UC Davis
UC Davis Electronic Theses and Dissertations

## Title

The System for Classification of Low-Pressure Systems (SyCLoPS): Development and Applications in Tropical Cyclone Research

## Permalink

https://escholarship.org/uc/item/77n265n9

## ISBN

9798193262480

## Author

Han, Yushan

## Publication Date

2026-09-08

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

The System for Classification of Low-Pressure Systems (SyCLoPS): Development and Applications in Tropical Cyclone Research

By

YUSHAN HAN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Atmospheric Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Paul A. Ullrich, Chair

Terrence R. Nathan

Matthew R. Igel

Committee in Charge

2026

© Yushan Han, 2026. All rights reserved

*This dissertation is dedicated to Typhoon Dujuan (2003), Typhoon Mankghut (2018), and Hurricane Francine (2024), who sparked my passion for weather and interest in tropical cyclone research.*

## Acknowledgements

I would like to express my first and foremost gratitude to my doctoral advisor, Paul Ullrich, who has always been supportive of my research interests and responsive to every question and concern throughout my Ph.D. journey at UC Davis. Without Paul's mentorship, I would not have developed the technical skills that have proven valuable within the atmospheric science community and beyond, nor would I have built the academic network that has benefited my current and future collaborations.

I would also like to thank my wife for her continuous support, care, and encouragement throughout my daily life, especially during my first year when I was still uncertain about the direction of my future research. I am also grateful to my parents, who have always supported my enthusiasm for meteorology since I was four years old. Their encouragement ultimately guided me toward my academic path in atmospheric science.

I would like to extend my thanks to the remaining members of my dissertation committee: Terry Nathan, Colin Zarzycki, Matt Igel, and Adele Igel. Their thoughtful feedback on my research proposal and manuscripts, as well as their dedication to the courses I took at UC Davis and Penn State, have been fundamental to the completion of my Ph.D.

My sincere thanks also go to the Chinese weather enthusiast community that I grew up with. Many preliminary ideas in my research were inspired by discussions and exchanges within this community. More importantly, I would like to thank the students involved in meteorology clubs across senior high schools in Shenzhen. Promoting atmospheric science and connecting weather and climate with the next generation were among the fundamental motivations behind my establishing of the first student meteorology club in Shenzhen and pursuing my academic career in atmospheric science.

Lastly, I would like to acknowledge my funding sources: the PCMDI project funded by the United States Department of Energy (DOE) and the 2025 Walter and Margaret Milton Award. Their generous support has enabled my Ph.D. research and facilitated collaborations with scientists around the world.

## Abstract

Objective identification of low-pressure systems (LPSs) in large-scale datasets, such as reanalyses and Earth system model outputs, is essential for understanding their physical characteristics, evaluating their representation in numerical models, and assessing their historical variability and projected future changes. Among all types of LPSs, tropical cyclones (TCs) have received the greatest attention because of their destructive impacts throughout their life cycles. Despite substantial advances in objective feature tracking, challenges remain in accurately identifying TC features in climate datasets, which can lead to biases in TC climatology and model evaluation. Contemporary TC research spans the entire TC life cycle—from precursor disturbances and genesis to post-tropical evolution. To facilitate such research using reanalyses and Earth system model outputs, an objective framework capable of tracking and classifying TCs throughout all stages of their life cycles is required.

To address these challenges, Chapter 2 of this dissertation presents the System for Classification of Low-Pressure Systems (SyCLoPS), an objective LPS tracking and classification framework built upon the TempestExtremes (TE) software. Unlike previous tracking algorithms, SyCLoPS first identifies every detectable LPS signal at a prescribed temporal frequency using fundamental atmospheric fields and subsequently classifies each detected system at both the node and track levels. The classification criteria are based on physically motivated definitions developed using ERA5 and are further refined with machine-learning techniques. Evaluation demonstrates that SyCLoPS improves tropical cyclone detection relative to the previous TE-based algorithm, increasing the hit rate (HR) by 2% while reducing the false alarm rate (FAR) by 6%. SyCLoPS also achieves 95% agreement with the subjective classifications contained in the International Best Track Archive for Climate Stewardship (IBTrACS). Furthermore, the framework is applicable to a broad range of LPS types, including monsoonal systems, polar lows, subtropical storms, extratropical cyclones, and thermal lows.

Chapters 3 and 4 demonstrate the application of SyCLoPS to advance the understanding of TC representation in both climate models and atmospheric reanalyses. Chapter 3 applies SyCLoPS to historical HighResMIP simulations to investigate the coupled evolution of TC intensity and size throughout the storm life cycle. By introducing an objective TC size retrieval algorithm that estimates TC quadrantal wind radii comparable to those in IBTrACS, the study enables the first direct comparison of TC size between climate models and observations. The results show that, despite increasing model resolution, most HighResMIP models and ERA5 remain unable to reproduce the observed intensity–size relationship, with simulated storms generally being too large for a given intensity and failing to capture the population of compact, intense TCs. Interestingly, the multi-model ensemble reproduces some aspects of TC climatology, as well as differences in intensity and size characteristics between ENSO phases, comparably to—and in some cases, better than—ERA5. Motivated by these findings, Chapter 4 evaluates the newly released JRA-3Q reanalysis and shows that it substantially outperforms ERA5 in TC representation, achieving an approximately 90% HR while reducing the FAR to only 6%. Error attribution analyses indicate that JRA-3Q's superior performance primarily results from its advanced TC bogussing scheme, whereas the degradation of ERA5 in earlier decades is largely attributable to increasing model error under sparser observational constraints. Together, these studies demonstrate the robustness and adaptability of SyCLoPS across diverse datasets while providing new insights into the strengths and limitations of current climate models and reanalyses in representing TCs.

# Table of Contents

*Acknowledgements* ..... *iii*

*Abstract* ..... *iv*

*Chapter 1. Introduction* ..... *1*

1.1 Motivation and Dissertation Overview ..... 1

1.2 References ..... 7

*Chapter 2. The System for Classification of Low-Pressure Systems (SyCLoPS): An All-in-One Objective Framework for Large-scale Datasets* ..... *9*

2.1 Abstract ..... 9

2.2 Introduction ..... 9

2.2 Data ..... 14

2.3 The Overall Workflow ..... 16

2.4 Justification for Classification Conditions ..... 21

2.5 Criteria Optimization for High-impact LPS Detection ..... 29

2.6 Results and Applications ..... 35

2.7 Summary and Conclusions ..... 49

2.8 Appendixes ..... 53

2.9 Supplements ..... 59

2.10 Software and Published Datasets ..... 69

2.11 References ..... 72

*Chapter 3. The Evolution of Tropical Cyclone Intensity and Size in Observations and HighResMIP Models ..... 77*

3.1 Abstract ..... 77

3.2 Introduction ..... 78

3.3 Data, Tracking, and TC Size Estimation ..... 80

3.4 Comparing ZU and SyCLoPS TC Trackers ..... 90

3.5 TC Intensity and Size in Observations ..... 94

3.6 TC Intensity and Size in HighResMIP Models ..... 97

3.7 Discussion ..... 105

3.8 Summary and Conclusions ..... 107

3.9 Appendixes ..... 110

3.10 Supplements ..... 114

3.11 Software and Published Datasets ..... 120

3.12 References ..... 121

*Chapter 4. Improvements in Tropical Cyclone Representation in JRA-3Q over ERA5 . 128*

4.1 Abstract ..... 128

4.2 Introduction ..... 128

4.3 Data and Methods ..... 130

4.4 Results ..... 132

4.5 Discussion and Conclusions ..... 139

4.6 Supplements ..... 144

4.7 References ..... 146

*Chapter 5. Conclusions* ..... *150*

5.1 Other Ongoing Applications of SyCLoPS ..... 150

5.2 Future Open Science Questions ..... 151

5.3 References ..... 154

# Chapter 1. Introduction

## 1.1 Motivation and Dissertation Overview

Cyclones, or low-pressure systems (LPSs), are mesoscale to large-scale atmospheric features capable of producing a wide range of hazards, including flooding, high winds, storm surge, and heatwaves. To identify and study their climatology, short-term forecasts, and future risks, it is essential to objectively identify, quantify, and characterize cyclones in reanalysis products, physical weather and climate models, and AI-driven models using automated algorithms.

Williamson (1981) was one of the first study to use an automated tracking algorithm to identify cyclonic features in ECMWF 10-day forecasts and their corresponding verification analysis. Since then, more than a dozen automated algorithms for detecting extratropical cyclones have been developed. However, differences in algorithmic design and the thresholds used to identify and filter cyclones have led to substantial spread and uncertainty in their climatological representation and life-cycle characteristics (Neu et al., 2013). Furthermore, evaluating the performance of extratropical cyclone trackers remains challenging because no consistent observational dataset exists to serve as a reference.

In contrast, tropical cyclones (TCs) are relatively sparse and exhibit well-defined dynamical and structural characteristics, making them well suited for objective feature detection. Moreover, the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al., 2010) provides a comprehensive observational archive of subjectively analyzed TC tracks compiled from meteorological agencies worldwide. The recent availability of high-resolution reanalyses and Earth system models has made objective identification of features as small as TCs increasingly practical. Consequently, several tracking algorithms have been developed specifically for TCs, including CNRM (Chauvin et al., 2006), OWZ (Tory et al., 2013), TRACK (Hodges et al., 2017; Strachan et al., 2013), and ZU (Ullrich et al., 2021; Zarzycki & Ullrich, 2017). Bourdin et al., (2022) compared these four TC trackers and found that ZU, which is built upon the state-of-the-art C++ feature detection software TempestExtremes (TE; (Ullrich et al., 2021; Ullrich & Zarzycki, 2017), achieved the highest detection skill as measured by hit

rate (HR) and false alarm rate (FAR). A more recent purely machine learning-based TC detection algorithm (Accarino et al., 2023) likewise achieved lower detection performance than the ZU tracker. Nevertheless, ZU still produces a considerable number of false alarms. To reduce the FAR, a post-processing procedure based on the Cyclone Phase Space (CPS; Hart, 2003) is applied to filter and truncate tracks, retaining only the tropical portion of each track, at an expense of lower HR.

This limitation reflects a broader deficiency of contemporary low-pressure system (LPS) tracking algorithms. Most existing trackers—including recent algorithms developed for monsoonal systems (Vishnu et al., 2020) and polar lows (Stoll, 2022; Stoll et al., 2018)—are designed to identify only a single type of LPS. Consequently, many filtering criteria and post-processing procedures are required to prevent contamination by other cyclone types. These restrictions often truncate tracks prematurely or exclude portions of an LPS life cycle, limiting the ability to investigate transitions among different cyclone types and the complete evolution of LPSs.

In Chapter 2, we present the System for Classification of Low-Pressure Systems (SyCLoPS), a new objective low-pressure system (LPS) tracking and classification framework built upon TE and developed using the fifth-generation European Centre for Medium-Range Weather Forecasts reanalysis (ERA5; Hersbach et al., 2020). ERA5 was selected as the development dataset because of its high horizontal resolution. Unlike previous objective trackers that target individual LPS types, SyCLoPS detects all identifiable surface LPSs and classifies them simultaneously. The framework assigns labels to each detected LPS node and subsequently classifies complete tracks for four high-impact LPS types: TCs, monsoonal systems, subtropical storms, and polar lows. The classification criteria are derived primarily from physically based definitions and meteorological intuition and are further refined through mathematical optimization, sensitivity analyses, and machine-learning techniques. Although machine learning is used to optimize classification thresholds, the framework contains no black-box components and remains fully interpretable, physically motivated, and readily adaptable to other reanalyses and Earth system model outputs. In subsequent studies and SyCLoPS software updates, we demonstrate that only

minimal tuning is required to apply SyCLoPS across multiple datasets with different horizontal resolutions.

Relative to the widely used ZU tracker, SyCLoPS substantially improves TC detection skill by reducing false alarms while simultaneously increasing the number of correctly detected storms. This improvement is partly attributable to a new method for distinguishing tropical from non-tropical systems without relying on the presence of a warm-core structure. We show that warm-core structure is neither unique to tropical systems nor persistent throughout the life cycle of all tropical LPSs. More importantly, SyCLoPS identifies and classifies the complete life cycle of surface LPSs, from genesis to lysis. For TCs, this includes objective identification of precursor disturbances, mature tropical stages, and post-tropical evolution, enabling accurate determination of transitions such as extratropical transition with minimal timing bias. SyCLoPS also provides the first globally applicable objective framework for identifying tropical-like cyclones—including subtropical storms and polar lows—that exhibit TC-like structures outside the tropics. The paper on the development of SyCLoPS has been published in the Journal of Geophysical Research: Atmospheres under the DOI: 10.1029/2024JD041287, and was selected as the editor's highlight for the Eos.

In Chapter 3, we apply SyCLoPS to investigate TCs in historical HighResMIP simulations (Haarsma et al., 2016). Unlike most previous studies, which primarily focus on TC frequency climatology and intensity representation in climate models, this chapter examines the coupled evolution of TC intensity and size throughout the complete TC life cycle in the two best-observed basins: the North Atlantic and the western North Pacific. The motivation for this study stems in part from the contrasting characteristics of storms during the 2024 North Atlantic hurricane season, which demonstrated that TCs with different combinations of intensity and size can produce markedly different societal impacts. For example, Hurricane *Helene* was relatively large and moderately intense, producing widespread inland flooding and extensive damage through heavy precipitation, whereas Hurricane *Milton*, which followed shortly afterward, was considerably more compact but much more intense, resulting in localized yet

severe wind damage. These contrasting evolutionary pathways highlight the importance of understanding both TC intensity and size when assessing storm hazards and developing disaster mitigation strategies.

Inspired by Musgrave et al. (2012), we draw an analogy between TC evolution and stellar evolution by categorizing storms into four structural regimes based on their intensity and size: super giants, representing large and intense storms near their peak intensity; strong dwarfs, representing compact but intense storms; weak giants, representing large but relatively weak storms; and weak dwarfs, representing storms that are both weak and compact. These categories are identified objectively using SyCLoPS together with a newly developed TC size retrieval algorithm that estimates quadrantal near-surface wind radii directly comparable to those reported in IBTrACS. This approach enables, for the first time, direct and objective comparisons of TC size between climate models and observations.

SyCLoPS plays several important roles in this study. First, as demonstrated for ERA5, SyCLoPS substantially reduces false alarms in HighResMIP simulations, resulting in more accurate assessments of TC representation and a cleaner sample for analyses of TC intensity and size. Second, SyCLoPS objectively identifies the end of the tropical phase of each storm, removing post-tropical segments from analyses of TC evolution. Third, the framework classifies TC tracks according to both their precursor disturbance types—including tropical depressions, monsoon depressions, and tropical transition—and their eventual outcomes, including landfall, extratropical transition, and dissipation over the ocean. We demonstrate that these track types are associated with distinct pathways of intensity–size co-evolution throughout the TC life cycle. An important enhancement to SyCLoPS introduced in this study is a resolution-dependent classification criterion in which one threshold is automatically adjusted for optimal TC detection skill according to the horizontal grid spacing of the input dataset. This adaptive threshold minimizes the reduction in detectable TCs arising solely from differences in model grid resolution and enables fairer comparisons among models with varying horizontal resolutions.

Perhaps the most important finding of Chapter 3 is that, despite horizontal resolutions approaching 0.25°, current climate models—including ERA5—still struggle to realistically reproduce the observed relationship between TC intensity and size. Relative to IBTrACS observations, simulated storms

generally become too large for a given intensity, and every HighResMIP model except MRI-S (Mizuta et al., 2019) fails to reproduce the entire population of strong dwarf TCs. Because MRI-S has a horizontal resolution comparable to several other HighResMIP models, its superior performance likely reflects differences in model physics and parameterization schemes rather than spatial resolution alone. We further find that coupled model simulations do not improve the simulated intensity–size relationship relative to their atmosphere-only counterparts.

