Point-by-point rebuttal letter

March 17, 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

The manuscript titled “ImacptETC1.0: Impact-orientated tracking of extratropical cyclones with global optimization and track reconciliation” introduces a new method of tracking extratropical cyclones in the Nordic region and linking them to storm surge impacts observed at water level recording stations in the region. The authors present a three step cyclone tracking process including an optimized global solution to connect storm centers in time, a BLOB method to connect storm tracks across mountainous terrain and land-sea barriers, and some post processing steps to identify cyclones of interest.

Overall, I think the manuscript presents some interesting ideas to the challenge of extratropical storm tracking, the most influential potentially being the so called ‘BLOB method’ used to track storms through discontinuous jumps across mountainous terrain. I think the manuscript could be greatly improved by providing a more detailed presentation of the results and reworking of certain thresholds defined in the method. I also think the results are lacking focus on the impacts of individual cyclones. Details of the magnitude of these events could be shown for example, and linking to specific storm or storms done through a wind and proximity analysis seems readily available. Therefore, I am recommending major revisions to the current work. Please find my detailed comments below.

We appreciate the reviewer’s careful evaluation of both the methodological framework and the presentation of the results. In particular, we value the reviewer’s comments regarding (i) the need for a clearer and more detailed presentation of the tracking results, (ii) the definition and justification of several key thresholds used in the algorithm (including their sensitivity and physical interpretation), and (iii) strengthening the connection between tracked cyclone events and the observed storm-surge impacts at tide gauge stations. We have revised the manuscript accordingly, including clarifications and additional discussion of parameter choices, improvements to figures and post-processing methods, and additional analyses/illustrations that better highlight the magnitude and characteristics of impactful cyclone events. Below, we respond to each comment point by point and describe the changes we have made in the revised manuscript.

R1C1 Line: 83: “impact-irrelevant” is dependent on the specific aim of each study. Dynamically strong systems out at sea could be very hazardous to those traveling by ship or coastlines vulnerable to large swell. I’d argue few strong storms are without any impact.

We agree with the reviewer that ”impact-relevance” is dependent on the scope of the study. Our use of the term was intended to distinguish between cyclones that trigger a specific higher-end stakeholder-relevant event (in this case, coastal storm surges) and those that do not.
We chose storm surges (i.e., sea levels) as our primary indicator of ”impact”

because they provide a physically consistent reference point to initialise the tracking algorithm. However, this framework is not limited to surges; if the focus were on inland wind damage or shipping hazards, the algorithm would function similarly provided a ”time of impact” is specified. We will clarify in the Introduction and the Discussion that ”impact” in this paper is exemplified by surges, but the framework is designed to be broadly applicable to other hazard types.
Notice also that we, as a result of this review stage, revise the algorithm and corresponding event list input data, to also include the location of impact so that the nearest storm track relevant for each event is detected.
We clarified in section 1 what is meant by ”impact-irrelevant”. Please see the description in lines 96-100 in section 1 and in lines 568-569 in section 5.3.

R1C2 Lines 82-85: I believe the method presented in this paper has many of the same issues described in these lines. The authors apply similar “simple” thresholds with respect to minimum duration, maximum MSLP center, and minimum track length, with a small addition of proximity to the impacted region. I would suggest softening this argument in the introduction.
We removed the word ”simply”. Please see the updated description in line 81.

R1C3 Figure 1: An additional marker or box showing the domain of panel on the left in right panel map would be helpful to the audience

Good suggestion. We added the domain of the left panel to the right side panel in Figure 1.

R1C4 Line 136: Expanding this time window could help limit the amount of storms filtered out for short duration time later.

We agree with the reviewer to the extent that expanding the time window could serve to make recognition of some tracks easier during post-processing, as it has for some cases been observed that the correct track begins at the first investigated time step. Increasing the time window could potentially highlight these correct tracks further. However, some tracks are not long-lasting in nature, and therefore expanding the time window could lead to additional, locally impact-irrelevant tracks appearing with the characteristics of a correct track, at least in terms of the time span of the track. Thus, expanding the time window around the timing of the impact-peak has the potential downstream ramification of filtering out correct tracks from other events during post processing. Please see the updated description in lines 146-148.

R1C5 Figure 2: Do all sea level events have photography? Why de-tided? Tides have large impact on flooding whether high or low, neap or spring. A weaker extratropical cyclone could have large impacts if it occurs during a spring high tide.

Fundamentally, the ETC tracking algorithm can be used on any type of event, not just storm surge events, so the choice of de-tiding is not so important for this paper. Here, we simply want to use the case of storm surge events in Denmark to present an application of the tracking algorithm. We have chosen to de-tide the water levels to ensure we examine significant surge events caused by severe storms rather than tidal influence, which varies significantly across Danish Coastlines. Therefore, the timing of storm surges, in e.g. the Wadden sea, in relation to the tidal phase, is very crucial for the severity of the peak. By de-tiding, this influence was removed from the algorithm's development. If one would want to study the conditions of storms that produce surges along Danish coasts (which we actually aim to do in future research), then a discussion of tides vs wind-induced surges is very relevant. Please see the rationale of de-tide in lines 117-122. The photography in the figure is simply a visualisation of an impact that has happened - it is only meant for this algorithm overview figure and not for any scientific use in this paper.

