

ImpactETC1.0: Impact-oriented tracking of extratropical cyclones with global optimisation and track reconciliation

Niels Agertoft¹, Jian Su¹, Jonas Wied Pedersen^{1,2}, Ida Margrethe Ringgaard¹, and Morten Andreas Dahl Larsen¹

¹Danish Meteorological Institute, Sankt Kjelds Plads 11, Copenhagen, 2100, Denmark

²DTU Sustain, Technical University of Denmark, Bygningstorvet, Building 115, Kgs. Lyngby, 2800, Denmark

Correspondence: Niels Agertoft (nie@dmi.dk)

Abstract. Extratropical cyclones (ETCs) play a critical role in shaping extreme weather events in the Nordic region, often driving storm surges, heavy precipitation, and high winds that can lead to significant socio-economic and environmental impacts. However, traditional cyclone-tracking methods focus primarily on large-scale atmospheric dynamics and do not explicitly link cyclone characteristics to their regional impacts. To address this gap, we introduce ImpactETC1.0, a novel framework that identifies and tracks ETCs, with a specific focus on their impacts, as illustrated here for storm surges. The framework includes several **novel** algorithmic features: global optimisation for the correspondence problem, BLOB analysis techniques to address track fragmentation arising from surface-level tracking over complex terrain, and three different post-processing options for identifying the impact-relevant ETCs. These are standard calibrated filtering, proximity-based selection of the geographically closest track, and a "gradient-tracing" approach that selects the storm that holds the impact location within its "low-pressure basin". Applied to the CERRA reanalysis dataset and focused on the Nordic region, ImpactETC1.0 successfully reconstructed ETC tracks across complex terrain and during periods of rapid storm evolution while keeping computational costs low. Compared with a standard nearest-neighbour heuristics-based approach, the global optimisation leads to more connections being formed at negligible additional runtime. The track reconciliation step was essential in preventing track fragmentation and premature track termination, producing storm tracks that were, on average, twice as long across complex land-ocean boundaries and mountain ranges. As the post-processing step is extremely quick to perform, a sensitivity analysis could be conducted, and a score (the Single Storm Score) was developed for automated calibration of filtering parameters. Together, these results demonstrate that ImpactETC1.0 enables **accurate** more continuous, less fragmented and impact-relevant ETC tracking.

1 Introduction

The tracking of extratropical cyclones (ETCs) has been a longstanding challenge in meteorology, with various methodologies developed to identify and analyse their evolution (see Walker et al. (2020), and references therein). Conventional ETC tracking methods rely on objective feature tracking algorithms that detect cyclones using observed or simulated atmospheric datasets (Hodges, 1995; Hodges et al., 2003; Neu et al., 2013; Raible et al., 2008; Walker et al., 2020). These methods have been widely applied in climate studies to analyse storm climatology and trends (Bengtsson et al., 2009; Feser et al., 2015; Hoskins and

Hodges, 2019; Lodise et al., 2022), they are also increasingly used in operational weather forecasting (Froude, 2010), extreme
25 weather detection (Ullrich et al., 2021), wind energy risk evaluation (Gonçalves et al., 2021), and risk impact assessment
(Hunter et al., 2016).

Tracking frameworks typically follow a conventional three-step storm-tracking process: identification, tracking, and post-
processing (Neu et al., 2013; Walker et al., 2020). Each of these steps introduces methodological decisions that influence
the final tracks. In the identification step, ETCs must be located geographically at each time step, which is usually done by
30 analysing atmospheric fields and identifying ETC characteristics, such as pressure minima or rotational motion around their
centres. The next step is to track ETCs through time by solving the so-called "correspondence problem". This is done by linking
identified ETC locations across time steps to form coherent tracks. The final post-processing step typically aims to filter out
irrelevant tracks and potentially smoothen noisy or abrupt ETC movement. Differences in algorithm design and the choice of
atmospheric dataset have resulted in significant variability in cyclone track statistics across different studies, making direct
35 comparisons challenging (Feser et al., 2015; Flaounas et al., 2023; Neu et al., 2013; Walker et al., 2020). Comparison studies
show that different tracking algorithms, even when applied to the same dataset, produce cyclone counts that differ significantly,
highlighting the sensitivity to identification criteria, correspondence strategies, and post-processing filters (Grieger et al., 2018;
Neu et al., 2013; Raible et al., 2008). These discrepancies are particularly pronounced in regions with complex terrain (Medina
and Houze Jr., 2016).

40 Many conventional tracking algorithms prioritise identifying storm centres and tracking their movement over time, without
explicitly linking cyclone characteristics to their impact-related consequences, such as flooding, storm surges, and wind damage
(Walker et al., 2020). This is reflected in the ETC identification step, where conventional definitions of ETCs often rely solely
on dynamical thresholds such as minimum mean sea level pressure (MSLP), and in the post-processing filter parameters, which
are defined without validating whether the resulting tracks correspond to relevant observed surface impacts (Zappa et al., 2013;
45 Catto, 2016).

The choice of atmospheric variable used in the ETC identification step strongly influences the resulting tracks (Walker et al.,
2020). Common choices are MSLP and its derivatives (Lodise et al., 2022; Pérez-Alarcón et al., 2024; Ragone et al., 2018;
Ullrich et al., 2021), relative vorticity (Flaounas et al., 2014; Lakkis et al., 2019; Hoskins and Hodges, 2019), and geopotential
height at various pressure levels (Aragão and Porcù, 2022; Hofstätter et al., 2016). Because the most severe impacts of ETCs
50 occur at the surface, tracking the cyclone centre at or near the surface is desirable from an impact-focused perspective. However,
surface-level variables can present significant challenges in regions with complex terrain. MSLP is not directly modelled in
reanalysis data, but instead derived from surface pressure via a hydrostatic approximation, which often breaks down over land,
especially in mountainous areas. Similarly, near-surface relative vorticity is affected by strong terrain gradients and land-ocean
boundaries. An alternative is to track ETCs using upper-level variables such as geopotential height or relative vorticity at levels
55 of 850 hPa or 700 hPa (Flaounas et al., 2014; Hofstätter et al., 2016; Priestley et al., 2020; Sanchez-Gomez and Somot, 2018).
However, the position and evolution of the upper-level features can diverge from the cyclone centre at the surface, and some
ETCs are shallow, making them difficult or impossible to track higher in the atmosphere (Lakkis et al., 2019). This divergence

can potentially cause a misrepresentation of the storm’s surface impacts. Hence, there is a need for methods that retain the impact-relevance of surface-level tracking while addressing the artefacts and discontinuities that arise near complex terrain.

60 Regarding the correspondence problem, most existing algorithms have implemented variations of a greedy nearest-neighbour heuristic where the closest available points in a sequence of time steps are connected (e.g., Aragão and Porcù, 2022; Flaounas et al., 2014; Pérez-Alarcón et al., 2024; Priestley et al., 2020; Ullrich et al., 2021). While nearest-neighbour heuristics are computationally efficient, they may lead to implausible or suboptimal connections, especially in dense storm fields or large, vague low-pressure areas. In such cases, there can be multiple possible ETC centres to connect to within a defined neighbour-
65 hood region. Some algorithms simply choose the nearest neighbouring point (Aragão and Porcù, 2022; Ullrich et al., 2021), while others predict where they expect the next point to appear and choose the options closest to that (Hofstätter et al., 2016; Sanchez-Gomez and Somot, 2018). Other algorithms make the choice dependent on the intensity of the ETC centre in the identification variable field and choose the option with the strongest extrema (Lodise et al., 2022; Pérez-Alarcón et al., 2024), or the option that yields the smallest change in value between time steps (Flaounas et al., 2014). It is common to deny connec-
70 tions above a certain distance (Aragão and Porcù, 2022; Lodise et al., 2022; Ragone et al., 2018), or if the difference exceeds a predefined threshold (Sanchez-Gomez and Somot, 2018). Some algorithms allow a small gap in the tracking procedure when the ETC centre cannot be located for a short period of time, to avoid the track breaking (Pérez-Alarcón et al., 2024; Ullrich et al., 2021). These additional features increase the likelihood of forming good connections, but do not guarantee that the correct one is formed. Moreover, in each correspondence problem, this issue gradually worsens as the set of possible connections
75 is exhausted, further reducing the likelihood of finding the optimal match.

Post-processing steps typically aim to filter out identified ETCs that are deemed insignificant to a particular study. This can be short-lived systems filtered using parameters such as minimum storm duration and minimum track length (Aragão and Porcù, 2022; Lodise et al., 2022; Ullrich et al., 2021), or systems deemed too low in intensity, with thresholds on maximum MSLP or minimum relative vorticity along tracks (Ragone et al., 2018; Sanchez-Gomez and Somot, 2018). Parameter values
80 for these post-processing filters are rarely calibrated against impact data. Instead, they are determined ad hoc for a given study purpose, often through round values such as a minimum duration of 24 hours and a minimum length of 1000 km. As a result, these methods may retain dynamically strong but impact-irrelevant systems or, conversely, discard slower-moving, small, or short-lived storms that drive local extremes. The lack of a consistent, impact-oriented filtering approach presents a major limitation to the application of storm track analysis to hazard assessment or long-term climate adaptation planning.

85 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 developments:

1. A global, yet computationally fast, optimisation solution for the correspondence problem with the Hungarian Algorithm (Kuhn, 1955; Munkres, 1957). This includes benchmarking against a typical nearest-neighbour heuristics to assess trade-offs between computational efficiency and tracking continuity.

- 90 2. Binary Large Object (BLOB) analysis techniques to address discontinuous jumps and fragmentation of storm tracks caused by MSLP issues, thereby improving surface-level tracking of the ETC centre over complex terrain and land-sea boundaries.
3. Impact-oriented calibration of common post-processing parameters, such as minimum track length, duration and spatial proximity to the impact location. This includes an assessment of the ability of the algorithm to capture the correct number
95 of ETCs per impact event, and multiple options for resolving track ambiguity.

The framework aims to distinguish between impact-relevant and impact-irrelevant ETCs. Whether an ETC is deemed impact relevant depends on the scope of the study. In this study, the ETC tracking framework is applied and evaluated for historical (1991-2020) storm surge events in Denmark, which constitutes a complex case due to the diverse nature of storms and ocean dynamics across the Atlantic-North Sea-Baltic Sea trajectory (Andrée et al., 2021, 2022; Andrée et al., 2023). Hence, ETCs
100 that do not cause these events are considered impact-irrelevant here. By providing the timing and location of other types of hazards, such as inland wind damage or shipping hazards, a different set of ETC tracks would be deemed impact-relevant, thus making the framework broadly applicable for various impact-oriented ETC tracking studies. 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.

105 2 Data and region of study

The CERRA dataset used in this study is a high-resolution regional reanalysis product developed by the Copernicus Climate Change Service (C3S) to support climate applications in Europe (Ridal et al., 2024). CERRA has a spatial resolution of 5.5 km and an hourly temporal resolution, making it particularly well suited to capture mesoscale features of ETCs across complex European terrains (see Figure 1). The dataset assimilates a wide range of observational inputs and is dynamically downscaled
110 from the ERA5 global reanalysis using the HARMONIE-AROME model configuration. For storm-tracking purposes, we employ two key atmospheric parameters: MSLP and the 500 hPa relative vorticity, ζ_{500} . Other tracking algorithms typically use ζ at 850 hPa or 700 hPa, but our experience in the study region showed that the Scandinavian mountains caused artefacts at these levels. Before calculating ζ_{500} , 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
115 derivatives. MSLP is used to identify low-pressure centres typically associated with ETC cores. ζ_{500} was calculated from the re-gridded wind components u and v using a centred finite-difference scheme,

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$$

To identify and evaluate impact-relevant storm tracks, a set of significant storm surge events was compiled based on observed water level time series from Danish tide gauge stations (see locations in Figure 1). We utilise detided water levels to ensure
120 the algorithm focuses on significant surge events driven by severe storms rather than tidal influences, which vary considerably

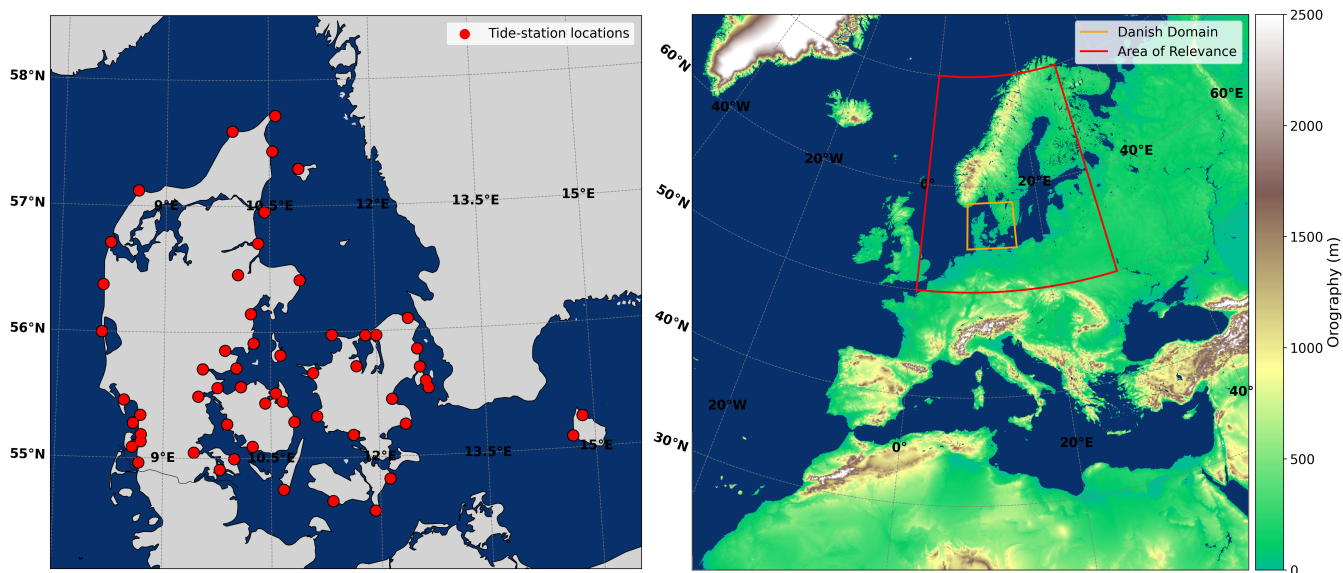


Figure 1. Map of Denmark with locations of tide gauge stations used to identify the high water level events (left) and full CERRA domain with surface topography (colour scale) and ocean mask (blue), as well as the "Area of Relevance" (AoR) and the Danish domain region in which water level impacts are sampled (right).

across the Danish coastline. Removing the tidal component enables a physically consistent and accurate identification of the surge peak timing, which is essential for initialising the storm-tracking window. While the relationship between tidal phases and wind-induced surges is critical for assessing total flooding severity, detiding isolates the meteorological driver for the purpose of algorithm development.

125 Storm surge events were selected using an empirical threshold corresponding to the 5-year return level at each station, ensuring that only significant extremes were included. To avoid double-counting of close extremes, a minimum time interval of 1.5 days between two events was imposed for both the same station and for events that occur across several stations. This selection process produced a set of dates in which coastal water levels exceeded thresholds indicative of high-impact events. We also performed a manual review of each case to exclude erroneous recordings and ensure that the elevated water levels
 130 were driven by extratropical cyclone activity rather than by other mechanisms, such as seiches or large anticyclones. As a result, six dates were excluded from the analysis due to the absence of an identifiable ETC driver. For all events, the time of impact was defined as the peak in the water-level time series, providing a physically consistent reference point for initialising the storm-tracking algorithm.

3 ETC tracking algorithm

135 The ETC detection and tracking algorithm of this paper begins with the user specifying a "time of impact", e.g., the date and hour of a local storm surge event. The algorithm then searches for and tracks a potential ETC within a user-defined time window around the time of impact. Figure 2 presents a schematic overview of the core components within the ImpactETC1.0 tracking framework. The method is designed to systematically identify and isolate extratropical cyclones responsible for local surface impacts within a specified time window. Each stage in the workflow builds on the previous one, progressively refining

140 the storm tracks from initial detection to final filtering. The figure outlines the logical flow between components, from storm centre identification to track linkage, spatial consolidation, and impact-based filtering. The sections are inherently independent in nature and thus can be seamlessly swapped for alternative approaches. The following subsections (3.1-3.4) detail the implementation and rationale behind each step, while subsection (3.5) provides an overview of the parameters involved.