Despite these deficiencies, the HighResMIP multi-model ensemble performs comparably to, and in some respects better than, ERA5 in reproducing TC track density, the composition of TC track types, and the observed differences in TC intensity and size between El Niño-Southern Oscillation (ENSO) phases. Although ERA5 is constrained by assimilated observations, it does not consistently outperform free-running climate models in representing both TC intensity and the intensity-size relationship. These findings naturally motivate the following question addressed in Chapter 4: To what extent can ERA5 be relied upon for TC research, and how can the representation of TCs in modern atmospheric reanalyses be further improved? Chapter 3 has gone through the first round of revision in Journal of Geophysical Research: Atmospheres.

With the release of the latest Japanese reanalysis, JRA-3Q (Kosaka et al., 2024), Chapter 4 evaluates its performance in representing TCs relative to ERA5, the reanalysis on which SyCLoPS was originally developed and trained. We find that JRA-3Q substantially improves TC representation, achieving an exceptionally high HR of approximately 90% while maintaining a FAR of only 6%, compared with 78% and 11%, respectively, for ERA5. More importantly, JRA-3Q exhibits much more stable TC detection skill over time in the well-observed basins (e.g., North Atlantic, Eastern Pacific, and western Pacific), whereas the performance of ERA5 deteriorates markedly in earlier decades, particularly during the pre-satellite era. Although previous studies have suggested using reanalyses to mitigate observational uncertainties and interagency inconsistencies in best-track datasets such as IBTrACS, our results demonstrate that caution is warranted when selecting a reanalysis product. Objective TC detection

in reanalyses can itself introduce artificial long-term trends in TC activity if the quality of the reanalysis varies over time.

The HR and FAR reported here likely represent the highest TC detection skill yet documented for a global atmospheric reanalysis. They also substantially exceed the performance reported by a recent machine-learning-based TC detection algorithm applied to JRA-3Q (Park et al., 2026). Together with the validation results presented in Chapter 3, these findings demonstrate the robustness and generalizability of SyCLoPS across multiple reanalysis datasets. An important update in SyCLoPS introduced for this study is the identification of a distinct class of "hybrid TCs" within the classification framework, which effectively addresses the previously low detection rate over the North Atlantic (Bourdin et al., 2022; Ye et al., 2026). Incorporating this class significantly increases the HR in both JRA-3Q and ERA5 while producing only a negligible increase in false alarms.

The substantial improvement achieved by JRA-3Q naturally raises the question: Why does JRA-3Q represent TCs more accurately than ERA5? To address this question, we decompose the sources of missed detections and false alarms in both reanalyses and investigates the contribution of TC bogussing to the superior performance of JRA-3Q. The results indicate that the TC bogussing scheme plays a dominant role in improving vortex initialization and accounts for a considerably larger fraction of true-positive detections (hits) in JRA-3Q than in ERA5 during the period when both reanalyses assimilated TC bogus data using different methodologies. In contrast, the decline in ERA5 detection skill—particularly the increase in FAR that is absent in JRA-3Q—can largely be attributed to model deficiencies becoming more apparent as observational constraints become sparser in earlier decades. We further show that uncertainties in IBTrACS, especially in the less-observed basins and during the pre-satellite era, also contribute to apparent detection errors. Consequently, the remarkably low FAR of JRA-3Q should not be interpreted as evidence that the model underlying JRA-3Q successfully eliminates all spurious TC vortices, as uncertainties in the observational record remain substantial, particularly for coastal and inland TCs in the North Indian Ocean. We hope this work can provide guidance for the appropriate use and

selection of reanalysis products in TC research and motivates future improvements in reanalysis systems.

Chapter 4 is now under preparation for publication in Geophysical Research Letters.

## 1.2 References


Accarino, G., Donno, D., Immorlano, F., Elia, D., & Aloisio, G. (2023). An Ensemble Machine Learning Approach for Tropical Cyclone Localization and Tracking From ERA5 Reanalysis Data. Earth and Space Science, 10(11), e2023EA003106. https://doi.org/10.1029/2023EA003106

Balaguru, K., Chang, C.-C., Leung, L. R., Ullrich, P. A., Han, Y., Rice, J. R., Hagos, S., Chavas, D., Taraphdar, S., Harrop, B., Sun, N., & Judi, D. R. (2026). Recent Tropical Cyclone Outer Size Increases in the Western North Atlantic. Earth's Future, 14(2), e2025EF007162. https://doi.org/10.1029/2025EF007162

Bourdin, S. (2026). RABTracks V1 (Version V1) [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.20426510

Bourdin, S., Fromang, S., Dulac, W., Cattiaux, J., & Chauvin, F. (2022). Intercomparison of four algorithms for detecting tropical cyclones using ERA5. Geoscientific Model Development, 15(17), 6759–6786. https://doi.org/10.5194/gmd-15-6759-2022

Chauvin, F., Royer, J.-F., & Déqué, M. (2006). Response of hurricane-type vortices to global warming as simulated by ARPEGE-Climat at high resolution. Climate Dynamics, 27(4), 377–399. https://doi.org/10.1007/s00382-006-0135-7

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., … von Storch, J.-S. (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. Geoscientific Model Development, 9(11), 4185–4208. https://doi.org/10.5194/gmd-9-4185-2016

Hart, R. E. (2003). A Cyclone Phase Space Derived from Thermal Wind and Thermal Asymmetry. Monthly Weather Review, 131(4), 585–616. https://doi.org/10.1175/1520-0493(2003)131%3C0585:ACPSDF%3E2.0.CO;2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hodges, K., Cobb, A., & Vidale, P. L. (2017). How Well Are Tropical Cyclones Represented in Reanalysis Datasets? Journal of Climate, 30(14), 5243–5264. https://doi.org/10.1175/JCLI-D-16-0557.1

Kosaka, Y., Kobayashi, S., Harada, Y., Kobayashi, C., Naoe, H., Yoshimoto, K., Harada, M., Goto, N., Chiba, J., Miyaoka, K., Sekiguchi, R., Deushi, M., Kamahori, H., Nakaegawa, T., Tanaka, T. Y., Tokuhiro, T., Sato, Y., Matsushita, Y., & Onogi, K. (2024). The JRA-3Q Reanalysis. Journal of the Meteorological Society of Japan. Ser. II, 102(1), 49–109. https://doi.org/10.2151/jmsj.2024-004

Mizuta, R., Yoshimura, H., Ose, T., Hosaka, M., & Yukimoto, S. (2019). MRI MRI-AGCM3-2-S model output prepared for CMIP6 HighResMIP [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1625

Musgrave, K. D., Taft, R. K., Vigh, J. L., McNoldy, B. D., & Schubert, W. H. (2012). Time evolution of the intensity and size of tropical cyclones. Journal of Advances in Modeling Earth Systems, 4(3). https://doi.org/10.1029/2011MS000104

Nordeng, T. E., & Rasmussen, E. A. (1992). A most beautiful polar low. A case study of a polar low development in the Bear Island region. Tellus A, 44(2), 81–99. https://doi.org/10.1034/j.1600-0870.1992.00001.x

Park, D.-S. R., Kim, D., Ko, H.-Y., Kim, H.-S., Cha, D.-H., Chang, M., Min, S.-K., Kwon, M., & Park, T.-W. (2026). TC2 ver. 1.0: An Objective Hybrid Tracker and Classifier for Tropical Cyclones version 1.0. Climate and Earth system modeling. https://doi.org/10.5194/egusphere-2026-3285

Stoll, P. J. (2022). A global climatology of polar lows investigated for local differences and wind-shear environments. Weather and Climate Dynamics, 3(2), 483–504. https://doi.org/10.5194/wcd-3-483-2022

Stoll, P. J., Graversen, R. G., Noer, G., & Hodges, K. (2018). An objective global climatology of polar lows based on reanalysis data. Quarterly Journal of the Royal Meteorological Society, 144(716), 2099–2117. https://doi.org/10.1002/qj.3309

Strachan, J., Vidale, P. L., Hodges, K., Roberts, M., & Demory, M.-E. (2013). Investigating Global Tropical Cyclone Activity with a Hierarchy of AGCMs: The Role of Model Resolution. 26(1), 133–152. https://doi.org/10.1175/JCLI-D-12-00012.1

Tory, K. J., Chand, S. S., Dare, R. A., & McBride, J. L. (2013). The Development and Assessment of a Model-, Grid-, and Basin-Independent Tropical Cyclone Detection Scheme. Journal of Climate, 26(15), 5493–5507. https://doi.org/10.1175/JCLI-D-12-00510.1

Ullrich, P. A., & Zarzycki, C. M. (2017). TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. Geoscientific Model Development, 10(3), 1069–1090. https://doi.org/10.5194/gmd-10-1069-2017

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., & Reed, K. A. (2021). TempestExtremes v2.1: A community framework for feature detection, tracking, and analysis in large datasets. Geoscientific Model Development, 14(8), 5023–5048. https://doi.org/10.5194/gmd-14-5023-2021

Vishnu, S., Boos, W. R., Ullrich, P. A., & O'Brien, T. A. (2020). Assessing Historical Variability of South Asian Monsoon Lows and Depressions With an Optimized Tracking Algorithm. Journal of Geophysical Research: Atmospheres, 125(15), e2020JD032977. https://doi.org/10.1029/2020JD032977

Williamson, D. L. (1981). Storm track representation and verification. Tellus, 33(6), 513–530. https://doi.org/10.1111/j.2153-3490.1981.tb01778.x

Ye, G., Leung, J. C.-H., Dong, W., Toumi, R., Xu, J., Li, W., Qian, W., Kong, H., & Zhang, B. (2026). A Reanalysis-Based Global Tropical Cyclone Tracks Dataset for the Twentieth Century (RGTracks-20C). Earth System Science Data, 18(5), 3587–3607. https://doi.org/10.5194/essd-18-3587-2026

Zarzycki, C. M., & Ullrich, P. A. (2017). Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. Geophysical Research Letters, 44(2), 1141–1149. https://doi.org/10.1002/2016GL071606

# Chapter 2. The System for Classification of Low-Pressure Systems (SyCLoPS): An All-in-One Objective Framework for Large-scale Datasets

## 2.1 Abstract

We propose the first unified objective framework (SyCLoPS) for detecting and clas-sifying all types of low-pressure systems (LPSs) in a given dataset. We use the state-of-the-art automated feature tracking software TempestExtremes (TE) to detect and track LPS features globally in ERA5 and compute 16 parameters from commonly found at-mospheric variables for classification. A Python classifier is implemented to classify all LPSs at once. The framework assigns 16 different labels (classes) to each LPS data point and designates four different types of high-impact LPS tracks, including tracks of trop-ical cyclone (TC), monsoonal system, subtropical storm and polar low. The classifica-tion process involves disentangling high-altitude and drier LPSs, differentiating tropical and non-tropical LPSs using novel criteria, and optimizing for the detection of the four types of high-impact LPS. A comparison of our labels with those in the International Best Track Archive for Climate Stewardship (IBTrACS) revealed an overall accuracy of 95% in distinguishing between tropical systems, extratropical cyclones, and disturbances. SyCLoPS produces a better TC detection skill compared to the previous algorithms, high-lighted by an approximately 6% reduction in the false alarm rate compared to the pre-vious TE algorithm. The vertical cross section composite of the four types of high-impact LPS we detect each shows distinct structural characteristics. Finally, we demonstrate that SyCLoPS is valuable for investigating various aspects of LPSs in climate data, such as the evolution of a single LPS track, patterns of LPS frequencies, and precipitation or wind influence associated with a particular LPS class.

## 2.2 Introduction

In response to the growing need for advanced impacts-relevant model and climate data analysis, objective feature detection has emerged as a key tool for detecting and tracking various meteorological

features in large-scale datasets. Researchers give considerable attention to detecting and tracking low-pressure systems (LPSs), or cyclones, which are often drivers for high-impact weather including high winds and extreme precipitation. Some more significant LPSs, such as tropical cyclones (TCs), monsoon lows (MLs) or monsoon depressions (MDs), subtropical cyclones (SCs), and extratropical cyclones (EXs), are commonly tracked using specialized tracking algorithms in reanalysis and climate model outputs to derive their climatology and perform climate projections (e.g., Guishard et al., 2009; Neu et al., 2013; Hurley & Boos, 2015; Roberts et al., 2020). Tropical-like cyclones (TLCs) refer to those storms in the subtropics and the polar regions that share many similarities with TCs, including Mediterranean hurricanes and polar lows (PLs). They are capable of producing significant coastal hazards (Toomey et al., 2022), but more rigorous automated tracking has only occurred relatively recently because of advances in model resolution and observations (e.g., Stoll et al., 2018; Zhang et al., 2021; Stoll, 2022; Flaounas et al., 2023).

TempestExtremes (TE; Ullrich & Zarzycki, 2017; Ullrich et al., 2021) is an all-inclusive, state-of-the-art automated Lagrangian feature tracking software package. It is designed to robustly and efficiently detect, track, and analyze any nodal or areal features in largescale datasets with user-friendly command lines and using parallelized C++. TE has been optimized for TC detection using geopotential thickness and mean sea level pressure closed contour criteria (Zarzycki & Ullrich, 2017). Bourdin et al. (2022) found that TC detection using TE outperforms other methods for the reanalysis dataset ERA5 (Hersbach et al., 2020). Vishnu et al. (2020) used TE to track monsoonal systems (MSs), which includes MLs and MDs, in the North Indian Ocean across different reanalysis products, and observed high success rates. TE's detector has also been combined with the cyclone phase space (CPS) of Hart (2003), which classifies storms based on thermal wind and thermal asymmetry parameters. For example, Zarzycki et al. (2017) used TE and CPS to track both TCs and post-TCs/EXs based on LPSs' thermal structure, and detect extratropical transition. Zhang et al. (2021) used a similar approach to detect Mediterranean hurricanes.

Although TE's algorithms are powerful for LPS detection, when applied standalone it generally needs to adopt strict criteria to avoid mischaracterizing features. Unfortunately, this issue is unavoidable when using any specialized method, and controls like seasonality, topographical masks, and latitudinal bounds are required to avoid polluting the dataset with incorrect detections (i.e., false alarms). For example, TE's TC detection algorithm enforces a 50º highest latitude and a 150 $m^2 s^{-2}$ highest surface geopotential limit on detected LPS tracks. Another popular TC track algorithm TRACK requires TC tracks to start between 30°S and 30°N (K. Hodges et al., 2017). Vishnu et al. (2020) tracked MSs only between June and September and in parts of the North Indian Ocean, further requiring detected LPSs to maintain at least 24 hours offshore. Stoll (2022) tracked PLs by first excluding tracks outside of the 30-80° band of both hemispheres and including features only over defined open water regions. Since the aforementioned algorithms are all designed for single-type LPS detection, they also need to filter other types of LPS tracks or other types of LPS nodes (data points) in a detected track that don't meet the specified criteria and thresholds. With these limitations, features tracked by the existing algorithms often end or break abruptly and are shorter in length than analogs in manually tracked datasets. Consequently, information from the complete LPS lifespan is not available. While the CPS approach can classify the thermal structure evolution of an LPS throughout its lifetime, it is not designed for broader LPS classification, as different types of LPS often share similar thermal structures. For example, mature TCs, some post-TCs, EXs experiencing warm seclusion and some subtropical or hybrid cyclones can all be categorized as shallow or deep symmetric warm-core systems according to the CPS (Hart, 2003). Conversely, cold-core systems are ubiquitous across latitudes, as evidenced by the prevalence of low-level cold cores in many weak tropical systems (Reed et al., 1977; Hopsch et al., 2010; Hunt et al., 2016). Hence, warm/cold core criteria are likely insufficient to effectively classify global LPSs.