R1C6 Line 145-147: How is the upscaling done? Are you left with a 350km grid afterward? If not, how much coarser is the grid compared to the original 5.5km? Is the upscaled grid directly related to the pruning radius? In essence this is down sampling your grid, which makes sense given the spatial resolution is far smaller than typical scales of ETCs. Where is the location of the final minima placed within the grid?

We acknowledge that our own choice of the word "upscaling" in the manuscript text can mislead the reader here. We do not "upscale" the grid in the sense that we coarsen the grid resolution. Instead, we start by splitting the full CERRA domain into non-overlapping square boxes with diagonal lengths of 350 km. We then perform our minima localisation on the original 5.5 km grid cells inside each box, which results in a single minimum within each box. Afterwards, we assess whether the minimum in each box is within a certain distance (the pruning radius parameter) of identified minima in nearby boxes. If the location of a point is within the pruning radius of another minimum, then the weaker of the two minima is eliminated as a potential candidate point. We save the latitude and longitude coordinates for the minima as well, so that the location of the final minimum in each spatial square is the 5.5 km grid cell that has the lowest value, thus remaining true to the original spatial resolution without upsampling. Please see the updated description in lines 152-154.

R1C7 Line 151: What is MSLPmax based on? Having a higher or lower background field of MSLP could affect the precision of this parameter. Could consider using the gradient or Laplace of MSLP to avoid this problem or as another check in your method.

It is correct that the background field will affect the precision of this parameter. However, our choice of a relatively simple maximum cap on MSLP values (MSLPmax) is to simply filter away a lot of non-relevant local minima that stem from either noise in the data or very weak minima that we are confident cannot produce significant impacts. The reviewer is right that it would be possible to implement a stronger filtering step through additional checks e.g. based on the gradient or Laplace of MSLP. Our choice of a simple maximum cap is motivated by the fact that this is extremely computationally efficient, which we deem important as the "Identify candidate points" step of the algorithm already is one of the computationally demanding steps in the algorithm. In the case of storm surges in this paper, we have intentionally set the value of MSLPmax at a conservative value of 1010 hPa. Please see the updated description in lines 166-170.

R1C8 Line 153: How do you know your relative vorticity intensity isn't driven by a strong frontal feature and possibly not associated with an ETC? The addition of a wind speed threshold in the same radius may be interesting as well. Also see Gramcianinov (2020) for tracking using relative vorticity. Gramcianinov, C. B., Campos, R. M., de Camargo, R., Hodges, K. I., Guedes Soares, C., & da Silva Dias, P. L. (2020). Analysis of Atlantic extratropical storm tracks characteristics in 41 years of ERA5 and CFSR/CFSv2 databases. Ocean Engineering, 216, 108111.

Thank you for the reference to Gramcianinov et al. We do not know if the relative vorticity is from a strong frontal feature, which might be true for a given case. As with the previous comment on the MSLPmax parameter, it would be possible to implement a stronger filtering step here than a simple vorticity threshold. Our choice is again motivated by computational speed. A wind speed threshold could potentially be interesting, but we believe it might be complicated to define a threshold value that is useful over both ocean and land. We are worried that it may filter away candidate points over land that we would consider interesting. Please see the updated description in lines 171-176.

R1C9 Figure 4: With this MSLP threshold you may miss storms if they're located in a higher background pressure field.

We believe this comment is similar to R1C7, see the full response up there. In short, we have chosen a quite conservative MSLP threshold value of 1010 hPa, which does not filter away impactful ETC's. We clarified this in the revised manuscript in line 170.

R1C10 Please add lat/lon ticks on all axes.

We added lat/lon to all relevant figures.

R1C11 Around the pressure minima shown in figure 4d, do you see closed isobars at 1mb intervals or do you see closed circulation in the wind field? I believe more validation of these cyclone minima is required.

When examining the need for validation of cyclone minima, we see two trade-offs that are relevant to discuss. One is about computational speed of the algorithm (the number of validation steps vs. the computations needed to perform these), and the other is about how conservative one wishes to be in terms of filtering potential candidate points (essentially the ratio between type 1 and type 2 errors). Based on several previous reviewer comments as well as this one, it seems that the reviewer generally prefers stronger filtering of minima than we have chosen to implement in this algorithm. The argument being that if we let too many minima "survive" this step, then it will be more difficult and potentially introduce errors when we connect points through time during tracking. We believe this to be a valid concern. However, we would like to refer to our analysis of computational requirements of each algorithmic step (presented in the original manuscripts Table 5). Here, "data load" and "identify candidates" already consume significant proportions of the total algorithmic run time. Additional validation checks, especially those that require computation of derived fields (e.g. Laplacian or gradients) or loading of additional reanalysis fields (such as wind speed), would significantly increase the total algorithmic run time. We therefore prefer to err on the side of making type 2 errors, i.e. allowing more minima to pass through the identification step. We are confident that the two super fast steps of the algorithm (tracking through time and post-processing) instead will handle and filter out these potentially non-relevant minima at a fraction of the computations it would require to handle them through more thorough validation early on. This discussion is highly relevant to the manuscript, and we added Figure 5 showing the closed isobars and wind field circulation for the specific event in Figure 4d. At this level (1 mb), we see closed isobars, closed circulation in the wind field, and significant wind speed levels for about half of the identified minima at this specific time step.

R1C12 Line 180: This implies your threshold for ETC speed is 600km/hr which is unrealistic compared to previous literature. How do you reconcile this or validate that these storm centers should be connected in time?