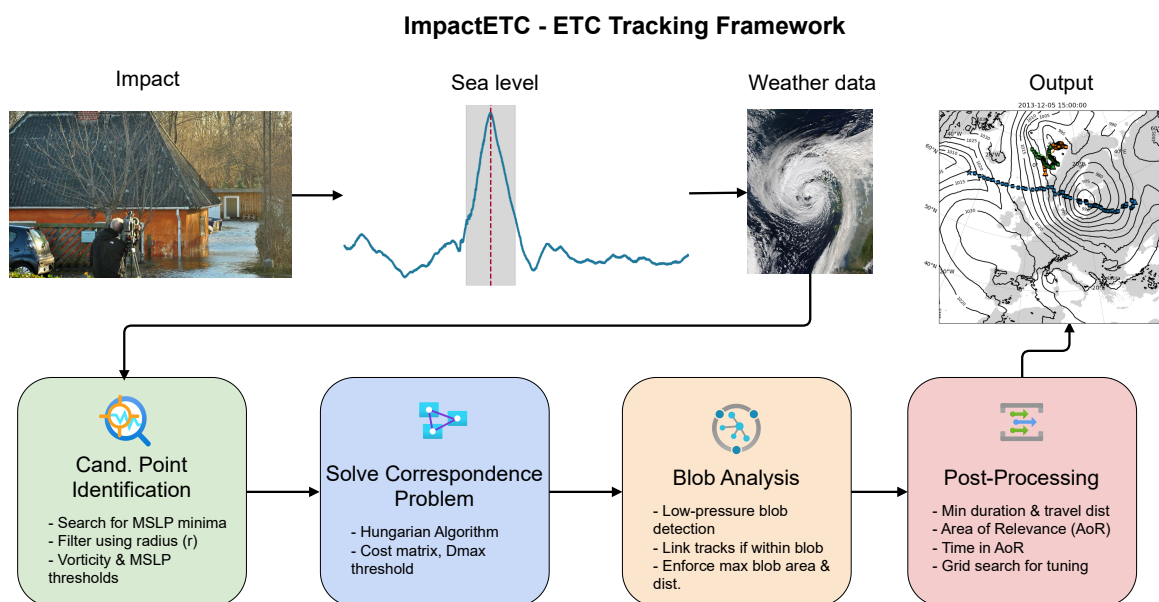


Figure 2. Overview of the ImpactETC1.0 framework for identifying ETCs linked to observed surface impacts (in order of arrows): Motivation of the study (impacts of Storm Bodil, Dec. 2013, Photo: Martin Stendel), detided water level at a coastal station (± 24 h impact window shaded), and associated ETC track. Bottom row (the four general stages of the storm tracking algorithm flow): (1) Candidate Point Identification (storm centres located by detecting MSLP minima and filtered using vorticity and spatial thresholds); (2) Solving the Correspondence Problem (storm centres linked across time using either a nearest-neighbour heuristic or the Hungarian Algorithm, with constraints on distance and cost); (3) BLOB Analysis—track discontinuities due to terrain or domain boundaries (resolved by reconnecting nearby low-pressure systems); (4) Post-Processing (tracks retained if they satisfy criteria for duration, travel distance, and overlap with a defined "area of relevance", with parameters tuned through grid search), and (algorithm output, top right) impact-relevant ETC tracks capable of explaining localised storm surge or flooding events.

3.1 Identifying candidate points

145 Since the aim is to identify the relevant ETC that caused a local impact at a specific time, we first define a time window around the time of impact with start and end timestamps t_s and t_e , respectively. For the application of the algorithm to storm surge events in this paper, a time window of ± 24 hours was chosen around the beginning and end peaks of each storm surge event, yielding a minimum of 49 time steps per event. This search window is specifically balanced to ensure sufficient track continuity for the primary driver while minimising the inclusion of secondary, impact-irrelevant tracks that could introduce
 150 ambiguity during the automated calibration of post-processing parameters. This search window is highlighted in Figure 3.

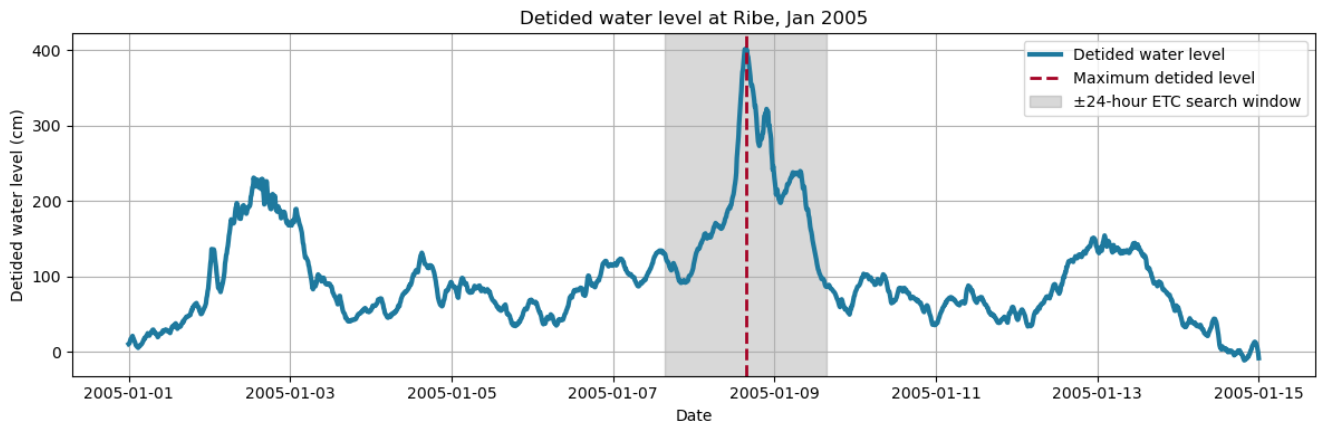


Figure 3. Time series of detided water level at Ribe, Denmark, during a storm surge event in January 2005. The dashed red line marks the time of maximum water level. The shaded region represents the ± 24 -hour window around the peak used to identify and track the associated ETC.

Within this time window, the algorithm identifies candidate locations for the centre of an ETC by searching for local MSLP minima in the entire CERRA domain (Ridal et al., 2024). Across large spatial domains with high resolution grids, there are two issues: 1) the number of grid cells is large and simple local neighbourhood searches will generate multiple local minima, and 2) there may be multiple ETCs present within the domain at any given time. To handle the high resolution of the 5.5 km grid
 155 and avoid detecting excessive local noise, we employ a spatial partitioning approach rather than traditional grid coarsening or "upsampling".

We first attempted using a kernel around each point to determine its status as a local minima, but this approach proved too computationally expensive. Instead, the goal is to upscale the original grid into coarser cells without missing any local minima due to overly large cells. The full domain is divided into non-overlapping square blocks where the diagonal length of each
 160 block is defined by the pruning radius, r . Using the Pythagorean theorem, we can derive the length of the squares from the diagonal, given by the pruning radius r as follows: $l = \sqrt{\frac{r^2}{2}}$. We proceed through the following steps (Figure 4 shows the remaining potential candidate points after each step):

1. The algorithm determines the local MSLP minima within each spatial block, specifically identifying the individual 5.5 km grid cell with the lowest value. This ensures that the final coordinates remain faithful to the original data resolution without downsampling (Figure 4 a).
165
2. Remove local MSLP minima if any neighbouring upscaled cells within the pruning radius r contain a lower MSLP minimum (Figure 4 b).
3. Remove local MSLP minima with values above "MSLP_{max}" (Figure 4 c). While background MSLP fields can vary, a simple maximum cap is used to efficiently filter out high-pressure noise and weak systems unlikely to drive significant surface impacts. This computationally efficient approach is prioritised because candidate identification is already one of the algorithm's most demanding components. For the storm surge applications in this study, MSLP_{max} is set to a conservative value of 1010 hPa.
170
4. Remove local minima that do not have a ζ_{500} intensity above $\zeta_{500\ min}$ within a radius of 500 km, an approach consistent with established vorticity-based feature tracking methods (Gramcianinov et al., 2020, Figure 4 d). While this threshold may occasionally retain features from strong frontal structures that are not directly associated with an ETC core, its implementation is prioritised for computational speed at this high-demand stage of the processing. We opted against additional wind speed thresholds, as defining a single value applicable to both land and ocean surfaces risks unintentionally filtering out relevant candidate points in complex terrain.
175

From these steps, we identify potential candidate points that represent the centre of an ETC. These will be used in the correspondence problem in the next step of the algorithm. The hyperparameters in these steps (r , maximum MSLP value at ETC centre, and minimum regional ζ_{500} intensity) can all be fine-tuned for a given application of the algorithm. For the storm surge examples in this paper, we use the values $r = 350$ km, $\zeta_{500\ min} = 1.5 \cdot 10^{-4} \text{ s}^{-1}$, and $\text{MSLP}_{max} = 1010$ hPa. The choice of r determines the size of the atmospheric structures that are preserved. It has to be large enough to filter away grid-scale noise and minor vortices if these are not of interest, but small enough to separate relevant features. If the user targets synoptic-scale systems, the pruning radius can reasonably be set to 500-1000 km, while smaller values are necessary to identify, e.g., secondary lows. The value of r can have a large impact on the computational efficiency of the algorithm, which will be investigated further in Section 5.2. In general, ~~lax pruning constraints lead to~~ smaller pruning radii yield more candidate points, which in turn affects the algorithm's computational efficiency.
180
185

In a few rare instances, we have observed that two nearby candidate points have identical MSLP values in the CERRA data, which means that the aforementioned step 2 cannot select a unique point within the pruning radius. In this case, we have chosen to keep both points as possible candidates for the correspondence problem in the following section, rather than making an arbitrary selection between them. In Figure 5, we show an example of MSLP minima points and 1 hPa isobar levels for the 1997-10-11 event, visualising the data in which the algorithm operates.
190

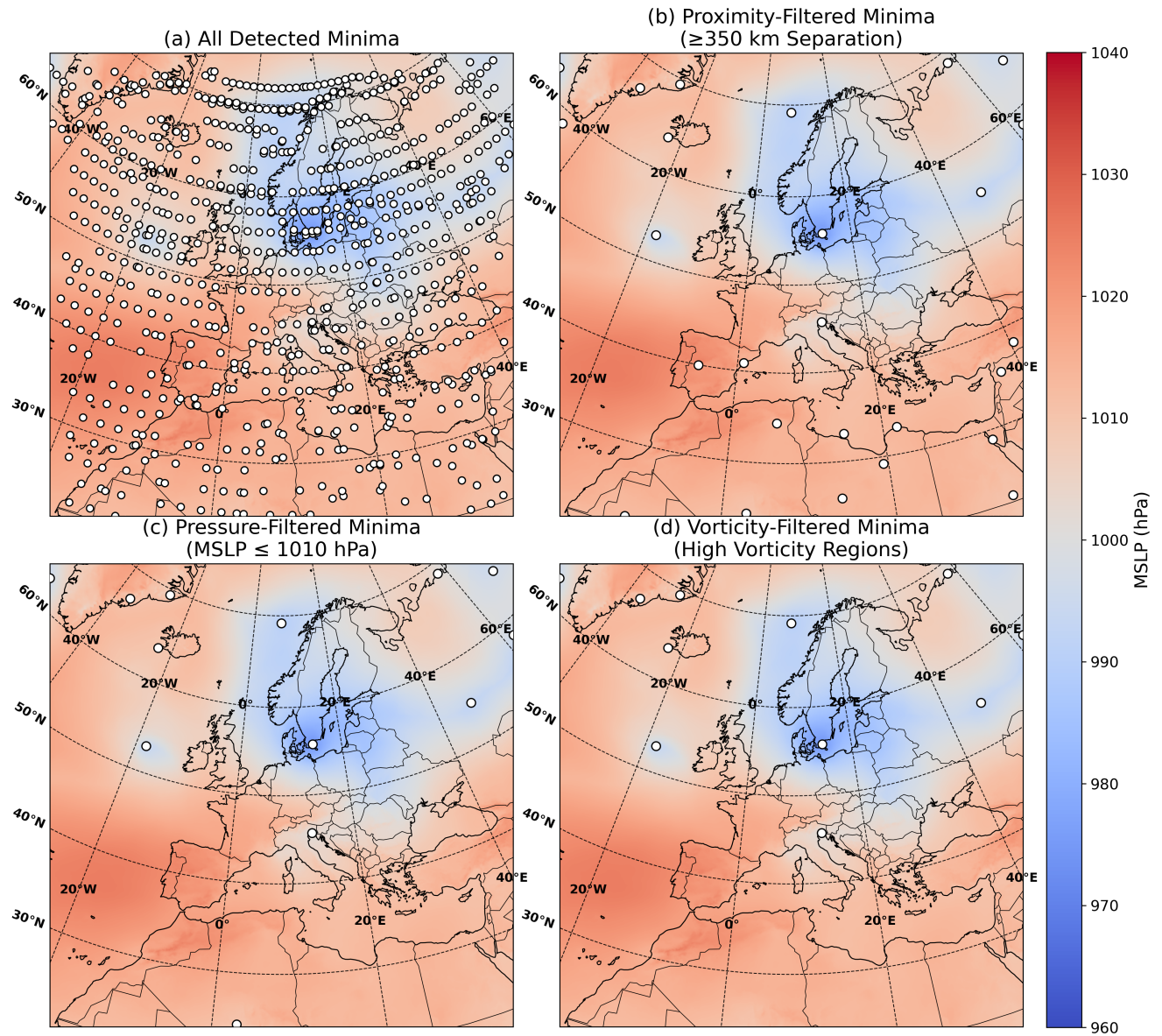


Figure 4. Visualisation of the progressive filtering of candidate cyclone centres using MSLP and 500 hPa vorticity criteria. From left to right, top to bottom: all local MSLP minima identified within upscaled grid cells. After applying radius-based culling, retaining only the lowest minimum within a defined neighbourhood. After removing candidates with MSLP values above the threshold $MSLP_{max}$. After further filtering out candidates lacking sufficient mid-tropospheric vorticity within a 500 km radius. The background shading depicts the MSLP field for context. The event is from 1997-10-11.

(d) Vorticity-Filtered Minima
(High Vorticity Regions) - Wind Speed & Direction with MSLP

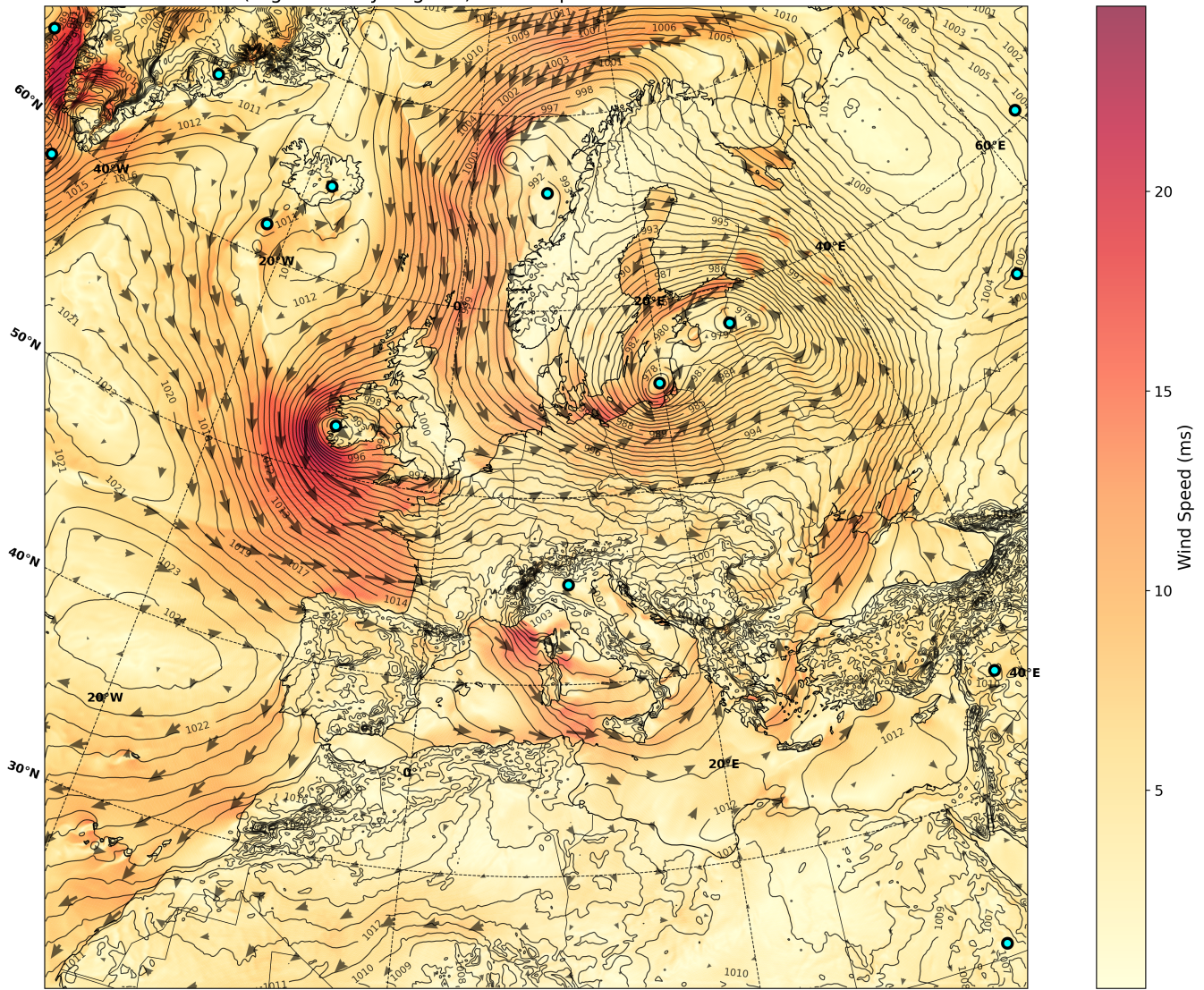


Figure 5. Joint plot of isobars (1 hPa scale), wind direction (arrows) and wind speed (color scale) for the 1997-10-11 event, to exemplify distinction of storm track centers.

