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.

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.

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.

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

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

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.

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?

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.

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.

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

Please add lat/lon ticks on all axes.

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.

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?

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.

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.

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 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.

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.

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?

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?

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

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?

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.

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?

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?

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

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.

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

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

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.