We agree that 600 km/h is not a physically realistic speed for the movement of an ETC. However, in the context of high-resolution reanalysis data (such as CERRA) and complex terrain, the D_{max} parameter functions more as a "search radius" for a diagnostic variable (MSLP minima) than as a physical speed limit. In regions with complex topography, the identified MSLP minimum can "jump" between grid cells, relative to the location of the synoptic system, from one hour to the next due to local pressure perturbations or artefacts, even if the synoptic system itself is moving at a more "normal" speed. While we also have a subsequent BLOB-based reconciliation step to fix these issues, that step is much more computationally demanding than both HA and NN, and we therefore aim to handle minor jump issues in this step here, as this is significantly more efficient. We have experimented more with the parameter values for D_{max} and checked the effects on the final tracks. Based on this, we will, in fact, implement a lower value of 300 km in the revised manuscript, where we will also improve our discussion on how to set this parameter for a given use case. 300 km is still a bit larger than what other tracking algorithm case studies often employ (which seems to be in the range of 100-150 km per hour), but that difference is related to the aforementioned problems with MSLP over complex terrain. Please see the updated description in Lines 210-218 (Section 3.2).

R1C13 Lines 199-200: The Hungarian Algorithm seems to only be given information at 2 time steps (t and t+1), then connects the local minima between these time steps according to the shortest distance summed across all connections. What are the implications of this decision affecting downstream connects in time? This framework could lead to the premature termination of tracks by limiting future options for connection. In other words, if you chose a point further away from your minima at time t, you have different minima to choose from within your 600km distance threshold at time t+1, t+2, t+3 and so on. Would be great to see an animation of tracks being formed by the method and how the NN functions in comparison over the hourly evolving MSLP and wind field.

The implemented HA and NN algorithms optimises connections with information from just two adjacent time steps, t_i and t_{i+1}. The limitation of such a one-step-at-a-time approach is that there are two ways this leads to a track breaking. First, if one of the time stamps along the ETC path is missing a candidate point, i.e. no local minima could be found. Second, even if a local minimum exists for all time steps, it can also have downstream implications for the connections that can be made later at t_{i+2} onwards. Tracks can break if the "optimal" connection from t_i chooses a point at t_{i+1} that is further away from the minimum in t_{i+2} than one of its competing points, thus potentially not allowing for a connection to be made between t_{i+1} and t_{i+2}. For such a scenario to happen, the user would have to specify a D_{max} value that is larger than the pruning radius, r, which creates the possibility of multiple candidate points to choose from within the D_{max} distance. Mitigating this second possibility, was the original main justification for introducing an extra algorithmic step to handle when tracks break due to the required connection being longer than what D_{max} allows, i.e. the BLOB step of the algorithm. This description is now inserted at Lines 195-203, Section 3.2. The request for animations is good and will improve understanding of the algorithm. In our Zenodo repository, we now provide an animation of the HA and NN algorithms connecting points through time for an illustrative event where the two solutions differ. In this specific example, the NN solution actually ends up with two tracks that cross each other in a single time step. See the revised description in Lines 407-409, Section 4.1.

R1C14 Line 218: Connecting candidate points that are 2500 km apart from on another seems excessive. The typical radius of an ETC is on the order of 500km. How sensitive is this process if you set the maximum distance between candidate points to 500km or 1000km? I'm concerned this will very likely lead to spurious connections across storms which should not be connected in time.

The "Max BLOB candidate point distance" parameter that the reviewer here refers to should be high enough for the algorithm to be able to fix track splitting and premature termination caused by, e.g., complex terrain, yet not be so large that spurious connections can be made. For the revised manuscript, we now perform three experiments with different pruning radius parameter values, which yield different problem complexities. With this we show that very few connections extend beyond 1000 km - even if they are allowed to do so. Instead, the vast majority of connections are below 600-800 km, which we deem to be a reasonable distance given that ETCs traverse significant mountainous terrain. Based on these extra experiments, we have updated our discussion for this parameter, and modified the recommended values to 600-800 km. See lines 271-274 in section 3.3, and lines 435-440 in section 4.2.

R1C15 Line 235-240: I think a criteria related to the local winds and water level would improve this framework. Why not include a criteria of the storm center needing to be within a certain distance (maybe 500-1000km) of the impacted area during the peak of the surge/wind event? Table 1. Why not define AoR using a radius of a certain distance around the impact location?

There are three separate suggestions in this comments (1: local winds/water levels, 2: max distance to between ETC center and impact location, 3: define AoR as a circle around impact location):

1. In our framework, we start from a known impact (here a storm surge) for which the user provides the location and time. We therefore do not need an additional step on local water levels (or winds) to identify if an impactful event has happened and if there is a relevant ETC that needs to be tracked for the event. The user has already made that decision.
2. We believe that the use of an AoR and associated requirements to how many time steps an ETC spends within the AoR is very related to including a criterion that the ETC centre should be within a certain distance at the time of impact. Rather than being within a certain distance (here 500-1000km), the ETC center needs to be within the AoR, which is essentially a user-defined "distance". Rather than being close (within the AoR) at the specific peak impact time, it has to be within the AoR during a user-specified time window. Requiring this at the exact peak impact time is essentially an equivalent, but more strict criterion than what we already have. Instead of implementing this as a post-processing step that filters out identified tracks based on a distance threshold, we would like to propose that we implement your suggestion in comment "R1C24", where we in the revised code automatically will choose the ETC track that is closest to the impact location at the impact time.
3. We prefer to use a rectangular AoR since this is easier and more efficient to work with programmatically, when subsetting and slicing many 2D fields. In many specific use cases (such storm surges in Denmark), there can be

> good arguments for not choosing an AoR that is symmetrical around the impact location (such as a circle with a certain radius). In this case, the geographical extent of the semi-enclosed Baltic Sea means that relevant ETCs can be located further to the east and north of the impact location (Denmark) than to the west and south. That is why we choose an AoR that extends further north and east in this study. See the updated version in lines 303-324.

