Point-by-point rebuttal letter - 2nd round of revisions

May 21, 2026

Pre-print manuscript for Geoscientific Model Development:
Agertoft, N., Su, J., Pedersen, J. W., Ringgaard, I. M., and Larsen, M. A. D.: ImpactETC1.0: Impact-oriented tracking of extratropical cyclones with global optimisation and track reconciliation, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-4466,2025.

Reviewer 1

I would like to thank the authors for making substantial improvements on the clarity of the methods and results in this manuscript. I think the gradient tracing method to link storms to storm surge impact is a great addition and the added animations enhance the readers ability to assess the methods shown. I think the paper would be improved overall from some minor corrections. One suggestion I have is making the conclusion less of a summary of the results/discussion section and instead each section focused on unique information.

We acknowledge this point and have tried to sharpen the conclusion by focusing less on a broader summary and more on specific quantitative and qualitative findings.

Please see specific comments:
R1C1: Line 100-102: It could be argued that this statement would be widely true of any dataset of tracked extratropical cyclones. I believe the framework here is applicable to other locations in terms of storm surge and other local effects, however there are other impacts which cause hazards far from the storm location including large swells which can travel across entire oceans or as mentioned later in the paper, precipitation events and inland flooding which will require a significantly different set of parameters to link to storm events. It would be good to add emphasis on the applicability to local events here.

We have emphasized the applicability to local events, at Line 102-104, by adding: The framework is in its current form mainly applicable to local events such as storm surges, i.e. impacts close to the ETC track. The challenge of dealing with impacts occurring far away from the ETC, i.e. inland flooding and large swells, is further discussed in section 5.3.

R1C2: Lines 574-576: This seems like a trivial update to the post processing step that probably does not deserve mentioning again here, as it is described earlier on in the paper.

Fair point, we have removed these three lines in the updated version of the manuscript.

R1C3: Lines: 617-619: This paragraph feels repetitive compared to the final points of the discussion. I don't think this reiteration is necessary. I would suggest saving the discussion on broader impacts for the conclusion section only.

In our revisions of the entire Discussion section following the comment above, this sentence has now been removed.

Reviewer 2

I appreciate the authors taking the time to consider the reviews and provide both a response and an updated manuscript. The revised manuscript is (in my eyes) materially improved. In particular, I believe the authors have strengthened the discussion around impact attribution and compound-event ambiguity, clarified the purpose and limitations of their BLOB reconciliation approach for terrain-driven track fragmentation, and improved the algorithmic framing of assignment choices (Hungarian vs greedy/nearest-neighbor) by focusing more on practical performance and behavior rather than implying meteorological "ground truth."

The added gradient tracing description is interesting; I like the authors exploring different ways for defining cyclones driving impacts. I could still see failure modes here (e.g., noisy sea level pressure fields which can occur in less diffusive models, sub-cyclone mesoscale pressure minima in very high-res simulations), but the authors do discuss some of these at the end of the manuscript, which I thought framed things well.

I still have some remaining comments below, mostly around further phrasing and clarification. These are primarily about claim discipline, a clear definition of what TA measures, a small amount of additional guidance for tuning Dmax, and some other small typos/phrasing tweaks. I broke them down in "broader comments" and "smaller ones" –I do think they should be considered, but assuming they are addressed in a reasonable fashion, I would expect the manuscript to be close to publishable.

** Broader comments

R2 - BC1: "Accuracy" framing

I still think some moderation of remaining claims of "accuracy" without independent track validation is warranted.

While several sections are now more careful, some headline statements still imply that the framework delivers "accurate" tracking in a general meteorological sense. The evaluation presented primarily demonstrates: - Differences in assignment behavior (Hungarian vs greedy), - Changes in continuity/fragmentation and track statistics (including reconciliation effects), - Improved event association in the sense of returning a plausible number of storms per impact event (your TA-like framing, see below).

These are valuable, but they do not constitute a direct validation of trajectory-level accuracy (i.e., position, intensity evolution, lifecycle) against an external reference (e.g., manual analysis, independent tracker comparison, or reanalysis-based synoptic verification). I would recommend replacing language such as "accurate tracking" with more precise claims consistent with what is demonstrated, for example, "more continuous / less fragmented tracks" or "improved practical association of impacts to candidate storms."

We acknowledge the point and have made an effort to catch phrases in the paper where the "accuracy" of the tracking is addressed, replacing with e.g.

"skill" or "more continuous/less fragmented" as nicely suggested. Also, for the Storm Count Accuracy score, we have added a note stating that it addresses counts of tracks per event, not the skill of reproducing the events themselves.

