analyses than the track statistics only.

Thank you for pointing this out, we added this to the code data availability.

Fig. 2: Why is the reduced smoothing necessary to visualize the merging and splitting? Does this not mean that the smoothing may not be necessary and that there could also be an advantage in retaining the fine-scale structures of the MCSs? Also, please check the description for panel a) since there is no dashed line, but instead only a solid line and a grey area. For panel b), would it not be more appropriate to choose a timestep that shows the actual merging with the overlap rather than the non-overlapping features that are about to merge? That would at least be more consistent with the text.

We selected this specific timestep because it illustrates a textbook Derecho event moving from eastern France toward Austria with exceptionally clear heavy precipitation cores. The mild spatial smoothing was applied to the tracking algorithm here specifically to keep the system unified for an additional timestep before it physically split and propagated southward toward Croatia.

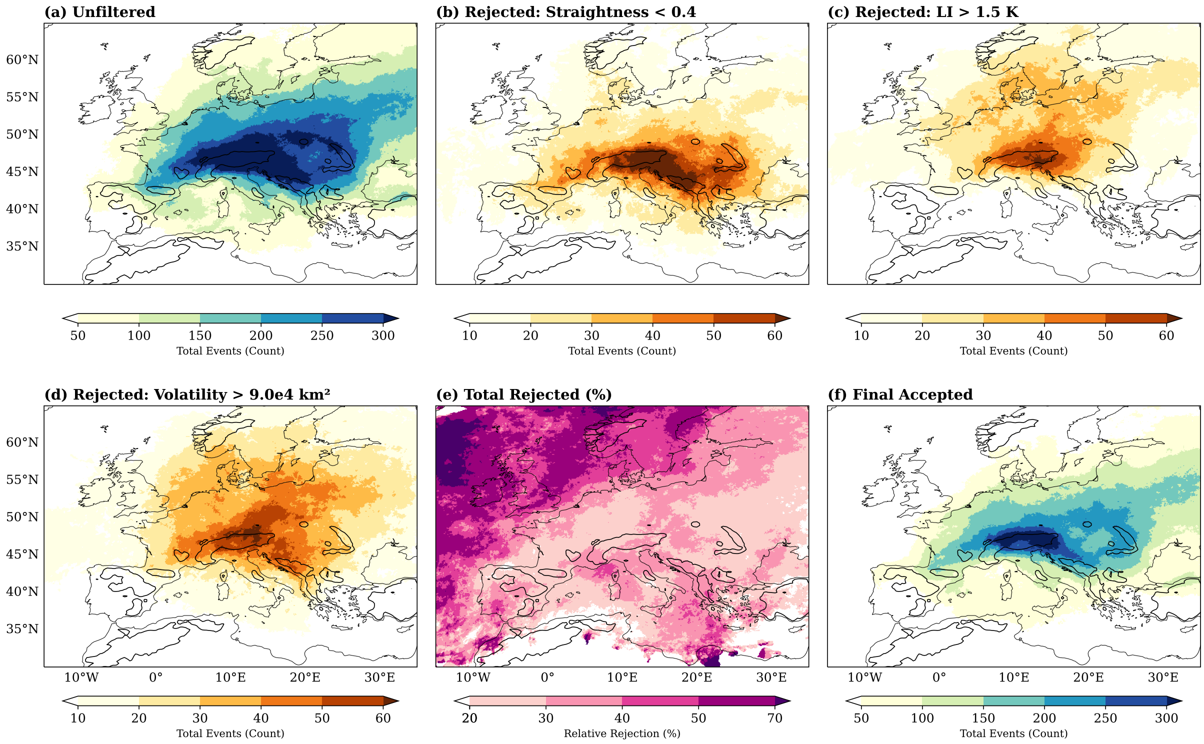

Figure 7: Panel (e) now shows the percentage of rejected candidates relative to the total number of candidates in panel (a). This now highlights the regions where the applied filters have the largest relative impact on the candidate population, which can be compared directly to the frontal activity shown in Fig. 4.

Finding a real-world example that consecutively demonstrates all three track behaviors—continuation (t_0), merging ($t_0 + 1$), and splitting ($t_0 + 2$)—in a single, clean sequence is quite rare. However, for maximum scientific transparency, we have explicitly noted in the text that the visualization in this specific figure relies on this spatial smoothing to demonstrate these consecutive tracking behaviors.