R1C16 There is also a question of the wind direction the impact location is experiencing. Depending on where you are in relation to the storm center being tracked you could be experiencing winds in opposite directions, which would directly impact the surge experienced. Adding such a criterion could help limit mis assigned storms.

It is correct that where you are relative to the storm's location will determine the wind direction you are locally experiencing. However, developing a criterion based on this is easier said than done, especially for complex coastal regions such as the Southern Baltic Sea with its semi-enclosed nature containing narrow entrances at the Danish straits and its many small islands. We imagine that this would require a lot of work by the users leading up to employing the algorithm. Users would need detailed information on the physical processes causing impacts, e.g. local, oceanographic knowledge of which wind directions a location is sensitive to. That would actually be something that storm tracking could help users figure out in the first place, so we would risk circular applications. (In fact, one of the applications that our research group intends to use this algorithm for in the future is to figure out which types of storms produce the impacts in different locations in complex coastal regions).

R1C17 Figure 7. Why are there storm centers that are within one 350km pruning radius of one another? Shouldn't these have been pruned or connected in time already given the 600km threshold of HA connections in time. Numbers 13,60,27, 73, 35.. all seem like their connecting 2 points in very close proximity to one another.

This issue relates to how we select and prune candidate points. During the candidate identification step, we have in a few cases observed that there can be multiple local minima with the exact same MSLP value that both survive all the filtering steps described in Section 3.1. In this case, there are several valid candidate points in the given time step, t, within the pruning radius distance of each other. We have chosen to keep both points as potential candidates rather than making an arbitrary decision between one of them. We now describe this in Lines 187-191 (Section 3.1). When the HA and NN algorithms then have to connect candidate points through time there is a chance that the track ends prematurely, if the connection from $t-1$ to t is to one of the points, but the continuation onwards from t to $t+1$ continues from the other point. We let the BLOB step of the algorithm handle this issue by connecting these two fragmented tracks. We now clarified this in Lines 441-447 (Section 4.2).

R1C18 Table 2. I'm not convinced the optimal solution from HA is physically correct with regards to real world tracking, therefore I find it inappropriate to use it as ground truth to the nearest neighbor approach. How would this method compare to a small subset of hand drawn tracks?

We agree with the reviewer, that the optimal solution from HA, while guaranteeing a mathematically optimal solution for the t to t+1 correspondence problem, does not necessarily result in physically correct tracks with regard to real-world tracking. We have now re-written the language regarding optimality and ground truth throughout the manuscript (especially in section 4.2 where the results of the two algorithms are compared), and instead frame our implementation of HA as a novel application to storm tracking, which we benchmark against a "standard" nearest neighbour approach. As stated earlier (in the reply to R1C13), we now provide an animation of an example event, so that readers can gain a better understanding of how the two algorithms work in practice.

R1C19 283: This may be true in terms of minimizing distance summed across all tracks, but I don't see clear evidence of this method producing "More accurate tracks." This implies the shortest track is always the most accurate which seems like it wouldn't be true in all cases. One specific example is the case of a storm center splitting into 2, this tracking method would miss a potential storm track if one center was slightly further away at one time step, but was longer lasting overall. Wouldn't it be better to track both or perhaps even the longer lasting segment that was initially further away?

This relates to the previous comment (R1C18). By itself, the HA simply minimizes the summed distance across all tracks, but importantly it only chooses the "shortest track" connection for each individual time step. In the specific example of a track splitting into 2, we do in fact only end up with a single track, which is a limitation of the algorithm. However, the BLOB step of the algorithm will actually select the overall longest lasting of the two split tracks as the final storm track, not the shortest one. Please see the updated version in lines 275-280 in Section 3.3, and the example event in Figure 7. We have removed the use of "most accurate", since this is only guaranteed from a mathematical optimization perspective.

R1C20 Line 298: What was the nature of the one event not tied to a storm track?

The impact was a storm surge on the island of Bornholm in the Baltic Sea. The surge was based on easterly winds, which were caused by strong pressure (MSLP) gradients over the Baltic Sea. The gradients formed due to the interaction between a minor ETC located over Ukraine and a high-pressure system that develops over Southern Norway. Ideally, the algorithm should have identified the minor ETC over Ukraine. When we inspect the candidate locations and tracks from the algorithm, they show that the track is not inside the AoR for long enough to exceed the calibrated post-processing parameter that requires a minimum number of time steps inside the AoR. As we discuss in the manuscript, such cases show that a single, unique parameter set will never perfectly capture all ETC's. In the revised manuscript, we have introduced two other post-processing methods for identifying the impact-inducing ETCs, i.e. choosing the nearest ETC (Section 3.4.2) and the gradient tracing method (Section 3.4.3). These methods accurately identify the correct ETC for this event.

