wThe authors are grateful to the editor and all reviewers for their time and energy in providing helpful comments that have improved the manuscript. <u>In our revised paper, we re-checked all revisions and performed grammatical corrections to help readers understand our manuscript easier.</u>

In this document, reviewer' comments have been answered point by point. Referee comments are shown in black italics and author responses are shown in blue regular text and revised version of the manuscript is shown in green text.

**Reviewer #2:**

**General comments:**

The manuscript describes a great data set and a laudable effort to construct such data base of PL and MPC tracks based on ERA5 and satellite data. However, the characteristics and hence value of the data set is scientifically unclear. For existing similar track data sets, it is investigated how these are matched. It occurs that only a marginal set of points in the data base is characterized in the manuscript by these existing sets. Moreover, these appear the easiest tracks to capture, hence the value of most of the tracks remains unclear. This is associated with the fact that I miss a critical scientific assessment of the tracks generated. The manuscript appears subjective, rather than rigorous. There are ways to verify PL and MPC tracks with observations of atmospheric dynamics, in particular wind scatterometers. The use of scatterometers in this manuscript is rather unclear from a dynamic perspective and poor. In the least, the manuscript should be scientifically clarified and the pros and cons of the methodology better stipulated. In addition, a section on future work appears appropriate as much remains unclear in my interpretation of the manuscript.

Dear Prof. Stoffelen, we deeply appreciate your rigorous and constructive critique. Your insights have been instrumental in refining our methodology and dataset validation. Below, we address each point systematically.

## Major revision

1. **Addition of hourly corrected ERA5 wind variables to the dataset**

To address the concern "The use of scatterometers... is unclear and poor", we adopted the recommended L4 product (WIND_GLO_PHY_L4_MY_012_006) to construct a new hourly 10-m vorticity dataset. Preliminary matching for January–April of 2001 indicates that ~90% of ERA5 vortices align with surface vorticity signatures. Full data processing is ongoing, as it will still take some time to download the data.. This approach will quantify ERA5's false alarm rate (e.g., Line 475)

Figure 1 exemplifies a matched case: the cyclone is embedded southeast of a synoptic-scale cold vortex moving southwestward from west of Svalbard to Greenland. The surface wind field, vorticity, and divergence distributions reveal a distinct frontal zone northwest of the cyclone, consistent with characteristics of typical PLs. Furthermore, the surface vorticity track aligns well with the 850 hPa vorticity track and exhibits greater smoothness, indicating that this dataset indeed provides finer dynamical features.

However, PL genesis mechanisms vary (Rasmussen & Turner 2003); surface vorticity validation may be insufficient for upper-level-triggered PLs (Blechschmidt et al. 2009; Yanase et al. 2016). Thus, we are integrating ERA5, scatterometer winds, AVHRR, and multi-source data to build a comprehensive validation framework.

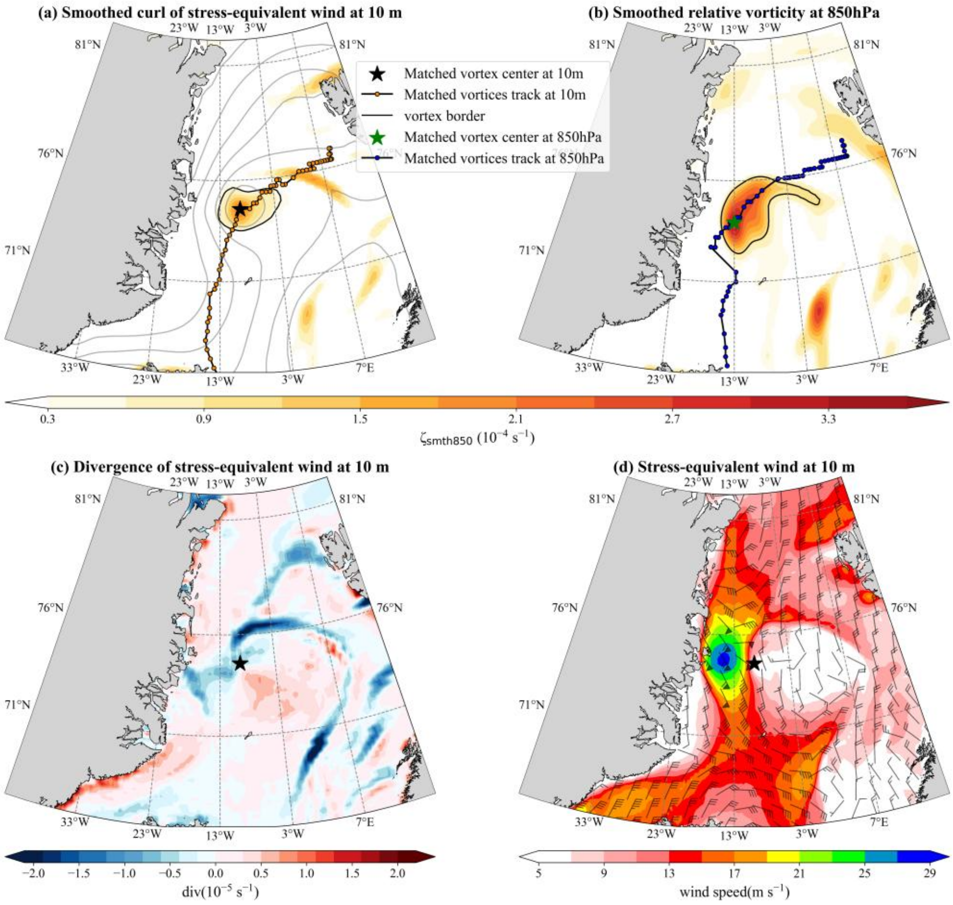


Figure 1 (a) 10 m relative vorticity from WIND_GLO_PHY_L4_MY_012_006, (c) divergence, (d) wind field, and (b) 850 hPa relative vorticity from ERA5. Uniform smoothing was applied with radii of (a) 45 km and (b) 60 km. Orange (10 m) and blue (850 hPa) dotted lines denote matched vortex tracks, while black and green stars mark their center positions at 21:00 UTC on 24 February 2001.