3.2 Solving the correspondence problem

195 The candidate points found in section 3.1 for each individual time step must now be connected through adjacent time steps to determine the storm tracks - a task known as the correspondence problem in the storm tracking literature (Walker et al., 2020). The general objective is to minimise the overall distance between the pairs of connected candidate points at time t_i and t_{i+1} . An added complication is that the number of candidate points in each time step is not guaranteed to be identical. If we define n and m as the number of candidate points at time t_i and t_{i+1} , respectively, then the problem can be either balanced if
 200 $n = m$ or unbalanced if $n \neq m$. A theoretical solution to this problem can be found through global optimisation. However, due to concerns about the computational requirements of global optimisation, it is common in current storm tracking algorithms to use a heuristic approach. This is often based on the nearest-neighbour approach, where one iterates through the candidate points at t_i and sequentially connects them to the candidate point closest in space at t_{i+1} . Usually, a threshold for the maximum possible travel distance of a storm centre between two time steps, D_{max} , is part of the heuristic algorithms. Heuristic-based
 205 approaches are not guaranteed to provide the correct solution, but they are computationally very fast.

In our tracking framework, we have implemented a global optimisation solution that avoids some of the pitfalls of nearest-neighbour heuristics. This type of correspondence problem is formally known as the "assignment problem" in optimisation theory, and it can be efficiently solved using the Hungarian Algorithm. The algorithm is well-studied in other fields, such as logistics (Seda, 2022) and scheduling (Zhang et al., 2024), and therefore we will only provide a general description here. The
 210 Hungarian algorithm begins with setting up a cost matrix, which here will be an $n \times m$ distance matrix, which represents the distance between all n candidate points at t_i and all m points at t_{i+1} . Equation 1 exemplifies how such a distance matrix is set up, with the leftmost matrix in Equation 1 being $n \times m$. The algorithm requires that the distance matrix be square, and since this is not guaranteed, the next step is to pad the matrix with phantom candidate points with a maximum cost (middle matrix in Equation 1). We then apply a maximum allowable distance for a connection to be "realistic", D_{max} . This parameter can
 215 be physically interpreted as a maximum allowable translation speed of the ETC centre, where reasonable estimates may be in the order of 100-150 km/h. However, the parameter can also be understood more conceptually as a "search radius" applied to the tracking variable (here MSLP). For the storm surge application presented in the Nordic region, we chose a value of 300 km, since complex terrain can cause the identified ETC centre to jump longer distances than what would be considered as physically reasonable. This maximum distance is implemented by censoring all matrix elements with values above D_{max} (see
 220 rightmost matrix in Equation 1).

$$\begin{bmatrix} 2588 & 105 & 83 \\ 6 & 2640 & 2510 \\ 2395 & 254 & 112 \\ 319 & 2694 & 2577 \end{bmatrix} \rightarrow \begin{bmatrix} 2588 & 105 & 83 & D_{max} \\ 6 & 2640 & 2510 & D_{max} \\ 2395 & 254 & 112 & D_{max} \\ 319 & 2694 & 2577 & D_{max} \end{bmatrix} \rightarrow \begin{bmatrix} D_{max} & 105 & 83 & D_{max} \\ 6 & D_{max} & D_{max} & D_{max} \\ D_{max} & 254 & 112 & D_{max} \\ D_{max} & D_{max} & D_{max} & D_{max} \end{bmatrix} \quad (1)$$

Once the distance matrix is set up, made square, and censored, the algorithm finds the optimal solution by minimising the sum of distances across all connections. This is done through row/column reduction, covering zeros with lines, and matrix adjustments in the following steps:

- 225 1. Identify the smallest element in the matrix and subtract it from all elements, thus obtaining a matrix with at least one zero.
2. For horizontal and vertical lines through the rows and columns of the matrix, identify the minimum number of lines that would be required to cover all zeros.
3. If the number of lines is exactly the size of the matrix, then an optimal solution has been found. If it is smaller than
230 the matrix size, an adjustment to the matrix is needed. The adjustment is to identify the smallest element h that is not covered by any line. Add h to all elements covered twice at the intersections of two lines and subtract h from all uncovered elements.
4. Repeat steps two and three until the number of lines required to cover the zeros is equal to the size of the matrix.
5. Select a set of connections that all have zeros. This is the optimal solution.

235 The connections found by the Hungarian Algorithm represent the solution that minimises the total sum of the connections between t_i and t_{i+1} . Computationally, in terms of big-O notation, it scales at $O(n^3)$. If multiple optimal solutions with the same objective value exist, the algorithm will return only one of them. In the case of storm tracking, with costs defined as the exact geographical distance between points, we deem it unlikely to occur.

The framework also contains a second option for solving the correspondence problem with a standard nearest-neighbour
240 (NN) heuristic, which we will benchmark HA against. For the NN approach, we employ a greedy algorithm, which iteratively selects the valid minimum distance connection until all pairings are exhausted or exceed a predefined distance threshold D_{\max} (see Algorithm 1):

Algorithm 1 Nearest neighbour heuristic for the correspondence problem

Require: Distance matrix $\mathbf{D} \in \mathbb{R}^{n \times m}$, maximum movement distance D_{\max}

Ensure: Assignment list \mathcal{A}

- 1: $\mathcal{A} \leftarrow \emptyset$
 - 2: **while** minimum value in \mathbf{D} is less than D_{\max} **do**
 - 3: Find $(i, j) = \arg \min \mathbf{D}$
 - 4: Add (i, j) to \mathcal{A}
 - 5: Remove row i and column j from \mathbf{D}
 - 6: **end while**
 - 7: **return** \mathcal{A}
-

The implemented HA and NN algorithms optimise 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 it can lead to a track breaking in two ways. First, if one of the timestamps 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 occur, the user would have to specify a D_{max} value larger than the pruning radius, r , which would create the possibility of multiple candidate points within the D_{max} distance. To mitigate this second possibility, we introduce an additional algorithmic step to handle cases where tracks break because the required connection exceeds D_{max} .

3.3 BLOB analysis for resolving storm track fragmentation over complex terrain

In this study, MSLP is used to identify candidate points for the centres of ETCs. We chose MSLP over other possible tracking variables because of its close relationship to surface winds that drive storm surges. However, MSLP is a diagnostic variable, computed from the modelled surface pressure by estimating the hypothetical sea level pressure under the assumption of hydrostatic balance (Pauley, 1998). Over complex terrain, this assumption often breaks down, e.g., over the Scandinavian mountains. As a result, artefacts and inaccuracies can arise in the MSLP field, potentially causing tracks to break into several fragments. A recurring example of this arises when an ETC crosses a mountain range: Candidate centres often appear to "stall" on the upstream side of the mountains, while a new candidate centre simultaneously appears downstream in the same time step. In this case, there are multiple alternative candidate points in the same time step for the same ETC system. This causes the track to break, resulting in a single ETC being misidentified as two separate systems. A tracking variable at higher atmospheric levels might be less prone to such issues, but it would also be less closely related to the surface impacts. We therefore introduce a novel track reconciliation step to allow for surface level tracking with MSLP while mitigating this type of track fragmentation. The reconciliation step is based on a BLOB analysis technique that performs synoptic-scale spatial continuity checks using the MSLP field.

We observed that when breaks occur, the upstream and downstream candidate points typically remain part of the same synoptic-scale low-pressure system, i.e. in close spatial proximity and with similar MSLP values. BLOB analysis leverages this by identifying contiguous low-pressure "regions" within a given MSLP range.

The method works as follows: for a given candidate point, we define a pressure range (e.g. ± 5 hPa around its MSLP value). The entire MSLP field is then binarised, with grid points within the range set to 1 (inside a BLOB) and those outside set to 0. The result is a BLOB that represents spatially coherent low-pressure areas. For each time step, we check whether other candidate points fall within the same BLOB as the one containing the current candidate point (see Figure 6). 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.

In certain edge cases, the BLOB's can become very large. To avoid false positives, we impose a maximum allowable BLOB bounding box of 3000 km and a threshold of 600 km for the maximum distance between two candidate points to be considered connected. When choosing the value of this second maximum BLOB candidate point distance parameter for a given use case, it is important to set it high enough to connect points within the same synoptic-scale system, e.g., in the range 500-1000 km. Its value should be larger than D_{max} to work as intended, but values much larger than 500-1000 km could risk merging distinct synoptic systems.

In simple cases, the discontinuity occurs at the end of one track fragment and the start of the next, which can then be merged. In more complex cases, two tracks may overlap temporally: the upstream track continues for a few time steps while the downstream track has already begun. The task then becomes choosing whether and when the jump from the first track to the start of the second track should occur. In these instances, we identify the first time step in which the two candidate points fall within the same BLOB. From that point onwards, we monitor both the continuations of the tracks and select the one that persists for the longest number of time steps as the final coherent track (see Figure 7).

In the results section, we demonstrate the value of this track reconciliation step by showing how frequently it is used for the storm surge events in this study, how it affects the length of the final tracks, and how it prevents broken tracks from being filtered out in the post-processing steps described in the next section.

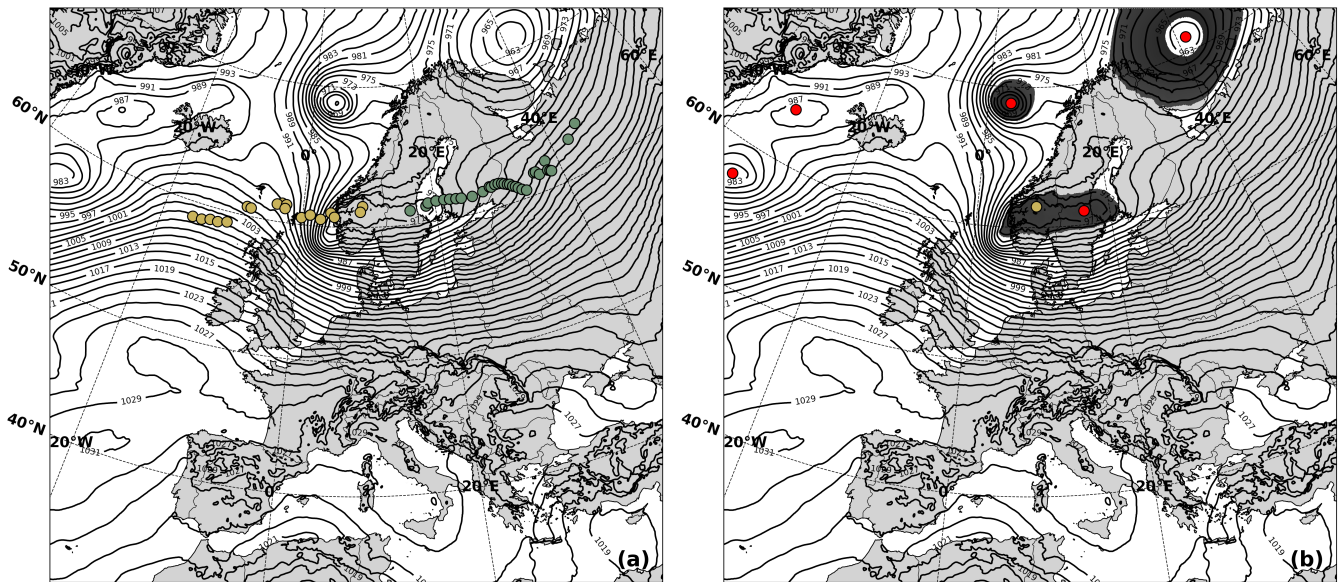


Figure 6. Illustration of the BLOB analysis step used for track reconciliation. Panel (a) shows an ETC crossing the Scandinavian mountains, where two valid MSLP minima appears in the same time step, which causes the track to break in two (yellow vs green). In (b), the BLOB analysis identifies spatially coherent low-pressure regions (BLOBs) and identifies candidate points within the same BLOB at the same time. Once identified, the method connects candidate centres upstream and downstream of the mountains to maintain a continuous storm track.

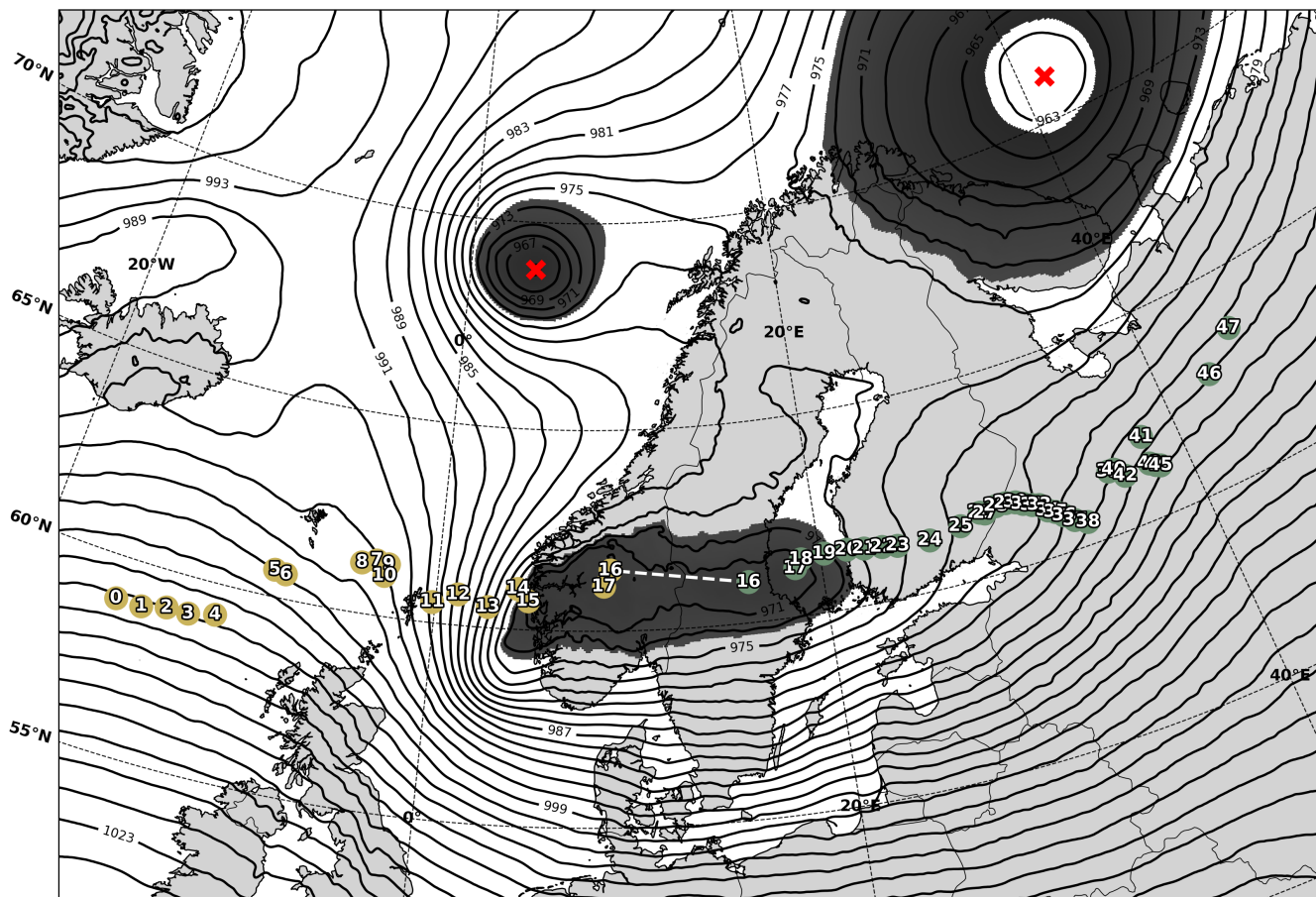


Figure 7. Example of track reconciliation between two fragmented tracks for the same case event as in Figure 6. Points indicate MSLP minima and their assigned numbers show the time step during the event. The ETC moved from left to right in the image. At time step 16, the ETC reaches the Scandinavian mountains and the track breaks, leaving two tracks that both have valid ETC centre at time steps 16 and 17. The method creates a BLOB with a ± 5 hPa range around the MSLP value at both points at step 16 (dark shaded area). Both points are within the threshold distance of the BLOB, and the algorithm jumps from the yellow to green track at point 16, because the white track continues on for a longer duration (green continues to time step 50, yellow only continues to time step 17). The red crosses denote other MSLP minima in the field, but since they are not within the same BLOB, they are ruled out as potential connections.