R1C21 Line 315: Varying AoR by degrees lacks physical meaning as distance between lines of longitude will vary substantially depending on where you are in the domain. Why not use a kilometers-based approach?

This is a valid point. However, due to our relatively small AoR and its location, this is not an issue in the present study. This has been clarified in section 4.3, including estimates of the variation in one degree between the northern and southern part of the domain. Further, the option to use kilometres-based distances instead of geographical degrees has been added to the list of future improvements in section 5.3 (lines 574-576).

R1C22 Line 371: Alignment with the underlying MSLP field is very difficult to observe in the still images covering a wide time/space range. I think an animation in the supplementary materials would be more affective.

We provided animations for the example events (Figure 13 in the revised manuscript) in the repository, highlighting the underlying MSLP field.

R1C23 Line 375-376: These large jumps could be due to the large search radius of 600km, as this occurs while the storm center is still offshore but seems to be getting distracted by noisier MSLP minima along the coast, or perhaps the merging of another low-pressure system. Would choosing a smaller initial search radius and allow the BLOB method to connect the tracks after improve the result?

The case event referred to here, "2000-01-29" in Figure 11 in the original manuscript, has a vaguely defined pressure minima, with a large area of similar MSLP values. The track jumps because the exact location of the smallest MSLP value in the vague field jumps long distances. There is nothing for us to do during the "stitching". This would have to be "fixed" by a new post-processing step that smoothes the tracks, which some other algorithms have decided to implement. Another fix would be to track areas rather than points in time, as e.g. implemented in the TempestExtremes tracking algorithm. However, we consider this outside the scope of our algorithm.

R1C24 Figure 11 (Lines 377-378): how were orange and green tracks determined to be irrelevant? I think different post processing criteria could be applied to improve performance of the algorithm. Why not add a step that chooses the storm track closest to the impact location at the height of the event or something similar?

Upon returning to this comment and the challenge on choosing the relevant vs. irrelevant tracks, we ended up developing two additional methods choosing the relevant tracks. The first one is, as suggested here, choosing the closest track, hereby assuming that this will always be the one that affects the impact location the most. As each impact location is assigned a track, multiple tracks may still be chosen for each event. However, as the closest track might not always be the culprit, we developed a second method using gradient-tracing. Here, a 'ball' is placed on the impact location at the time of peak impact and the MSLP that the ball rolls to will be the one that caused the impact. The assumption here being that the impact is caused by the ETC that contains the impact location within its low-pressure "basin". Only locations associated with pre-identifed tracks are possible end-locations for the ball. As a further development of this, we introduced the option to include randomness in the initialisation of this process, resulting in a percent-wise division of the most likely tracks to cause the impact. This allows for several tracks to be associated with each impact location, i.e. indicating a potential compound event.
To maintain the clarity of our core results and avoid distracting the audience with secondary analyses, we have opted to include these new features within the Methods (Section 3.4.2 and 3.4.3) and Discussion (Section 5.3) sections rather than performing an extensive re-analysis in the Results. This allows users to understand and apply these optional tools without losing focus on the calibration of the ImpactETC1.0 framework.

R1C25 Line 384: how is the relationship between steep gradient and water level quantified? Is timing of the ETC and water level considered?

With this sentence, we simply mean that steep pressure gradients cause strong winds, which cause the storm surge. This was just a descriptive analysis of the event, and no quantitative analysis was done (Line 486).

R1C26 Line 430: I agree that more work should be done to assess the interaction of multiple ETCs and I think some of the conclusions drawn about storms being "impact irrelevant" may be premature here. Since the manuscript is focused on impact on sea level, I think some of these steps should be addressed here. It should be relatively low effort to apply one or two post processing steps to the current work.

We agree with this comment and have addressed in the response to R1C24.

R1C27 431-432: Wind speed could also be investigated to limit the filtering out of impactful storms.

We agree, with the caveat that filtering based on additional variables (here wind speed) would increase computational demand as discussed earlier. But we added a discussion of this to the future improvement options (lines 560-562).

R1C28 449: As previously stated, I think suboptimal is poorly defined and the global optimal solution should not be used as benchmark in this manner.

We agree with the reviewer, and have changed the wording accordingly (Line 591).

R1C29 452-453: ETCs exist on timescales more than adequately resolved by the model being used here. One could even argue the high resolution is already unnecessary for resolving these large atmospheric features and could result in more confusion and runtime for algorithms as the number of small pressure perturbations increase. As resolution increases it is likely that model output would be down sampled spatially, much in the same way that the upscale gridding is performed here. I don't see the practicality of ETC tracking with higher resolution.

We agree that not all applications might benefit from continuously refining model grids, and fully acknowledge the points of both reviewers. We see two separate issues here: (1) whether higher resolution is unnecessary for resolving large atmospheric features, and (2) whether tracking is beneficial in higher resolution.

As a general rule, finer-scale models are able to better resolve not just atmospheric patterns themselves, but also processes and physics that in themselves lead to better results, also for storm tracks specifically (see e.g. Polichtchouk et al. (2025): "Effects of Atmosphere and Ocean Horizontal Model Resolution on Tropical Cyclone and Upper-Ocean Response Forecasts in Four Major Hurricanes". Monthly Weather Review). Moving to finer scales in geoscience has been the general trend, in line with advances in computational power, for decades. It is also likely the case that associated hazards (precipitation, wind fields, etc.) may be better resolved with higher resolution. The reviewer is likely right that there will not be major benefits to tracking ETCs in very high resolution grids. However, there may be benefits in terms of tracking other atmospheric objects, such as convective rainfall cells. Please see the revised manuscript in lines 593-596.