## 2. Clarification on AVHRR data availability

Regarding the low proportion of vortices exhibiting cyclonic cloud features in AVHRR (also noted by Reviewer #3), we quantified AVHRR availability for vortex points and tracks. The results show that:

After excluding vortex tracks with >60% land presence (~20% reduction), 47,167 tracks remained for AVHRR matching. Matching required: (1) full 200-km radius coverage for individual points, (2) ≥2 matched points within ±3h of peak vorticity and ≥6 points per track lifetime. Figure 2 shows wintertime (Nov-Apr) matching statistics: 43% of points and 61% of tracks matched on average. Only ~3% of matched tracks were incorporated into the IMPMCT dataset. This low inclusion rate stems from cloud obstruction, cloud-ice contrast limitations, temporal resolution constraints, and detection methodology (e.g., higher 2001 inclusion reflects meticulous manual identification, while 2023's lower rate resulted from post-publication incidental discoveries). Crucially, IMPMCT's cyclone proportion underestimates true PMC prevalence, as many low-cloud even no-cloud PMCs lack discernible features.

While AVHRR covers relatively few cases, our dataset aims to provide a multi-source, high-accuracy collection—particularly those with clear cloud features—to aid users in understanding these phenomena (e.g., for model studies of PL-related clouds).

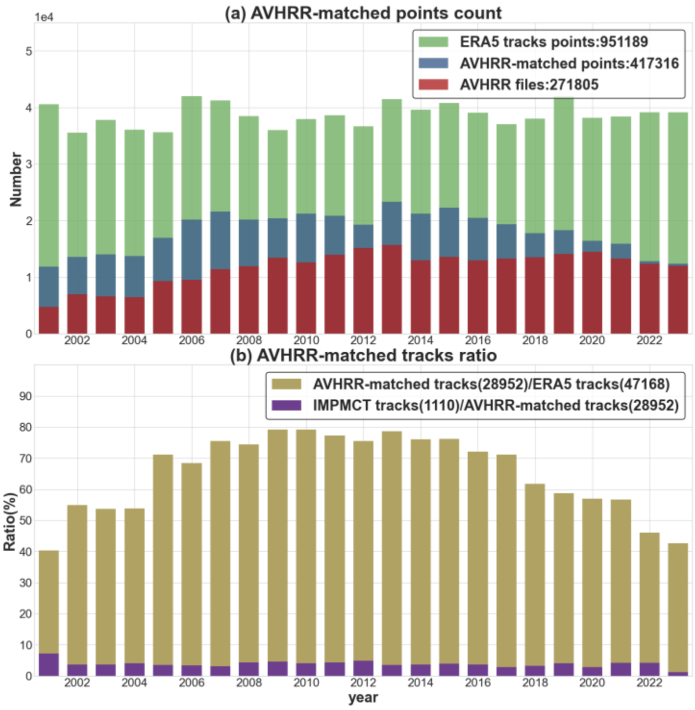


Figure 2: Annual winter (November-April) time series: (a) ERA5-derived vortex points (green), available AVHRR files (red), and AVHRR-matched vortex points (blue). (b) Ratio of AVHRR-matched vortex tracks to ERA5-derived tracks (yellow), and ratio of IMPMCT tracks to AVHRR-match tracks (purple). Note: Bars represent distinct categories (not stacked).

## 3. Explanation of mismatches with existing datasets

We thoroughly compared IMPMCT with existing PL datasets and added specific analyses of mismatches with the Rojo list and Stoll's PL tracks:

To further investigate mismatches between the reanalysis-based tracks and existing PL datasets, we implemented a nearest-point matching analysis (Table 2). A successful nearest-point match was recorded when a PL center from any list had at least one co-temporal vortex center within 120 km (60 km for the Stoll dataset). The track-level mismatches primarily stemmed from these point-level discrepancies. Crucially, the methodological differences between datasets explain the variation: While the Noer list derives from numerically modeled and AVHRR-assimilated hourly positions (typical of operational forecasting systems), the Rojo list relies on direct AVHRR identification at irregular temporal intervals, resulting in greater deviation from ERA5 representations. Furthermore, the Rojo compilation includes numerous secondary PL centers—features inherently less resolved by reanalysis data (Stoll, 2022)—whereas Noer focuses primarily on dominant PLs of operational significance. This distinction is clearly reflected in our analysis: Major PL centers (n=2,527) exhibited an 80% matching rate, while secondary centers (n=1,115) showed significantly lower alignment (54%), thereby reducing Rojo's overall match rate.

For the Stoll dataset, we additionally calculated a vortex matching rate (Table 2), counting a match when a Stoll center fell within the spatial domain of its nearest co-temporal vortex. This metric primarily addresses positional offsets caused by vorticity peak misalignment, which appears attributable to differences in smoothing algorithms (illustrated in Fig. S2). Our implementation seems to employ stronger uniform smoothing compared to Stoll's methodology, explaining why more lenient identification thresholds yield superior track matching with Stoll's dataset. This provides a new insight for applying the algorithm. Although the algorithm is not highly sensitive to the specific input vorticity fields, provided their grid spacing is sufficient to capture mesoscale vortices, the smoothing method also constitutes a significant factor contributing to variations in the identification results, alongside the identification parameters discussed in the sensitivity experiments (sect. 3.1.1). The smoothing approach should be specifically adapted to the assimilation noise and effective resolution of the input vorticity field. For instance, Gaussian smoothing may be preferable for model data with lower noise levels, as it more effectively preserves the positions of vortex cores.