> Fig. 3: Would it make sense to show panel (e) as a percentage of panel (a), similar to Fig. 4 Panel (e) indicates that the regions where most storm candidates are rejected correspond to regions with the highest MCS activity, which is not surprising. However, as the authors note, frontal systems may play a proportionally larger role in areas influenced by the North Atlantic storm tracks. This relative importance is not currently apparent. Presenting panel e) as a fraction of panel a) would provide a clearer picture of the regions where the largest relative fractions of storms are removed by the applied filters. This would also facilitate a direct comparison with Fig. 4, allowing readers to assess whether the systems filtered out, presumably to remove fronts, are indeed captured in the alternative frontal dataset.

We thank the reviewer for this excellent suggestion. Plotting panel (e) as a relative percentage significantly improves the clarity of the figure and allows for a much more direct comparison with the frontal activity in Figure 4. We have updated the manuscript accordingly.

l. 215: “rather than relying solely on instantaneous properties available during the initial detection phase” sounds like the convention of MCS trackers is to apply filters only to the MCS initiation phase. Do you have a reference for this to which tracker you compare this to? In my experience, many trackers apply filters across the lifecycle. So, I am wondering if the novelty of EMMA is maybe that the criteria are applied to EACH timestep instead of a criterion that only must be true for a given period of time, such as a certain precipitation volume during at least four hours in Feng et al. (2024)?

We completely agree that the original phrasing was imprecise; state-of-the-art algorithms (e.g., PyFLEXTRKR, TAMS, MOAAP) do indeed apply duration and lifecycle thresholds, such as requiring a system to maintain a minimum area for a continuous period—a criterion that EMMA also employs.

We have updated the manuscript to clarify our intended meaning. The core novelty of EMMA's post-processing is not that it evaluates the lifecycle, but how it evaluates it. While standard convention often defines a successful track by checking if static size and precipitation thresholds are maintained during a contiguous ”mature” phase, EMMA goes a step further. It applies dynamic, behavioral filters—specifically trajectory straightness, step-by-step area volatility, and mean environmental instability—evaluated across every single timestep of the storm's entire existence, from initial detection to final dissipation. This approach ensures the system physically behaves, propagates, and evolves like a genuinely organized mesoscale system throughout its whole lifespan.

> A core novelty of the EMMA-Tracker is how it evaluates the system's entire lifecycle. Rather than just checking if a system meets static thresholds for a predefined number of hours, EMMA assesses the storm's physical behavior, such as its track straightness and step-by-step area changes, at every single timestep.

l. 222, 234, 242: Add a colon after each criterion

Done.

l. 243: Do you have a reference for the statement that the area of frontal systems changes more quickly than those of MCSs?

We thank the reviewer for this insightful question. We have clarified this section to show that this 'volatility' is not necessarily a fundamental physical property of the fronts themselves, but rather an algorithmic signature of how their precipitation is detected.

Because synoptic-scale systems (like fronts) are often large and fragmented in precipitation, a tracking algorithm may intermittently merge or split separate precipitation clusters as they evolve. This causes sudden, large 'jumps' in the detected area—what we refer to as detection artifacts—which are physically implausible for the typically smooth upscale growth and decay of a single, distinct MCS.

We use this area volatility as a heuristic filter to separate these non-coherent synoptic features from distinct MCSs. This choice is supported by our validation against an independent frontal database (Fig. 4), which confirms that systems rejected by this filter are overwhelmingly associated with frontal activity (often exceeding 80% overlap). We have revised the text to make this distinction between physical evolution and detection behavior more transparent. We tried to formulate this now more clearly in the manuscript:

> While distinct MCSs typically follow a smooth upscale growth and decay lifecycle, the precipitation fields of large-scale frontal systems often appear fragmented to a tracking algorithm. This fragmentation leads to rapid, physically implausible fluctuations in the detected area as separate clusters are merged or split. To distinguish these diagnostic tracking artifacts from the natural evolution of organized convective systems, we apply an area volatility filter.

References

Peter Berg, Christopher Moseley, and Jan O. Haerter. Strong increase in convective precipitation in response to higher temperatures. *Nature Geoscience*, 6(3):181–185, Mar 2013.