Reviewer 2

In this paper, the authors introduce ImpactETC1.0, software that tracks extratropical cyclones (ETCs). They specifically focus on linking detected/tracked storms to observed regional impacts such as storm surges. The authors argue three key novelties to this work: (1) the application of the Hungarian Algorithm to globally optimize the "correspondence problem" of connecting storm centers across timesteps, which reduced "suboptimal connections" compared to traditional greedy algorithms; (2) a "BLOB" analysis technique for track reconciliation that merges broken tracks when storms cross orographically complex regions, and (3) an automated calibration procedure for post-processing parameters using a new metric they termed "Single Storm Score." Applied to historical storm surge events in Denmark using the high-resolution CERRA reanalysis dataset (1991-2020), the code successfully identified a series of impact-relevant ETCs.

In general, the paper is interesting and well-written (some minor typos are noted below). The material is well-suited for GMD. I would contend that the authors slightly overemphasize certain aspects of the work, and there is some additional room for contextualization and framing. More details regarding that are below. However, this critique aside, the ideas are worth considering, and the field of storm tracking more broadly lacks papers that discuss the "nuts and bolts" algorithmic performance and combine that with a discussion of tradeoffs, which this paper does. I would suggest some form of revisions based on the itemized feedback below. However, pending those revisions, I think the paper could be suitable for publication (and relevant for interested researchers) in GMD with some additional work.

Note, I have not tested the software/data, but I have verified that the DOI linking to the Zenodo is valid, and it contains code that appears sound, so this would adhere to EGU's data requirements.

We sincerely thank the reviewer for the constructive and encouraging review, and appreciate the positive assessment of the manuscript's suitability for GMD. We will revise the manuscript to strengthen contextualization and framing, clarify the physical motivation for key hyperparameters (e.g., pruning radius), the role of global optimization, refine the description and scope of the BLOB reconciliation and tracking-variable choices, expand the discussion of impacts and compound events, and moderate the use of terms such as “novel” and “innovative.” We have also addressed all minor comments and improved several figures and text passages accordingly. Below we provide a detailed point-by-point response.

Major Theme #1, tracking mechanics during ETC trajectory building

The Hungarian algorithm implementation is interesting, and the argument for potential use cases for it over a greedy nearest neighbor is compelling computationally. As the authors do note, most of the wall clock time of their software is tied up in data ingestion, so even an order of magnitude slower during the "stitching" step is unlikely to have deleterious consequences on the overall workflow.

Regarding the physical motivation for the algorithm application. From a synoptic meteorology perspective, where the greedy algorithm seems to fail (Table 2) is for small pruning radii. This makes intuitive sense because there are far more potential "local minimum" low-pressure centers with a smaller radius with which to merge/eliminate nearby ones.

However, the realized physical spatial scales for ETCs are 1000km (if not larger, this can be deduced from observation or scale analysis of the Navier-Stokes equations). Given this, I would struggle to find a practical reason why an ETC tracker would need to have such a small pruning radius. If we agree on this, the Hungarian algorithm doesn't objectively buy a lot of skill, because with larger pruning radii (physically consistent with ETC scales), the deviation between the nearest neighbor and the Hungarian Algorithm narrows significantly. However, as noted by the authors, it doesn't cost a lot more, and I think they do a good job of arguing that, without the cost being prohibitive, it is a physically defensible choice. The action item here may be for the authors to discuss how the scales of motion in weather phenomena dictate hyperparameter settings such as the pruning radius. I could imagine a situation where such an algorithm may be more beneficial for "noiser" fields (e.g., tracking individual thunderstorms, cloud tracing, etc.), so perhaps this can be noted as a target for future application.

The reviewer makes good points. We agree that physical atmospheric processes should inform parameter selection. In our case study of storm surge impacts, we have observed that events can be caused by smaller systems that would have been filtered away with a pruning radius of ~1000 km (e.g. secondary lows). We also note that, in addition to the physical atmospheric processes, the dataset/model under investigation and the variable on which tracking is performed are relevant to parameter selection. ETCs may be on scales of 500-1000 km, but if there are artefacts in the reanalysis or MSLP issues due to topography, then this will have to be handled in the parameter selection as well, especially for parameters such as D_{max}. Reviewer 1 also commented on the need for more justification and discussion of parameter values and their implications. In the revised manuscript, we have implemented this for all steps of the algorithm. We now described how to set the pruning radius (Section 3.1), D_{max} (Section 3.2) and the BLOB parameters (Section 3.3) in the revised manuscript.

We have added a discussion of how the Hungarian Algorithm may be more beneficial for tracking smaller objects in the atmosphere and noisier fields, and that future research should address this. See lines 520-524 in Section 5.2

Finally, the authors discuss "higher-resolution for ETCs" in the final paragraphs. While local scale impacts (e.g., precipitation, specific winds in coastal channels, etc.) may be better resolved, I'd argue that ETC tracking doesn't benefit from high-resolution (and may be better served by coarsening data anyways).