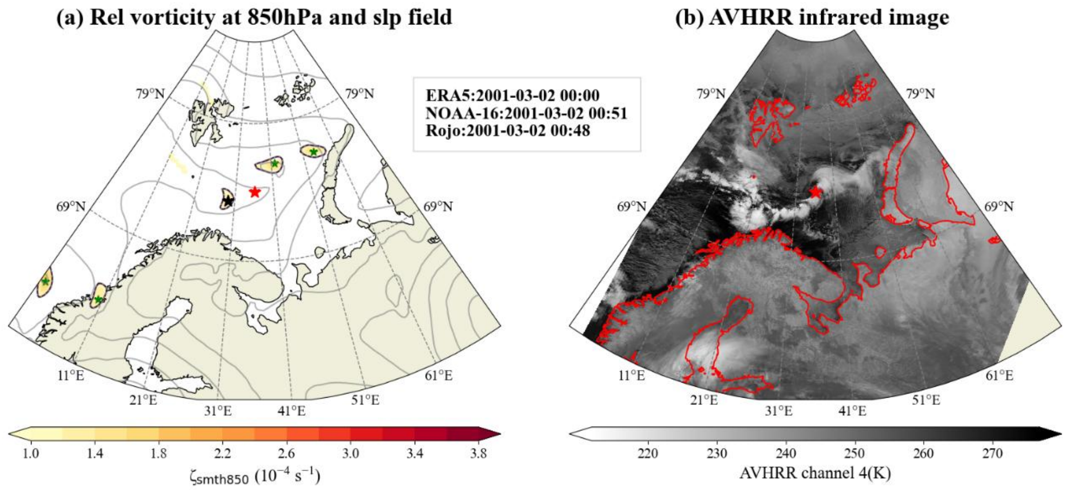


Figure 1: (a) 850 hPa relative vorticity field obtained by ERA5 data. (b) AVHRR infrared imagery concurrent with the time step in (a). The shading represents 850 hPa relative vorticity smoothed over a uniform 60 km radius and local vorticity maxima are identified by green star symbols, while regions enclosed by solid black contours denote their borders. The red star symbol marks a mismatched cyclone center from Rojo's PLs list, while the black star symbol marks the nearest local vorticity maxima from the cyclone center (227 km).

Table 2: the matching rate of the reanalysis-based track dataset for IMPMCT generation compared to other PL track datasets.

| PL tracks | Time period | Tracks in Nordic Sea (>3hr) | Track matched fraction(%) | Points | Nearest points matched fraction(%) | Vortex matched fraction(%) |
|---|---|---|---|---|---|---|
| Noer | 2002-2011 | 114 | 87 | 1670 | 85 | - |
| Rojo | 2000-2019 | 370 | 69 | 3642 | 71 | - |
| Stoll | 2000-2020 | 3179 | 93.68 | 75650 | 93 | 99 |

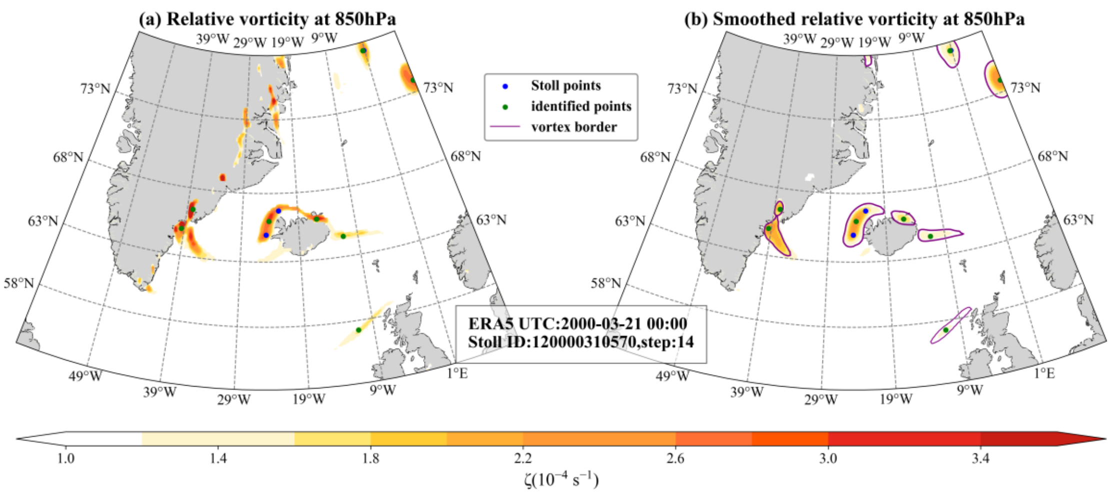


Figure S2: ERA5 850-hPa fields: (a) Relative vorticity. (b) Uniform 60-km smoothed vorticity. Vorticity field comparison showing center displacement between Stoll (blue points) and our detection (green points).

## 4. Comprehensive validation of the dataset

Since all tracks follow identical generation procedures, unvalidated PMC tracks share consistency with verified ones. Nevertheless, validation remains essential. We added an overall track characterization:

For most newly identified mesoscale cyclones not present in other PL lists, a direct validation approach involves applying objectively derived PL identification thresholds from prior studies to independently verify three key characteristics: polar environment, mesoscale size, and cyclonic intensity:

1) Polar-front criterion: Since PMCs are defined as mesoscale cyclones forming north of the polar front (Rasmussen and Turner, 2003), we employ two indicators to distinguish polar air masses from extratropical air masses: Tropopause Potential Temperature ($\theta_{trop}$) and the Maximum poleward value of 200 hPa wind speed ($U_{200,p}$). For each cyclone, we compute the track-averaged $\theta_{trop}$ averaged within a 250 km radius of the cyclone center and the track-averaged $U_{200,p}$ within ±1.0° great-circle distance longitude. Stoll (2022) defined $\theta_{trop} < 300.8$ K as indicative of polar air mass origin for PLs, effectively distinguishing them from extratropical cyclones with a high retention rate (76%) across subjective archives while preserving 90% of known PLs. Han and Ullrich (2025) used $U_{200,p}$ (WIND200MAX) < 25 m $s^{-1}$ to position PLs north of the polar jet, achieving an ~80% hit rate for PL classification with a miss rate of only 11.9%.
2) Mesoscale-size criterion: Vortex radius calculated from the vorticity field is used to exclude extratropical cyclones penetrating polar regions and large-scale frontal structures. In Stoll (2022), a maximum vortex diameter of 430 km (representing the 90th percentile across all PL lists) was applied, excluding approximately 24% of non-PL vortices. As we employ the same vorticity boundary threshold ($1.0\times10^{-4}$ $s^{-1}$) for vortex definition, this criterion remains valid for our dataset.
3) Cyclonic intensity criterion: An effective metric for characterizing mesoscale cyclone intensity is the Pressure anomaly ($p_{def}$), defined as the difference between the mean Sea Level Pressure (SLP) within a 110 km radius and the SLP at the cyclone centre ($p_{def} = \overline{SLP}_{110km} - SLP$). Stoll (2018) demonstrated that high $p_{def}$ values (90% of PLs > 0.4 hPa) highlight the anomalous intensity of the local low-pressure centre relative to its environment, signifying a steep pressure gradient near the core, indicative of small, deep low-pressure systems typical of PLs. We calculate the maximum $p_{def}$ based on the SLP centre for each vortex track. For tracks where no SLP centre is identified, $p_{def}$ is set to 0.

All discriminatory features for IMPMCT tracks are computed from ERA5 data. The quantiles of these features and the proportion of tracks meeting each criterion are presented in Table 4. Notably, 88.4% of tracks satisfy the polar-front criterion, 90% meet the mesoscale criterion, and 84% fulfill the cyclonic intensity criterion. It is important to note that these thresholds were developed based on the more intense subset of PLs. For the broader spectrum of PMCs, the thresholds for $\theta_{trop}$ and $p_{def}$ are inherently stricter, as they correspond to the cold air outbreak environments and stronger destructive potential typically associated with PLs. Consequently, the vast majority of tracks in IMPMCT satisfy these validation criteria. Furthermore, the hourly time series of these discriminatory features are included in the dataset as auxiliary information to facilitate targeted case selection for user research.

Table 1: Quantiles of discriminatory features and proportion of IMPMCT tracks meeting validation criteria.

| criterion | Track feature | percentage 50% | 75% | 90% | Proportion meeting the criterion (%) |
|---|---|---|---|---|---|
| **Polar front** $\theta_{trop} < 301$ K or $U_{200,p} < 25$ m s$^{-1}$ | $\theta_{trop}$ (K) | 298.9 | 304.1 | 310.0 | 88.4 |
| | $U_{200,p}$ (m s$^{-1}$) | 18.4 | 23.7 | 29.7 | |
| **Mesoscale** r < 215 km | r (km) | 137.1 | 176.9 | 213.5 | 90 |
| **Cyclonic** $p_{def} > 0.4$ hPa | $p_{def}$ (hPa) | 1.41 | 2.26 | 3.18 | 84 |

## Minor Revision

1. Line 95: These images are not so clear. In a): Could a PMC also be in (8,74), (36,77) or (36,77)? Why not? In b): Could the PL also be in (34,76)? Why not?

**Re:** Thank you for highlighting this. We acknowledge that cyclone center identification involves a degree of subjectivity. The revised manuscript now explicitly states: The centers of comma cloud and spiral cloud configurations were determined visually following Forbes and Lottes (1985), based on the characteristic curvature and convergence of cloud bands surrounding the circulation core in satellite imagery.

Consistency with Rojo's centers is high (90% within 60 km; see Minor Revision #28).

2. Line 96: The ERA5 grid distance is 31 km, hence good dynamical representation will at most be 150 to 300 km following typical dynamical closure procedures. Is that good enough for PL/PMCs?

**Re:** We appreciate this clarification. The text has been revised to: With the improved resolution of reanalysis datasets, their ability to capture PLs has progressively advanced (Laffineur et al., 2014; Smirnova and Golubkin, 2017) …

3. Line 109: Belmonte Rivas and Stoffelen also suggest some other reasons for poor PL/PMC representation in ERA5: lack of transient variability, lack of divergence, lack of resolution; it appears of interest to mention these aspects.

**Re:** We have incorporated your insight to better describe ERA5's shortcomings: However, ERA5 significantly underestimates near-surface wind speeds within PL-affected regions (Gurvich et al., 2022; Haakenstad et al., 2021), attributed in part to insufficient representation of transient wind variability, surface divergence, and unresolved mesoscale features (Belmonte Rivas and Stoffelen, 2019). This limits its ability to objectively capture PLs' high-wind characteristics, thereby introducing notable limitations.

4. Line 113: Having looked at many collocated IR and scatterometer wind vector fields (e.g., here below), I have some problem with the terminology "cyclonic cloud feature". Cyclonic cloud features might occur due to closed surface circulation (cyclone definition) indeed, while wind shear conditions may also generate clouds in circles shapes on the mesoscales. Moreover, a cyclone may also exist in an abundance or lack of clouds in which a cyclone is not recognized in an IR image. In the image below (from today) circular cloud patterns are present on the left hand side, while the streamlines of the vector winds do not coincide with the cloud streaks. On the other hand, a cyclonic wind feature appears on the right side

of the plot, but where high clouds cover the wind structure below. This is today's example, while examples of apparent IR cloud mismatch with ocean vector winds occur almost every day on this site, in particular at high latitude.

**Re:** Thank you for your suggestion. The cyclonic clouds identified in this work are primarily based on: 1) typical PMC cloud morphologies described in Forbes and Lottes (1985) and PL cases from the Noer list; 2) corresponding ERA5 vortex tracks (>6 hr); 3) statistical validation showing >80% of cyclones exhibit strong surface lows ($p_{def}$ > 0.4 hPa; Table 4). Therefore, the cyclonic clouds in the dataset do possess cyclonic characteristics. We further note: Crucially, IMPMCT's cyclone proportion underestimates true PMC prevalence, as many low-cloud even no-cloud PMCs lack discernible features.