Instead of trying to develop multiple specialized trackers, we propose a new objective framework, called the System for Classification of Low-Pressure Systems (SyCLoPS), to detect, track, and classify all detectable LPSs worldwide at once, without any spatial or seasonal limitations. We test our framework in ERA5, and focus exclusively on surface-level LPSs (so upper-level disturbances and lows are out of our

scope). The detection and tracking are completed using TE commands, and the classification is done in a separate Python classifier that assigns 16 different types of LPS labels/classes (TC, ML, EX, SC, PL, etc.) and four types of high-impact LPS tracks, which are TC Track, MS Track, and two types of TLC Tracks (subtropical TLC and PL tracks). The classification process is based primarily on conventional definitions and physical (dynamical/thermodynamic) nature of different LPSs, and is simplified to the extent possible. The atmospheric variables used for classification are commonly found in reanalysis and climate model outputs and the parameters used to classify LPSs can be directly computed by TE for each detected LPS node. Basic machine learning techniques and mathematical optimization are used to refine our thresholds against archives of observed and subjectively identified LPSs. The resulting framework involves only thresholds on basic meteorological fields and includes no "black box" elements.

The new framework produces considerably longer LPS tracks because of the low detection threshold we choose. The labeled LPS nodes are also directly comparable to the subjective LPS status (labels) in the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010). Labeled tracks are also comparable to other subjective TC, MS, and TLC track catalogs. When compared to these catalogs, the framework maintains or improves on the LPS detection skill of the specialized trackers (under a variety of metrics) without implementation of the aforementioned restrictions. For example, other TC detection frameworks will often pick up some stronger warm-core extratropical or subtropical systems that were not recorded in IBTrACS even if latitudinal bounds were enforced (Bourdin et al., 2022). The new algorithm addresses this problem, leading to a noticeable increase in TC detection skill without further post-processing.

The manuscript is structured as follows. Chpater 2.2 summarizes all the datasets we use to verify our classification thresholds and quantify LPS detection skill. Chapter 2.3-2.5 explain the general detection and classification processes and justify each of the conditions we set for classification. In Chapter 2.6, we include some highlights from the classified LPS catalog and discuss potential applications of the SyCLoPS framework. Chapter 2.7 concludes the paper and addresses known limitations. For the readers' convenience, Table 2.1 summarized all the initialisms and definitions of LPS

types (classes) and detection skill metrics we use in this article. Appendixes and supplementary can be found in Chapter 2.8 and 2.9. A description of the SyCLoPS software available on GitHub is presented in Chapter 2.10.

| Initialism | Full Term | Definition |
|---|---|---|
| HAL | High-altitude Low | LPSs found at high altitudes without a warm core |
| THL | Thermal low | Shallow systems featuring a dry and warm lower core |
| HATHL | High-altitude Thermal Low | LPSs found at high altitudes with a warm core |
| DOTHL | Deep (Orographic) Thermal Low | Non-shallow LPSs featuring a dry and warm lower core driven by topography |
| TC | Tropical Cyclone | LPSs that would be named in IBTrACS |
| TD | Tropical Depression | Defined in Chapter 2.4.2 |
| TLO | Tropical Low | Non-shallow tropical systems that fall short of TD requirements |
| MD | Monsoon Depression | TDs developing in monsoonal environment. Also denoted as TD(MD) |
| ML | Monsoon Low | TLOs developing in monsoonal environment. Also denoted as TLO(ML) |
| MS | Monsoonal System | Monsoon LPSs (MDs plus MLs) |
| TLC | Tropical-Like Cyclone | Non-tropical LPSs that resemble TCs |
| SS (STLC) | Subtropical Storm (Subtropical Tropical-Like Cyclone) | Type of TLC. Defined in Chapter 2.2 |
| PL (PTLC) | Polar Low (Polar Tropical-Like Cyclone) | Type of TLC. Defined in Chapter 2.2 |
| SC | Subtropical Cyclone | Defined in Chapter 2.4.3 |
| EX | Extratropical Cyclone | Most typical non-tropical cyclones |

| | | |
|---|---|---|
| DS | Disturbance | Shallow LPSs or waves with weak surface circulations. DSD, DST and DSE are dry, tropical and extratropical DSs |
| QS | Quasi-stationary | LPSs that stay relatively localized |
| HR/FAR/IR | Hit Rate/False Alarm Rate/Infrequency Rate | Defined in Chapter 2.5.1-2.5.3 |
| HRMFAR | Hit Rate Minus False Alarm Rate | TC detection skill metric (Chapter 2.5.1) |
| CSI | Critical Success Index | MS detection skill metric (Chapter 2.5.2) |
| HRMIR | Hit Rate Minus Infrequency Rate | TLC detection skill metric (Chapter 2.5.3) |

**Table 2.1**. Initialism table for LPS types and detection skill metrics.

### 2.2 Data

We detect, track, and classify LPSs from 1979 to 2022 in ERA5 at 3-hourly frequency. ERA5 is used for developing SyCLoPS since we include detection of small or mesoscale features like early-stage TCs and TLCs, which require reanalysis with finer data resolution. Although SyCLoPS is developed using ERA5, it uses a small number of commonly output meteorological fields and so is intended to be applicable to other global or regional meteorological datasets. A list of ERA5 variables used in SyCLoPS is given in Appendix 2A. Since climate model data often have a different format and frequency, and may be missing some of the ERA5 variables used to develop SyCLoPS, we perform adaptation experiments on SyCLoPS using alternative ERA5 fields to demonstrate how SyCLoPS could be adapted to climate model outputs (Appendix 2B). The results show that SyCLoPS can be slightly modified to adapt to a variety of data formats while maintaining good performance.

By design, SyCLoPS uses TE's closed contour criteria to compute many of its classification parameters since the closed contour criteria is largely insensitive to model resolution (Ullrich & Zarzycki, 2017). TE's closed contour criteria uses graph search to ensure that all paths along the unstructured grid from a field's local maximum (or minimum) decrease (or increase) before reaching a specified great-

circle distance (GCD). For example, a positive closed contour in mean sea level pressure (MSLP) with delta MSLP (ΔMSLP) uses graph search to ensure that all paths along the unstructured grid from a local minimum (MSLP0) pass through a node with value greater than or equal to MSLP0+ ΔMSLP before reaching the specified great-circle distance (GCD).

Four types of high-impact LPS we detect, which includes TCs, MSs and two types of TLCs, are verified against four subjective datasets to evaluate detection skill. TCs are verified against the widely used IBTrACS dataset. Specifically, we verify against 3510 main-type tracks from 1979-2021 inclusive that meet the minimum requirement for a TC to be considered formed and named by global agencies—with peak wind speeds of at least 34 knots (17.5 m $s^{-1}$). Hence, TC detection in SyCLoPS follows the same definition as the agencies. We refer to this subset of IBTrACS as IB-TC. For MSs, we use the Sikka archive Sikka (2006) for verification, which provides manually-identified North Indian Ocean monsoonal system positions on historic surface weather maps at daily frequency, digitized by Hurley & Boos (2015).

In SyCLoPS, we regard subtropical storms (SSs) and polar lows (PLs) as TLCs that develop in the subtropics and the polar regions. Hence, they can also be denoted as subtropical TLCs (STLCs) and polar TLCs (PTLCs). Few subjective datasets are available for these two types of TLCs. SSs are regarded as the more intense subtropical or hybrid cyclones that carry at least gale-force surface sustained winds (National Hurricane Center, n.d.; Guishard et al., 2009). Mediterranean TLCs (hurricanes) may be viewed as being among global SSs, and they are perhaps most well studied, with relatively more observational data available. However, existing subjective datasets for those storms are still very incomplete. To reconstruct a more credible subjective archive for Mediterranean TLCs (representing SSs), we obtain the subjectively tracked data from Flaounas et al. (2023), wherein trained meteorologists identify and track Mediterranean cyclones, including TLCs and other noticeable systems in the region, using ERA5's mean sea level pressure (MSLP) field. We then use the cyclone information table provided in Flaounas et al. (2023) and two online documentation sources to select tracks that can be potentially classified as SSs (as defined above), while avoiding weaker systems. We also observe that IBTrACS contains a few subtropical storm records, which agencies tend to record when they have the potential to transform into

TCs. However, these records are largely incomplete and can often be confused with weaker TCs, so we do not separate them from the TC archive and include them in the SS archive.

We verify PL detection against the well-known and widely used STARS (Sea Surface Temperature and Altimeter Synergy for Improved Forecasting of Polar Lows) dataset (Noer et al., 2011). This subjective PL dataset is constructed based on the frequently cited PL definition from Rasmussen & Turner (2003)— "A polar low is a small, but fairly intense maritime cyclone that forms poleward of the main baroclinic zone (the polar front or other major baroclinic zone). The horizontal scale of the polar low is approximately between 200 and 1000 km and surface winds near or above gale force." STARS contains 185 polar lows subjectively identified by the Norwegian Meteorological Institute from 2002-2011 in parts of the Nordic Seas, based on synoptic-scale analysis in their weather model and satellite infrared data. There exists other less popular subjective PL datasets that record substantially more PLs, including many LPS sub-centers in multi-center PL cases, found in the same region using different satellite data (Stoll, 2022). This implies there is ambiguity in the PL definition, likely aggravated by a lack of near-surface observations for accurate evaluations. It should also be noted that a model/data resolution as coarse as ERA5 may simply be incapable of "seeing" some small or multi-centered TLC cases, regardless of the detection criteria applied. For example, Stoll (2022) reported that only 61-81% of PLs in the existing PL datasets have a matched LPS track in ERA5.

In our objectively tracked LPS dataset, we match our tracks to the tracks in each subjective dataset using different matching algorithms, described in Text 2.S1 in the Supplements (Chapter 2.9). To avoid misclassification of MSs, we also use an objectively tracked North Atlantic easterly wave (EW) dataset (Lawton et al., 2022) to construct a corresponding surface LPS dataset (also see Text 2.S1). These matched TC, MS, SS, PL, and EW datasets are later used in the classification and data analysis process.

### 2.3 The Overall Workflow

MSLP is the starting point for our LPS node detection. Although prior research has also employed 850 hPa vorticity or streamfunction fields to detect monsoon LPSs and other cyclones (e.g., K.

I. Hodges, 1994; Hurley & Boos, 2015; Vishnu et al., 2020), we consider MSLP a preferable variable for two reasons: (1) local minima/maxima of MSLP are widely used in meteorological agency operations to locate surface pressure systems, including LPSs in IBTrACS; and (2) there is a global agency consensus on the definition of MSLP (Knapp et al., 2010). Further, low-level vorticity or streamfunction data may not be directly available in many datasets when intersecting with the surface, and they have opposite signs for cyclonic systems in the Northern and Southern Hemispheres, which could lead to tracks across equatorial regions becoming disconnected if they are not detected and tracked twice using different signs. These factors all contribute to the computational burden of detecting LPSs globally without spatial limitations.

MSLP signals may be generally weaker in the tropics, which makes it difficult to differentiate some weaker LPSs associated with tropical waves. To capture these features in the tracker, we have aimed for the lowest reasonable detection threshold. With this in mind, an LPS node (center) is flagged at a local minimum MSLP if the MSLP value is lower than its surrounding by at least 10 Pa (0.1 hPa) within 5.5° GCD from the node (a 10-Pa delta positive closed contour criterion). Although we find this is an extremely low threshold, very weak EWs (particularly those over land) that don't have the minimal surface closed MSLP contour will not be detected.

The first step taken by TE is pointwise detection of LPS nodes. Besides the MSLP closed contour criterion, we also require that nodal candidates located within 6º GCD of each other merge into a single node, preserving the one with the lowest MSLP value, which aligns with the specification in Zarzycki & Ullrich (2017). The second step in TE connects consecutive nodes in time and form LPS tracks, which requires tuning argument specifications in TE to meet the following criteria:

1. LPSs must at least sustain for a time span of 18 hours to avoid too many weak, short-lived, diurnal lows.
2. Nodal candidates in two consecutive time steps must be located within 4º GCD distance of each other. This specification is chosen based on the fact that the translation speed of the fastest EXs

rarely go over 40 m $s^{-1}$, or about 140 km per hour (Bernhardt & DeGaetano, 2012; Lodise et al., 2022).

3. A maximum allowable gap of 12 hours within a track is implemented. A longer gap time is not preferred because it may include too many weak diurnal lows.
4. To exclude negligible LPSs, a track must contain at least five time steps during which the MSLP at the LPS node is 100 Pa lower than its surrounding (within a 5.5° GCD from the node).

A total of 7,781,105 data points (nodes) and 379,301 distinct tracks are identified in the 44-year period from 1979 to 2022. The whole tracking process takes about two hours using two NERSC Perlmutter nodes (each with 128 threads). During the first two steps, TE also conveniently computes 15 parameters (data columns) from input variable fields and outputs them to each node in the detected LPS tracks. In the third step of TE, we prepare the data for computing the LPS size parameter (LPSAREA) by running additional TE commands. After completing TE procedures, a Python classifier script is implemented to first compute LPSAREA and form the input (LPS) catalog for classification, structured in accordance with our catalog column table in Appendix 2C. The input catalog includes a list of 16 parameters that are used throughout the classification process explained below. The Python classifier then classifies each LPS node based on conditions on those input parameters and assigns them labels from the 16 distinct classes, following the workflow illustrated in the accompanying flowchart (Fig. 2.1). We call the output of this classification workflow the classified (LPS) catalog. The structure of this catalog can also be found in Appendix 2C. Please refer to Text 2.S5 for a brief explanation for the selection of the specifications in the parameters used in the input catalog. The required TE and Python script files, both LPS catalogs, and a manual for using SyCLoPS data are all uploaded to an open-source repository (see Chapter 2.10).

Most of the criteria chosen for classification of these systems are largely insensitive to climate change because they are based on conventional definitions (the nature of LPSs) and subjective datasets. Namely, a named tropical cyclone will always have a deep warm core, develop in a low-shear

environment, and satisfy a certain intensity threshold in all climates. Other important criteria that are potentially more sensitive to global warming are justified in the next section.

Note that we use the LPS catalogs mainly for the purpose of testing our classification conditions for classifying all 16 LPS classes, with an emphasis on TCs (since they are the most influential and well-documented type of LPS). Hence, the provided catalogs may not be the most comprehensive global LPS dataset for certain features, as we filter out numerous very weak systems and do not include the smallest and most short-lived systems that are likely to be poorly represented in ERA5 data. Our detection specifications should be enough for general high-impact LPSs; however, some very small and short-lived but detectable high-impact systems, such as certain PLs, can be missed because of the rather large node merging distance (6° GCD), the 18-hour minimum time requirement, or the 3-hourly detection rate. For these cases, users may select their own regions and LPS features of interest and run TE commands with alternative specifications before performing classification. We provide one such example TE command alternation for detecting PLs in the Nordic Seas in Text 2.S7.

Figure 2.1 shows the LPS classification flowchart. The chart flows from top to bottom through five major branches to disentangle each major class of LPS. Boxes with red texts indicate the final LPS classification full names and short labels. Details of the conditions applied in the flowchart can be found in Appendix 2D. The first branch (green) is the high-altitude branch, where we apply the high-altitude condition to extract those LPSs with a surface elevation (ZS) higher than the 850 hPa geopotential height (Z850; typically around 1500 m) from the input catalog, given that most of the more influential LPSs occur at a lower altitude. Although not as important, some of these high-altitude lows can be major contributors to precipitation over or near plateaus (Tucker, 1999; Li et al., 2019). The second branch (yellow) is the dry branch. In this branch, we segregate those LPSs that hardly produce any precipitation due to their dry low-level circulation. This generally includes thermally-driven shallow thermal/heat low systems, which can affect local cold fronts and heat waves (Reeder et al., 2000; Spengler et al., 2005). Third is the tropical branch (blue), which contains several impactful features such as TCs, MDs and MLs. At this level, the remaining unclassified LPSs will be determined to be tropical or non-tropical based on

the tropical condition described in Chapter 2.4.2. The fourth branch (purple) is the extratropical branch, where we segregate TLCs (SSs and PLs) and then differentiate SCs from the most typical EXs. Details of the conditions used in the classification process will be discussed further in the next section.