3.4 Post-processing of tracks

295 Within the impact time window, multiple ETC tracks are often present throughout the data domain. These tracks vary in length, duration, and spatial trajectory. Therefore, a crucial final step in any storm tracking algorithm is the filtering of tracks that are either insignificant or unrelated to the local impact.

In this study, we implement three options for post-processing tracks to identify the storm(s) that caused the local impact.

1. Impact-oriented calibration of standard post-processing parameters
- 300 2. Nearest storm selection
3. Gradient tracing

3.4.1 Impact-oriented calibration of standard post-processing parameters

The first option allows the user to apply four criteria that an ETC track must satisfy to be considered relevant. These are standard post-processing criteria that are used in many different tracking algorithms in the literature:

- 305 – Minimum travel distance: The ETC must travel a minimum distance, measured as the total length of the track in kilometres.
- Minimum duration: The ETC must persist for a minimum duration, defined as the time between the first and last detected points along its track.
- Area of Relevance (AoR): The storm track must enter a predefined spatial area around the impact location.
- 310 – Minimum time in AoR: The ETC must spend a minimum amount of time within the AoR.

The first two criteria primarily eliminate short-lived or spatially limited tracks across the domain, while the latter two filter out tracks unlikely to have contributed to the local impact. While users can implement a circular AoR centred on a specific impact location, this framework utilises a rectangular AoR for programmatic efficiency when slicing 2D atmospheric fields. In cases like Danish storm surges, a non-symmetrical AoR is often physically preferable; for example, relevant ETCs may
315 be located farther north or east in the Baltic Sea than to the west or south. Consequently, the AoR serves as a flexible, user-defined distance criterion, requiring the ETC to remain within a relevant proximity for a specified time window. Because the user provides the initial "time of impact" based on observed local water levels or winds, additional identification steps for the hazard itself are not required within the tracking loop.

In order to evaluate the sensitivity of the post-processing parameters and define a set of optimal parameters for our case
320 study, a full grid search exploration of the parameter space is carried out, and the post-processing results are recorded. The parameter ranges applied for the grid search are as follows:

- Minimum travel distance: [0 km, 100 km, ..., 1400 km, 1500 km]
- Minimum duration: [0 h, 1 h, ..., 39 h, 40 h]
- Δ AoR: [-5° , -4° , ..., 9° , 10°]
- 325 – Minimum time in AoR: [0 h, 1 h, ..., 29 h, 30 h]

We note that the AoR is centred on ~~60° lat, 15°N~~60°N, 15°E and initially spans from ~~50—70° lat~~50—70°N and 0—30°E. Here, the AoR change is in degrees with negative values decreasing the AoR and vice versa. The initial AoR encompasses the Baltic region and the North Sea (see the area highlighted in Figure 1), since an ETC must enter this region to affect the water levels on the Danish coasts. In this study, we chose to define the AoR and the incremental changes in geographic degrees. 330 Alternatively, this could have been defined in kilometres. This choice is possible for our AoR, as the effect from the varying length of a $\Delta 1^\circ$ is small (1-2 km in the zonal direction) due to the relatively small size and location of our AoR. Hence, we took a practical approach of using degrees, as this is how the CERRA data is given. However, for larger AoR and/or AoR located close to the poles, this is not a feasible solution and will be discussed further in section 5.3.

To explore the parameter sensitivity, we use two metrics. The first metric assesses whether the post-processing filters yield 335 the ~~true correct~~ number of ETC tracks per ~~event in our case data~~impact event. To do this, we have undertaken the labour-intensive task of manually labelling the ~~correct number true count~~ of ETC tracks within the AoR ~~for all impact events and the impact time window for all events in our case data~~. We then define a simple accuracy score for the evaluation, which we call the "True Storm Count Accuracy" (TASCA):

$$\text{TASCA} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(y_i = \hat{y}_i) \quad (2)$$

340 where y is the true number of ETCs, \hat{y} the predicted number given a set of post-processing parameter values, $\mathbf{1}(y_i = \hat{y}_i)$ is an indicator function equal to 1 if correct and 0 if incorrect, and n is the total number of events. Please note that this metric only considers the accuracy of the counts of ETC tracks per impact event, and should not be confused with an accuracy assessment of the individual ETC track paths or a causal hazard attribution. Users of ImpactETC1.0 can do the same for their case implementations if the number of events to investigate is relatively small and manual labelling is deemed possible.

345 We recognise that developing a labelled dataset may not be desirable or possible for all applications of the algorithm, especially when evaluating many impact events. To still calibrate post-processing parameters rather than simply choosing values on an ad hoc basis, we tested a second metric that requires no manual work. If this second metric can calibrate parameter values to match those of the True Storm Count Accuracy, it can be used for large-scale automated analyses. We define this second metric as the "Single Storm Score" (S):

$$350 \quad S = \frac{N_1 - N_0 - N_{3+}}{N_0 + N_1 + N_2 + N_{3+}} \quad (3)$$

where N_i represents the number of events observed with i associated ETC tracks. This score rewards events with exactly one detected ETC (interpreted as a successful and unambiguous match), penalises missed detections (N_0) and excessive ambiguity (N_{3+}). It treats cases with two ETCs as neutral, since we have observed this in our case example. The result is a simple, interpretable score ranging between -1 and 1, where higher values indicate better post-processing performance.

355 3.4.2 Nearest storm selection

The framework presents an alternative approach to calibrating filtering parameters using the Single Storm Score (S), which relies on the simple logic that the impact-relevant ETC was likely the one geographically closest to the impact location at the time of impact. This logic ensures that only a single ETC will be identified for each impact location. This simplifies post-processing by rendering the AoR redundant, thereby reducing the number of post-processing criteria from 4 to 2. The user is
360 free to set the relevant parameters, and for this case study, we set “minimum duration” to 8h and “minimum travel distance” to 300 km. If there are multiple eligible ETC tracks within the model domain, at the time of impact, the method calculates the distance from the impact location to each eligible track and selects the nearest.

3.4.3 Gradient tracing

A limitation of the "Nearest storm selection" approach is that it may not always be the track closest to the impact location that
365 causes the event. Or it might be an interplay of multiple ETCs that cause it. As a third option, we therefore developed a "Gradient tracing" method based on the premise that the spatial pattern of the MSLP field contains important directional information about where the cause of the impact may originate. It applies the logic that the ETC that caused the impact probably holds the impact location within its "low-pressure basin". Similar to the Nearest storm selection method, the “minimum duration” and “minimum travel distance” criteria are used to filter out micro-tracks, and in this case, are set to $8h$ and $300km$, respectively.
370 The gradient tracing method works by analysing the instantaneous MSLP field at the exact time of impact. It begins in the grid box where the impact location is and evaluates the 8 surrounding grid boxes. It then iteratively moves in the direction of steepest descent until it reaches a local minimum in the MSLP field. Once it has reached a local minimum, it evaluates whether this local minimum is a valid end point, defined as a point on one of the identified ETC tracks at the exact time of impact. If it is not a valid end point, it multiplies the MSLP value at the local minimum by a factor of 1.1, effectively raising the point above
375 its neighbours, ensuring that a step into a neighbouring location is possible, and then moves in the direction of steepest descent again. It continues to iterate like this until it reaches a valid end point. The pseudo-code for the Gradient tracing method is defined in Algorithm 2. An example showcasing the method is highlighted in Figure 8(a), where the blue circle represents the starting location and the red squares represent a selection of the locations tracing has passed through towards termination at an ETC track, which are represented by the yellow stars.

380 However, for some events, we observed that the initial impact location was located on a “ridge” in the MSLP field, where even small changes in the MSLP field determining which side of the ridge was stepped downward. We therefore introduced an additional option to include a measure of uncertainty to the identification of the impact-relevant storms. This is done by introducing spatial uncertainty in the impact location, from where the gradient tracing starts. Instead of only tracing from the exact impact location, we randomly sample 100 starting locations within a 250 km radius around the impact location, and
385 perform gradient tracing for all of these. Doing so returns a percentage distribution of traces that terminate in each eligible ETC track, indicating how certain a specific ETC is to be the cause of the impact. A benefit of this is that this result can serve as an indicator of potential compound events when multiple ETCs interact to cause an event. Figure 8(b) shows an example

where the exact impact location is represented by the blue circle, and the small squares surrounding the blue circle represent the randomly sampled starting points for the tracing exercise, with their colours representing the ETC track location where they terminated. As the legend shows, 88 % of traces end at one ETC, while 12 % end at another.

Algorithm 2 Gradient tracing for filtering of tracks

Require: MSLP matrix $\mathbf{M} \in \mathbb{R}^{n \times m}$, maximum step iterations $Iter_{\max}$, starting location radius S_{radius} , number of starting locations Z_{\max} , storm center objects SC , each with attributes {loc, storm number}

Ensure: Storm centers reached list \mathcal{A}

```

1:  $\mathcal{A} \leftarrow \emptyset$ 
2: Sample  $Z_{\max}$  starting locations  $\mathcal{Z}$  uniformly within radius  $S_{\text{radius}}$ 
3: for each  $Z_{\text{start}} \in \mathcal{Z}$  do
4:    $Z_{\text{cur}} \leftarrow Z_{\text{start}}, \quad \text{step} \leftarrow 0$ 
5:   while  $\text{step} \leq Iter_{\max}$  do
6:      $N_{\min} \leftarrow \arg \min_{n \in \mathcal{N}(Z_{\text{cur}})} \mathbf{M}[n]$ 
7:     if  $\mathbf{M}[N_{\min}] < \mathbf{M}[Z_{\text{cur}}]$  then
8:        $Z_{\text{cur}} \leftarrow N_{\min}$ 
9:     else if  $Z_{\text{cur}} \in SC$  then
10:       $\mathcal{A} \leftarrow \mathcal{A} \cup \{s.\text{storm number}\}$  {where  $s \in SC$  s.t.  $s.\text{loc} = Z_{\text{cur}}$ }
11:      break
12:     else
13:        $\mathbf{M}[Z_{\text{cur}}] \leftarrow 1.1 \cdot \mathbf{M}[Z_{\text{cur}}]$ 
14:     end if
15:      $\text{step} \leftarrow \text{step} + 1$ 
16:   end while
17: end for
18: return  $\mathcal{A}$ 

```

3.5 Overview of parameters

The full tracking algorithm contains many parameters and hyperparameters, some of which can be reasonably selected from knowledge of the physical system, while others need to be tuned for a given case area. In the literature on tracking algorithms, we often find a lack of (hyper-)parameter documentation, so here we present Table 1, which provides a full and transparent overview of the parameters in this algorithm.

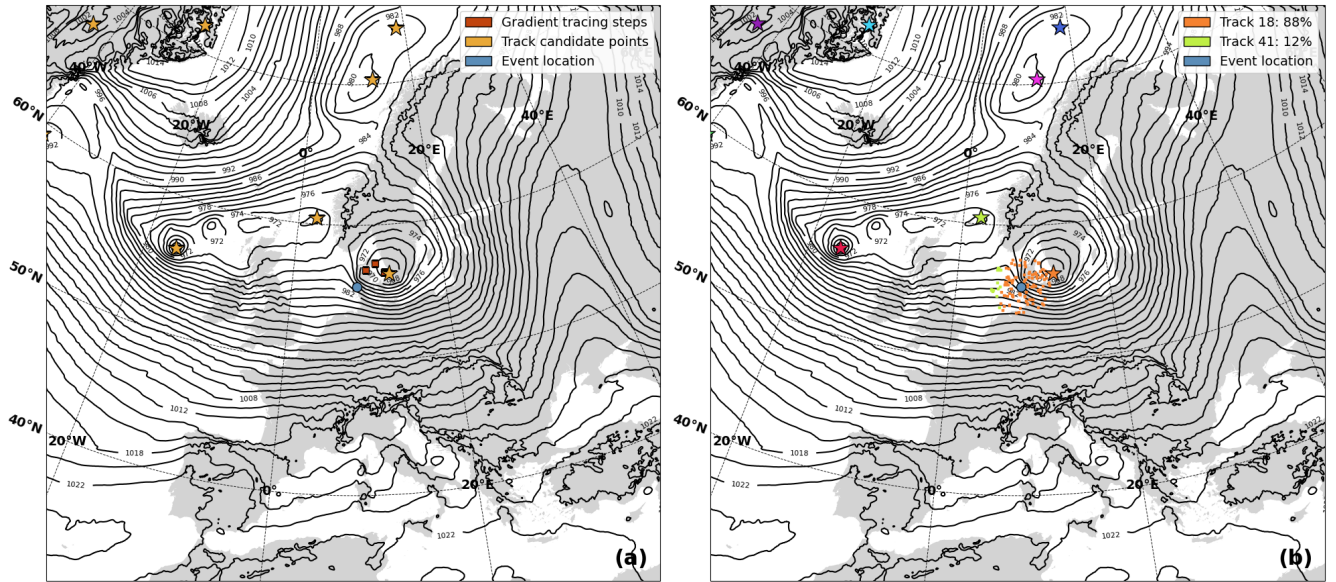


Figure 8. Visualisation of gradient tracing algorithm for the 2002-02-26 event where a storm surge happened at the city of Esbjerg. Contour lines show the MSLP field at the time of impact, and the star symbols represents identified ETC centres points. (a) The left panel shows the steps (red dots) that the gradient tracing method makes as it iteratively moves from the location of impact (blue dot) to one of the ETC centres. (b) The right panel shows how 100 starting points are sampled within a 250 km radius of the impact location. Each ETC centre (star) now has a distinct colour, and the colour of the starting points indicates which ETC centre, each starting point is assigned to with the gradient tracing method. The percentage numbers in the legend shows the proportion of starting points that end up at each ETC, here 88 % vs. 12 % for two nearby ETCs.

4 Results

4.1 Correspondence problem: global optimisation vs heuristics

Here we present results on the trade-off between solution quality and computational efficiency of the HA algorithm presented in Section 3.2 over a range of complexities of correspondence problems.

400 To explore the influence of problem complexity on NN and HA performance, we vary the pruning radius r used in the candidate point identification step. Smaller r -values yield denser point clouds and larger correspondence problems. We tested six radii: 100, 175, 250, 350, 500, and 700 km, evaluating the run time and three additional metrics that explore solution differences between HA and the greedy NN solution:

1. Number of correspondence problem solutions that yield different connections.
- 405 2. Number of additional connections in the global HA solution relative to the greedy NN solution.
3. Mean difference in the objective value of the correspondence problem, when the two solutions differ (in %).

During this analysis of the performance of HA and NN for the 37 impact events, a total of 1966 correspondence problems were solved, as summarised in (Table 2). Dates of the impact events are included in Table A1 in the Appendix.

As the radius decreases, the number of candidate points and assignment options increases, which leads to greater problem complexity. Below a radius of 250 km, there is a sharp increase in different matches between HA and NN as well as extra HA connections. This happens because the pruning radius is significantly lower than the D_{max} value of 300 km, leading to multiple candidate points becoming eligible for connection. For larger radii and smaller problem sizes, the algorithms have very few legal moves, leading to similar solutions. At a 100 km radius, 13.9 % of correspondence problems yield different solutions using the NN method, and 3.97 % of the problems result in fewer valid storm track connections compared to HA.

Although the HA is significantly slower at small radii (e.g. 4.88 s at 100 km), it consistently achieves more connections between candidate points, which can be important for avoiding premature track termination. In applications where the domain is large and time steps are longer than the one-hour steps in CERRA (e.g. global reanalyses), similar complexity issues could potentially arise even with larger radii, making global optimisation even more relevant.

In our repository (Agertoft et al., 2026), we show 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.

The D_{max} parameter is important for how the correspondence problem is solved with both HA and NN, as it controls the maximum connection distance between two candidate points in time. We have carried out a minor sensitivity analysis for D_{max} in order to assess its impact on the tracks that are formed. This can be found in the appendix in Figure A1. As expected, we observed that small D_{max} -values led to an increased number of tracks and generally shorter track durations, as fewer connections could be made across time steps. The sensitivity analysis showed that there was a smooth trend in the resulting number of tracks and length of individual tracks as D_{max} was varied. There was thus no obvious cut-off value to choose. We expect this to generally be the case for other areas and reanalysis time step lengths. However, to give some practical guidance for how to tune this parameter, we recommend that a good lower estimate for it would be a physically realistic value of how far the object being tracked can potentially travel in a single time step. The user can then increase the D_{max} value to pragmatically account for other effects, such as noise in the tracking field, terrain-driven discontinuities, etc. This is inherently a tuning problem that requires some trial and error testing.