5. Line 170: remove "resolution"; Skamarock (2004) defines effective resolution as 5-10 times the grid distance of an atmospheric circulation model, due to the necessary dynamical closure for numerical stability of the model.

**Re:** The term "resolution" has been removed as suggested.

6. Line 172: Note that in particular the initiation of PMCs and PLs in ERA5 is brought by wind scatterometers as can be observed in time sequences at https://scatterometer.knmi.nl/tile_prod/index.php. Hence ERA5 PMCs/PLs may be biased to the availability of the satellite data used, which could be problematic in time series analyses of PMCs/PLs. As readers may not be generally aware of this dependency, it is better to state it.

**Re:** We emphasize this dependency in the Discussion: It is noteworthy that while this study demonstrates ERA5 reanalysis data's enhanced capability in capturing PMCs and PLs, it does not reflect ERA5's predictive skill for such systems. This predictive capability should be evaluated by testing ERA5 background states in characterizing PLs/PMCs, thereby isolating the influence of real-time assimilated data—particularly scatterometer measurements (Furevik et al., 2015).

7. Line 173: with a spatial resolution -> on a spatial grid

**Re:** Revised to: "on a spatial grid".

8. Line 182: To refer to scatterometer accuracy, one may use Vogelzang and Stoffelen (2022).

**Re:** Added scatterometer accuracy citation (Vogelzang and Stoffelen 2022): These advanced instruments are specifically engineered to deliver accurate(e.g., ASCAT-A zonal/meridional wind component error standard deviations of ~0.37/0.51 m $s^{-1}$ and ASCAT-B of ~0.39/0.44 m $s^{-1}$, Vogelzang and Stoffelen 2022), high-resolution, continuous wind vector measurements under all weather conditions, offering comprehensive global coverage of near-surface wind patterns.

9. Line 192: ASCAT-A, -B and -C have been operational since 2007.

**Re:** Corrected ASCAT operational timeline: QuikSCAT operated from 1999 to 2009, whereas ASCAT start operating since 2007.

**10.** Line 197: with stable spatiotemporal resolution -> exploiting the evolving global observing system; I.e., not necessarily of stable spatiotemporal resolution effectively, since depending on the initialization of small scales by observations, when available.

**Re:** Revised to "exploiting the evolving global observing system": To establish a more comprehensive cyclone track dataset in the Nordic Seas, we first utilize ERA5 reanalysis data exploiting the evolving global observing system to obtain all vortex tracks.

**11.** Line 208: Scatterometers measure the surface wind vector field and hence curl and divergence. See, e.g., Belmonte Rivas and Stoffelen (2019). King et al. (2022) found that tropical divergence as measured by scatterometers is closely related to moist convection. Similarly, one would expect that cyclonic disturbances are very well depicted in curl and divergence. These are furthermore available at https://data.marine.copernicus.eu/product/WIND_GLO_PHY_L4_MY_012_006/description. It also provides hourly corrected ERA5 wind variables for reference. Why not put them in the database? They provide a stable reference over time as each instrument product does not change over time.

**Re:** Thank you for your suggestion. As outlined in Major Revision #1 regarding the process of adding hourly corrected ERA5 wind variables to the dataset, these dynamical characteristics will be included in the dataset once the data download is complete.

**12.** Line 232: The vorticity field appears noisy as I understand the text. Nevertheless, no observations exist to initialize 4D dynamical structures well on scales below 100 km over the ocean, hence 60-km filtering may not be too problematic. The noise may be due to the fact that you use analyses, rather than more consistent dynamical model fields, i.e., background (first guess) ERA5 data as in Belmonte Rivas and Stoffelen (2019) for example. Reanalyses fields are affected by the observations being assimilated, using spatial structure functions, which are posed as stream function and velocity potential "blobs", defined based on forecast ensemble statistics. These increments may not treat vorticity fields well and produce noise. Another reason may be in interpolation of the vorticity fields, but where no details are provided.

**Re:** Thank you for your correction. It is necessary to introduce a potential source of vorticity perturbations in the text. The original text has been revised to: A uniform 60-km smoothing radius is applied to hourly 850-hPa relative vorticity to disconnect weak continuity zones and eliminate minor perturbation maxima, which may arise from assimilation increments (Belmonte Rivas and Stoffelen, 2019).

**13.** Line 314: All steps appear rather ad hoc, but together they define a vortex isolation and data procedure. Moreover, it appears as a community procedure, as others elaborated similar procedures. Does the procedure work similarly well for other reanalyses, mesoscale models or the operational ECMWF analysis? To me, it appears tuned to the characteristics of your input ERA5 fields. Perhaps mention that other meteorological model fields may require further tuning of the vortex detection procedure.

**Re:** Thank you for your comment. This algorithm is indeed a general procedure, with specific code written by ourselves and made public in the data and code availability section. In our response to Reviewer #1, we supplemented sensitivity experiments on the algorithm parameters (https://doi.org/10.5194/essd-2025-186-AC1). In addition to the above parameters, the selection of smoothing algorithms also has a certain impact on the identification results. In

Major Revision #3, we added a description of the algorithm's applicability: This provides a new insight for applying the algorithm. Although the algorithm is not highly sensitive to the specific input vorticity fields, provided their grid spacing is sufficient to capture mesoscale vortices, the smoothing method also constitutes a significant factor contributing to variations in the identification results, alongside the identification parameters discussed in the sensitivity experiments (sect. 3.1.1). The smoothing approach should be specifically adapted to the assimilation noise and effective resolution of the input vorticity field. For instance, Gaussian smoothing may be preferable for model data with lower noise levels, as it more effectively preserves the positions of vortex cores.

**14.** Line 316: established -> constructed

**Re:** Term replaced as suggested.

**15.** Line 318: Terrain-induced flows are normally tied to the terrain and not to the wind, hence presumably they'd typically not produce vortex tracks according to your criteria?

