



Supplement of

IMPMCT: a dataset of Integrated Multi-source Polar Mesoscale Cyclone Tracks in the Nordic Seas

Runzhuo Fang et al.

Correspondence to: Jinfeng Ding (dingjinfeng@nudt.edu.cn)

The copyright of individual parts of the supplement might differ from the article licence.

S1. Performance of YOLOv8-obb-pose model

We specifically provide the YOLOv8-obb-pose model weights to enable other researchers to replicate the model, alongside a small validation dataset for performance evaluation and to facilitate its implementation (Fang, 2025). The validation dataset comprises 1334 Vortex-Centered Infrared (VCI) images from the Nordic sea region spanning the years 2001 and 2023, with 500 cyclone-containing images of the dataset. None of these images are involved in any training process of the model. The model's performance is evaluated using three common metrics—precision, recall, and mean average precision (mAP)—for both keypoint detection and oriented bounding box prediction tasks on this validation set.

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{predictions} \quad (S1)$$

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{ground\ truths} \quad (S2)$$

$$L_{oks} = \exp\left(-\frac{d^2}{2s^2k^2}\right) \quad (S3)$$

In Eq. (S1) and (S2), for the oriented bounding box task, a true-positive (TP) is defined as a correctly predicted class when the predicted bounding box overlaps sufficiently with the ground-truth bounding box, with the overlap degree measured by the Intersection over Union (IoU , *Intersection/Union* between the predicted and ground-truth boxes). For the keypoint detection task, a TP occurs when the predicted keypoint is sufficiently close to the ground-truth keypoint, with proximity quantified by the Object Keypoint Similarity (OKS) metric defined in Eq. (S3) (Maji et al., 2022), in which d denotes the Euclidean distance between the predicted and true keypoint locations, k represents the keypoint importance weight, and s corresponds to the area of the object's bounding box. Additionally, false positives (FP) are instances where predicted bounding boxes (or keypoint locations) fail to meet the required IoU (or L_{oks}) threshold with any ground-truth object bounding. False negatives (FN) occur when ground-truth objects (bounding boxes/keypoints) are undetected or fail to meet the IoU/L_{oks} threshold with predictions.

The model's performance on this validation dataset with IoU (L_{oks}) threshold set as 0.5 and confidence threshold set as 0.3 is summarized in Table S1. The mAP50-95 metric represents the mean average precision across IoU or L_{oks} thresholds from 0.5 to 0.95, calculated by dynamically optimizing confidence thresholds to balance precision and recall at each threshold level. The model achieved notable precision and recall values in both prediction tasks. The mAP50-95 for bounding box prediction is significantly lower than mAP50, indicating that the model performs better on well-defined cloud feature samples compared to ambiguous ones. In contrast, the keypoint prediction for cyclone center locations demonstrates consistently high accuracy regardless of sample complexity, suggesting that boundary box prediction for cyclones is more challenging than localizing their centers. Overall, comma-shaped clouds exhibit significantly higher prediction accuracy than spiral clouds. This discrepancy may stem from class imbalance in the dataset or the model's incomplete ability to distinguish between spiral and comma-shaped cloud structures, implying substantial room for improvement in the

35 model's generalization capabilities.

Table S1 The Yolov8-obb-pose model's performance on this validation dataset

Class	Instances	BOX(P	R	mAP50	mAP50 -95)	Pose(P	R	mAP50	mAP50 -95)
comma	285	0.94	0.87	0.92	0.59	0.97	0.90	0.94	0.94
spiral	215	0.83	0.88	0.85	0.45	0.88	0.96	0.93	0.93
all	500	0.88	0.88	0.88	0.52	0.92	0.93	0.94	0.94

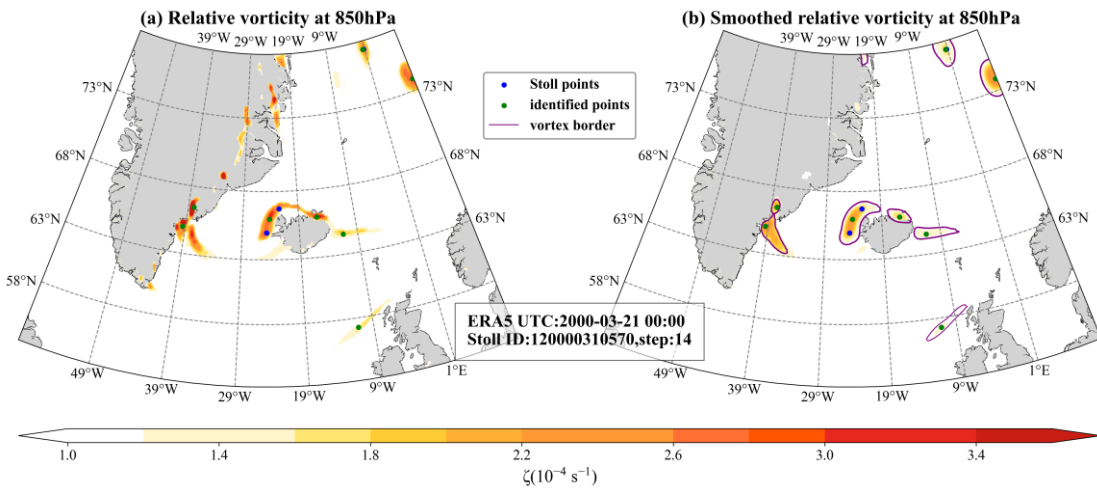
S2. The matching rate of the reanalysis-based track dataset with different vortex identification parameters compared to other PL track datasets.