We agree that not all applications might benefit from continuously refining model grids, and fully acknowledge the points of both reviewers. We see two separate issues here: (1) whether higher resolution is unnecessary for resolving large atmospheric features, and (2) whether tracking is beneficial in higher resolution.

As a general rule, finer-scale models are able to better resolve not just atmospheric patterns themselves, but also processes and physics that in themselves lead to better results, also for storm tracks specifically (see e.g. Polichtchouk et al. (2025): "Effects of Atmosphere and Ocean Horizontal Model Resolution on Tropical Cyclone and Upper-Ocean Response Forecasts in Four Major Hurricanes". Monthly Weather Review). Moving to finer scales in geoscience has been the general trend, in line with advances in computational power, for decades. It is also likely the case that associated hazards (precipitation, wind fields, etc.) may be better resolved with higher resolution. The reviewer is likely right that there will not be major benefits to tracking ETCs in very high resolution grids. However, there may be benefits in terms of tracking other atmospheric objects, such as convective rainfall cells.

While not necessary for the paper, it would be interesting to see how the algorithm compares to a more established ETC tracker such as TRACK (Hodges) or TempestExtremes (Ullrich). Given the synoptic scales mentioned above, I would assume there is likely minimal difference, particularly for well-defined tracks, between the methods.

We agree that a comparison with another existing tracking algorithm would be interesting. However, such an undertaking would be quite laborious and require a thorough assessment and discussion of the other tracking algorithms pros and cons. We therefore consider it outside the scope of the paper - as also suggested by the reviewer. But we agree with the reviewer that for well-defined tracks, e.g. those over the ocean, there is likely minimal difference. Potential differences would be more likely to occur over complex terrain. We added a discussion of this in Section 5.1.

Major Theme #2, tracking variable choices and broken track evaluation For the BLOB analysis, I interpret the code to work such that if I have a point that exists at $t = t_1$ and another point exists at $t = t_3$ (or thereabouts) with a large spatial gap, the BLOB area operator allows them to be "glued" together as a single track based on an overlap strategy. Frankly, in many ways, this seems to behave somewhat like Ullrich et al. (2021) and Peréz-Alarcón et al. (2024), both of which the authors mention. Can the authors comment as to the specific, unique benefits of applying such analysis versus allowing time-varying gaps in the "gluing" stage? I.e., does the BLOB method provide materially different results than the simpler logic during stitching of ”any ETCs within nhrs of model time and (nhrs * max-travel-dist/hr) in space are considered the same system?”

This is not correctly understood by the reviewer, which is, of course, our responsibility to convey clearly. The BLOB implementation does not address gaps in time, but instead focuses on what to do when a track splits into several downstream fragments that overlap in time. The reviewer explains an example of a gap in time along the track, i.e. when there are no candidate points at time t_2, but a connection from t_1 to t_3 is desired. What the BLOB operator actually handles is when there are more than one valid candidate points at time t_2 that actually belong to the same ETC, which can lead to an unwanted break and fragmentation of the track. This is especially necessary for spatial jumps across mountain ranges and rough terrain, both when the D_{Max} parameter is set to a small value (and connections thus cannot be made due to large spatial jumps), and when multiple candidate points are available e.g. on each side of a mountain range. We have improved the description of the BLOB step in the revised manuscript to avoid potential misunderstanding. Please see lines 252-262 in Section 3.3.

The reasoning behind the BLOB logic is to deal with the common challenge of the land surface and orography influencing near-surface quantities (particularly problematic for mean sea level pressure (MSLP), as it’s a derived quantity based on the elevation and surface pressure in the model). It is my understanding that the design choices of alternative tracking fields (i.e., see the contributions to the Neu BAMS paper that the authors cite) are chosen in large part because of some of the orography-induced challenges the authors point out here. The authors then go on to use MSLP as a core tracking variable (in conjunction with 500mb vorticity). It appears this was done because low-level vorticity fields are noisy (aside: this is not unexpected in high-resolution data, although this can be smoothed as in many of Hodges’ papers). However, this does introduce a bit of what I would consider semi-circular logic (”we have a problem of broken tracks to solve, but we choose a tracking variable that has been shown to be prone to broken tracks”). I’ll admit I’m playing a bit of devil’s advocate here since I prefer tracking on MSLP myself (and MSLP is probably most tied to surface wind/wave forcing relevant for surge, which may be worth pointing out more aggressively in the manuscript), but some context in the manuscript might help smooth out the rationale.

These are good considerations by the reviewer. We have now improved our explanation and the context for why we chose MSLP (which is because it is most important for surface winds, as the reviewer also explains). The choice of tracking variable and trade-offs between tracking surface or upper air variables has a full paragraph devoted to it in the introduction (lines 46-59) with reference to other algorithms that make different tracking variable choices than us. We have furthermore elaborated on this issue in the justification for why we need the BLOB step (Lines 252-262, Section 3.3.)

Major Theme #3: Impacts and compound events

It is my understanding that the core of the ETC (i.e., the sea level pressure minimum) must be in the AOR for it to be classified as an impactful storm. However, it is well known that mid-latitude cyclones can have impacts extending far from the core of the storm (e.g., cold fronts). While I assume the authors are most concerned with the surge associated with the wind field near the comma head, it would be good to discuss this a bit more. For example, such a technique may struggle if applied to ETCs that produce precipitation impacts, as such impacts can be rather disconnected from the storm's dynamical core (e.g., see atmospheric rivers tied to mid-latitude cyclones).