**Re:** Thank you for your suggestion. Terrain disturbances may still generate PLs (Kristjánsson et al., 2011), but our work does not separately classify such polar lows. The original expression may have been inaccurate, so we have revised it to: … including not only cyclonic systems but also low-pressure troughs, and small-scale atmospheric disturbance.

**16.** Line 320: established -> comparison; recall that AVHRR are not a direct measurement of PMC, cf. comment 113.

**Re:** Revised to: To assess whether these vortices represent PMCs, AVHRR infrared imagery is used for comparative validation.

**17.** Line 455: The concept of environmental wind speed bis not clear. What is its use? The 10m wind vector around a moving vortex is rather variable, depending on steering flow and vortex strengths. The baroclinic nature of these high-latitude vortices makes their surface appearance usually asymmetrical. I can understand you'd like to capture this, but this is not clear from the text. Please clarify what relevant dynamical characteristics can be extracted. Fig. 11b appears a vortex interacting with land and hence surface winds are distorted?

**Re:** Thank you for your suggestion. The calculation of environmental wind speed is primarily used to capture larger-scale atmospheric motions near cyclones, such as frontal zones and large-scale strong winds during cold air outbreaks. In the framework for constructing surface cyclonic circulation characteristics outlined in Major Revision #1, the concept of environmental wind speed will be replaced by surface cyclonic vorticity.

**18.** Line 458: To first order, the destructive force goes with the third power of the wind speed, irrespective of it is generated by the environmental flow, vortex contribution or related to local convection, all count. In open sea, the waves, build by the wind, are of course very important as well, as the dimensions of the structures at sea may resonate with long and forceful waves.

**Re:** Thank you for your suggestion. It is true that wind speed is positively correlated with destructive force. However, determining whether such strong winds are driven by the cyclone itself is also a key issue. Nevertheless, as mentioned in Major Revision #1, using surface vorticity will be a good indicator to describe the intensity of the cyclone itself.

**19.** Line 470: The scatterometer section is rather poor as scatterometers, in particular ASCAT, reveal detailed dynamical PL characteristics. Wind vectors fields reveal the exact surface position, structure and divergence and curl and with high coverage. See also comment 208. Unfortunately, not much has been published on active satellite surface winds and PLs, while Furevik et al. (2015) provide some overview.

**Re:** We acknowledge this gap and highlight future use of L4 winds (Major Revision #1).

**20.** Line 475: You find many tracks that are not in AVHRR. Following the comment above, this could well be because the vortical structure is not well expressed in the clouds. Observed dynamics at the surface may prove a better way to verify these vortices. A problem here is that scatterometer winds are only consulted after a imperfect AVHRR filter, rather than before this filter. This can be done by exploiting collocated model and scatterometer data and their spatial gradients, which are available. When only one scatterometer is available (up to 2007), then track cannot be well verified, but every occasion a vortex appears in a scatterometer swath verification may be done. That would results in hits and misses of ERA5 vortices, which verify your product more substantially in my view.

**Re:** Thank you very much for your suggestion. We clearly recognize that due to the spatiotemporal resolution of AVHRR and the challenge of polar remote sensing, AVHRR-based identification methods will miss a large number of PMC cases. However, for existing case studies, cloud images remain essential indicators of cyclone development and position. Therefore, to ensure the comprehensiveness of tracks, we still hope to retain PMC tracks verified by AVHRR comparison. However, ASCAT/QuikSCAT-based identification methods show great potential in our view. Therefore, as shown in the process outlined in Major Revision #1, the matching results between near-surface vortices and ERA5 vortices will be added to the revised manuscript, and this method will be mentioned in Future Work.

**21.** Line 476: “measurable wind patterns”? My experience in scatterometry for PL/MPC is that tracking is very well feasible and measurable. I copy below a slide I show in nowcasting training using https://scatterometer.knmi.nl/tile_prod/index.php . For a description, see the figure above. Several things to note here: 1) Many scatterometer acquisitions exist over a day to verify both model dynamics (green arrows) and IR images (grey-scale). 2) IR clouds follow the dynamics seen at the surface, i.e., the dynamics produce clouds in upward motion and dissolve clouds in downward motion, i.e., the clouds follow the winds. 3) Initially, a through appears in the scatterometer winds below a cloud shield, where the green arrows are not informed by it initially. As scatterometer winds are assimilated at ECMWF the disturbance appears in the model data over the day. As mentioned earlier, L3 and L4 products are produced with scatterometer information, model information, incl. ERA5, and fields of spatial derivatives. These appear more ideal to “measure” model and, after collocation, AVHRR characteristics in PL/MPC than the rather unfavorable diagnostics presented here.

**Re:** Thank you very much for your suggestion. Major Revision #1 shows a surface vortex track obtained using the tracking algorithm, which matches the 850hPa vorticity track. This is indeed highly feasible and measurable. Its effectiveness will be verified after processing all surface wind field data.

**22.** Line 486: 3 hours implies three points, right? 50% in these cases implies only 2 of 3 points and 80% 3 of 3 points and 4 of 5 hits for longer tracks for example. It is clear that adding more lenient vortex criteria will improve apparent skill as the Stoll data set is fixed. It does not necessarily imply better performance though as Stoll. How much false tracks/points do you add?

**Re:** Thank you for your comment. Table 2 shows the proportion of nearest matching points between vortex centers and other cyclones. For the PL lists of Stoll and Noer, the matching rate of vortex points in the reanalysis track dataset is sufficiently high because their datasets are assimilated by models and have hourly resolution.

However, for the Rojo list, which only uses AVHRR for identification, the temporal resolution of cyclone tracks can be several hours. Therefore, for the Rojo list, three points are often more than six hours apart, and it is more likely that vortices and cyclone centers are far apart, as shown in Figure 1. Hence, a 50% matching ratio threshold is appropriate. More importantly, the selection of a 50% matching ratio and a 150km separation threshold is intentionally set to be the same as in Stoll (2022) to facilitate testing the effectiveness of the vortex algorithm.