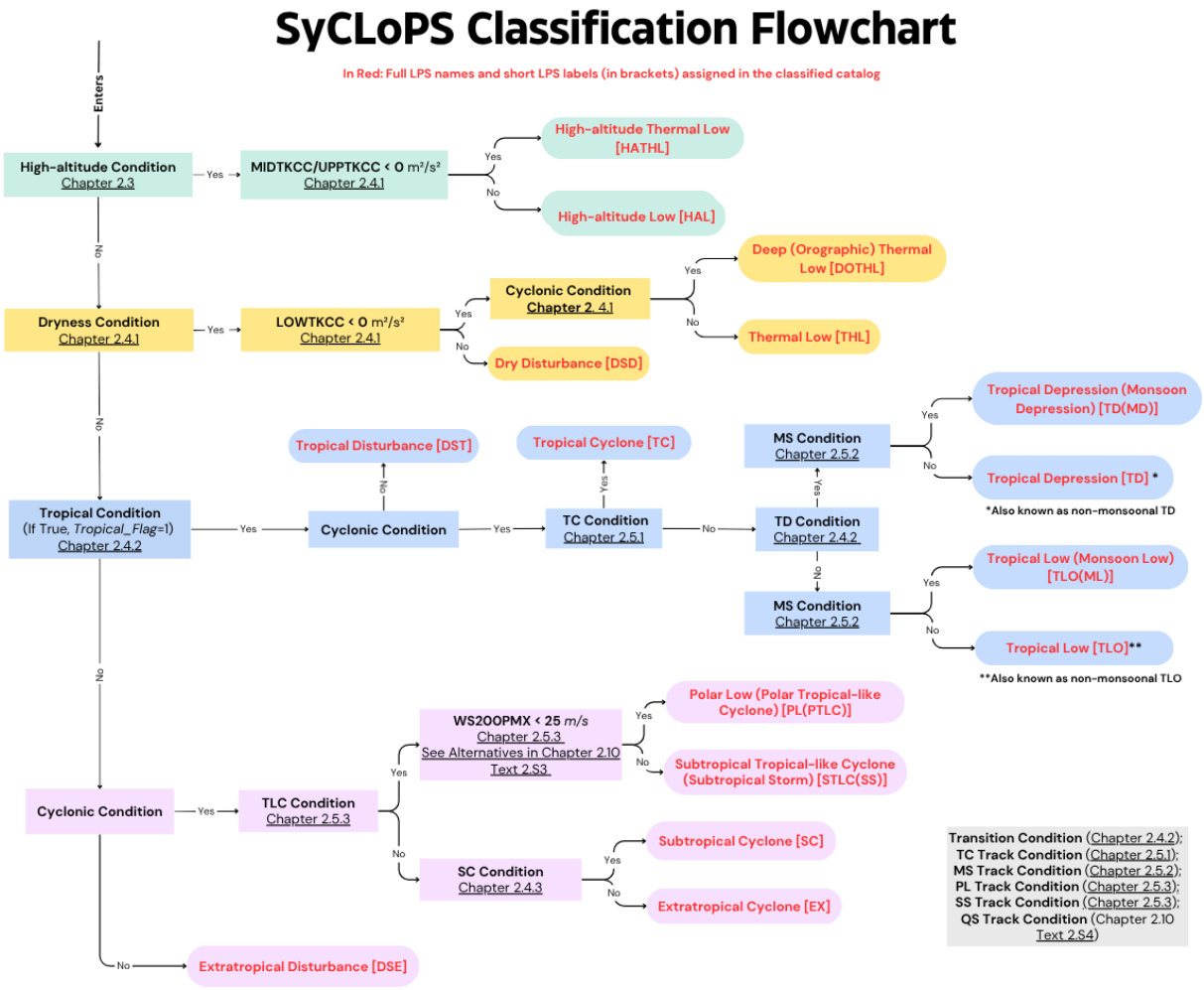



**Figure 2.1**. A depiction of the LPS classification workflow. The workflow begins in the top-left. A complete list of parameters is given in Appendix 2C, and a complete list of conditions is provided in Appendix 2D. Chapter numbers are noted in the figure below each condition to indicate where details can be found in the text. LPS full names and short labels given by SyCLoPS's output (the classified catalog) are denoted in red.

## 2.4 Justification for Classification Conditions

### 2.4.1 High-altitude and Dry Branch Conditions

In the high-altitude branch, we separate two classes based on the mid-level/upper-level warm core criterion. In this study, we use geopotential thickness to detect the existence of warm cores, similar to the criteria used in Zarzycki & Ullrich (2017) for detecting the upper-level warm core of TCs. Zarzycki & Ullrich (2017) suggested that geopotential thickness is more effective at detecting TCs in climate data than a single-layer temperature criterion because it better captures the full vertical structure of the upper-level warm core. Geopotential thickness is also a smoother meteorological field that is subject to less data noise than single-level temperature. LOWTKCC, MIDTKCC, and UPPTKCC are the negative closed contour criteria of the geopotential thickness between 700 hPa and 925 hPa, 500 hPa and 700 hPa, and 300 and 500 hPa, respectively, over a 6.5º GCD from the maximum thickness within 1.0º GCD of an LPS node. Hence, a threshold of less than 0 $m^2 s^{-2}$ for these three criteria means that the thickness near the detected location is greater than its surroundings, indicating the presence of a warm core of an LPS. In this case, if MIDTKCC or UPPTKCC is less than 0 $m^2 s^{-2}$, it indicates that a high-altitude LPS is warm-cored at those levels and may be thermally driven. If this condition is met, the label "High-altitude Thermal Low (HATHL)" is used in the classified catalog; otherwise, the label "High-altitude Low (HAL)" is used.

The dryness condition determines whether an LPS node will enter the dry branch: It requires that average 850 hPa relative humidity over a 2.5º GCD (RH850AVG) is less than 60%. This threshold is determined by the lowest track-maximum RH850AVG (the maximum RH850AVG of all nodes within a track) in the matched PL and SS dataset, chosen to prevent misclassification of significant non-tropical systems in the relatively drier subtropical/extratropical regions. Additionally, 850 hPa relative humidity (RH) thresholds are also deemed insensitive to climate change as RH at this level generally responds weakly to surface warming, especially in polar and tropical regions (Sherwood et al., 2010; Lau & Kim, 2015; Merlis et al., 2024). Therefore, we consider RH850AVG at 60% a safe threshold to separate dry

convective systems from moist convective systems. Next, the LOWTKCC < 0 $m^2 s^{-2}$ criterion is used to examine LPSs' low-level warm cores. If the condition is not met, the node will be classified as a "Dry Disturbance (DSD)" in the catalog; otherwise, we check the cyclonic condition.

The cyclonic condition uses average 500 hPa relative vorticity over a 2.5° GCD radius (VO500AVG) to determine if an LPS has cyclonic circulation beyond mid-level (i.e., is shallow). A typical heat low is considered to have a dry and warm low-level (between 925 hPa and 700 hPa) core and is shallow (not extending over 500 hPa) in nature (e.g., Smith, 1986; Hoinka & Castro, 2003; Lavaysse et al., 2009). Hence, if an LPS node does not qualify for the cyclonic condition, it will be labeled as a "Thermal Low (THL)" in the catalog. However, some deeper THLs still emerge near elevated topography in the daytime, such as the type II southwest vortex in southwest China (Feng et al., 2016), so the remaining LPSs in the dry branch are labeled "Deep (Orographic) Thermal Lows (DOTHL)."

### 2.4.2 Tropical Branch Conditions

The next step in the classification framework focuses on tropical systems. Traditionally, tropical systems such as TCs have been identified using a warm-core criterion (e.g., K. Hodges et al., 2017; Zarzycki & Ullrich, 2017; Roberts et al., 2020). However, in the course of this work we found that this criterion is often satisfied outside of the tropics and leads to many false alarms in the classification. This observation motivated us to examine other fields. We do not consider temperature parameters since they can be thermally-forced by local topography, and various types of LPSs can exist over similar sea surface temperatures (SSTs) in the subtropical oceans in different dynamical and thermodynamic environments. Consequently, we found that maximum relative humidity at 100 hPa within 2.5° GCD of LPS node (RH100MAX) is more reliable and flexible for disentangling tropical and extratropical systems as a proxy of "tropicality." There are two reasons that physically ground this choice. First, RH at 100 hPa is distinctly higher in the tropics. This is because only in the tropics is the tropopause often found above 100 hPa, as a result of active moist convection. RH is high there because of the low tropopause temperature and presence of upper-level moisture. RH100MAX also decreases sharply in the subtropics, reflecting the dynamics of the troposphere and the transition between the tropics and subtropics near the edge of the

Hadley cells (see Fig. 2.S1 for an illustration of the 1979-2022 global mean RH100MAX). Second, higher RH100MAX values indicate the presence of deep convection associated with a tropical system, and so this parameter is sensitive to extratropical transition scenarios during which it decreases rapidly while TCs gradually lose their deep convective cores and become post-tropical.

To illustrate the behavior of RH100MAX during extratropical transition, we examine a recent extratropical transition case (2023 hurricane "Lee") plotted with RH at 100 hPa in Fig. 2.2. In Fig 2.2a, the system was embedded in a region of high-level 100 hPa RH two days before transition completion (between 00-06 UTC, Sep 15) as defined by the National Hurricane Center. In Fig. 4.2.2b, less than one day before transition completion, the surrounding RH had dropped as the hurricane enters the subtropics. However, a belt of higher RH remains stretched out from the deep tropics, indicating the system's remaining "tropicality." RH100MAX is still over 90% at this point, as indicated by the bluish color within the 2.5° GCD circle. In Fig. 2.2c, hours after transition completion, we can see that the 100 hPa environment near the hurricane had become warm and dry, leading to a dramatic decrease in RH100MAX.

In addition to RH100MAX, average environmental deep-layer wind speed shear between 200 hPa and 850 hPa over a 10.0° GCD (DEEPSHEAR) is also used to distinguish tropical systems in the subtropics, especially during extratropical transition. Deep-layer shear is a good physics-related indicator of baroclinicity or an unfavorable environment for tropical deep convection. Post-TCs and general SCs/EXs primarily derive energy from baroclinic sources and are often surrounded by much more intense wind shear compared to tropical systems.

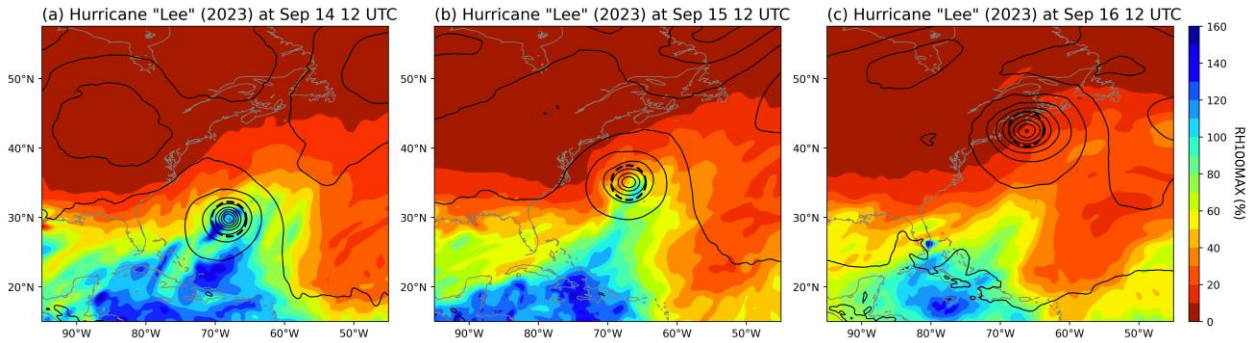


**Figure 2.2**. ERA5 Relative Humidity (RH) at 100 hPa during an example of an extratropical transition case (Hurricane "Lee" of 2023) at (a) one day before transition (b) during transition, and (c) soon after transition. MSLP is shown using black contours. Note that RH at 100 hPa can exceed 100% in some datasets, reflecting supersaturation.

The tropical (and transition) condition with RH100MAX and DEEPSHEAR is constructed as follows. First, we use two LPS node clusters that are hard to distinguish if we were to use an SST-based or warm-core criteria. One node cluster consists of all the matched tropical systems recorded in IBTrACS in the subtropics (the tropical cluster), and the other consists of potential subtropical systems over relatively warm SSTs that are not recorded anywhere in IBTrACS (the subtropical cluster). Note that both clusters (especially the unverified subtropical cluster) will inevitably include some misclassified or transitional LPSs. Details of how we select these two clusters can be found in Text 2.S2. We then apply the decision tree classifier over RH100MAX and DEEPSHEAR using Gini index splitting criteria to the nodes in the two clusters with a tree depth of two. Results in Fig. 2.3 show that the tropical and subtropical clusters are best distinguished by a minimum RH100MAX threshold of about 20% (rounded off to the nearest 5%) and a maximum DEEPSHEAR of 10 m $s^{-1}$, with a best accuracy score of nearly 80%. The accuracy scores mentioned in this article are the best classification accuracy scores a decision tree model can achieve to correctly classify two sets of LPSs given a ground truth. We perform a sensitivity test as demonstrated in Text 2.S2 and Fig. 2.S2. The result shows that these two thresholds are relatively stable and insensitive to the SST threshold used for selecting these two clusters. The elongated outer contours of the tropical cluster towards the left in Fig. 2.3 are likely made up of some LPSs near or after extratropical/tropical transition (for reference, about 5% of labels in IBTrACS are "Extratropical"), but also some "drier" tropical systems in drier or less convective basins. For example, while only 6 or 0.6% of western North Pacific TC tracks have a track-maximum RH100MAX under 50%, 94 or 17% of TC tracks in the North Atlantic fall into this range, with 24 tracks falling under 20%. According to these results, the 20% threshold will serve as the minimum RH100MAX requirement in the tropical condition, and the 10 m $s^{-1}$ threshold will be the minimum DEEPSHEAR requirement in the transition condition, as

stronger tropical systems can tolerate a much greater DEEPSHEAR value, such as in many extratropical transition cases. We also notice that DEEPSHEAR of some weak LPSs closer to the equator can slightly exceed the 10 m $s^{-1}$ threshold due to the tropical easterly jet. The cores of tropical easterly jets at 200 hPa are most commonly found near 5° N to 15° N (Lu & Ding, 1989). Hence, we impose that the transition condition will not be triggered when an LPS is within 15° latitudes of the equator.