J. L. Catto, N. Nicholls, C. Jakob, and K. L. Shelton. Atmospheric fronts in current and future climates. *Geophysical Research Letters*, 41(21):7642–7650, 2014. doi: https://doi.org/10.1002/2014GL061943. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL061943`.

Jennifer L Catto and Stephan Pfahl. The importance of fronts for extreme precipitation. *Journal of Geophysical Research: Atmospheres*, 118(19):10–791, 2013.

Nicolas A. Da Silva and Jan O. Haerter. The precipitation characteristics of mesoscale convective systems over europe. *Journal of Geophysical Research: Atmospheres*, 128(23):e2023JD039045, 2023. doi: https://doi.org/10.1029/2023JD039045. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023JD039045`. e2023JD039045 2023JD039045.

Z. Feng, J. Hardin, H. C. Barnes, J. Li, L. R. Leung, A. Varble, and Z. Zhang. Pyflextrkr: a flexible feature tracking python software for convective cloud analysis. *Geoscientific Model Development*, 16(10):2753–2776, 2023. doi: 10.5194/gmd-16-2753-2023. URL `https://gmd.copernicus.org/articles/16/2753/2023/`.

Zhe Feng, L. Ruby Leung, Nana Liu, Jingyu Wang, Robert A. Houze Jr, Jianfeng Li, Joseph C. Hardin, Dandan Chen, and Jianping Guo. A global high-resolution mesoscale convective system database using satellite-derived cloud tops, surface precipitation, and tracking. *Journal of Geophysical Research: Atmospheres*, 126(8):e2020JD034202, 2021a. doi: https://doi.org/10.1029/2020JD034202. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JD034202`. e2020JD034202 2020JD034202.

Zhe Feng, Fengfei Song, Koichi Sakaguchi, and L. Ruby Leung. Evaluation of mesoscale convective systems in climate simulations: Methodological development and results from mpas-cam over the united states. *Journal of Climate*, 34(7):2611 – 2633, 2021b. doi: 10.1175/JCLI-D-20-0136.1. URL `https://journals.ametsoc.org/view/journals/clim/34/7/JCLI-D-20-0136.1.xml`.

Zhe Feng, Andreas F. Prein, Julia Kukulies, Thomas Fiolleau, William K. Jones, Ben Maybee, Zachary L. Moon, Kelly M. Núñez Ocasio, Wenhao Dong, Maria J. Molina, Mary Grace Albright, Manikandan Rajagopal, Vanessa Robledo, Jinyan Song, Fengfei Song, L. Ruby Leung, Adam C. Varble, Cornelia Klein, Remy Roca, Ran Feng, and John F. Mejia. Mesoscale convective systems tracking method intercomparison (mcsmip): Application to dyamond global km-scale simulations. *Journal of Geophysical Research: Atmospheres*, 130(8):e2024JD042204, 2025. doi: https://doi.org/10.1029/2024JD042204. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024JD042204`. e2024JD042204 2024JD042204.

T Fiolleau and R Roca. A new methodology for the detection and tracking of mesoscale convective systems in the tropics using geostationary infrared data. *Journal of Applied Meteorology and Climatology*, 52(6):1327–1344, 2013.

E Galanaki, K Lagouvardos, V Kotroni, E Flaounas, and A Argiriou. Thunderstorm climatology in the mediterranean using cloud-to-ground lightning observations. *Atmospheric Research*, 207:136–144, 2018.

Klaus Haslinger, Korbinian Breinl, Lovrenc Pavlin, Georg Pistotnik, Miriam Bertola, Marc Olefs, Marion Greilinger, Wolfgang Schöner, and Günter Blöschl. Increasing hourly heavy rainfall in austria reflected in flood changes. *Nature*, 639(8055):667–672, March 2025.

Riccardo Hénin, Alexandre M. Ramos, Sebastian Schemm, Célia M. Gouveia, and Margarida L. R. Liberato. Assigning precipitation to mid-latitudes fronts on sub-daily scales in the north atlantic and european sector: Climatology and trends. *International Journal of Climatology*, 39(1):317–330, 2019. doi: https://doi.org/10.1002/joc.5808. URL `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5808`.