**23.** Line 493: demonstrates? Clearly, wind variability is high in cold air outbreaks near the surface and upper air interaction more fierce. Allowing more noise in ERA5 vorticity or more lenient vortex criteria will reveal more tracks, but are they reliable? If some of them appear in the proximity of observed tracks, it appears insufficient to demonstrate capability. How many unverified tracks are produced (false alarms)? Could these accidentally be added to the hit

list? In that case skill is not enhanced, but rather PL/MPC noise is added.

**Re:** Thank you for your comment. The following aspects demonstrate that the increase in matching rate is not primarily due to increased matching of cyclones by transient disturbances:

1) As shown in Major Revision #3, our algorithm exhibits stronger uniform smoothing than Stoll, which suppresses noise to a considerable extent.
2) Only vortex tracks (>6hr) exceeding 60% of the reference PL trajectory's lifespan are included in the matching process. We apologize for not elaborating on the third point in the text. In the revised manuscript, this condition has been added: Furthermore, to avoid incidental matches of transient spurious tracks to PLs, only vortex tracks with lifespans exceeding 60% of the reference PL trajectory's duration are included in the matching process.

**24.** Line 495: So, ERA5 finds about 10 times more PLs/MPCs than the most extensive observational data set (Rojo). Is this noise? Looking at your AVHRR score, noise appears indeed manifest; 57,688 ERA5 vortex tracks, only 1,184 or 1 in 50 are confirmed. This may be related to the fact that AVHRR is a rather indirect measure of vortical activity, while you appear to appreciate the skills of AVHRR. What are the >90% misses in your data set? As these amounts appear rather overwhelming, it appears very relevant to understand their characteristics if these are used for geophysical analyses or trend analyses. The difference with Stoll's 3179 tracks from the same ERA5 is also rather overwhelming. What are the differences? I further understand less than 700 (only about 1%) remain for further comparison. I'm concerned what the other 99% represent?

**Re:** Thank you for your comment. In Major Revision #2, we show that the proportion of vortex tracks verifiable by AVHRR is approximately 61%, and due to the image resolution issues of AVHRR, the proportion of truly verifiable vortices is likely even lower. For tracks not included in the IMPMCT track set, in addition to the lack of AVHRR track verification, there are several possible reasons:

1) Extratropical types, which appear too uniformly bright or large in cloud images and do not conform to the common cloud appearance of PLs/PMCs, thus being excluded;

2) No corresponding cyclonic cloud appearance or insufficiently obvious cyclonic cloud features, which occurs in cases of cloud obstruction， or cloud-free cyclones formed due to low water vapor;

3) High separation from ERA5 vortex tracks, with only cyclone tracks that maintain good consistency with the movement of vortex tracks being retained.

The above three situations indicate that the tracks not selected by the dataset do not represent false alarms; in other words, we only strictly retain samples with clear and unambiguous cloud images, and exclude those that are unclear or cannot be fully confirmed.

Additionally, regarding the false alarm rate of the ERA5 vortex track set, we plan to verify it using wind scatterometer data as described in Major Revision #1, which will be the focus of our next manuscript revision.

**25.** Line 507: demonstrate that such discrepancies are not errors -> characterize such discrepancies; they are errors as ERA5 uniquely represents PLs/MPCs.

**Re:** Revised to: “characterize these discrepancies”**.** In Major Revision #3, we calculated the vortex matching rate with Stoll (99%), demonstrating that such mismatches are more likely caused by peak misalignment due to different smoothing algorithms.

**26.** Line 509: stable -> negligible

**Re:** Term replaced as suggested.

**27.** Line 512: remove “stable”; the choice of this word is a bit concerning, does it imply that you favor a smooth representation of disturbances? Spatial smoothing is applied, but it can obviously kill PLs/MPCs, which is a negative effect. If Stoll uses data from ERA5 that is less interpolated, it may in fact be a good thing that it represents more variability? Please elaborate in your manuscript.

**Re:** Thank you for your comment. It is indeed arbitrary to assume that a lower standard deviation indicates greater correctness. The word “stable” may imply a preference for smoothness, while greater variability may also be reasonable. In the revised manuscript, this description has been replaced with: Additionally, vortex property differences increase with distance, indicating that discrepancies between IMPMCT and Stoll tracks stem from differing identification thresholds. To further demonstrate that such discrepancies are not errors but reflect slightly different tracking paradigms, we calculate the standard deviation of vortex properties across three consecutive time steps for each track and computed track-wide averages. Consistently low-amplitude variations in vortex properties along a track suggest coherent feature identification by the respective method. Fig. 13 (b), (c) and (d) show the track-averaged local standard deviations of the three vortex properties for IMPMCT and Stoll datasets. Crucially, the magnitudes of these local variabilities in IMPMCT tracks are generally comparable to those in Stoll’s tracks. This alignment in variability indicates that the increasing property differences observed at larger separation distances arise from small-scale peak misalignments inherent to each method's detection logic, rather than representing erroneous identification by either approach. In fact, the IMPMCT variabilities are often slightly smoother, consistent with its specific algorithmic choices.

**28.** Line 532: Please indicate in the figure legend what percentage of the most favorable (matched) cases it represent. The non-matched cases are less detectable and probably have much less favorable verification.

**Re:** Thank you for your suggestion. Although we allow a 120 km separation, in fact, 90% of cyclones matched with Rojo are within 60 km, and this information has been added to the manuscript. For unmatched cyclones, in addition to cases where AVHRR data is unavailable, situations where ERA5 fails to capture them as shown in Major Revision #3 may also be reasons.

A total of 1432 cyclone centers from the Rojo list (corresponding to 140 cyclone tracks) were matched to the IMPMCT cyclone tracks. Notably, although the maximum allowable matching distance was set to 120 km, the 90th percentile of the matching distances was 55.97 km. This suggests minimal cyclone center identification errors when scan times could not be strictly aligned.