R2 - BC2: Clarify what "True Accuracy" (TA) measures and what it does not

The TA-style metric (as described in the revision) appears to quantify whether the method returns the correct number of storms per impact event (or within an AoR and time window), based on manual labeling of storm count. This is a useful metric for impact association. However, most readers (see above) will interpret "accuracy" as track-path correctness or physical attribution of the forcing mechanism. I suggest adding a few lines at the first introduction of TA, explaining being very explicit that it evaluates storm-count attribution (number of relevant tracks per event) rather than track-path accuracy or causal hazard attribution. Also, ensure subsequent discussion uses terminology consistent with this meaning (for example, "count correctness" or "event association accuracy" rather than "tracking accuracy").

We used the term "true accuracy" because of its common use in binary, contingency table analysis within broader data science application, but we do see the potential confusion around the word "accuracy" in the context of ETC tracking. We have therefore decided to rename the metric to "Storm Count Accuracy" and made the explanation at the first introduction of the metric more explicit (see lines 335-342, in Section 3.4.1). In the subsequent discussion section (Section 4.3), we have made the requested terminology adjustments, so that we explicitly talk about storm count accuracy and not tracking accuracy.

R2 - BC3: Provide clearer guidance or minimal sensitivity regarding Dmax (and other key hyperparameters)

The revision improves the narrative around parameter choices and how they might impact results (i.e., the "tuning" problem). However, I still have some questions about Dmax... it is simultaneously justified as a pragmatic allowance for terrain-driven discontinuities but later shown to be insufficient in some cases. This is not a flaw, as the authors note, but it means the manuscript should offer readers practical guidance for tuning Dmax in other datasets and time steps. Frankly, this will greatly increase the reproducibility and portability of the framework if people download the code and use it themselves for their applications. My gut tells me just acknowledging this is worthwhile, although a minimal sensitivity analysis (even for a subset): show fragmentation rates or track continuity metrics for Dmax = 200, 300, 400 km (or similar) would be interesting.

We have carried out a minor sensitivity analysis by varying both D_{max} and the pruning radius together. Since this step of the full framework happens before the BLOB-based track reconciliation and the post-processing filtering of small tracks, we cannot directly assess "fragmentation rates" and "continuity"

for the final resulting tracks metrics. We chose to only look at the "stitching" step of the framework and assess the resulting number of tracks being produced and the duration/length of these tracks. We describe the key take aways from this analysis as well as some additional practical guidance for how to define this parameter in the main manuscript (see the revised Lines 421-433 in Section 4.1). We have placed a figure (Figure A.1) with the results in Appendix A alongside more details on how the experiment was conducted.

** Typos and formatting suggestions.

R2C1: Line 6: "includes several novel" – again, I am not sure I'd call these features "novel". Might just say "... includes several algorithmic..."

We do believe that the combination of the Hungarian algorithm, BLOB detection, post-processing options such as the gradient tracing and storm selection as well as the single storm score parameter tuning are indeed novel. However, as the previous sentence also calls the ImpactETC1.0 a "novel framework", we can agree to remove the novelty issue here. This has been implemented accordingly.

R2C2: Lines 93-94: Similarly, novel is used twice in succession: "Motivated by these challenges, we introduce a novel ETC tracking framework designed to enhance the relevance of ETC tracks for on-the-ground impact assessments. The new framework contains several scientific novelties:" – I might replace the first novel with "new," maybe the second one could be "developments".

Good point. Revised.

R2C2: Line 111: Is the native CERRA grid something e.g., Lambert Conformal and not regular lat/lon? Might be worth just adding what type of grid is being interpolated from.

I think it's probably also worth pointing out that the method (as described) requires a Cartesian grid. The authors may feel this is self-evident, but with the growing adoption of unstructured meshes in the climate modeling community, it is worth noting. Developing a method that performs well on unstructured meshes (i.e., without the need for regular latitude-longitude data) might be a useful target for future work.

We added the specific type of grid to "Section 2: Data and region of study" (line 113, see the addition in red font in this sentence):
"..., all variables were re-gridded from the native CERRA Lambert conformal conical grid to a regular latitude–longitude grid using bilinear interpolation to simplify subsequent processing and ensure consistency in spatial derivatives."

We have also added the application to unstructured grids to the end of Section 5.3 "Future improvements" lines 609-611 , which now reads:
"The framework currently only works for data on structured grids but future

work could explore how to adapt it to data on unstructured meshes, where especially the candidate identification and BLOB reconciliation steps would need improvements."

R2C3: Line 155: For this step, would it be possible to make the feature detection stage embarrassingly parallel since the correspondence problem is only solved after all timesteps have been analyzed? While I can imagine 5km cells to be "expensive," with a standard HPC for the current year, this might be more feasible.

You are right that this step could be made embarrassingly parallel. This question triggered a discussion in the author group about how we would optimize the code for parallel processing in general. We have therefore decided to include a brief description of this in the existing discussion paragraph on *computational efficiency* in Section 5.3, see lines 556-563, that now reads:
The current implementation of the framework is serial but future developments could exploit parallelisation on modern computing infrastructure. Since the framework is event-focused by design, it would be straightforward and efficient to analyse independent impact events in parallel. Specific algorithmic steps could further benefit from parallelisation, especially the candidate point identification and the BLOB construction, since they treat individual time steps independently. The track reconciliation based on the constructed BLOBs would still have to remain serial to avoid conflicts during merging of track fragments. The grid search employed for the impact-oriented calibration of post-processing parameters is embarrassingly parallel. The "nearest storm selection" and "gradient tracing" post-processing options treat each impact location independently and could similarly be parallelised.