40 Three sensitivity experiments were conducted with the following configurations:

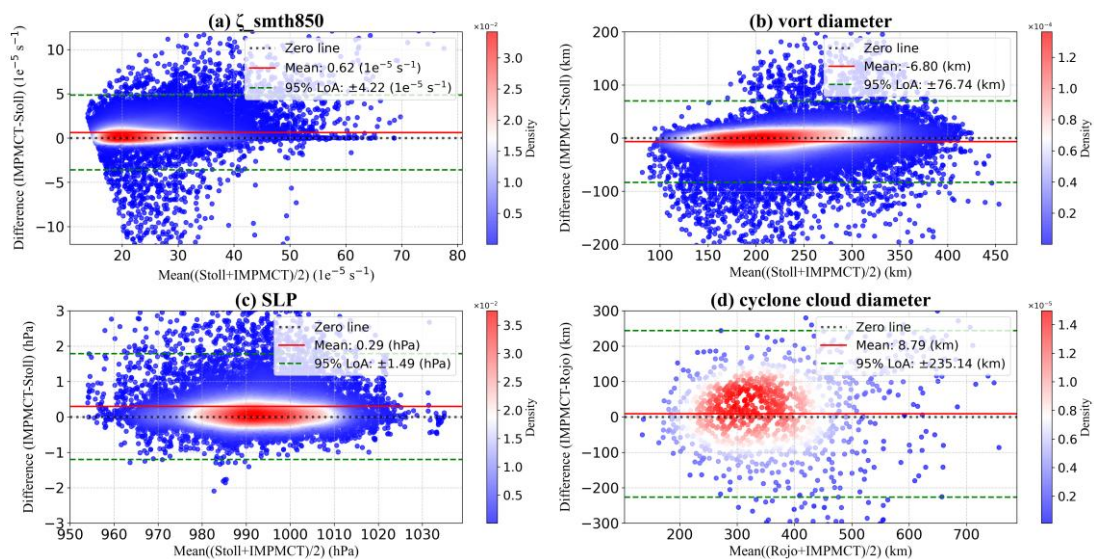
- 1) Experiment a (lenient thresholds): $\zeta_{max0} = 1.2 \times 10^{-4} \text{ s}^{-1}$, $\zeta_{min0} = 1.0 \times 10^{-4} \text{ s}^{-1}$, $\gamma = 0.15$;
- 2) Experiment b (intermediate thresholds): $\zeta_{max0} = 1.2 \times 10^{-4} \text{ s}^{-1}$, $\zeta_{min0} = 1.0 \times 10^{-4} \text{ s}^{-1}$, $\gamma = 0.25$;
- 3) Experiment c (strict thresholds, following Stoll et al. 2021): $\zeta_{max0} = 1.5 \times 10^{-4} \text{ s}^{-1}$, $\zeta_{min0} = 1.2 \times 10^{-4} \text{ s}^{-1}$, $\gamma = 0.25$

45 **Table S2 the matching rate of the reanalysis-based track dataset with different vortex identification parameters compared to other PL track datasets.**

Experiment	Track counts	Matching rate(%) with		
		Stoll	Rojo	Noer
a	59975	93.68	69.73	87.72
b	52708	92.04	68.11	86.84
c	33622	87.39	61.35	80.70



50 **Figure S1 Comparasion 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)**



55 **Figure S2 Bland-Altman analysis of Property Differences Between IMPMCT and Other PL list.**(a) 850 hPa relative vorticity, (b) vortex equivalent diameter, (c) SLP, (d) cyclone cloud diameter. The x-axis represents the mean property value of IMPMCT and the other dataset; the y-axis represents the difference in properties (IMPMCT minus PL list). Point color indicates Gaussian kernel density. The black dashed line denotes the zero line. The red solid line indicates the mean difference of the sample properties. The upper and lower green dashed boundaries represent the limits of agreement (LoA), defined as the mean difference ± 1.96 standard deviations of the differences. *Note: Differences for properties (a), (b), and (c) are comparisons between IMPMCT and the Stoll (2022) dataset, whereas (d) uses the Rojo list. The difference analysis for track-max near-surface wind speed is not shown due to insufficient sample size.

References

- 65 Fang, R.: validation dataset for yolov8-obbb-pose cyclone-detect-model, Zenodo[data set], <https://doi.org/10.5281/zenodo.15119534>, 2025
- Maji, D., Nagori, S., Mathew, M., and Poddar, D.: YOLO-pose: Enhancing YOLO for multi person pose estimation using object keypoint similarity loss, 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 2636–2645, <https://doi.org/10.1109/CVPRW56347.2022.00297>, 2022.
- 70 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.