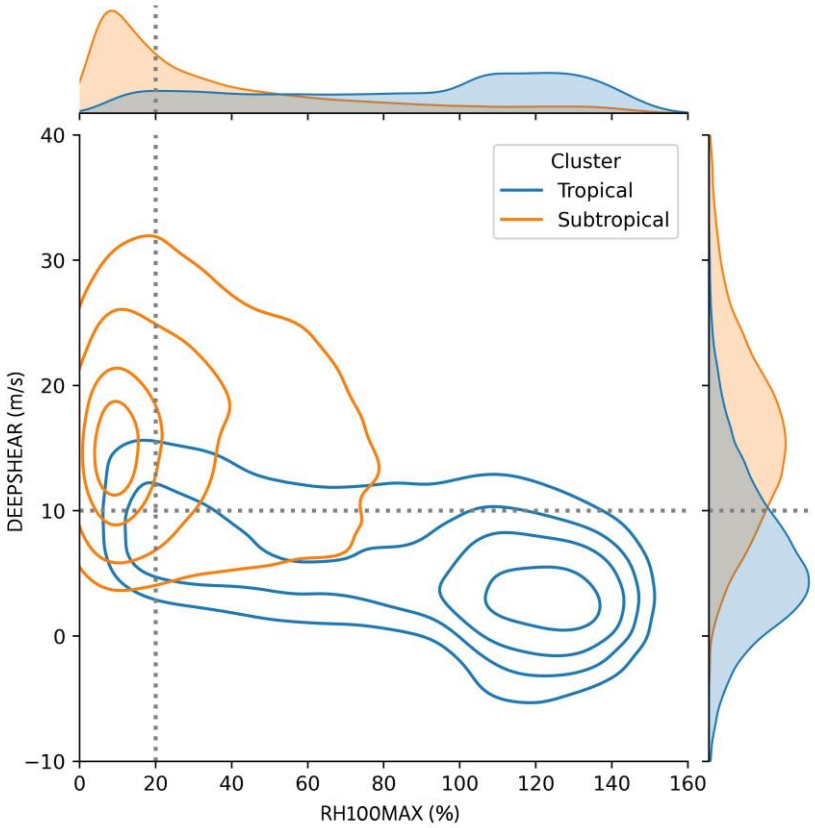



**Figure 2.3**. Kernel Density Estimate (KDE) on the RH100MAX-DEEPSHEAR plane for the tropical cluster and the subtropical cluster. The KDE levels are 0.25, 0.5, 0.75, and 0.9. Grey dotted lines indicate the classification thresholds determined by the decision tree classifier.

To find the upper limit of DEEPSHEAR for the tropical condition, we select 883 extratropical transition tracks from the matched TC datasets whose pairs in IBTrACS have a "ET (Extratropical)" or "MX (Mixture, contradicting nature reports from different agencies)" label following the last "TS (Tropical System)" label in the "NATURE" column, and here we define the time of the last "TS" label as the extratropical completion time of each extratropical track. The pre-transition cluster is defined by those

nodes that are 3 to 24 hours before the transition completion, and the post-transition cluster is made up of those that are 3 to 24 hours after transition completion. We apply the decision tree classifier based on DEEPSHEAR to find the boundary between the two clusters. With an accuracy score of 65%, the results show that the optimal DEEPSHEAR threshold to distinguish the two clusters is around 18 m $s^{-1}$. The accuracy score is not high, but it is to be expected—most TCs gradually transform into EXs, suggesting potential subjectivity in their classification. We do not round off the DEEPSHEAR threshold to the nearest 5 m $s^{-1}$ since environmental wind shear typically changes slowly in magnitude along an LPS track. This result is also stable to small changes in the selection of the time range for each cluster.

Finally, we consider 55% as the upper limit of RH100MAX for the transition condition for three reasons: (1) If we apply the decision tree classifier based solely on RH100MAX to separate the tropical and non-tropical cluster, the threshold for RH100MAX will be about 55% with a 74% accuracy score; (2) the median RH100MAX is about 55% at the time of extratropical transition completion as defined above, and (3) the median track-minimum RH100MAX in the matched EW dataset (the matched tropical LPS dataset with the lowest average RH100MAX) is also about 55%. In summary, the tropical condition refers to RH100MAX > 20% and DEEPSHEAR < 18 m $s^{-1}$. Upon fulfillment of the tropical condition, the transition condition is satisfied when RH100MAX < 55% or DEEPSHEAR > 10 m $s^{-1}$, and the latitude is poleward of 15°.

If only the RH100MAX and DEEPSHEAR thresholds are included in the tropical condition, we find that a small number of polar systems could also satisfy the tropical condition. As shown in Fig. 2.S1, polar regions can also feature a relatively higher 100 hPa RH that potentially exceeds the RH100MAX threshold in our tropical condition. This is mostly the result of persistent darkness during polar wintertime, which allows the upper air temperature to fall to exceptionally low values despite a lack of moisture. On the other hand, DEEPSHEAR also tends to be quite low in polar regions, as they are not in the main baroclinic zone. It should be fairly easy to further separate these false alarms, since polar regions are not adjacent to the tropics and are therefore distinctly colder. A plot of the air temperature at 850 hPa at the node (T850) distribution for all systems satisfying the RH100MAX and DEEPSHEAR thresholds

indicates that tropical systems and polar systems are separate from each other by a ~15 K (270 K to 285 K) gap (see Fig. 2.S3a). Hence, an additional T850 criterion (T850 > 280 K) is included to further distinguish the two systems. Given that the gap that separates the T850 of the vast majority of the polar systems and the T850 threshold is large (~ 10 K), and the fact that most of these polar systems reside near the Antarctic continent, where climate cooling has been observed over the past decades (Doran et al., 2002; Oliva et al., 2017), we expect this condition to hold even under worst-case scenarios of global warming, in which the global average temperature is predicted to increase by 4-5 K by the end of the century (Masson-Delmotte et al., 2021).

We also expect the proposed RH100MAX threshold to remain valid under the more extreme global warming scenarios. Merlis et al. (2024) showed that climate model simulations with a uniform +4K SST perturbation results in tropical RH at 100 hPa increasing at a modest rate of 0.5-1% per 1 K SST warming. This is mainly due to the expansion of the Hadley cells (Sherwood et al., 2010). A fixed RH100MAX threshold will therefore realistically reflect tropical expansion under global warming. RH at 100 hPa in the subtropics and midlatitudes shows even smaller changes (~0-0.5% $K^{-1}$), most likely due to the rising extratropical tropopause (from about 200-250 hPa) at a rate of ~-5 hPa per 1 K surface warming (Lu et al., 2007). In another experiment using 33 climate models, subtropical and mid-latitudinal RH at 100 hPa shows an average increase of less than 2% toward the end of the 140-year simulation period during which $CO_2$ is prescribed to increase by 1% per year (Lau & Kim, 2015). This change is considered minimal given the uncertainties inherent in the chosen threshold.

Further down the tropical branch, the cyclonic condition determines whether an LPS is shallow and so should be tagged as a "Tropical Disturbance (DST)" in the classified catalog. The next step involves the TC condition, which identifies tropical cyclones (TCs) and labels "TC" in the catalog. The conditions for this step are obtained by criterion (parameter threshold) optimization and are discussed in Chapter 2.5.1. Tropical depressions (TDs) are sometimes referred to as the weakest TCs before their potential formation. Therefore, the TD condition requires an LPS to at least have weak upper-level warm cores (UPPTKCC < 0 $m^2 s^{-2}$). We do not require a low-level warm core for TDs as many weaker tropical

systems develop an upper-level warm core before a stable low-level warm core is established (Reed et al., 1977; Hopsch et al., 2010; Hunt et al., 2016). We additionally require the greatest positive closed contour delta of MSLP over a 5.5° GCD (MSLPCC55) to exceed 160 Pa, determined by the median LPS node's MSLPCC55 at the IBTrACS track start time of each matched TC track, as agencies tend to start recording LPSs when they are reaching TD wind speed intensities (typically around 20 to 25 knots). Regardless of whether an LPS satisfies the TD condition, the MS condition is also applied to separate monsoonal and non-monsoonal LPSs. The MS condition is obtained by optimization and discussed in Chapter 2.5.2. After both the TD and MS conditions have been checked, the classifier assigns one of the four TD and tropical low (TLO) labels accordingly, as shown in Fig. 2.1. Specifically, monsoonal systems that satisfy both the MS and TD conditions and only the MS condition are considered MDs and MLs and are labeled "TD(MD)" and "TLO(ML)", respectively, in the classification catalog. Similarly, (non-monsoonal) TDs and TLOs that do not satisfy the MS condition are labeled "TD" and "TLO" in the catalog.

### 2.4.3 Extratropical Branch Conditions

LPS nodes that do not satisfy the tropical condition are non-tropical (extratropical) systems in the extratropical branch. The cyclonic condition separates "Extratropical Disturbance (DSE)" in the classified catalog before they are examined under the TLC condition obtained by optimization. The criteria for identifying TLC labels in the catalog, which include 'SS(STLC)" and "PL(PTLC)", will be discussed in Chapter 2.5.3. The remaining LPS nodes will go through the subtropical cyclone (SC) condition which follows the general definition of a typical SC—a non-frontal LPS that features a shallow warm core and an upper-level cold low isolated/detached from the midlatitude westerlies extending its circulation to the surface in the subtropics (U.S. Navy, 1994; Evans & Braun, 2012). Our SC condition states that an LPS must: (1) have a positive closed contour delta of geopotential at 500 hPa over a 3.5° GCD referenced to the minimum value within 1.0° GCD (Z500CC) greater than 0 $m^2 s^{-2}$ to satisfy the upper-level cold low characteristic; (2) have a LOWTKCC less than 0 $m^2 s^{-2}$ to guarantee that the low-level is warm-cored; and (3) have the maximum poleward 200 hPa wind speed within 1.0° GCD longitude (WS200PMX) greater than 30 m $s^{-1}$ (an effective minimum wind speed for identifying jet streams, see Koch et al. (2006)) to

increase the likelihood of the system being equatorward of the polar jet. Since WS200PMX might not be reliable in some regional models, alternatives to WS200PMX criteria used in SyCLoPS are listed and explained in Text 2.S4. We do not require EXs to be cold-cored since many Shapiro-Keyser EXs can be warm-cored due to the warm seclusion in their mature stage (Schultz & Keyser, 2021).

## 2.5 Criteria Optimization for High-impact LPS Detection

As discussed in Chapter 2.4, the criteria for TCs, MSs (MDs and MLs), and TLCs all rely on criterion optimization. In this section, we describe the optimization procedure for these four classes of LPS: Specifically, we optimize each type of LPS's detection skill against different skill metrics discussed below to find the best selected parameter threshold combination contingent upon conditions upstream of the workflow. The selection of these parameters is primarily based on physical intuition and previous studies. Since our optimization criteria are based on comparison to subjectively labeled LPS tracks, the optimization procedure also labels LPS tracks by track conditions using the node count parameter (the count of nodes with a specific label within a track) to more stably define an LPS track. Considering that many climate datasets have a 6-hourly temporal resolution (instead of the 3-hourly resolution we are using in this version of the catalog), we start with a minimum node count of two during the optimization.

### 2.5.1 TC Condition Optimization

The TC condition follows the cyclonic condition in the tropical branch. To identify variables for the optimization procedure, we require the greatest positive closed contour delta of MSLP over a 2.0° GCD (MSLPCC20) to satisfy some minimum value, LOWTKCC to be less than 0 $m^2 s^{-2}$ (indicating a low-level warm core), and UPPTKCC to satisfy some maximum value in magnitude, since TCs are generally characterized by compact MSLP contours and deep warm cores. We use a MSLPCC20 criterion instead of a maximum wind speed criterion because the latter is much more sensitive to model resolution and can be more easily distorted by complex topography. We also demand the node count of TC-labeled nodes within a track to have some minimum value to define a TC track. Evenly spaced values of these three parameters (over 3000 combinations) are considered to find the maximum detection skill.

A “test” TC dataset is constructed based on each possible 3-parameter combination for the 1979-2021 period, and it is compared to the reference dataset IB-TC. A “hit” occurs if the test dataset is matched to a track in the reference dataset by appearing within 2° GCD from a reference dataset data point at the same timestamp. A “miss” occurs if a track in the reference dataset does not have a match in the test dataset. A “false alarm” is a track found in the test dataset but is not matched to any tracks in the reference dataset. The TC detection skill metric used here is the hit rate (HR) minus false alarm rate (FAR), expressed as HRMFAR. HR is defined as the ratio of hits to the total number of hits plus misses, and FAR is defined as the ratio of false alarms to the total number of detected/selected tracks.

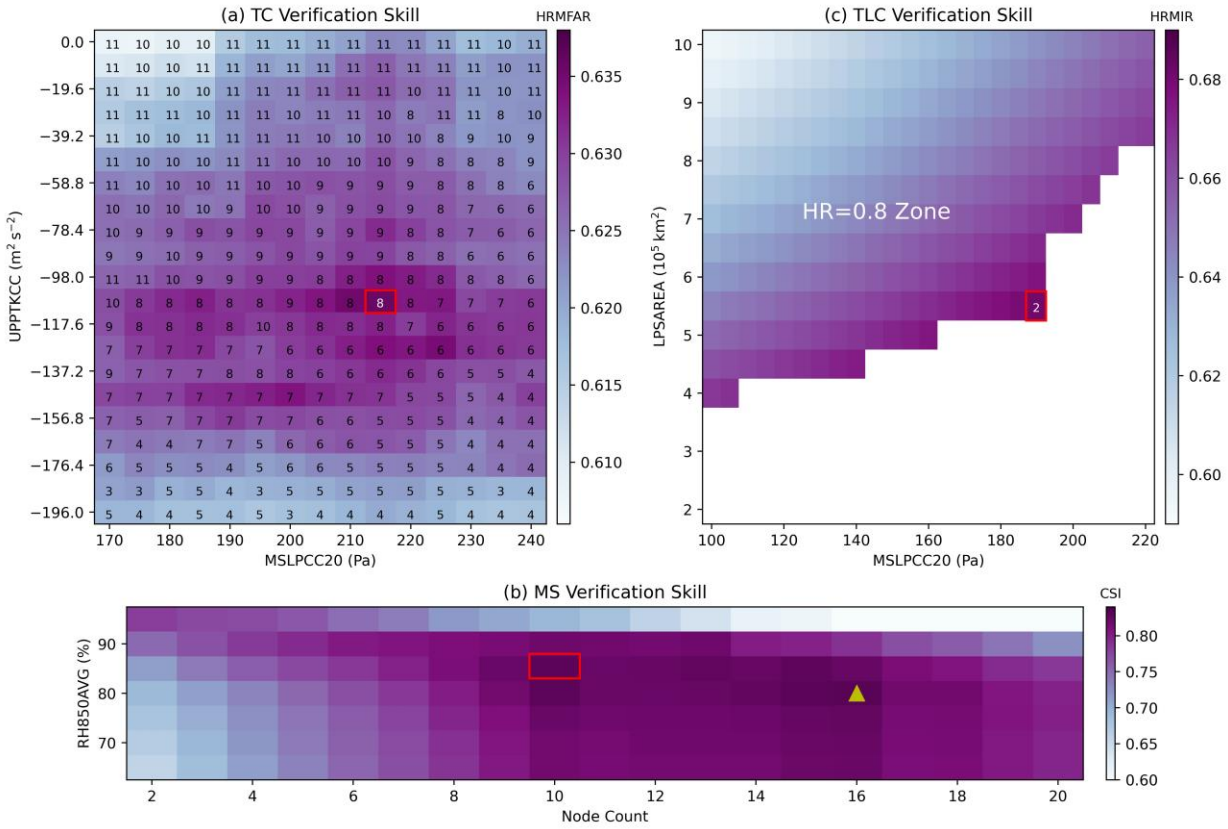



**Figure 2.4**. Detection skill optimization for (a) TC, (b) MS, and (c) TLC using different parameter threshold combinations and detection skills metrics shaded by their metric scores. The numbers in (a) and (c) represent the optimized time step for each combination (the optimal node count in (c) is uniformly two for every combination). The red rectangle in (a) and (c) indicate where the metric scores are maximized and consequently the final thresholds chosen. In (b), the yellow triangle indicates the maximized score,

and the red rectangle marks the final thresholds chosen. In (c), the metric score shading is only shown for the HR = 0.8 zone.