4.2 Effect of track reconciliation

To evaluate the effect of the BLOB analysis technique for track reconciliation, we compare the full tracking system with and without the reconciliation step (see Table 3). This comparison does not assess the physical accuracy-of-skill-of-reproducing the reconciled tracks as such. Instead, it quantifies how often reconciliation is applied and how it influences properties and survivability in the final track dataset after post-processing filters. The evaluation metrics are as follows:

1. The number of impact events in which at least one storm track survived post-processing out of the 37 total events.

2. The number of events with at least one surviving track that have used reconciliation out of the total number of events
440 with at least one surviving track.
3. The number of tracks where reconciliation was used out of all surviving tracks in all events.
4. The number of reconciliations across all the surviving tracks versus the number of surviving tracks where reconciliation was used.
5. The average increase in track duration for tracks where reconciliation was used.
- 445 6. The average increase in track length for tracks where reconciliation was used.

Using the pruning radius of 350 km as an example, we notice that out of 37 total events, we have 37 events where at least one track survived the post-processing when utilising gradient tracing (see Table 3). The BLOB utilisation played a critical role in 19 events, affecting 19 surviving tracks. Across those 19 post-processed tracks, 28 reconciliations have occurred, meaning multiple surviving tracks have undergone more than one reconciliation. In summary, the use of BLOB analysis has ensured
450 that a significantly higher number of tracks have survived post-processing. All 19 events in which BLOB analysis has been utilised have been manually investigated on an individual basis, and it has been found that the vast majority of the tracks in these events would otherwise not have represented reality satisfactorily.

Figure 9 shows the geographical locations where BLOB reconstructions have occurred, as well as the distribution of the distances between candidate points connected by BLOB reconstructions. How active the BLOB step is depends heavily on
455 the user-specified value of the pruning radius parameter. We therefore present BLOB reconstructions for the three different pruning radius values of 700 km (a, b), 350 km (c, d), and 100 km (e, f). In all cases, *Max BLOB candidate point distance* is set to a large value of 2500 km to analyse how connections are formed. The results generally show that smaller pruning radii lead to more blob connections because they leave many candidate points at each time step. The histograms of all three experiments show that there are a few reconstruction distances in the range of 1000-1600 km, which intuitively seems high.
460 The vast majority of distances lie between the applied pruning radius (vertical, dashed line) and around 800 km, which is more reasonable given its intent to fix jumps across mountainous terrain. For the Nordic countries, we recommend using values of 600-800 km for this parameter, but users will need to re-evaluate this for other areas where these jumps may be smaller or larger.

Figure 9 also shows cases where reconstructed connections happen over very short distances. This shows up as reconstruction
465 distances below the pruning radius value in the histograms (see Figure 9(b, d, f)). These are formed during time steps t when multiple candidate points have the exact same MSLP value within each other's pruning radius. As explained earlier, we keep both points as candidates in such cases, but 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. The BLOB reconstruction handles this issue and connects the two broken
470 track fragments.

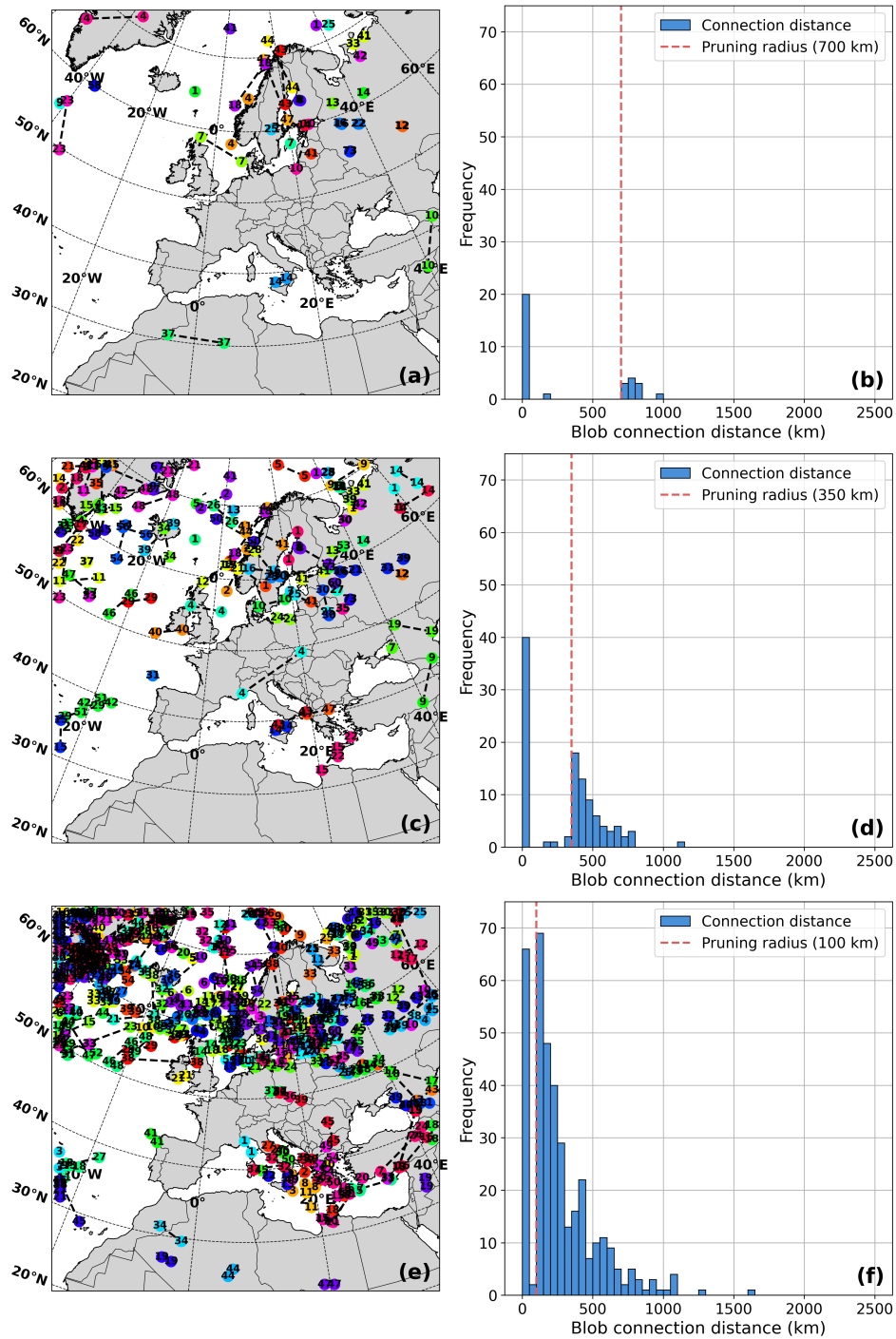


Figure 9. Overview of the BLOB track reconciliation connections for three pruning radius values of 700 km (a,b), 350 km (c,d), and 100 km (e,f). The left column of plots shows the geographic location of the reconstructed connections (a,c,e), where similar numbers at the point indicate the start and end of a connection. The right column shows a histogram of the distances of these connections (b,d,f) with a dashed vertical line that represents the pruning radius value.

4.3 Sensitivity and optimisation of post-processing parameters

Figure 10 shows the results of the parameter sensitivity analysis with both metrics. Each subplot in the figure shows the sensitivity of ETC detection to one of the four post-processing parameters. Given the discretisation of the chosen grid search for each parameter, there are a total of 325,376 possible combinations of the four parameters. The plots show the median, the 5th and 95th percentiles of the score values across all combinations. The results show that ETC detection is most sensitive to the size of the AoR (Figure 10a), time in AoR (Figure 10b), and the minimum travel distance (Figure 10c). Notably, there is the special case where minimum time in AoR equals 0, meaning that an ETC does not need to enter the AoR for detection, which leads to a large drop in accuracy-performance (Figure 10b). The large distance between the 5th and 95th percentiles, as indicated by the shaded areas, shows that the same parameter value can lead to very different post-processing performance depending on the values of the other post-processing parameters. Thus, there is a strong dependence between parameters; e.g., the size of the AoR and the minimum time in the AoR are closely related.

Figure 11 shows the marginal and joint distributions of the True Storm Count Accuracy and Single Storm Score after the values of both scores have been min-max normalised to fit the same 0-1 range. The marginal distributions follow each other closely with small deviations in some ranges. The joint distribution confirms that the Single Storm Score yields normalised values similar to the True Storm Count Accuracy. This, combined with the general agreement of the scores in Figure 10, shows that the Single Storm Score is good at approximating the True Storm Count Accuracy and thus suitable for impact-focused calibration when manual labelling is not possible.

For the presented case of ETCs that produce storm-surge events, we select the parameter set that yields the best Single Storm Score. Given the grid search discretisation, there are 325,376 possible combinations of the four post-processing parameters. In the case presented, 20 unique parameter sets yielded the same score. We selected the parameter set that was closest to the median of each parameter (see values in Table 4), thus avoiding extremes or accidentally fortunate local minima. We note that this method for selecting parameter combinations is only suitable if the possible combinations are centred in nature. If they had, e.g. been clustered in two different areas of the parameter space, the user would have to devise another way to select the final set. It is worth noting that even though all 325,376 parameter combinations were simulated in a brute-force manner, this still only took $\bar{3}$ minutes to complete on a standard laptop. This highlights that calibrating post-processing parameters is a simple, fast step that should be performed.

Table 4. Optimal post-processing parameter values identified for the storm surge case in this study.

| Minimum travel distance (km) | Minimum duration (h) | Minimum time in AoR (h) | Δ AoR (degrees) |
|------------------------------|----------------------|-------------------------|------------------------|
| 1000 | 26 | 16 | 4 |

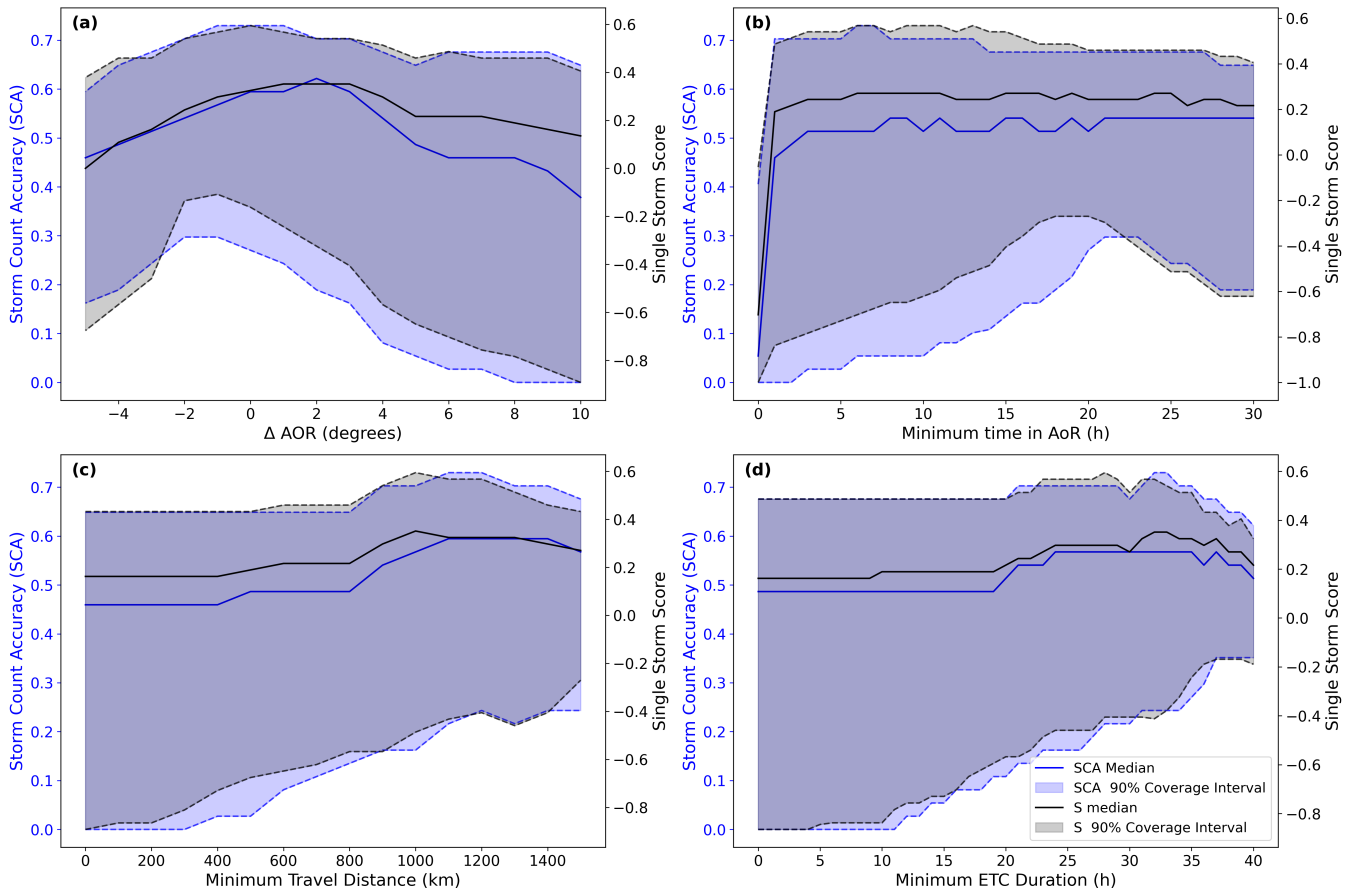


Figure 10. Sensitivity of the ImpactETC1.0 algorithm's performance to key post-processing parameters. Each panel shows the impact of varying a specific parameter on two metrics: the S score (black) and FA-SCA relative to historical impact records (blue). (a) Sensitivity to Δ AoR in degrees, (b) number of time steps required within the AoR, (c) minimum travel distance of detected ETC tracks in km, and (d) minimum ETC duration in hours. These experiments highlight trade-offs in tuning filtering parameters.

5 Discussion

5.1 Strengths and limitations of the tracking algorithm

To assess the quality of the identified ETC tracks, we present a set of representative storm surge cases here. Figure 12 shows six examples of well-defined ETC tracks that exhibit consistent spatial progression, relatively smooth trajectories, and alignment with the underlying MSLP field. These cases demonstrate the algorithm's ability to capture the temporal and spatial evolution of ETCs, even as they traverse complex regions with steep topography and land-ocean boundaries.

In contrast, Figure 13 highlights six events that illustrate the challenges of the algorithm. In these subplots, tracks that are coloured red were deemed impact-irrelevant despite manual inspection of the events suggesting that this was the wrong

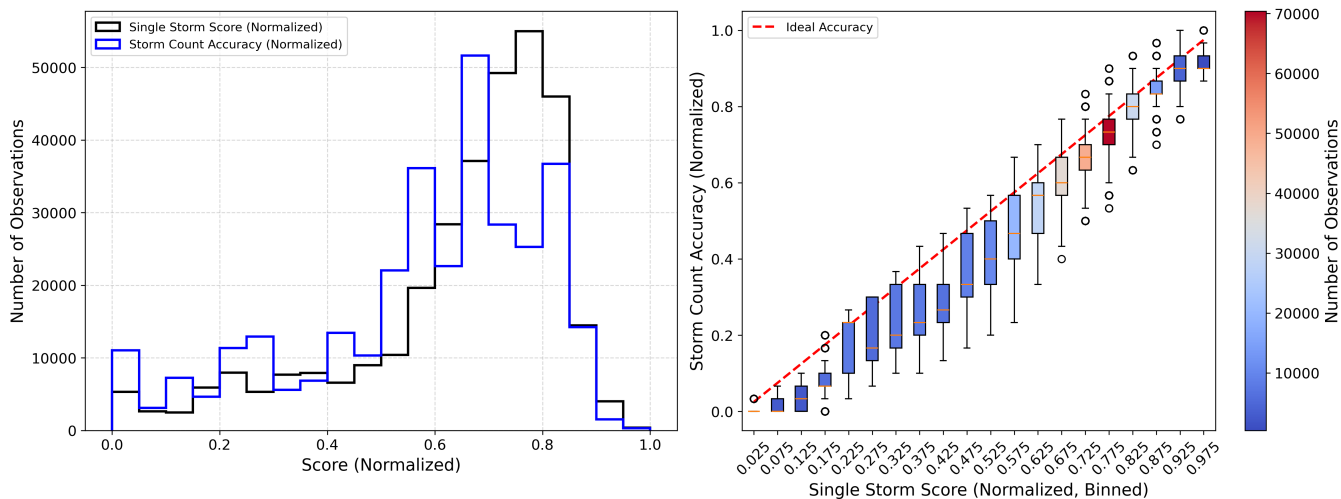


Figure 11. Comparison of marginal (left) and joint (right) distributions of the True Storm Count Accuracy and the Single Storm Score after both scores have been min-max normalised to fit 0-1 ranges. The joint distribution is shown as corresponding True Storm Count Accuracy values for discretised bins of Single Storm Score values. The red dashed line indicates the ideal 1:1 relationship. Colour shading in box plots indicates the number of observations per bin.