Elizabeth J. Kendon, Nikolina Ban, Nigel M. Roberts, Hayley J. Fowler, Malcolm J. Roberts, Steven C. Chan, Jason P. Evans, Giorgia Fosser, and Jonathan M. Wilkinson. Do convection-permitting regional climate models improve projections of future precipitation change? *Bulletin of the American Meteorological Society*, 98(1):79 – 93, 2017. doi: 10.1175/BAMS-D-15-0004.1. URL `https://journals.ametsoc.org/view/journals/bams/98/1/bams-d-15-0004.1.xml`.

Guangxing Lin, Christopher R. Jones, L. Ruby Leung, Zhe Feng, and Mikhail Ovchinnikov. Mesoscale convective systems in a superparameterized e3sm simulation at high resolution. *Journal of Advances in Modeling Earth Systems*, 14(1):e2021MS002660, 2022. doi: https://doi.org/10.1029/2021MS002660. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002660`. e2021MS002660 2021MS002660.

Petra Mikuš, Maja Telišman Prtenjak, and Nataša Strelec Mahović. Analysis of the convective activity and its synoptic background over croatia. *Atmospheric Research*, 104:139–153, 2012.

C. Morel and S. Senesi. A climatology of mesoscale convective systems over europe using satellite infrared imagery. ii: Characteristics of european mesoscale convective systems. *Quarterly Journal of the Royal Meteorological Society*, 128(584):1973–1995, 2002. doi: https://doi.org/10.1256/003590002320603494. URL `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/003590002320603494`.

Christopher Moseley, Peter Berg, and Jan O. Haerter. Probing the precipitation life cycle by iterative rain cell tracking. *Journal of Geophysical Research: Atmospheres*, 118(24):13,361–13,370, 2013. doi: https://doi.org/10.1002/2013JD020868. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JD020868`.

Kelly M Núñez Ocasio and Zachary L Moon. Tams: a tracking, classifying, and variable-assigning algorithm for mesoscale convective systems in simulated and satellite-derived datasets. *Geoscientific Model Development*, 17(15):6035–6049, 2024.

G. Pacey, S. Pfahl, L. Schielicke, and K. Wapler. The climatology and nature of warm-season convective cells in cold-frontal environments over germany. *Natural Hazards and Earth System Sciences*, 23(12):3703–3721, 2023. doi: 10.5194/nhess-23-3703-2023. URL `https://nhess.copernicus.org/articles/23/3703/2023/`.

Andreas F. Prein, Wolfgang Langhans, Giorgia Fosser, Andrew Ferrone, Nikolina Ban, Klaus Goergen, Michael Keller, Merja Tölle, Oliver Gutjahr, Frauke Feser, Erwan Brisson, Stefan Kollet, Juerg Schmidli, Nicole P. M. van Lipzig, and Ruby Leung. A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, 53(2):323–361, 2015. doi: https://doi.org/10.1002/2014RG000475. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014RG000475`.

A. Schaffer, T. Lichtenegger, A. Ossó, and D. Maraun. Resolution dependence and biases in cold and warm frontal heavy precipitation over europe in cmip6 and euro-cordex models. *Weather and Climate Dynamics*, 6(4):1815–1830, 2025. doi: 10.5194/wcd-6-1815-2025. URL `https://wcd.copernicus.org/articles/6/1815/2025/`.

Armin Schaffer, Tobias Lichtenegger, Heimo Truhetz, Albert Ossó, Oscar Martínez-Alvarado, and Douglas Maraun. Drivers of cold frontal hourly extreme precipitation: A climatological study over europe. *Geophysical Research Letters*, 51(20):e2024GL111025, 2024. doi: https://doi.org/10.1029/2024GL111025. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024GL111025`. e2024GL111025 2024GL111025.

Russ S Schumacher and Kristen L Rasmussen. The formation, character, and changing nature of mesoscale convective systems. *Nature Reviews Earth & Environment*, 1(6):300–314, 2020.

Wei-Ming Tsai, Suqin Duan, Travis A. O'Brien, Jennifer L. Catto, Paul A. Ullrich, Yang Zhou, L. Ruby Leung, Zhe Feng, William R. Boos, D. L. Suhas, Fiaz Ahmed, and J. David Neelin. Co-occurring atmospheric features and their contributions to precipitation extremes. *Journal of Geophysical Research: Atmospheres*, 130(5):e2024JD041687, 2025. doi: https://doi.org/10.1029/2024JD041687. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024JD041687`. e2024JD041687 2024JD041687.