Figure 2.4a shows the detection skill of all chosen combinations of the MSLPCC20 and UPPTKCC thresholds, with the optimal node count shown at the top of each combination. The best detection criteria combination found is UPPTKCC < -107 $m^2 s^{-2}$ (-11 m), MSLPCC20 > 215 Pa (210 Pa is also acceptable since it yields a near-identical score), and TC-labeled node count > 8 with HRMFAR reaching 64%. More Details about the TC detection performance are discussed in Sec. 6. In summary, UPPTKCC < -107.8 $m^2 s^{-2}$ and MSLPCC20 > 215 Pa are used to define the TC condition along with LOWTKCC < 0 $m^2 s^{-2}$, and the node step threshold specifies that there must be at least eight 3-hourly time steps of TC-labeled nodes within a track for the track to be a TC track (the TC track condition).

### 2.5.2 MS Condition Optimization

For MS detection optimization under global detection and without seasonal constraints, we demand criteria that could separate MSs from other weaker tropical LPSs, such as EWs. MSs are considered to be born within monsoon troughs as opposed to the intertropical convergence zone (ITCZ) or upper-level easterly waves. According to National Hurricane Center (n.d.)'s definition, monsoon troughs are characterized by their westerly flow south of the trough, compared to easterly trade winds on both sides of the ITCZ. Furthermore, onsets of regional summer monsoons are often defined as a pattern of a change in wind speed and direction toward stronger westerlies (e.g., Qian & Lee, 2000; Gan et al., 2004). Thus, we develop the U850DIFF parameter, which is the difference between the weighted area mean of the positive and negative values of 850 hPa U-component wind over a 5.5° GCD. This allows us to determine whether westerly winds (positive U-component wind magnitudes) or easterly winds (negative U-component wind magnitudes) dominate the local 850 hPa environment of a system. The plot of the U850DIFF distribution for the MS and EW matched datasets shows that U850DIFF = 0 m $s^{-1}$ effectively segregates the two clusters (see Fig. 2.S3b). Hence, we select U850DIFF > 0 as a minimal requirement for the MS condition.

As implied by the matched MS and EW datasets, MSs in the North Indian Ocean usually have a higher RH850AVG than North Atlantic EWs. This is not surprising, as monsoon regions are generally considered to have more convective activity and larger moisture transport. Given that Vishnu et al. (2020) also used a averaged 850 hPa RH criterion to exclude non-monsoonal systems, we decide to require a minimum RH850AVG threshold in the MS condition.

Critical Sucess Index (CSI) as defined below by Vishnu et al. (2020) is the metric for evaluating MS detection skill.

$$CSI = \frac{hits}{hits + (misses + false\ alarms)/2} \tag{2.1}$$

A "hit" here is defined as a track in the test dataset having at least one node that is within 3.0° great-circle distance (GCD) of a track point in the Sikka dataset on the same date, and the track must also exist in the matched MS dataset. Since TCs are included in the Sikka dataset and are considered the most intense monsoon LPSs by the Indian Meteorological Department, tracks that already satisfy the TC track condition and are in the matched MS dataset are automatically considered matched (hits). We apply the detection optimization over the same domain and season as Vishnu et al. (2020), except that we incorporate the entire available data period rather than just a portion of it. We compute CSI of all the selected threshold combinations to find where CSI is maximal. The result shown in Fig. 2.4b suggests that the maximum CSI reaches 0.83 for RH850AVG = 80% and MS time step = 16. Our best CSI is thus identical to the value found in Vishnu et al. (2020), in support of the framework's ability to detect weaker tropical systems. We finally choose the second highest CSI (also over 0.83) combination, RH850AVG = 85% and the node count = 10, for the desired thresholds because we would like to include shorter MS tracks. Following these results, the MS condition is set as RH850AVG > 85% and U850DIFF > 0 m $s^{-1}$, and a track is considered to be a MS track only if 10 or more of its 3-hourly nodes satisfy the MS condition and are labeled as "TD(MD)" or "TLO(ML)" in the classified catalog. The MS track label highlights those weaker tropical LPS tracks that are more stably labeled as monsoonal systems, although

they can also coincide with TC-labeled tracks (i.e., TC tracks detected by SyCLoPS) per our standards. EWs and other tropical waves are likely to be included in the non-monsoonal TDs/TLOs that fall short of the MS condition, as well as other weak LPS classes (i.e., tropical/dry disturbances). One may assume that westward-propagating disturbances and non-monsoonal TDs/TLOs that meet the tropical condition (indicated by "Tropical_Flag = 1" in the classified catalog) and are not in MS-labeled tracks are mainly easterly waves.

### 2.5.3 TLC Condition Optimization

"Tropical-like cyclones" refers to non-tropical LPSs that resemble "real" TCs in certain ways. For instance, a mature Mediterranean hurricane may have a distinct eyewall and a deep warm-core structure despite lower SSTs and greater baroclinity in a non-tropical environment (Pytharoulis et al., 2000). PLs (sometimes referred to as intense polar mesoscale cyclone or Arctic hurricanes) and Mediterranean hurricanes (a branch of SSs), although still vaguely defined, may all be described as a group of mesoscale (small), intense, and short-lived (in terms of their TLC-stage lifespan) LPSs that can be classified as "tropical-like." Based on the definitions mentioned in Chapter 2.2, the most noticeable difference between PLs (PTLCs) and SSs (STLCs) might be that PLs develop north of the polar front or the main baroclinic zone in cold air masses, compared to SSs emerging from the subtropics. The term "hurricane-like extratropical cyclone" is also used in Romero & Emanuel (2017) to group Mediterranean hurricanes and North Atlantic PLs together. Here, we adopt a similar view that SSs and PLs are comparable to one another but different from typical extratropical/subtropical cyclones and could be flagged under the same TLC condition. The conventional definition of PLs as being north of the polar front can then be used to distinguish between them. Under this framing, SyCLoPS offers a means for the first objective global identification of all TLC systems in climate data.

Since a minimum requirement of near gale-force wind intensity similar to TCs is required to define SSs and PLs, similar parameters such as MSLPCC20 and LOWTKCC are used to detect TLCs as we did to detect TCs. We expect TLCs to have, on average, a shallower/weaker warm-core structure compared to TCs. Hence, we first impose a minimum requirement for the two warm core criteria

(LOWTKCC < 0 $m^2 s^{-2}$ and MIDTKCC < 0 $m^2 s^{-2}$). Static-stability or open-water criteria used in previous PL detection studies (Zappa et al., 2014; Stoll et al., 2018; Stoll, 2022) are not considered here as they appear too restrictive to global TLC detection. For example, PLs may also appear closer to the baroclinic zone in a more sheared environment (Montgomery & Farrell, 1992; Terpstra et al., 2016), and intense storm activity can often occur over Antarctic sea ice (Hepworth et al., 2022). The other significant distinction of TLCs is their small or mesoscale sizes. Thus, we generate LPS size blobs and compute the LPS size (extent) measured by the LPSAREA parameter as described in Appendix 2E, in addition to the parameters directly computed by TE to evaluate the defined two-dimensional size (area) of LPSs.

We use the combined matched SS and PL dataset, which consist of 174 tracks, as our reference dataset for optimization. We concede that it's difficult to evaluate or compare global TLC detection skills because credible TLCs' records are limited and regional in scope, and their definition inexact. To overcome this, we first remove 12 tracks in the reference dataset that have a track-maximum MSLPCC20 lower than 215 Pa (the MSLPCC20 standard for TCs) to further avoid including tracks that are too weak to be considered TLCs. Second, we acknowledge that some TLCs could be embedded within a synoptic scale circulation or trough in the background, sometimes with a twin low nearby (see Fig. 2.S4 for an example), and thus will appear large (or be zero if embedded in a system with lower MSLP) using our size computation method. To work around this observation, we determine that LPSs with MSLPCC20 > 420 Pa (90% percentile of detected non-tropical, non-shallow LPSs' MSLPCC20) and a ratio of MSLPCC20 to MSLPCC55 greater than 0.5 (reflecting the presence of a dominant and more compact LPS core within the larger system) satisfy the embedded TLC alternate criterion and are exempt from the LPSAREA requirements. Third, since the scope and quality of the reference dataset is constrained, FAR becomes rather meaningless and is replaced by the infrequency rate (given that TLCs are infrequent), defined as the fraction of selected TLC tracks among all detected tracks that have at least one node that passes the cyclonic condition in the extratropical branch (a 0.1 infrequency rate as defined would correspond to a 45 occurrence per month in this SyCLoPS LPS dataset). Hence, the detection skills metric we use for the TLC condition is the HR minus infrequency rate, or HRMIR. Here, HR is simply defined

as the fraction of tracks that are detected (hits) in the reference dataset. Given the limited sample size, HR is rounded to the nearest tenth (i.e., 0.750 and 0.849 will be rounded to 0.8) to roughly reflect its 90% confidence interval (CI) and potential sampling errors.

We iterate the selected range of MSLPCC20 and LPSAREA threshold combinations for TLC condition optimization. The best HR attained is at the 0.8 level as shown by the shading in the upper left zone of Fig. 2.4c. Within this zone, the TLC condition is optimal when IR is the smallest (11.9%) at MSLPCC20 > 190 Pa and LPSAREA < $5.5\times10^5$ km$^2$ (given that LPSAREA is nonzero) on top of the other thresholds we mentioned above. The LPSAREA threshold chosen here agrees with the meso-$\alpha$ scale range (i.e., roughly a 400-500 km LPS radius), which aligns with the upper size range of many studied TLCs (e.g., Holland et al., 1987; Rasmussen & Turner, 2003; Fita et al., 2007; Noer et al., 2011). Due to the short-lived nature of TLCs, HRMIR in all combinations maximizes when the node count equals two. LPS nodes that have been tagged as TLC will then be further classified as PLs or SSs depending on whether they are located further north to the polar jet (WS200PMX < 25 m s$^{-1}$). Tracks with two 3-hourly TLC-labeled nodes (PL or SS) and at least one PL or SS-labeled node are then assigned "PL(PTLC)" or "SS(STLC)" track labels. For an alternative test, we remove the embedded TLC criterion and perform the optimization. The results show an identical HR, a slightly lower IR, and 72% overlapped detected TLC tracks (i.e., LPS tracks picked up by both TLC detection algorithms) when MSLPCC20 > 145 Pa and LPSAREA < $7.0\times10^5$ km$^2$. Thus, it may be treated as an alternative TLC condition, although it risks excluding many embedded TLC nodes.

## 2.6 Results and Applications

### 2.6.1 Main Results

In Fig. 2.5a, we plot the kernel density estimate (KDE) on the RH100MAX-DEEPSHEAR coordinate of all 6-hourly sampled LPS nodes that have passed the second (dry) branch and meet the cyclonic condition to verify the efficacy of our tropical and transition conditions. Our results demonstrate the validity of using RH100MAX and DEEPSHEAR thresholds as the foundation for these conditions.

The KDE clearly depicts two main clusters, separated by RH100MAX and DEEPSHEAR. The solid-line and dash-line boundaries delineate the tropical and transition conditions. The RH100MAX transition threshold cuts through the narrowest part of the KDE. The cluster centered inside the tropical condition bounds is the tropical system cluster, while the cluster centered outside the box is the non-tropical system cluster, which is apparent from Fig. 2.5b, where the KDEs of the five matched datasets are placed on the RH100MAX-DEEPSHEAR coordinate. Within the tropical system cluster, the TC cluster spans the widest range as it includes LPSs undergoing extratropical transition and at post-TC stage, whereas the MS and EW clusters have the highest and lowest mean RH100MAX values respectively. Most of the matched tropical LPSs are within the transition boundaries (which may be deemed as the deep tropics). Inside the non-tropical system cluster, the SS cluster has a higher mean DEEPSHEAR value than the PL cluster. The red filled contours in Fig. 2.5a depict the KDE of warm-core systems, defined as the previously selected LPS nodes that meet the criterion of UPPTKCC < -58.8 $m^2 s^{-2}$ (-6 m). The KDE shows that systems with a upper-level warm core can exist in both tropical and non-tropical clusters, and thus, the warm-core criteria may not be ideal for classifying LPSs across the spectrum. According to the classified catalog, the vast majority of the labeled tropical systems are confined within 40 degrees of the equator.

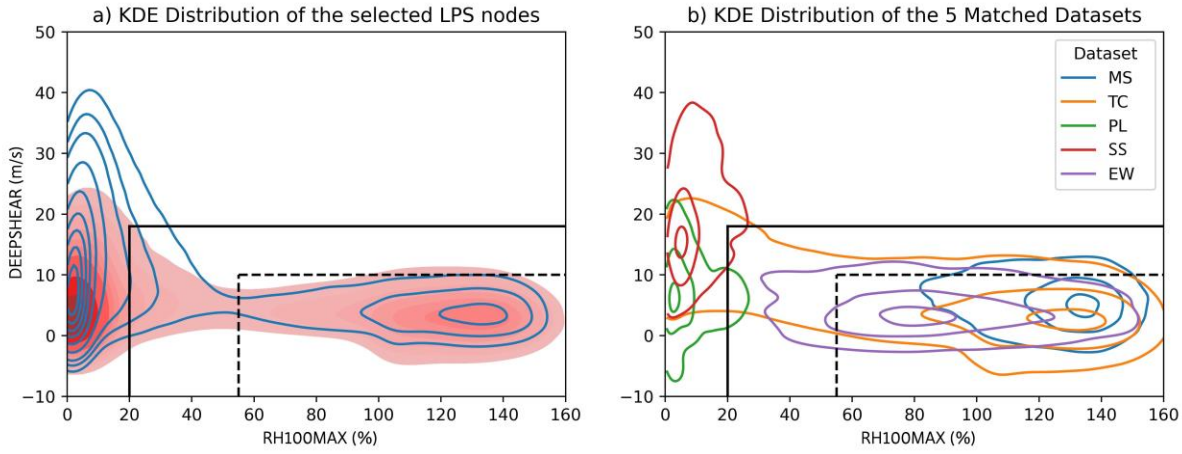



**Figure 2.5**. (a) The 10 KDE levels evenly distributed between 0 and 1 of all the selected detected LPS nodes (blue contours) and the warm-core LPS nodes (red filled contours); and (b) the three KDE levels set at 0.1, 0.5, and 0.9 of the five matched datasets on the RH100MAX-DEEPSHEAR coordinate.

SyCLoPS LPS labels are generally in good agreement with the labels in IBTrACS. Two types of labels are provided in IBTrACS: first, the labels in the "NATURE" column assigned by global WMO agencies, and second, the labels assigned by US meteorological agencies in the "USA_STATUS" column. The WMO labels are more general than the USA labels, as the USA labels include more classes based on LPS intensity. Miscellaneous labels that are vaguely defined and have a small sample size are not included in the comparison. Information about the two agencies' labels can be found on the IBTrACS website and in Landsea & Franklin (2013). Labels are compared when LPS nodes in our dataset and IBTrACS track points lie within 2.0º GCD of each other at the same timestamp. Fig. 2.6 shows the probability of success (POS) for correctly labeling an LPS node of a particular class (setting IBTrACS labels as ground truth). Overall, 95% of the matched nodes are in agreement with the WMO labels, mainly contributed by the high "TS" POS of 97% (WMO's "TS" label refers to "tropical system"). Since WMO's "DS (disturbance)" label also exists, we regard "TS" as all non-shallow tropical systems, which is equivalent to all tropical system labels but "DST" in our labeling system. This result suggests that very few tropical systems are mistakenly labeled as non-tropical systems by our classification. The extratropical system POS is at 78% when we compare labels in the extratropical branch to the "ET (extratropical)" label of WMO. The majority of the extratropical records in IBTrACS are post-TCs immediately after extratropical transition. Therefore, it suggests a rather small error in the extratropical transition completion time justified by our classification when compared to IBTrACS. Breaking down the tropical systems, our TC POS remains at a relatively high level of 74% against TC labels given by USA agencies, while TD (which includes TD and MD) has a much lower POS of 47%.