505 decision. Other colours than red show tracks that were deemed impact-relevant and has an uncertainty estimate associated with it in the legend of each subplot. For the storm surge case of this paper, we chose 300 km as the value of D_{max} but several of the events in Figure 13 ends up with fragmented tracks because this value is too low. This is true for the two red track fragments in the 2001-01-29 event and the two red fragments in the 2004-11-18 event. The 2005-01-08 consists of two small red track fragments and two larger ones coloured in blue and orange, however, manual inspection revealed that these all belong the same

510 ETC. It was again a too small D_{max} value that caused the break between the second red fragment and the blue fragment, and again between the blue and orange fragments. If a user of the framework is particularly interested in a single-event analyses of these events, it would be straightforward to experiment with new D_{max} values until the desired track is obtained. However, for batch processing of many events at once where manual single-event inspection is deemed unfeasible, it is a limitation of the framework that the user has to specify a single D_{max} value for all cases. Based on performing many experiments with the

515 algorithm ourselves, we find that no single parameter set performs perfectly at all events.

During the 2007-01-12 event, three separate ETC were present at the time of impact (coloured red, orange, and blue). While the red and orange tracks may look two fragments of a single track because the red ends close to where the orange starts, they were in fact two separate systems. At the time of impact, the red track was located west of the impact location, the orange track to the east, and the blue track far north between Iceland and Norway. Manual evaluation of the event suggests that the impact is

520 caused by the combined effect of the red and orange tracks that together produce strong MSLP gradients (and related westerly winds) in the North Sea. The MSLP field in the still image of the figure shows that the impact location sits on a ridge in the field. 86 % of the initialised gradient tracing attempts traces the impact northwards on the eastern side of the Scandinavian mountains

all the up to the blue ETC track, and not westwards towards the red track, which there is left undiscovered as an impact-relevant ETC. This example shows that the gradient tracing method in some special cases can be vulnerable to small-scale variations in the MSLP field, even when it initialises many tries in a 250 km radius around the impact location.

In the 2016-12-26 event shows a unique failure of the framework. Here, manual inspection of the event revealed that the two red fragments and the green fragment belonged to the same ETC, which originated in the North Atlantic and crossed the Scandinavian mountains in a northwest-southeast movement, terminating in Eastern Europe. However, the framework was not able to connect these fragments because there are several time steps where no ETC centre candidate points could be found as the ETC crossed the Scandinavian mountains. This happened because the MSLP field became very vague without a well-defined minima. Between the second red and the green fragments there was a gap of four time steps without a minima, which incidentally was right at the time of impact of several of the impacted locations shown in the inset at the lower left corner of the 2016-12-26 event. Because the combined red-green fragmented track did not exist in the four time step gap, the gradient tracing algorithm traced the impacts to the orange and blue ETCs that were located further away.

5.2 The value of global optimisation to solve the correspondence problem

A central question when choosing between a heuristic and a mathematically exact solution to the correspondence problem is whether the additional computational cost of global optimisation meaningfully affects overall algorithm performance. Table 5 shows the total computational run times for serial execution of ImpactETC1.0 on all 37 events. The runtime is divided into the main steps of the framework and tested across several pruning radii (i.e., r -values), which lead to different problem complexities to solve. This analysis shows that solving the correspondence problem with the HA accounts for only a small fraction of the total computational time, especially compared to more time-consuming components such as data loading and candidate point identification. Even at the highest level of correspondence complexity (100 km pruning radius), solving the correspondence problem remains a minor contributor to total run time compared to candidate point processing and BLOB analysis. The analysis in Table 2 shows that the differences between HA and NN drastically increase with the complexity of the problem (small pruning radii), and that HA enables additional connections. While we here investigate tracking of ETCs, which exist at large spatial scales, it is likely that HA would be even more suitable for tracking smaller atmospheric objectives and for tracking in noisier fields. This is something future research could target.

5.3 Future improvements and considerations for broader application

While the ImpactETC1.0 framework demonstrates robust performance in identifying impact-relevant ETC tracks, there are several avenues to further improve its [accuracy-performance](#) and computational efficiency.

Computational efficiency: Data loading consistently dominates runtime across all pruning radii, and while optimised data handling could reduce this component, it will likely remain a significant fraction of total runtime. The candidate identification step is already quite scalable, but it might be possible to avoid excessive block comparisons with further developments. For track reconciliation, there are likely gains to be made, e.g. by adding a distance-based pre-check before BLOB initialisation, which would reduce unnecessary BLOB constructions. [The current implementation of the framework is serial but future developments](#)

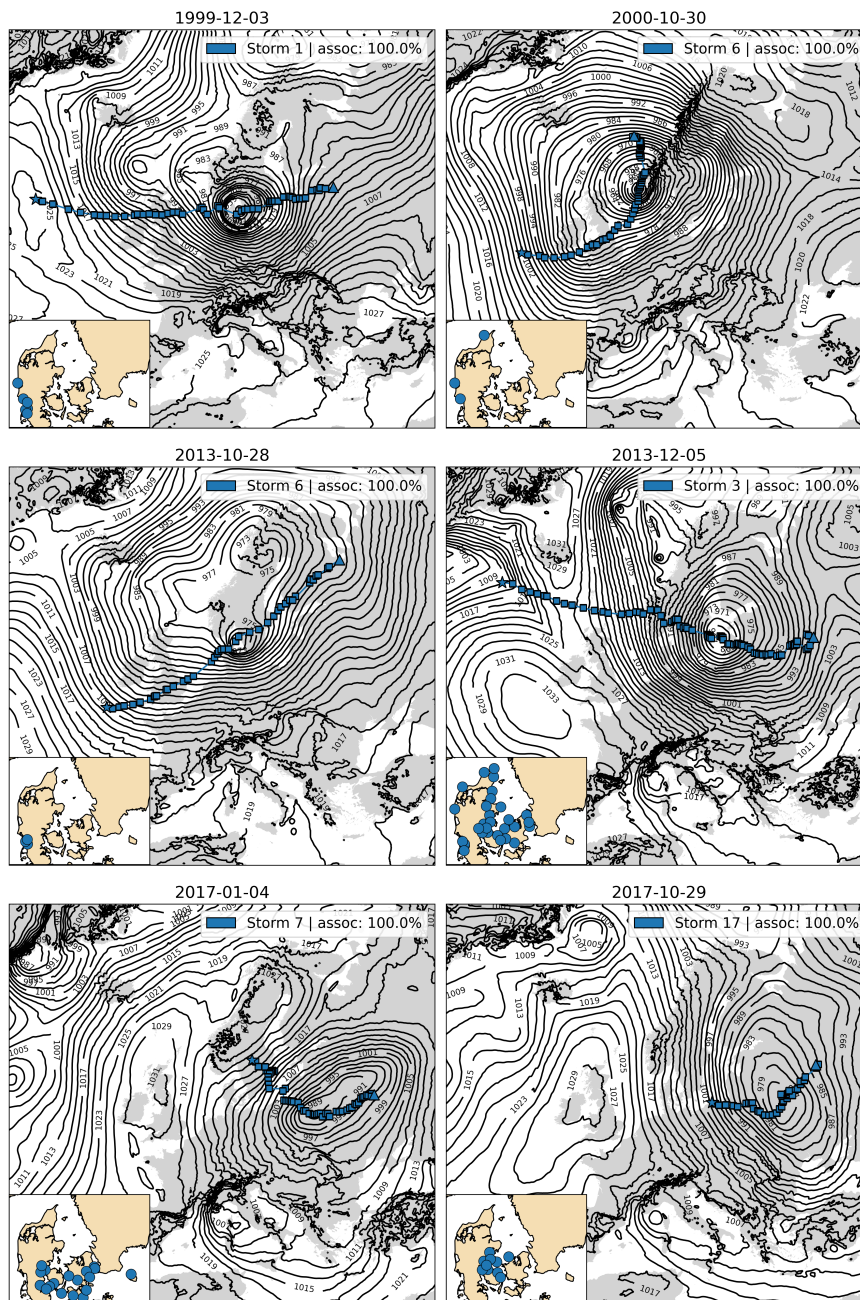


Figure 12. Six examples of well-defined, successfully identified ETC tracks by gradient tracing post-processing. Blue dots mark the time evolution along each track, and stars indicate the initial position of each track. The background shows the MSLP field at the median time step of each track, providing synoptic context for storm positioning and structure. For each example, the insert maps show locations of the water-level gauges where storm surges were found.

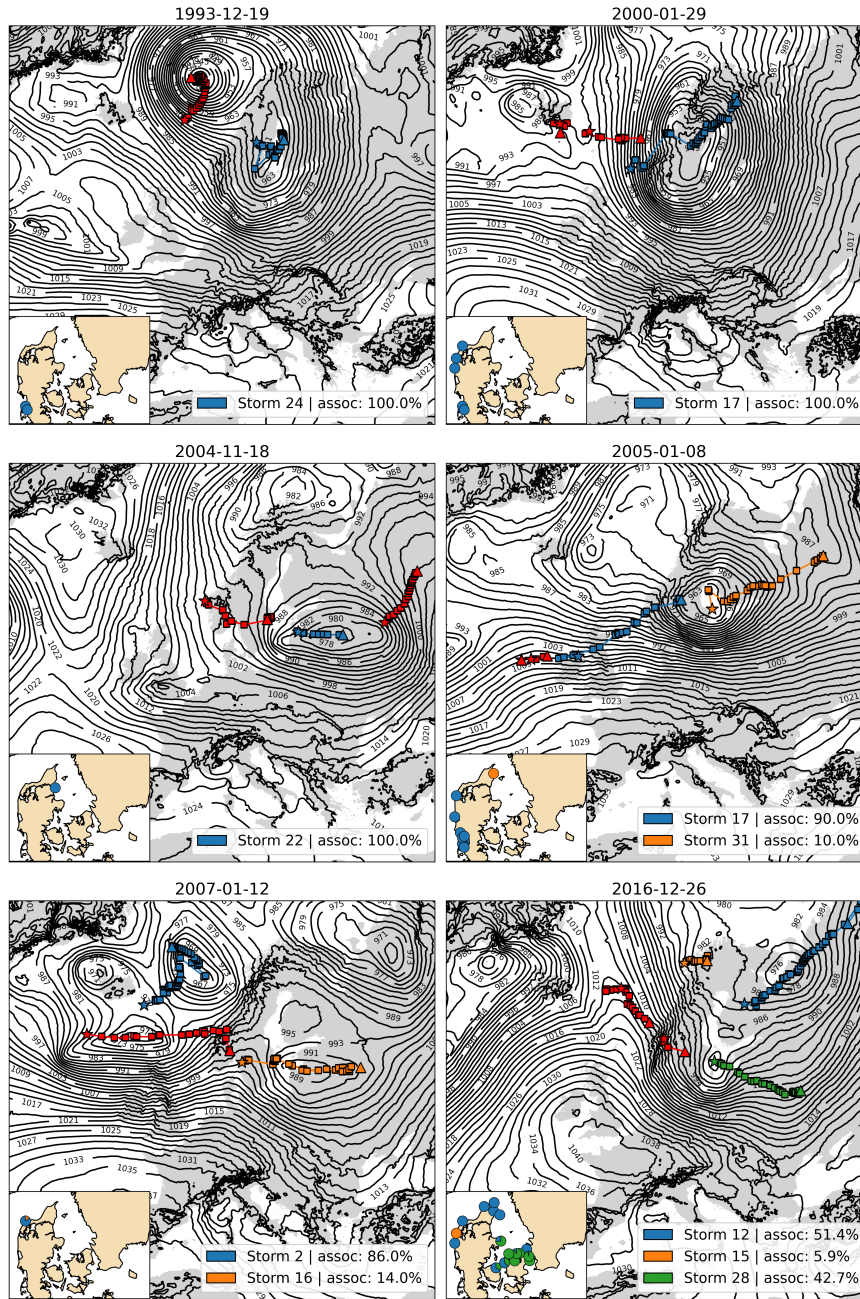


Figure 13. Six examples of ETC tracks that show the limitations of the algorithm, e.g. track splitting, short-lived centres, non-smooth movements, or secondary impact-irrelevant tracks that survive post-processing. Blue tracks are successfully identified ETCs, orange and green denote secondary impact-irrelevant tracks that survive post-processing, while red shows tracks that are impact-relevant but were filtered out. The gradient tracing approach has been used to quantify, in percentages, the storm tracks relation to impact locations for each ETC event (lower-left insert maps, coloured by storm). In our repository (Agertoft et al., 2026), we provided the animations of these six events.

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.

Improved tracking: In terms of the quality of the final tracks, more research could be done on how to refine the track reconciliation step, for example, as seen in Figure 13 for the 2016-12-06 event, where the BLOB analysis makes the wrong connection because there is a single time step where no candidate point is found on the correct track. It might be worth exploring how to incorporate multiple adjacent time steps into the BLOB analysis, rather than limiting reconciliation checks to simultaneous candidate points. Or alternatively to allow for temporal gaps during the HA/NN solutions to the correspondence problem (Ullrich et al., 2021; Pérez-Alarcón et al., 2024). It could also be worth testing the implementation of user-defined masks of high-uncertainty areas, such as mountain ranges, where the BLOB analysis is given more freedom in space and time to make the right connections. This extension would enable a more realistic identification of track jumps across complex terrains or during rapid ETC evolution. There may also be improvements by using ETC tracking information at higher pressure levels, such as 700 hPa, to improve the continuity of surface-level tracks during reconciliation. While the BLOB-based reconciliation step is crucial for avoiding premature track termination, it sometimes causes irregular jumps between points on the track. It would be interesting to explore ways to smooth tracks around reconciliation jumps. This could potentially also be guided by information at higher pressure levels.

Post-processing: The presented framework introduced three post-processing methods intended to filter out impact-irrelevant ETC tracks. Future research could further investigate the differences between these methods, e.g., comparing under which conditions the gradient tracing and nearest storm selection fail. There is also scope for more research on improving the robustness of the gradient tracing method against small-scale effects in the MSLP field. Future research could further investigate the version of the gradient tracing method with many initialisations to identify compound events where multiple ETCs interact to produce the impact. In this study, we investigate storm surge impacts driven by surface winds, which is a phenomena that is closely related to the local MSLP gradients. Other types of impacts, such as extreme precipitation events where, e.g., atmospheric rivers can cause impacts to extend far away from the ETC centre, further research will have to develop additional methods to attribute an impact to a specific nearby ETC.

Additional features: To deepen the physical interpretation of ETC evolution, associating additional properties - such as MSLP at ETC centre, vorticity, wind speed, or gusts — along identified tracks would allow classification of different stages of the storm life cycle, including genesis, intensification, mature phase, and occlusion. Even though these additional features could help limit the unintended removal of short-lived yet impactful systems that drive significant extremes, investigating such variables would likely increase the algorithm's computational demand. Future iterations could also incorporate local wind

590 direction criteria. In complex coastal regions like the Danish straits, specific wind directions are primary drivers of surges; however, defining these thresholds requires significant a priori local oceanographic knowledge.

Limitations of 1-D tracking and Lagrangian analysis: The main feature of the framework presented here is that it is impact-oriented, thus distinguishing between impact-relevant and irrelevant ETCs. This distinction depends on the type of impact studied, defined by the timing and location of the maximum impact, and it is the main purpose of the post-processing steps. 595 We present three post-processing methods, all based on the assumption that the ETC responsible for the impact is close to the impact location and at the time of maximum impact. This assumption is softened in the Gradient Tracing method, as a stronger ETC farther away may be chosen instead. However, this still implicitly assumes that the impact is caused by phenomena close to the core of the ETC. Certain types of impacts, such as pluvial flooding from precipitation, may occur far from the core. For those types of impacts, the ETCs identified with our framework will most likely be the impact-relevant ETC, but without 600 explicitly associating the tracks with the atmospheric phenomenon (i.e., precipitation) causing the impact type (i.e., pluvial flooding), it cannot be determined for certain. This requires a deeper analysis of drivers of the hazards for each individual event and is beyond the scope of the present study. This lack of information about the spatio-temporal evolution of the atmosphere is a limitation not only for the framework presented here but also, in general, for point-wise Lagrangian analysis and 1-D tracks. This could be incorporated in a future version of the algorithm.

605 ~~*Adding the option of kilometres-based distances:* Due to our relatively small AoR and its location, it is feasible in this case to use geographical degrees to define AoR. However, this might not be suitable for all AoR sizes and locations. Therefore, a future improvement could be to allow users to choose whether to use geographic degrees or kilometre-based distances.~~

Broader applications: We believe that ImpactETC1.0 is broadly applicable to other case areas. Users applying the framework to other reanalysis datasets or to different geographic regions should consider recalibrating key parameters, especially candidate 610 detection thresholds, correspondence distance limits, and AoR size, to account for differences in data resolution, domain size, and regional storm climatology. 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.