R2C4: Line 185: I might call this "small pruning radii" instead of "lax pruning."

Good point - revised accordingly.

R2C5: Line 187: I am surprised the local minima values are exactly the same, which is difficult even with single precision. I suppose keeping both in these cases is fine, but functionally (and from a meteorological perspective), I do not see how it is different than applying a random choice.

We double checked this, and as an example we see that just for the first of the 37 ETC impact event in our study, it happened 15 times that a local minimum had a nearby local minimum (within the pruning radius distance) with an identical MSLP value. So it is not that uncommon actually.

Technically, keeping both is not the same as a random choice, since keeping both means that two tracks that start and end in each of the respective time steps can be reconciled via our BLOB step. This is not be possible in all cases

if only one candidate point is kept, since the point that is kept might barely be out of range of the candidate point in the previous or next time step, thus leading to a fractured track.

R2C6: Line 243: The first track-breaking mechanism can sometimes be mitigated by allowing temporal gaps during stitching; see Ullrich et al. (2021), which is already cited.

You are right. In the author group, we have been debating whether to handle this track-breaking mechanism during "stitching" or inside our BLOB reconciliation operator. The future work section in the discussion already includes a mention of the BLOB option for such a fix (Section 5.3, under "*Improved Tracking*"), but we have now also added references there to Ullrich et al. (2021) and also Perez-Alarcon et al. (2024) as examples of how to handle this during stitching. See lines 568-569 that now reads:
Or alternatively to allow for temporal gaps during the HA/NN solutions to the correspondence problem (Ullrich et al., 2021; Peréz-Alarcón et al., 2024).

R2C7: Line 253: I am not sure the word "hypothetical" is needed here, since this is commonly how sea level pressure correction is applied operationally.

We removed the word "hypothetical".

R2C8: Line 254: Consider adding a reference supporting why/when SLP reduction breaks down over complex terrain (physical reasoning and prior documentation).

We added a reference at line 256, Pauley 1998, on the problems with SLP reductions.

R2C9: Line 316: "We note that the AoR is centred on 60^o lat, 15^oN and initially spans from 50 - 70^o lat and 0 - 30^oE." This seems wrong for a few reasons. One, I think they mean 60N, 15E, but also it should be "N" and not "lat."

You are of course right, this was a typo. It is corrected now. See line 326.

R2C10: Figure 5. Consider slightly reducing the contour density, which makes a lot of noise over the Alps, Turkey, N. Africa, etc. I would also suggest the storm track centers be made a different color (blue? purple?) to better stand out against the underlying shading. That or the points should be larger with a bolder outline.

We have updated the colours of the storm track centres to cyan so that they better stand out against the underlying colour palette of wind speeds. The high contour density of 1 hPa intervals were a request from Reviewer 1 in the previous round of revisions - with the purpose of being able to see if the contours are

closed around the candidate points. We experimented with lowering the density to 2 hPa intervals, but this made it impossible to see the closed contours for several of the candidate points. We have therefore decided to keep the high density of contour lines, even if this makes the plot a bit noisy over the Alps, Turkey, N. Africa, etc., and hope that this is OK with the reviewer.

R2C11: Figs 6b and 7: Why is there a white patch in the middle of the ETC in the top right (northeast) corner?

We expanded the description of the event that is shown in Figure 6 and changed the colour of one of the candidate points in Figure 6(b) to improve the explanation, see the new lines 272-278:
"Figure 6(a) shows an example of a single ETC track that has broken into two fragments as it passed over the Scandinavian mountains. Figure 6(b) shows a snapshot at the time step where the track broke. Here, the yellow point is the candidate point under investigation, the red points are other ETC candidate points in the same time step, and the shaded areas are the computed BLOBs where MSLP values in the domain sit within a range of ± 5 hPa of the MSLP value of the yellow point. In this case, there are three separate BLOBs in the domain but the algorithm only looks for red candidate points within the BLOB that contains the yellow point. The rest are disregarded."

For the ETC in the top right corner, its MSLP values around its center is simply below the ± 5 hPa range of the MSLP value of the point over Norway (the candidate point under investigation in this example).

R2C12: Table 5: Consider reducing precision (fewer decimal places) to improve readability.

We agree, and have reduced the precision for readability. During handling of this comment, we realised that we had made a mistake in the column "Total run time" where we had put in wrong numbers during the last round of revision to the manuscript. We have therefore re-run the timing experiments and updated all the numbers. This re-running did not change any of the conclusions and the text that discusses the results in the table have therefore not changed.