We find that TDs are almost equally likely to be misclassified as TCs and tropical lows (TLOs and MLs), which reflects ambiguity in their definitions and inevitable biases in LPS intensity evaluations by agencies, reanalysis, and our classification. If TD is considered a category of TC for both our and IBTrACS labels, the POS of "TC+TD" rises to 85%. The weakest system labels in IBTrACS, including "LO (low)" and "DS/DB (disturbance)," are more vaguely defined. They are often used at the start of TC tracks, and the labeled LPSs may not have a discernible surface center (Landsea & Franklin, 2013;

National Hurricane Center, n.d.). Hence, we treat them as the same label, which is equivalent to TLO systems (TLO and ML), disturbances (DST, DSD and DSE), and thermal low systems (THL, DOTHL, and HATHL) in our labeling system. POS of about 50% is realized for this category compared to labels of WMO and USA agencies. Similarly, they are subject to biases in intensity evaluations and their exact definitions. When TDs are included in this class for both our and IBTrACS labels, the POS increases to 78%.

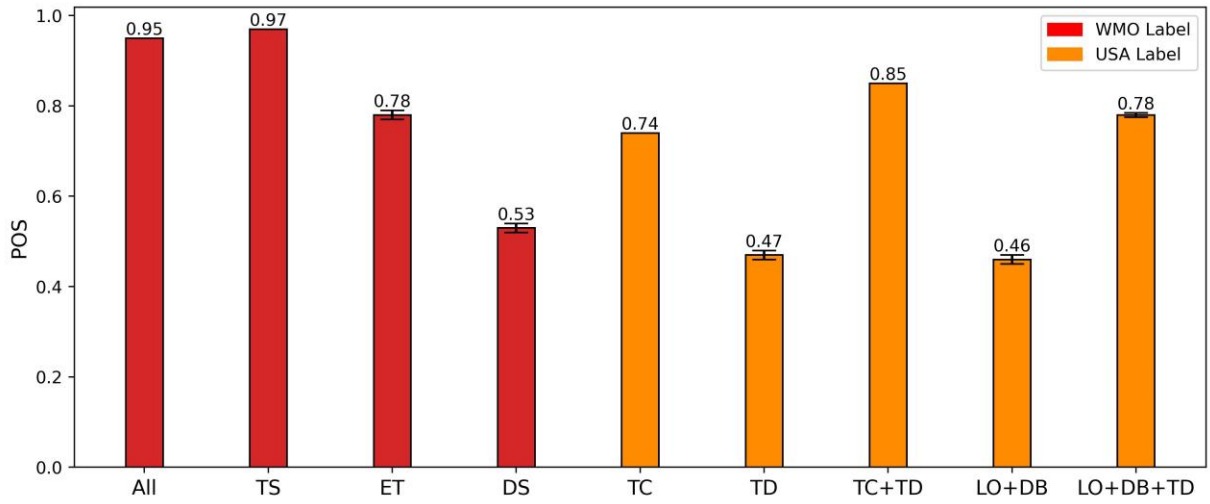



**Figure 2.6**. POS of our labels when compared to labels given by WMO (red bars) and USA agencies (orange bars). See text for details. POS values are shown on the top of bars with 95% confidence level error bars denoted.

TC detection skill is improved using SyCLoPS's TC track labels when compared to the previous TE algorithm (the Zarzycki & Ullrich method; Zarzycki & Ullrich, 2017). For the Zarzycki & Ullrich method, TC tracks are identified when nodes in a track that satisfy UPPTKCC < -58.8 $m^2 s^{-2}$ (-6 m), MSLPCC55 > 200 Pa, WS (maximum wind speed at 10 m within 2.0◦ GCD) > 10 m $s^{-1}$, and ZS < 150 $m^2 s^{-2}$ are detected for at least 54 hours equatorward of 50◦ latitude. We resample the detected TC dataset to match the 6-hourly data frequency of IBTrACS. The HRMFAR of Zarzycki & Ullrich is computed against the same IB-TC dataset for the period of 1979-2021 using the same definition of hits and false alarms mentioned in Chapter 2.5.1. Table 2.2 summarizes the TC detection skill metrics of both methods.

The mean start (end) time differences in the table refer to the time differences between the start (end) time of the detected TC track and the corresponding IB-TC track's start (end) time.

| Method | HR | FAR | HRMFAR | Mean start time difference (hr) | Mean end time difference (hr) |
|---|---|---|---|---|---|
| Zarzycki & Ullrich | 76.2% | 20.1% | 56.1% | 28 | -30 |
| SyCLoPS | 78.2% | 14.6% | 63.6% | -49 | 28 |

**Table 2.2**. Detection skill comparison between SyCLoPS and the Zarzycki & Ullrich method

To summarize, the detection skill improvements are: (1) HRMFAR is increased by 7.5% due to a 5.5% decrease in FAR and a 2% increase in HR; and (2) the early detection of the pre-TC stage and late detection of the post-TC stage are significantly improved, extending TC track length by an average of 136 hours. The effects of these improvements are revealed in Fig. 2.7. Tracks detected using SyCLoPS are visibly longer at both ends (the pre-TC stage and the post-TC stage) compared to those tracked by the Zarzycki & Ullrich method. Notably, the new approach more closely matches IBTrACS observations in the South Atlantic and the Southeast Pacific, among other subtropical oceans, by largely reducing false alarms in those regions. We also notice that many official wind data in IBTrACS's tracks are missing in earlier years in basins of the Indian Ocean (so they are not included in IB-TC) due to the fact that some agencies did not accept their regional responsibility until the early 1990s. Hence, many false alarms in the tropical Indian Ocean for both methods could actually be real TCs (hits). Discrepancies in wind measurement standards, observations, and operational procedures among agencies for different basins are also noted in Schreck et al. (2014), suggesting the presence of a "TC gray zone" due to these biases — i.e., a range of parameter values where different experts would draw different conclusions on the classification of a feature. Therefore, perfectly matching a subjective TC dataset is likely impossible. The blue and red dots show the extratropical and tropical transition completion positions of applicable TC tracks indicated by the classified catalog. About 40% of the identified TC tracks undergo extratropical transition, which is comparable to the global extratropical transition fractions using a CPS method reported in Datt et al. (2022). We compare the detected TC tracks of SyCLoPS that have a match to an

extratropical transition case recorded in IBTrACS and find that SyCLoPS detects 87% of all such extratropical transitions with an average (median) error of +10 (+6) hours in the transition completion time (i.e., the timestamp of the first labeled extratropical node after the TC stage). The HRMFAR of this new algorithm may be further elevated through post-processing operations such as eliminating QS tracks or marginal TC tracks that primarily reside in the transition zone. We advise being cautious when eliminating any marginal TCs since they can reside in the "TC gray zone." As an example, the 2001 Australia "Duck" is a classic marginal TC (see Garde et al., 2010). Although this storm was not recorded by the agency and it does not satisfy our TC track condition, it is labeled as "TC" at four time steps under our classification. See Fig. 2.S8 for a labeled track map of this special case.

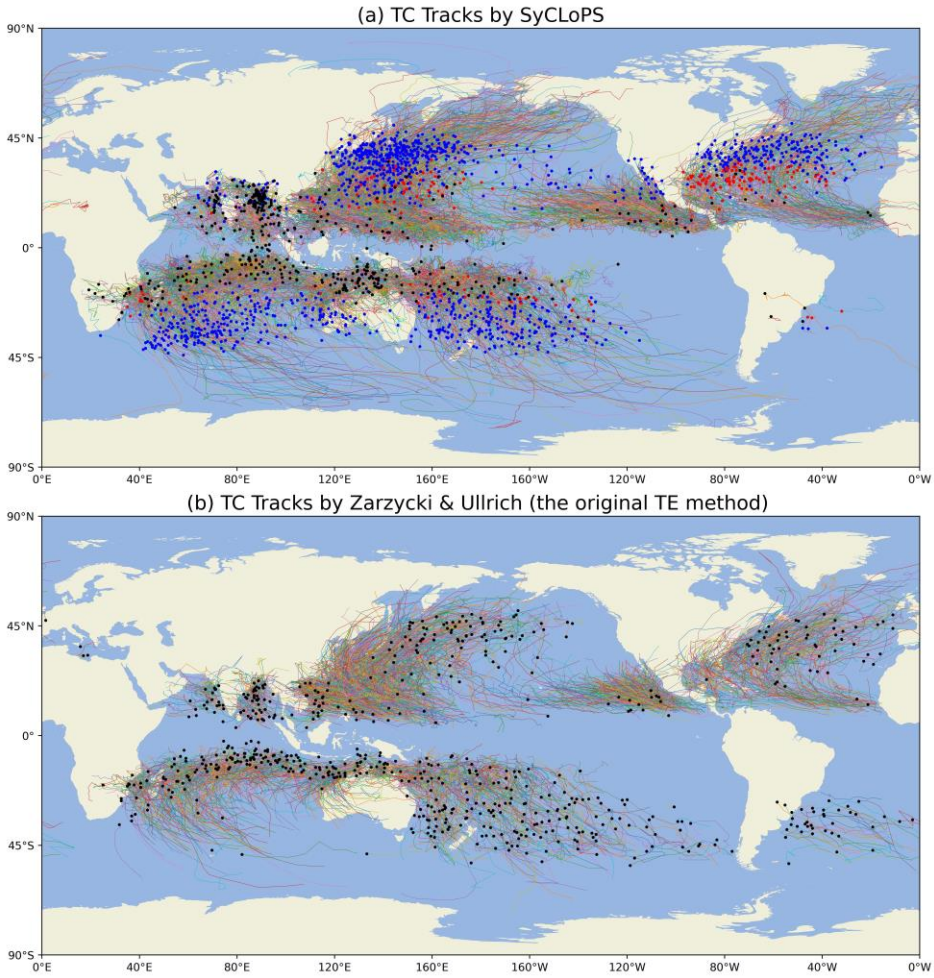



**Figure 2.7**. 1979-2021 TC tracks as tracked by (a) SyCLoPS, and (b) the Zarzycki & Ullrich method. Black dots are the first locations of false alarm tracks. Blue and red dots are extratropical and tropical

transition completion locations indicated by SyCLoPS. 1337 extratropical transition cases and 195 tropical transition cases are detected.

Major globally detected LPS annual frequencies for the different classes of LPSs are shown in Fig. 2.8. In general, the frequencies of these systems are in accordance with observations. Please refer to Fig. 2.S5 for a frequency bar plot of all LPS classes. The first row of Fig. 2.8 contains the least frequent LPS classes, followed by MS in the second row, which are all high-impact LPSs that can be considered extremes. TC frequencies are consistent with their track activity in the tropics and during extratropical transition (Fig. 8a). In Fig. 2.8b, SSs are more frequent in the Mediterranean Sea, the most studied hotspot for these features. They are also commonly found in the storm-track regions (western North Pacific and the northwestern Atlantic, as defined in Blackmon et al. (1977)), the southwest Atlantic, the southeast Pacific, the Japan Sea, and the Tasman Sea close to Southeast Australia. Those regions are all well known for their intense or tropical-like LPS activities, which include Australian east-coast cyclones, Chilean storms, Japanese south-coast explosive cyclones, TLCs/mesocyclones in the Sea of Japan and the Yellow Sea, and subtropical storms across the Atlantic (see e.g., Heo & Ha, 2008; Guishard et al., 2009; Iwao et al., 2012; Winckler et al., 2017; Gozzo et al., 2014; Shimada et al., 2014; Cavicchia et al., 2018). We expect that the successful classification of subtropical storms is effective for reducing TC false alarms in the SyCLoPS framework. PL activity reaches as far south as the Sea of Japan, and they are most prevalent in the Nordic Seas, the Gulf of Alaska, and over or near the sea ice of the Southern Ocean (Fig. 2.8c). Intense post-TCs are sometimes classified as TLCs in our framework, and removing them has only a minor impact on the frequencies of SSs and PLs. MSs are mainly constrained in the tropical monsoon region defined in Li & Zeng (2003) and have two evident hotspots in the North Indian Ocean and near the Gulf of Tonkin in the South China Sea (Fig. 2.8d). Non-monsoonal TLOs and TDs are found throughout the tropics, with some overlap with MS activity and evidence of QS tracks shown by localized high frequencies mainly near rainforest regions (Fig. 2.8e). After filtering those QS tracks labeled by the QS track condition, strong LPS occurrences largely disappear, leaving other features mostly untouched. SCs are more widespread but less concentrated compared to SSs (Fig. 2.8g). EX is the most common type of

LPS labeled, and it is omnipresent outside of the tropics (Fig. 2.8h). Disturbances are found globally across latitudes, and THLs and DOTHLs are located primarily on arid lands (Figs. 2.8i-k). Finally, high-altitude LPSs occupy mountainous areas, including parts of the Antarctic continent (Fig. 2.8l).

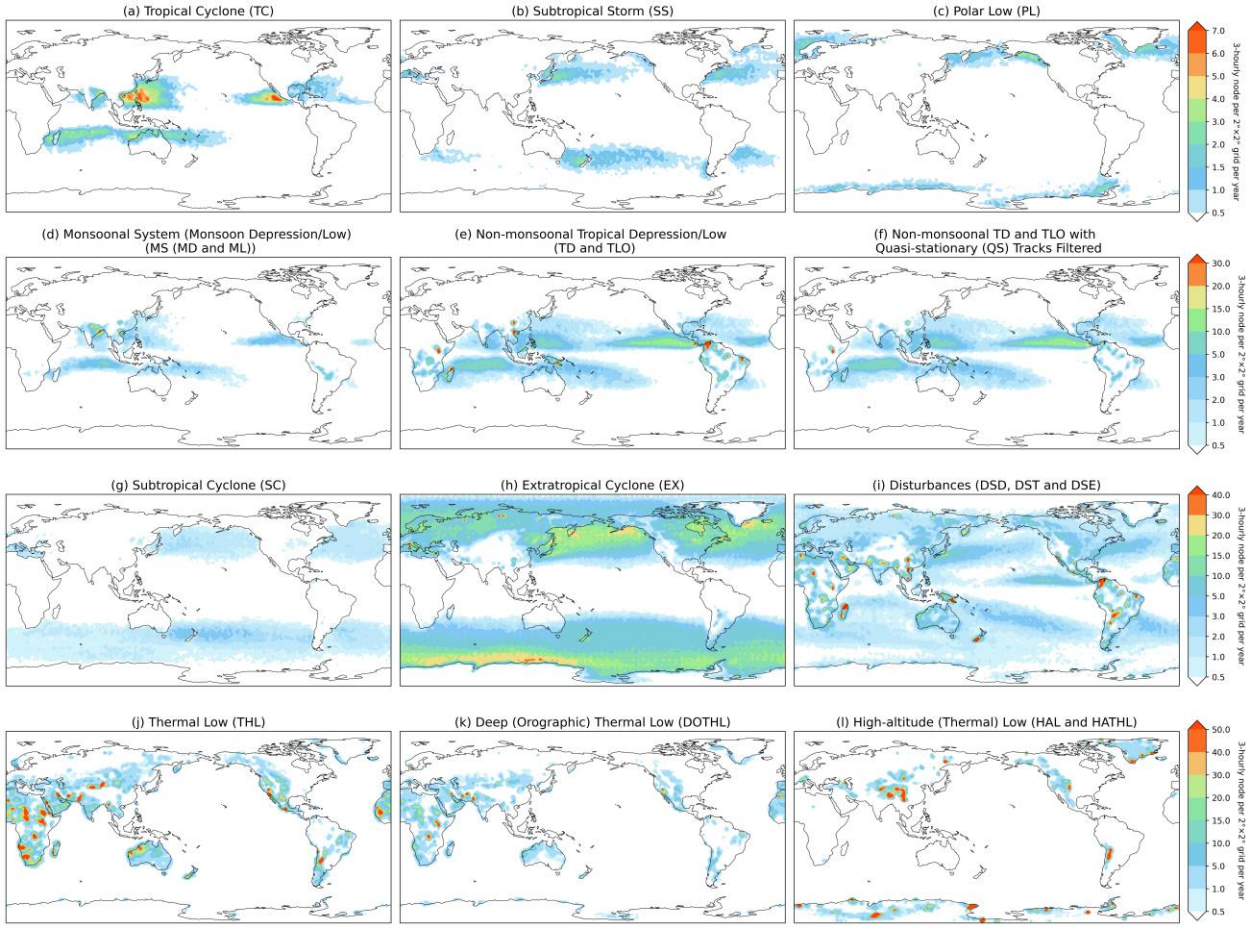



**Figure 2.8**. 1979-2022 annual frequency of (a) tropical cyclone (TC), (b) subtropical storm (SS), (c) polar low (PL), (d) monsoonal system (MS (MD and ML)), (e) non-monsoonal tropical depression (TD) and tropical low (TLO), (f) same as (e), but with quasi-stationary (QS) tracks filtered, (g) subtropical cyclone (SC), (h) extratropical cyclone (EX), (i) disturbances (DSD, DST and DSE), (j) thermal low (THL), (k) deep (orographic) thermal low (DOTHL), and (l) high-altitude LPSs (HAL and HATHL) as measured by 3-hourly nodes per 2º×2º grid.