6 Conclusion

615 ~~This study introduces~~

In this study, the ImpactETC1.0, a novel framework for tracking extratropical cyclones (ETCs) with a specific emphasis on linking storm tracks to observed storm tracking framework is introduced, bridging the gap between large-scale dynamics and localised hazards by linking extratropical cyclone (ETC) tracks directly to surface impacts. The framework incorporates impact-based event selection, Novelties of the paper include the use of global optimisation via the Hungarian Algorithm, 620 BLOB-based track reconciliation, and (efficient, e.g., over complex terrain), several post-processing methods that help, and an automated validation metric that helps the user identify the storm(s) that caused the impact. ~~These methodological~~

advances enable accurate reconstruction of the ETC tracks, even in regions with complex topography and during rapid storm development, while maintaining relatively low computational demands

625 The framework is exemplified using the 5.5 km hourly CERRA reanalysis data for the 1991-2020 period, and the impact-based selection was validated against historical storm surges that exceeded 5-year return levels (here, 37 events), using a ± 24 -hour search window around peak sea levels. On this basis, the algorithm isolates the primary storm track(s) relevant to the storm surge while effectively filtering out irrelevant ones.

630 Using a variant of a nearest-neighbour heuristic to solve the correspondence problem is standard practice in ETC tracking algorithms. Such heuristics are fast, but come with the risk of making wrong connections. In this study, a comparison between the framework uses the Hungarian Algorithm and the (HA) for global optimisation, and validated this approach against the commonly used nearest-neighbour heuristic shows that, in scenarios approach. The HA approach provided more connections, especially under conditions with many candidate points that must be connected over time, the nearest-neighbour heuristic yields different solutions in 14 % of cases. The Hungarian Algorithm's global optimisation requires an order-of-magnitude more computation time but is still very fast and accounts, and, despite its increased computational complexity, remains highly efficient, accounting for less than 1 % of the total algorithm's run time, which is dominated by candidate point selection and track reconciliation. Our results suggest that these issues of accuracy and speed could become increasingly significant for the broader scientific field of ETC tracking if higher-resolution reanalysis datasets are used in the future, and perhaps even more so for tracking other smaller and more dynamic meteorological phenomena such as convective precipitation cells system runtime.

640 As it can be difficult to perform surface-level tracking of ETCs To address storm-track fragmentation over complex terrain, this study developed and tested a BLOB-analysis-based step to reconcile fragmented tracks. The results show that such a reconciliation step is necessary when ETCs traverse complex terrain such as the Scandinavian mountains and land-ocean boundaries. The issue of fragmented tracks becomes more pronounced when many candidate points need to be connected over time. In the tested case, reconciling fragmented tracks results in final ETC tracks that are, on average, twice as long because it prevents premature termination. The BLOB analysis used in this study represents a novel contribution to the storm-tracking literature, indicating potential for further improvements and applications in future research. A BLOB-based reconciliation identifies contiguous low-pressure regions to reconnect broken tracks. At the chosen pruning radius of 350 km, the BLOB reconciliation increased the average track duration by 70 % and the total length by 62 %, preventing, e.g., premature termination over the Scandinavian mountains.

650 Additionally, an impact-oriented calibration of post-processing parameters enables the removal of unnecessary tracks while increasing the correspondence between identified ETC tracks and impact events, thereby enhancing applications in impact studies. The framework offers three distinct Two post-processing options : standard-calibrated filtering, a proximity-based selection of the track geographically closest to the impact, and a "gradient-tracking — nearest storm selection, and gradient-tracking — were introduced and evaluated alongside automatic calibration with the " approach that identifies which ETC that contains the impact location within its low-pressure basin. The study investigated two metrics to calibrate the parameters. One of these was based on labour-intensive manual labelling of the correct number of ETCs per impact event. The other used automatic scoring in the form of a metric named the Single Storm Score, which penalises both under- and overestimation of the resulting number

of ETC tracks. The results show that the Single Storm Score produces calibrated parameter values similar to those obtained by manual labelling. These results demonstrate that it is possible to calibrate post-processing parameters for impact-focused tracking without performing extensive manual event inspection.

660 Proposed future improvements include using vertical storm information to improve the tracking and reconciliation of fragments, applying additional post-processing steps to better discriminate tracks related to and unrelated to the impact, and enhancing the speed and efficiency of key algorithmic steps" which can be used to avoid manual tuning or ad-hoc parameter selection. While data loading currently accounts for a substantial share of processing time, the framework's modular design enables future parallelisation and operational scaling.

665 For future application, ImpactETC1.0 has the potential to be used more widely for storm tracking applications in other geographic regions, provided it is calibrated to those conditions. This offers a powerful tool for advancing, e.g. climate hazard assessments and operational forecasting capabilities offers a scalable tool beyond storm surges alone for assessing additional applications such as wind damages, providing a robust, automated methodology for historical hazard analysis and future climate adaptation planning. All together, these methodological advances enable skilful reconstruction of the ETC tracks,
670 even in regions with complex topography and during rapid storm development, while maintaining relatively low computational demands.

7 Appendix A

Figure A1 shows the D_{max} sensitivity analysis, where D_{max} was varied for two different pruning radius scenarios. The metrics selected for studying the effect on the formed tracks were (a) the number of tracks, and (b) the track duration. The sensitivity
675 analysis was conducted on the correspondence problem-step of the algorithm, which is executed before the BLOB-based track reconciliation and the post-processing steps, which also have significant influence on the final track properties. These two steps connects small track fragments and filters minor, impact-irrelevant tracks away, respectively. These steps were not executed for this minor sensitivity analysis to isolate the influence of D_{max} , and this resulted in many small tracks. Most of which would be filtered out later. We therefore chose to plot the 90th quantile of the track duration rather than the mean or median track length
680 to better see the D_{max} influence.

For good measure, we note that neither an increased number of tracks nor an increased q90 track duration is inherently desirable by themselves, as a high number of tracks implies many smaller tracks with severed connections and an abnormally high q90 track duration would imply tracks with connections that are not representative of the underlying system.

Code and data availability. The codes and the dataset used in this paper are available at Zenodo: <https://doi.org/10.5281/zenodo.20309449>
685 (Agertoft et al., 2026). The CERRA sub-daily regional reanalysis data used in this study are freely available from the Copernicus Climate Change Service Climate Data Store (C3S CDS): <https://doi.org/10.24381/cds.622a565a> (Schimanke et al., 2021).

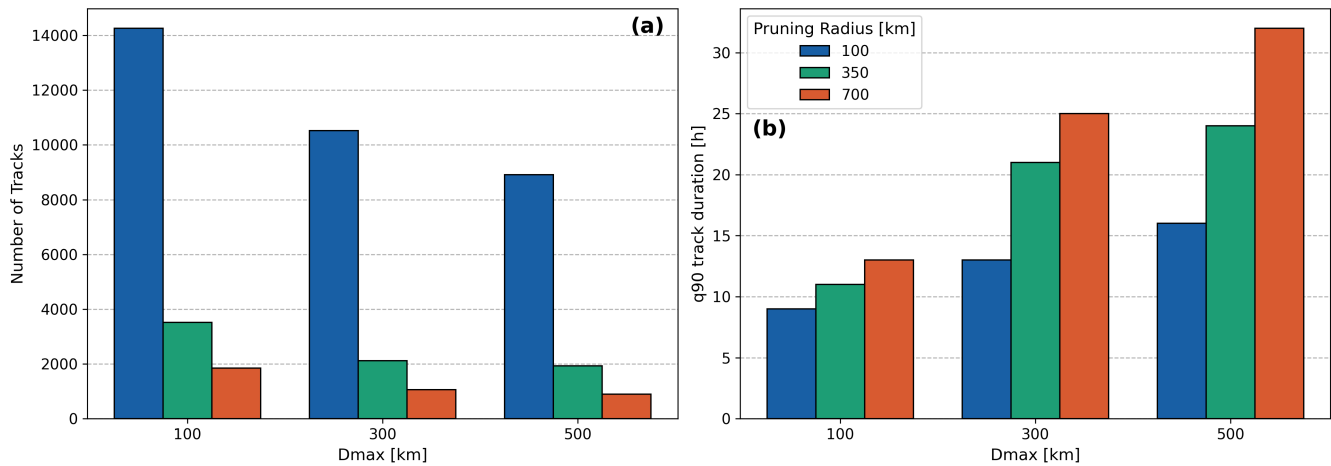


Figure A1. Sensitivity of D_{max} and Pruning radius represented through two track metrics. (a) shows the number of tracks under different D_{max} and Pruning radii combinations, (b) shows the q90 track duration in hours under different D_{max} and Pruning radii combinations. These experiments highlight trade-offs in tuning filtering parameters.

Author contributions. NA conceptualized the tracking framework, developed the methodology and software, performed the analysis, made visualizations and wrote the initial draft. JS handled data acquisition, data analysis, and visualizations. JP conceptualized the framework, data analysis and interpretation. IR performed data analysis. ML acquired funding and was the project coordinator for study conception and design. All authors revised and edited the manuscript. All authors have read and approved the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. Niels Agertoft worked for the ECCO project, which received funding from Innovation Fund Denmark and the European Union's Horizon Europe Programme under the 2022 Joint Transnational Call of the European Partnership Water4All (Grant agreement ID: 101060874). Jonas Wied Pedersen was financially supported by the Innovation Fund Denmark through the industrial postdoc programme [grant number: 0197-00005B]. The authors thank ECMWF and Copernicus for making the CERRA dataset available.

References

- Agertoft, N., Su, J., Pedersen, J. W., Ringgaard, I. M., and Larsen, M. A. D.: ImpactETC1.0 release, <https://doi.org/10.5281/zenodo.20309449>, 2026.
- 700 André, E., Su, J., Dahl Larsen, M. A., Drews, M., Stendel, M., and Skovgaard Madsen, K.: The role of preconditioning for extreme storm surges in the western Baltic Sea, *Natural Hazards and Earth System Sciences*, 23, 1817–1834, <https://doi.org/10.5194/nhess-23-1817-2023>, 2023.
- André, E., Su, J., Larsen, M. A. D., Madsen, K. S., and Drews, M.: Simulating major storm surge events in a complex coastal region, *Ocean Modelling*, 162, 101 802, <https://doi.org/10.1016/j.ocemod.2021.101802>, 2021.
- 705 André, E., Drews, M., Su, J., Larsen, M. A. D., Drønen, N., and Madsen, K. S.: Simulating wind-driven extreme sea levels: Sensitivity to wind speed and direction, *Weather and Climate Extremes*, 36, 100 422, <https://doi.org/https://doi.org/10.1016/j.wace.2022.100422>, 2022.
- Aragão, L. and Porcù, F.: Cyclonic activity in the Mediterranean region from a high-resolution perspective using ECMWF ERA5 dataset, *Climate Dynamics*, 58, 1293–1310, <https://doi.org/https://doi.org/10.1007/s00382-021-05963-x>, 2022.
- Bengtsson, L., Hodges, K. I., and Keenlyside, N.: Will Extratropical Storms Intensify in a Warmer Climate?, *Journal of Climate*, 22, 2276 – 2301, <https://doi.org/10.1175/2008JCLI2678.1>, 2009.
- 710 Catto, J. L.: Extratropical cyclone classification and its use in climate studies, *Reviews of Geophysics*, 54, 486–520, <https://doi.org/10.1002/2016RG000519>, 2016.
- Feser, F., Barcikowska, M., Krueger, O., Schenk, F., Weisse, R., and Xia, L.: Storminess over the North Atlantic and northwestern Europe—A review, *Quarterly Journal of the Royal Meteorological Society*, 141, 350–382, <https://doi.org/https://doi.org/10.1002/qj.2364>, 2015.
- 715 Flaounas, E., Kotroni, V., Lagouvardos, K., and Flaounas, I.: CycloTRACK (v1.0)—tracking winter extratropical cyclones based on relative vorticity: sensitivity to data filtering and other relevant parameters, *Geoscientific Model Development*, 7, 1841–1853, <https://doi.org/https://doi.org/10.5194/gmd-7-1841-2014>, 2014.
- Flaounas, E., Aragão, L., Bernini, L., Dafis, S., Doiteau, B., Flocas, H., Gray, S. L., Karwat, A., Kouroutzoglou, J., Lionello, P., Miglietta, M. M., Pantillon, F., Pasquero, C., Patlakas, P., Picornell, M. A., Porcù, F., Priestley, M. D. K., Reale, M., Roberts, M. J., Saaroni, H., Sandler, D., Scoccimarro, E., Sprenger, M., and Ziv, B.: A composite approach to produce reference datasets for extratropical cyclone tracks: application to Mediterranean cyclones, *Weather and Climate Dynamics*, 4, 639–661, <https://doi.org/10.5194/wcd-4-639-2023>, 2023.
- 720 Froude, L. S. R.: TIGGE: Comparison of the Prediction of Northern Hemisphere Extratropical Cyclones by Different Ensemble Prediction Systems, *Weather and Forecasting*, 25, 819 – 836, <https://doi.org/10.1175/2010WAF2222326.1>, 2010.
- Gonçalves, A., Liberato, M. L. R., and Nieto, R.: Wind Energy Assessment during High-Impact Winter Storms in Southwestern Europe, *Atmosphere*, 12, <https://doi.org/10.3390/atmos12040509>, 2021.
- 725 Gramscianinov, C., Campos, R., de Camargo, R., Hodges, K., Guedes Soares, C., and da Silva Dias, P.: Analysis of Atlantic extratropical storm tracks characteristics in 41 years of ERA5 and CFSR/CFSv2 databases, *Ocean Engineering*, 216, 108 111, <https://doi.org/https://doi.org/10.1016/j.oceaneng.2020.108111>, 2020.
- 730 Grieger, J., Leckebusch, G. C., Raible, C. C., Rudeva, I., and and, I. S.: Subantarctic cyclones identified by 14 tracking methods, and their role for moisture transports into the continent, *Tellus A: Dynamic Meteorology and Oceanography*, 70, 1–18, <https://doi.org/10.1080/16000870.2018.1454808>, 2018.

- Hodges, K. I.: Feature Tracking on the Unit Sphere, *Monthly Weather Review*, 123, 3458 – 3465, [https://doi.org/10.1175/1520-0493\(1995\)123<3458:FTOTUS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<3458:FTOTUS>2.0.CO;2), 1995.
- Hodges, K. I., Hoskins, B. J., Boyle, J., and Thorncroft, C.: A Comparison of Recent Reanalysis Datasets Using Objective Feature
735 Tracking: Storm Tracks and Tropical Easterly Waves, *Monthly Weather Review*, 131, 2012 – 2037, [https://doi.org/10.1175/1520-0493\(2003\)131<2012:ACORRD>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<2012:ACORRD>2.0.CO;2), 2003.
- Hofstätter, M., Chimani, B., Lexer, A., and Blöschl, G.: A new classification scheme of European cyclone tracks with relevance to precipitation, *Water Resources Research*, 52, 7086–7104, <https://doi.org/https://doi:10.1002/2016WR019146>, 2016.
- Hoskins, B. J. and Hodges, K. I.: The annual cycle of Northern Hemisphere storm tracks. Part I: Seasons, *Journal of Climate*, 32, 1743–1760,
740 <https://doi.org/https://doi.org/10.1175/JCLI-D-17-0870.s1>, 2019.
- Hunter, A., Stephenson, D. B., Economou, T., Holland, M., and Cook, I.: New perspectives on the collective risk of extratropical cyclones, *Quarterly Journal of the Royal Meteorological Society*, 142, 243–256, <https://doi.org/10.1002/qj.2649>, 2016.
- Kuhn, H. W.: The Hungarian method for the assignment problem, *Naval Research Logistics Quarterly*, 2, 83–97,
<https://doi.org/10.1002/nav.3800020109>, 1955.
- 745 Lakkis, S. G., Canziani, P., Yuchechen, A., Rocamora, L., Caferra, A., Hodges, K., and O’Neill, A.: A 4D feature-tracking algorithm: A multidimensional view of cyclone systems, *Quarterly Journal of the Royal Meteorological Society*, 145, 395–417,
<https://doi.org/10.1002/qj.3436>, 2019.
- Lodise, J., Merrifield, S., Collins, C., Rogowski, P., Behrens, J., and Terrill, E.: Global climatology of extratropical cyclones from a new tracking approach and associated wave heights from satellite radar altimeter, *Journal of Geophysical Research: Oceans*, 127, e2022JC018925,
750 <https://doi.org/https://doi.org/10.1029/2022JC018925>, 2022.
- Medina, S. and Houze Jr., R. A.: Kelvin–Helmholtz waves in extratropical cyclones passing over mountain ranges, *Quarterly Journal of the Royal Meteorological Society*, 142, 1311–1319, <https://doi.org/10.1002/qj.2734>, 2016.
- Munkres, J.: Algorithms for the Assignment and Transportation Problems, *Journal of the Society for Industrial and Applied Mathematics*, 5, 32–38, <https://doi.org/10.1137/0105003>, 1957.
- 755 Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., Coccozza, A., Dacre, H. F., Feng, Y., Fraedrich, K., Grieger, J., Gulev, S., Hanley, J., Hewson, T., Inatsu, M., Keay, K., Kew, S. F., Kindem, I., Leckebusch, G. C., Liberato, M. L. R., Lionello, P., Mokhov, I. I., Pinto, J. G., Raible, C. C., Reale, M., Rudeva, I., Schuster, M., Simmonds, I., Sinclair, M., Sprenger, M., Tilinina, N. D., Trigo, I. F., Ulbrich, S., Ulbrich, U., Wang, X. L., and Wernli, H.: IMILAST: A Community Effort to Intercompare Extratropical Cyclone
760 Detection and Tracking Algorithms, *Bulletin of the American Meteorological Society*, 94, 529 – 547, <https://doi.org/10.1175/BAMS-D-11-00154.1>, 2013.
- Pauley, P. M.: An Example of Uncertainty in Sea Level Pressure Reduction, *Weather and Forecasting*, 13, 833 – 850,
[https://doi.org/10.1175/1520-0434\(1998\)013<0833:AEOUIS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0833:AEOUIS>2.0.CO;2), 1998.
- Peréz-Alarcón, A., Coll-Hidalgo, P., Trigo, R. M., Nieto, R., , and Gimeno, L.: CyTRACK: An open-source and user-friendly python toolbox for detecting and tracking cyclones, *Environmental Modelling and Software*, 176, 106027,
765 <https://doi.org/https://doi.org/10.1016/j.envsoft.2024.106027>, 2024.
- Priestley, M. D., Ackerley, D., Catto, J. L., Hodges, K. I., McDonald, R. E., and Lee, R. W.: An overview of the extratropical storm tracks in CMIP6 historical simulations, *Journal of Climate*, 33, 6315–6343, <https://doi.org/https://doi.org/10.1175/JCLI-D-19-0928.1>, 2020.