We agree with the reviewer. We introduced a gradient-tracing method as an option in the post-processing step to solve the problem as reviewer described (Section 3.4.3). Here, a 'ball' is placed on the impact location at the time of peak impact and the MSLP that the ball rolls to will be the one that caused the impact. The assumption here being that the impact is caused by the ETC that contains the impact location within its low-pressure "basin". Only locations associated with pre-identifed tracks are possible end-locations for the ball. As a further development of this, we introduced the option to include randomness in the initialisation of this process, resulting in a percent-wise division of the most likely tracks to cause the impact. This allows for several tracks to be associated with each impact location, i.e. indicating a potential compound event.

Along the same line, I was a bit disappointed that an impact-focused paper didn't really discuss the hazards themselves. E.g., in this high-resolution model, one could easily analyze the wind field that is the driver of the surge in the region and see how this is tied to the ETC centers that are detected. The authors wouldn't necessarily need to run a hydrodynamical model, but just evaluate the magnitude and spatial extent of the wind impinging on the shoreline. In fact, I would argue that this is a key drawback of pointwise Lagrangian analysis; 1-D tracks lack information about the spatiotemporal evolution of the 3D atmosphere, land, and ocean during relevant extreme weather events. I think it would be a nice addition to the paper (a qualitative case analysis, not unlike what was taken with a handful of events) and could improve the "impact" (no pun intended) the work makes on the community. However, if the authors do not wish to undertake such an endeavor, it should be discussed more at the end of the manuscript.

We understand the concerns of the reviewer, and agree with the stated limitations of pointwise Lagrangian analysis and 1-D tracks. However, we believe that hazard analysis for individual events is outside the scope of this paper, where the main focus is to start from a known on-the-ground impact date and location, and then deliver an algorithm that can identify the main driver behind that impact (here and ETC and its associated track). It would be a natural next step for users of this algorithm, to start to associate the tracks with the hazard properties (here strong winds and storm surges) in order to further understand the impacts. This is in fact something we intend to utilize the algorithm for in the future ourselves, where we will analyse the properties of different tracks and the impacts they have on different coastal regions of the Baltic and North Sea. We added this discussion in lines 561-574 in Section 5.3.

Major Theme #4: Semantical accuracy

I have slight issues with phrasing like "several innovative components." A definition of "innovative" is "introducing new, valuable ideas, methods, products, or services that are original, solve problems, and improve upon the status quo, often by creatively combining existing elements or challenging norms to deliver fresh solutions and create impact." I will admit, I find the Hungarian optimization the most novel component of the work. I consider the BLOB analysis (see above) to be a distinctive slant on previous strategies aimed at addressing track gaps in space and time. The post-processing optimization can broadly be considered a basic grid-search hyperparameter tuning that prioritizes having a singular event in the impacts domain (i.e., it's computationally efficient because it's only modifying a small subset of hyperparameters). I'll also point out that, tied to the above, the hypothesis implicit in the optimization is to enforce the notion that a singular event leads to specific hazards. For very local scales, that may be the case, but there is a large volume of literature surrounding compound extremes, including multiple events/stressors amplifying a hazard (e.g., flooding).

While I think all of the above are worth publishing, I think the authors should take care not to deemphasize existing work in the space via the use of what I would consider somewhat loaded terms like "novel" and "innovative."

We certainly do not wish to deemphasize existing work. In the revised manuscript, we make sure that our choice of words addresses these concerns. We also improved the discussion of the post-processing optimization and the limitations of the Single Storm Score as objective function as pointed out by the reviewer (Section 5.3).

—

Minor comments:

R2C1 In general, I found the paper to be relatively free of typographical errors. I did see some spellings like "artefacts" that I think are less common usage, but Google assures me they are valid. As always, I encourage the authors to do a thorough read-through after revision, but I commend them on the attention to detail in the initial submission.

The commendation is appreciated, we did another round of grammar checks.

R2C2 Figure 1. Some more geographic information (e.g., lat/lon ticks) would help in the left panel.

We agree with the reviewer. We add lat/lon ticks in the left panel.

R2C3 Line 137. Figure 3 should be capitalized.

We capitalized it.

R2C4 Is Fig. 5a missing some of the inland points over Norway? (e.g., black 16/17 from Fig. 6)

Yes - we corrected accordingly.

R2C5 There are many examples where the possessive is applied to plural acronyms/words (e.g., ETC's in lines 380-384). These should not have an apostrophe (apostrophe only used for possessive or contractions).

We double checked the grammar throughout the manuscript.

R2C6 In Figs. 5 and 6, the colormap is inappropriately washed out on the strong cyclone in the NE corner. Adjust the color scale.

Yes, we revised it - see also next comment.

R2C7 For Figs. 5 and 6, I also wonder if a different colormap would be beneficial (perhaps a perceptually uniform colormap to avoid color-blind issues). Red-blue implies biases or differences visually.

We removed the colormaps and visualized it through contours instead.

R2C8 Fig 10. The inset is a bit small; I would try to make it larger. Also, it may be worth trying to add some more color contrast (e.g., coloring the land mass in the inset brown instead of gray or something to highlight it's a distinct figure from the parent it overlays.

We made the inset bigger in both Figure 12 and 13 and highlighted the land mass of the inset through the use of a different color.