**29.** Line 541: "Despite" or "Due to"? Less favorable cases may not match well?

**Re:** Thank you for your suggestion. Owing to the inclusion of secondary PL centers (30% in Rojo; Major Revision #3), which exhibit lower detectability (54% match rate), overall consistency decreases.

**30.** Line 544: "reasonable"; you allow a 120 km separation and then one gets separations with a SDD of about 120 km, which implies little skill. Do you reason for little skill? Presumably, further work is needed to explain the lack of skill? Better explain to the users what further work would be appropriate in this discussion.

**Re:** Under the condition that the identification error of the center is low, such differences in cyclone scale are more likely due to inconsistent methods for measuring cyclone diameter. In the original text, we explain these differences: When identification errors of the cyclone center are low, the differences in diameter compared to the Rojo list stem not only from inconsistent measurement methods but also significantly from subjectivity. Due to the frequent presence of frontal cloud bands associated with cyclones, consistent measurement of the cyclone's long axis proves challenging. Furthermore, when a cyclone is adjacent to other cloud systems, its extent is often subject to ambiguity.

In addition, these characteristics are provided as reference quantities, and users can adjust them according to the given cloud image examples.

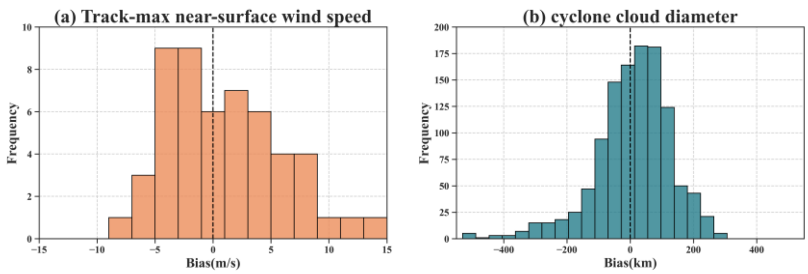


Figure 2: Frequency distribution of bias in (a) Track-max near-surface wind speed and (b) diameter between matched cyclones in the Rojo and IMPMCT datasets (Rojo minus IMPMCT). The cyclone diameter in IMPMCT is calculated as the average of the width and length of the bounding box enclosing the cyclone.

**31.** Line 559: How do you know what these cases are? They have not been verified, at least not in the manuscript. Could they not be numerical artefacts? Are they associated with real features or are these ERA5 simulated features?

**Re:** Thank you for your comment. Our tracks do include longer segments relative to Stoll's track set, which is due to more lenient thresholds. However, in our individual observations, we confirm that these segments are sufficiently closely connected; otherwise, there would definitely be abrupt changes in vortex properties, which would be reflected in the local standard deviation mentioned in Minor Revision #27. All tracks are matched with cyclone images. For track segments beyond the coverage of cyclone images, there seems to be no good ways to confirm their authenticity. After obtaining the matching of near-surface vorticity as shown in the process in Major Revision #1, these segments may be better verified.

**32.** Line 565: Is the point not how reliable ERA5 is to represent PLs and PMCs? One could test that using the cases where verification is available and determine and not yet used in ERA5 (by data assimilation). Therefore, testing ERA5 background states, winds are independent of any new observations, one could establish the capability of ERA5 to predict PLs and MPCs. Only after this, ERA5 can be used with confidence for associated geophysical studies in my view. Would you agree?

**Re:** Thank you for your suggestion. We agree that testing ERA5's ability to capture PLs and PMCs using background fields rather than assimilated fields can verify ERA5's predictive capability. As shown in Minor Revision #6, we emphasize this fact in the discussion section.

**33.** Line 580: As explained above, observations directly associated with PL/MPC dynamics may be further exploited to characterize these systems and the fidelity of reanalyses to represent them.

**Re:** Thank you for your suggestion. As shown in Major Revision #1, dynamical characteristics will be added to the IMPMCT dataset soon as an important supplement to surface circulation characteristics.

## Reference

Blechschmidt, A. -M., Bakan, S., and Graßl, H.: Large-scale atmospheric circulation patterns during polar low events over the nordic seas, J. Geophys. Res., 114, 2008JD010865, https://doi.org/10.1029/2008JD010865, 2009.

Yanase, W., Niino, H., Watanabe, S. I., Hodges, K., Zahn, M., Spengler, T., and Gurvich, I. A.: Climatology of polar lows over the sea of japan using the JRA-55 reanalysis, Journal of Climate, 29, 419–437, https://doi.org/10.1175/JCLI-D-15-0291.1, 2016.

Belmonte Rivas, M. and Stoffelen, A.: Characterizing ERA-interim and ERA5 surface wind biases using ASCAT, Ocean Sci., 15, 831–852, https://doi.org/10.5194/os-15-831-2019, 2019.

Furevik, B. R., Schyberg, H., Noer, G., Tveter, F., and Röhrs, J.: ASAR and ASCAT in polar low situations, Journal of Atmospheric and Oceanic Technology, 32, 783–792, https://doi.org/10.1175/JTECH-D-14-00154.1, 2015.

Forbes, G. S. and Lottes, W. D.: Classification of mesoscale vortices in polar airstreams and the influence of the large-scale environment on their evolutions, Tellus A, 37A, 132–155, https://doi.org/10.1111/j.1600-0870.1985.tb00276.x, 1985.

Kristjánsson, J. E., Thorsteinsson, S., Kolstad, E. W., and Blechschmidt, A.: Orographic influence of east greenland on a polar low over the denmark strait, Quart J Royal Meteoro

Soc, 137, 1773–1789, https://doi.org/10.1002/qj.831, 2011.
Stoll, P. J., Graversen, R. G., Noer, G., and Hodges, K.: An objective global climatology of polar lows based on reanalysis data, Quart J Royal Meteoro Soc, 144, 2099–2117, https://doi.org/10.1002/qj.3309, 2018.
Stoll, P. J., Spengler, T., Terpstra, A., and Graversen, R. G.: Polar lows – moist-baroclinic cyclones developing in four different vertical wind shear environments, Polar lows, 2021.