- Ragone, F., Mariotti, M., Parodi, A., von Hardenberg, J., and Pasquero, C.: A climatological study of Western Mediterranean Medicanes in numerical simulations with explicit and parameterized convection, *Atmosphere*, 397, <https://doi.org/https://doi.org/10.3390/atmos9100397>, 2018.
- 770 Raible, C. C., Della-Marta, P. M., Schwierz, C., Wernli, H., and Blender, R.: Northern Hemisphere Extratropical Cyclones: A Comparison of Detection and Tracking Methods and Different Reanalyses, *Monthly Weather Review*, 136, 880 – 897, <https://doi.org/10.1175/2007MWR2143.1>, 2008.
- Ridal, M., Bazile, E., Le Moigne, P., Randriamampianina, R., Schimanke, S., Andrae, U., Berggren, L., Brousseau, P., Dahlgren, P., Edvinsson, L., El-Said, A., Glinton, M., Hagelin, S., Hopsch, S., Isaksson, L., Medeiros, P., Olsson, E., Unden, P., and Wang, Z. Q.: CERRA, the Copernicus European Regional Reanalysis system, *Quarterly Journal of the Royal Meteorological Society*, 150, 3385–3411, <https://doi.org/https://doi.org/10.1002/qj.4764>, 2024.
- 775 Sanchez-Gomez, E. and Somot, S.: Impact of the internal variability on the cyclone tracks simulated by a regional climate model over the Med-CORDEX domain, *Climate Dynamics*, 51, 1005—1021, <https://doi.org/https://doi.org/10.1007/s00382-016-3394-y>, 2018.
- 780 Schimanke, S., Ridal, M., Le Moigne, P., Berggren, L., Undén, P., Randriamampianina, R., Andrea, U., Bazile, E., Bertelsen, A., Brousseau, P., Dahlgren, P., Edvinsson, L., El Said, A., Glinton, M., Hopsch, S., Isaksson, L., Mladek, R., Olsson, E., Verrelle, A., and Wang, Z.: CERRA sub-daily regional reanalysis data for Europe on single levels from 1984 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), <https://doi.org/10.24381/cds.622a565a>, [Accessed Jan 31, 2024], 2021.
- Seda, M.: The Assignment Problem and Its Relation to Logistics Problems, *Algorithms*, 15, <https://doi.org/10.3390/a15100377>, 2022.
- 785 Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., and Reed, K. A.: TempestExtremes v2. 1: A community framework for feature detection, tracking, and analysis in large datasets, *Geoscientific Model Development*, 14, 5023–5048, <https://doi.org/https://doi.org/10.5194/gmd-14-5023-2021>, 2021.
- Walker, E., Mitchell, D., and Seviour, W.: The numerous approaches to tracking extratropical cyclones and the challenges they present, *Weather*, 75, 336–341, <https://doi.org/https://doi.org/10.1002/wea.3861>, 2020.
- 790 Zappa, G., Shaffrey, L. C., Hodges, K. I., Sansom, P. G., and Stephenson, D. B.: A Multimodel Assessment of Future Projections of North Atlantic and European Extratropical Cyclones in the CMIP5 Climate Models, *Journal of Climate*, 26, 5846 – 5862, <https://doi.org/10.1175/JCLI-D-12-00573.1>, 2013.
- Zhang, S., Xue, Y., Zhang, H., Zhou, X., Li, K., and Liu, R.: Improved Hungarian algorithm-based task scheduling optimization strategy for remote sensing big data processing, *Geo-spatial Information Science*, 27, 1141–1154, <https://doi.org/10.1080/10095020.2023.2178339>, 2024.
- 795

Table 1. Overview of parameters in the full ETC tracking algorithm including explanation and the values used in the final setup for the case application.

| Parameter | Algorithm Step | Explanation | Value for this case |
|-----------------------------------|---|---|------------------------------------|
| Impact window size | Identify candidates | Time window around the start (t_s) and end (t_e) of peaks for that event. | ± 24 h |
| t_s | Identify candidates | Start date of event | - |
| t_e | Identify candidates | End date of event | - |
| Pruning radius | Identify candidates | A minima is pruned if another minima with lower pressure is found within a given range | 350 km |
| $MSLP_{max}$ | Identify candidates | A minima is pruned if it has a pressure above $MSLP_{max}$ | 1010 hPa |
| $\zeta_{500 min}$ | Identify candidates | A minima is pruned if it exceeds a ζ_{500} vorticity threshold within the $\zeta_{500 min}$ Range | $1.5 \cdot 10^{-4} \text{ s}^{-1}$ |
| $\zeta_{500 min}$ Range | Identify candidates | Allowed range for a minima to associate vorticity. | 500 km |
| D_{max} | Correspondance <u>correspondence</u> | Maximum allowed distance for connection between minimas in adjacent timesteps. | 300 km |
| Pressure range | BLOB analysis | Pressure range surrounding the local minima for thresholding the MSLP slice prior to BLOB detection | ± 5 hPa |
| Max BLOB bounding box | BLOB analysis | Maximum height or length of the bounding box surrounding the identified BLOB | 3000 km |
| Max BLOB candidate point distance | BLOB analysis | Maximum distance between two points found to be in the same BLOB | 600 km |
| Minimum travel distance | Post-processing | Minimum distance an ETC must travel to not be filtered out | 200 km |
| Minimum duration | Post-processing | Minimum duration of an ETC for it to not be filtered out | 32 h |
| AoR Latitude Min | Post-processing | Minimum latitude of the AoR | 50° |
| AoR Latitude Max | Post-processing | Maximum latitude of the AoR | 70° |
| AoR Longitude Min | Post-processing | Minimum longitude of the AoR | 0° |
| AoR Longitude Max | Post-processing | Maximum longitude of the AoR | 30° |
| Δ AoR | Post-processing | Change in size of the AoR that an ETC must enter | 0° |
| Minimum time in AoR | Post-processing | Minimum amount of time steps spent inside the AoR | 12 h |
| Search Radius | Post-processing | Radius surrounding impact location to initialise Gradient Tracing from | 250 km |
| Number of Points | Post-processing | 38 Number of points initialised within Search Radius | 100 |
| Steps Before Timeout | Post-processing | Maximum number of steps towards minima be- | 100000 |

Table 2. Comparison of Hungarian Algorithm (HA) and Nearest Neighbour (NN) for different pruning radii. The presented run times are to solve all 1966 correspondence problems across all 37 events.

| Pruning radius [km] | Median problem size | Run time for HA [s] | Run time for NN [s] | # Different solutions (%) | # Extra HA connections (%) | Mean objective difference [%] |
|---------------------|---------------------|---------------------|---------------------|---------------------------|----------------------------|-------------------------------|
| 700 | 9 | 0.11 | 0.12 | 1 (0.05) | 0 (0.0) | 0.01 |
| 500 | 12 | 0.13 | 0.14 | 0 (0.0) | 0 (0.0) | 0.0 |
| 350 | 16 | 0.18 | 0.18 | 4 (0.2) | 0 (0.0) | 0.02 |
| 250 | 21 | 0.26 | 0.21 | 4 (0.2) | 2 (0.1) | 0.0 |
| 175 | 29 | 0.56 | 0.3 | 23 (1.17) | 8 (0.41) | 0.08 |
| 100 | 52 | 4.88 | 0.55 | 273 (13.89) | 78 (3.97) | 0.94 |

Table 3. Summary of track reconciliation performance across different pruning radii, utilising gradient tracing post-processing. Metrics include the number of events with detected tracks, events and tracks where reconciliation is active, total number of reconciliations, and improvements in track duration and distance.

| Metric \ pruning radius [km] | 700 | 500 | 350 | 250 | 175 | 100 |
|---|-------|-------|-------|-------|-------|-------|
| 1. Events with a track | 36/37 | 37/37 | 37/37 | 37/37 | 37/37 | 37/37 |
| 2. Events where reconciliation is active | 8/36 | 11/37 | 19/37 | 20/37 | 25/37 | 32/37 |
| 3. Tracks where reconciliation is active | 8/51 | 11/52 | 19/54 | 23/55 | 34/57 | 39/59 |
| 4. Total number of reconciliations | 9/8 | 13/11 | 28/19 | 46/23 | 61/34 | 85/39 |
| 5. Average increase in track duration (%) | 32 | 57 | 70 | 99 | 81 | 105 |
| 6. Average increase in track length (%) | 36 | 52 | 62 | 75 | 66 | 81 |

Table 5. Breakdown of algorithm run time (s) across individual framework components for different candidate point pruning radii, r . Columns show total run time vs times spent on data loading, candidate identification, solving the correspondence problem, BLOB construction, and post-processing utilising gradient tracing. Values are averages over all events, highlighting how pruning radius affects computational demands and the relative contributions of each algorithm component.

| Pruning radius [km] | Total run time [s] | Data load [s] | Identify candidates [s] | Solve correspondence [s] | BLOB construction [s] | Post-processing [s] |
|---------------------|-------------------------|------------------------|-------------------------|--------------------------|-----------------------|---------------------|
| 700 | 1068.44 803 | 698.4 640 | 25.78 25 | 0.12 0.11 | 83.2 83 | 59.21 54 |
| 500 | 1521.99 855 | 1054.08 647 | 32.04 35 | 0.13 0.14 | 106.59 118 | 65.52 55 |
| 350 | 1416.25 909 | 780.97 652 | 47.93 48 | 0.19 0.18 | 151.03 152 | 67.26 57 |
| 250 | 1666.09 968 | 787.44 641 | 71.07 72 | 0.27 0.26 | 196.65 197 | 55.79 58 |
| 175 | 2197.53 1104 | 897.47 642 | 119.35 119 | 0.56 | 290.04 287 | 55.84 56 |
| 100 | 3185.46 1627 | 727.79 673 | 322.79 337 | 4.88 5.64 | 475.87 556 | 65.42 56 |

| start_date | end_date |
|---------------------|---------------------|
| 1991-01-09 07:00:00 | 1991-01-09 12:15:00 |
| 1991-12-20 11:00:00 | 1991-12-20 19:00:00 |
| 1992-01-17 09:00:00 | 1992-01-17 09:00:00 |
| 1993-01-12 23:15:00 | 1993-01-12 23:15:00 |
| 1993-02-21 16:00:00 | 1993-02-21 22:45:00 |
| 1993-12-19 23:30:00 | 1993-12-20 00:00:00 |
| 1996-11-06 07:30:00 | 1996-11-06 07:30:00 |
| 1997-10-11 18:45:00 | 1997-10-11 18:45:00 |
| 1998-10-25 22:00:00 | 1998-10-25 22:00:00 |
| 1999-12-03 15:45:00 | 1999-12-03 18:30:00 |
| 2000-01-29 19:45:00 | 2000-01-30 04:00:00 |
| 2000-10-30 17:00:00 | 2000-10-30 23:00:00 |
| 2002-01-28 22:00:00 | 2002-01-28 23:10:00 |
| 2002-02-21 00:10:00 | 2002-02-21 10:50:00 |
| 2002-02-26 18:50:00 | 2002-02-26 19:40:00 |
| 2003-12-06 15:16:00 | 2003-12-06 21:30:00 |
| 2004-11-18 16:10:00 | 2004-11-18 16:10:00 |
| 2005-01-08 14:40:00 | 2005-01-09 01:40:00 |
| 2006-11-01 16:40:00 | 2006-11-02 09:50:00 |
| 2007-01-12 08:00:00 | 2007-01-12 08:00:00 |
| 2007-01-19 06:40:00 | 2007-01-19 06:40:00 |
| 2007-01-25 02:00:00 | 2007-01-25 02:00:00 |
| 2008-03-01 20:00:00 | 2008-03-02 03:30:00 |
| 2011-02-11 22:10:00 | 2011-02-11 22:10:00 |
| 2011-11-27 18:20:00 | 2011-11-27 20:50:00 |
| 2011-12-09 18:40:00 | 2011-12-10 04:40:00 |
| 2012-01-14 06:10:00 | 2012-01-14 08:20:00 |
| 2013-10-28 15:30:00 | 2013-10-28 16:00:00 |
| 2013-12-02 03:20:00 | 2013-12-02 03:20:00 |
| 2013-12-05 15:00:00 | 2013-12-06 23:50:00 |
| 2015-01-10 11:40:00 | 2015-01-11 10:20:00 |
| 2015-11-29 18:50:00 | 2015-11-29 18:50:00 |
| 2016-12-26 19:10:00 | 2016-12-27 04:40:00 |
| 2017-01-04 16:50:00 | 2017-01-05 01:00:00 |
| 2017-10-29 20:00:00 | 2017-10-29 23:50:00 |
| 2019-01-02 04:30:00 | 2019-01-02 17:40:00 |
| 2019-01-09 02:10:00 | 2019-01-09 07:10:00 |

Table A1. Start and end dates of events, if only one station was hit then start date equals end date, if not then multiple stations were hit sequentially and start date represents the peak of the earliest station hit, and end date represents the peak of the last station hit in that event. Timestamp highlights peaks of event.

Table A2. Summary of track reconciliation performance across different pruning radii, utilizing Single Storm Score post processing. Metrics include the number of events with detected tracks, events and tracks where reconciliation is active, total number of reconciliations, and improvements in track duration and distance.

| Metric \ pruning radius [km] | 700 | 500 | 350 | 250 | 175 | 100 |
|---|-------|-------|-------|-------|-------|-------|
| 1. Events with a track | 35/37 | 36/37 | 34/37 | 37/37 | 35/37 | 35/37 |
| 2. Events where reconciliation is active | 10/35 | 12/36 | 17/34 | 23/37 | 21/35 | 29/35 |
| 3. Tracks where reconciliation is active | 10/37 | 12/42 | 18/38 | 26/47 | 23/41 | 32/40 |
| 4. Total number of reconciliations | 11/10 | 14/12 | 24/18 | 51/26 | 45/23 | 79/32 |
| 5. Average increase in track duration (%) | 38 | 58 | 73 | 99 | 95 | 116 |
| 6. Average increase in track length (%) | 38 | 54 | 60 | 80 | 69 | 101 |

Table A3. Summary of track reconciliation performance across different pruning radii, utilizing Closest Track post processing. Metrics include the number of events with detected tracks, events and tracks where reconciliation is active, total number of reconciliations, and improvements in track duration and distance.

| Metric \ pruning radius [km] | 700 | 500 | 350 | 250 | 175 | 100 |
|---|-------|-------|-------|-------|-------|-------|
| 1. Events with a track | 37/37 | 37/37 | 37/37 | 37/37 | 37/37 | 37/37 |
| 2. Events where reconciliation is active | 7/37 | 9/37 | 16/37 | 18/37 | 20/37 | 28/37 |
| 3. Tracks where reconciliation is active | 7/41 | 9/41 | 16/42 | 19/43 | 23/41 | 29/45 |
| 4. Total number of reconciliations | 9/7 | 10/9 | 22/16 | 40/19 | 45/23 | 67/29 |
| 5. Average increase in track duration (%) | 13 | 46 | 68 | 97 | 84 | 112 |
| 6. Average increase in track length (%) | 41 | 47 | 54 | 76 | 71 | 90 |