We show the vertical cross section composites at the latitude of LPS's center for the six selected LPS classes labeled by SyCLoPS in Fig. 2.9. TCs feature a classic dumbbell-like structure resembling the shape of a cumulonimbus, as indicated by the two RH maxima at the lower- and the upper-level (Fig. 9a). Diabatic heating or latent heat release in TCs, as suggested by the cyclonic potential vorticity (PV) contours, is evident throughout the lower-level and upper-level. The deep warm-core structure suggested by the potential temperature contours is most evident in the TC composite. Fig. 2. 9b shows that a typical THL features a classic warm and dry low-level core, which is largely constrained to the boundary layer. As shown in figures 2.9c and d, weaker tropical systems have far less developed convection and warm cores compared to TCs. MSs have comparatively higher RH at each level and a slightly more developed lower-level circulation than other typical TDs/TLOs. An eastward tilt of the RH field below 300 hPa is also noticeable in the non-monsoonal TD and TLO composite in Fig. 2.9d. Cyclonic PV contours stretching down from the subtropical tropopause in the SS composite (Fig. 2.9e) imply that some SSs undergo a downward development pathway, extracting cyclonic PV from upper-level PV anomalies or PV streamers, which agrees with the Mediterranean hurricane development mechanisms described in Flaounas et al. (2022). The warm core and the diabatic heating are more constrained to the lower level for PLs, as depicted in Fig. 2.9g. The global patterns (Figs. 2.9e and g) of SS and PL also appear to have less noise (induced by local topography) and an enhanced PV signal in the vertical structure of SS and PL compared to the patterns (Figs. 2.9f and h) represented by regional SSs and PLs associated with the records in the subjective SS and PL datasets. Despite the fact that both TCs and TLCs (SSs and PLs) have relatively deep warm cores, the upper-level RH of SSs and PLs (as shown in Figs 2.9e-f) is significantly lower than that of TCs and other tropical systems. This distinction supports our choice of the RH100MAX criterion in the tropical condition.

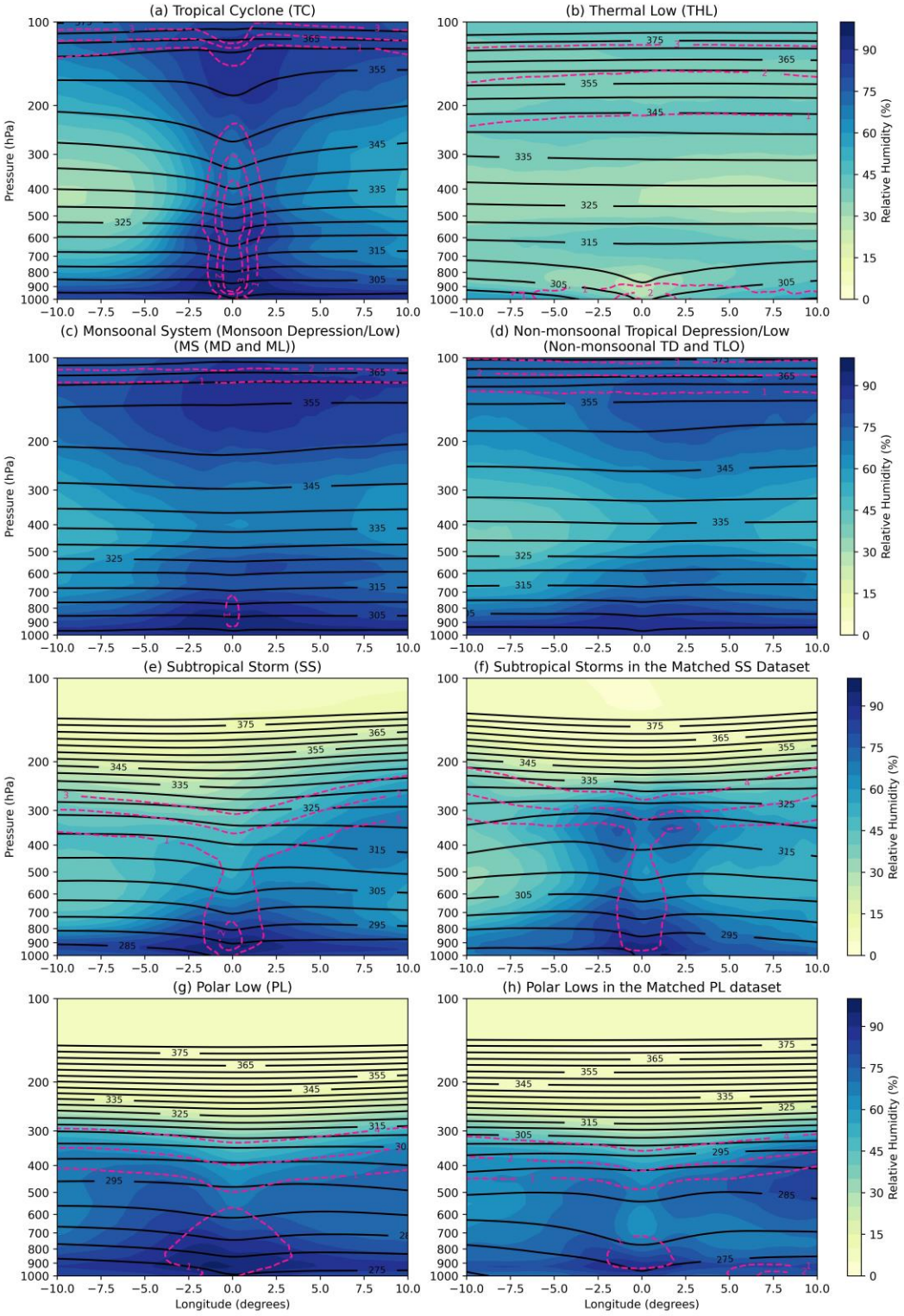



**Figure 2.9**. Vertical cross section composites of LPS node classified as (a) tropical cyclone (TC), (b) thermal low (THL), (c) monsoonal system (MS (MD and ML)), (d) non-monsoonal tropical depression

(TD) and tropical low (TLO), (e)-(f) SS and SSs in the matched-SS dataset, and (g)-(h) PL and PLs in the matched-PL dataset. Dark pink dashed lines are contours of cyclonic potential vorticity (unit: PVU (1.0×10−6 m2 s−1 K kg−1)), and black contours are potential temperature (K). The TC, MS, SS and PL composites are each based on 1000 randomly chosen nodes tagged with the specific type of label in the specific type of LPS track (i.e., 1000 TC-labeled nodes in TC-labeled tracks). The non-monsoonal TLO/TD composite is based on 1000 randomly chosen nodes labeled "TLO" or "TD", except those in MS or QS tracks. The THL composite is based on 1000 randomly chosen THL-labeled nodes.

### 2.6.2 Other Applications

We now show some simple applications based on the classified catalog produced by SyCLoPS. One major benefit of SyCLoPS is that it can reveal a fairly complete history of each LPS track, so that the evolution of an LPS can be effectively traced. Thus, a useful application is showing a track along with its labeled nodes, such as the example in Fig. 2.10. The example depicts the track history of the western North Pacific typhoon Mindulle in 2021. Mindulle is first detected as a disturbance near the equator, then gradually intensifies as a non-monsoonal TLO/TD before it becomes stably labeled as TCs. It completes its extratropical transition around 40° N and later develops into a SS and PL before dissipating as a EX in Alaska. The genesis time (the first TD label time), the first TC record time, and the transition completion time are all within 12 hours of the corresponding IBTrACS records, while the record given by SyCLoPS further extends the IBTrACS track length. A phase diagram displayed by the DEEPSHEAR-RH100MAX coordinate is attached to the figure. The phase evolution shows that the RH100MAX of the TC stays at a high level while the environmental wind shear gradually increases. The system's RH100MAX decreases sharply during extratropical transition, which is completed when the TC no longer satisfies the DEEPSHEAR criteria from the tropical condition. In its final stage, the system enters a lower-sheared environment with very low RH100MAX. More examples like this of different LPS classes (including North Atlantic hurricanes, the "Duck", an MS, and TLCs) can be found in Figs. 2.S7-9 in SI.

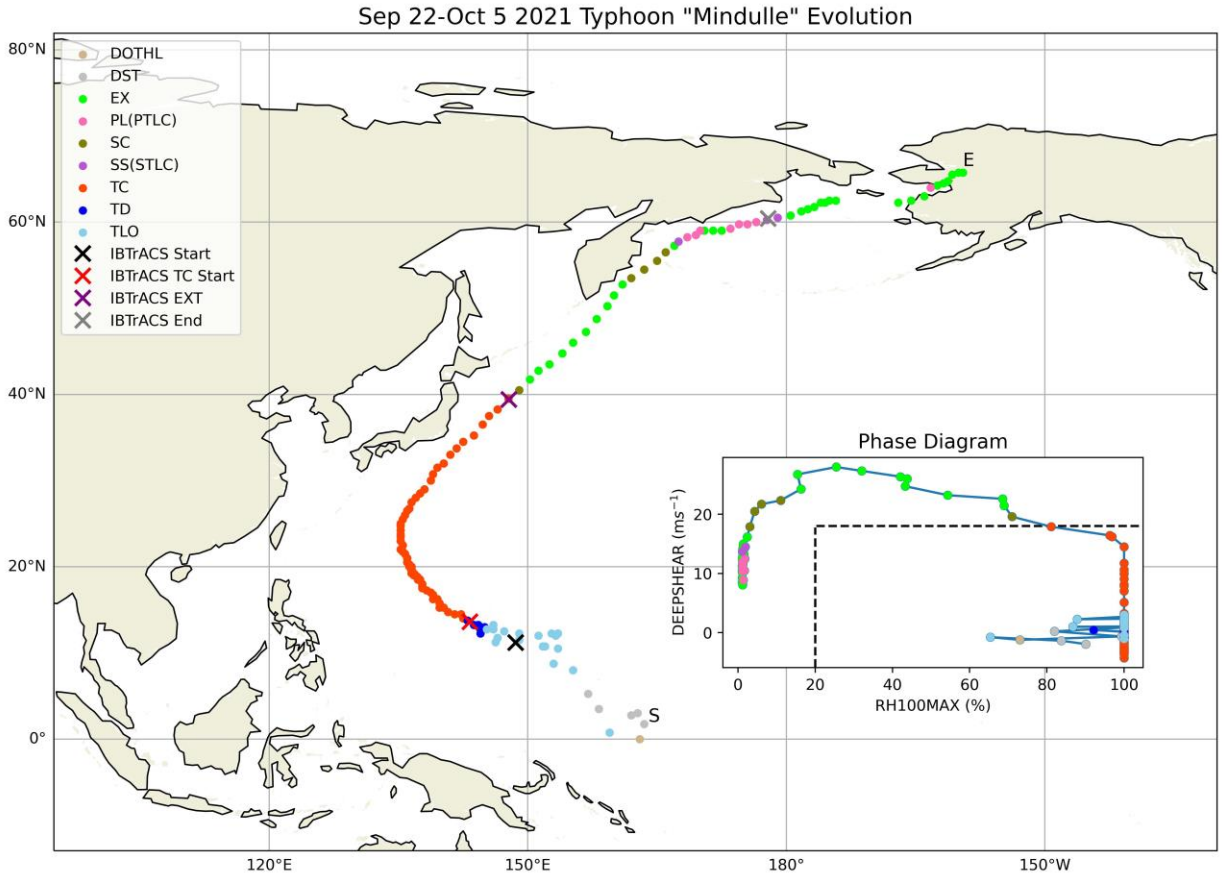



**Figure 2.10**. An example of different LPS labels in a TC (2021 Typhoon "Mindulle") lifetime. The phase diagram shows its evolution on the RH100MAX-DEEPSHEAR coordinate with the tropical condition threshold outlined in dashed lines. We convert all instances of supersaturation of RH100MAX to 100% in the phase diagram. S and E indicate the start and the end of the track. The cross marks indicate the position of the start of IBTrACS record (black), the first IBTrACS TC record (red), IBTrACS extratropical transition completion (purple), and the end of IBTrACS record (gray).

The labeled nodes can also be combined with the LPS size blobs we generated when computing LPSAREA to derive the accumulated integrated kinetic energy (IKE; Powell & Reinhold, 2007) of targeted LPSs. Fig. 2.S6 shows an illustration of the labeled LPS size blobs. IKE is directly correlated with the potential destructiveness of LPSs, as it takes the size parameter into account. The IKE of an LPS is defined as the 1 mdeep mean kinetic energy at the surface level (here approximated by the 925 hPa level) within the LPS extent we define for LPSAREA. Accumulated IKE of an LPS can be useful to study

trends in LPS activity (Kreussler et al., 2021). In Fig. 2.11, we show the accumulated IKE (in trillion joules, TJ) of the four types of high-impact LPS nodes from 1979 to 2022. Specifically, blobs associated with TCs (TC nodes) in TC-labeled tracks,

TDs and TLOs (MD, ML, TD and TLO nodes) in MS-labeled tracks, SSs in SS-labeled tracks, and PLs in PL-labeled tracks are selected, respectively, for their IKE accumulations. The results indicate that TCs have the most widespread and severe wind impact over land, while the kinetic energy of MSs accumulates the most along the coast of the Bay of Bengal. SSs are kinetically active in several hotspots globally. The influence from their winds extends to places including the east coast of the United States, southeast Australia and New Zealand, southern Chile, northern Japan, and the Mediterranean coasts, among other. Besides the Antarctic region, PLs pose the greatest threats to the coasts of the Nordic Seas and the Gulf of Alaska. IKE's spatial distribution patterns may appear different from the LPS frequencies because IKE is storm size-sensitive. For example, even though high TC frequencies are found concentrated in the Eastern Pacific basin (Fig. 8a), TC IKE is far more prominent in the western North Pacific basin due to its largest mean observed TC size among all major basins (Chavas & Emanuel, 2010). Similarly, IKE for SS in the Mediterranean Sea appear much smaller compared to other hotspots like the western North Pacific and the northwest Atlantic, as TLCs in an open ocean basin can be relatively larger without topographic constraints. Those larger TLCs can possibly be embedded TLCs or "twin-cyclones" like the one shown in Fig. 2.S4, and their existence is documented in many case studies in the two basins (e.g., Yamamoto, 2012; Fu et al., 2018; Yokoyama & Yamamoto, 2019).

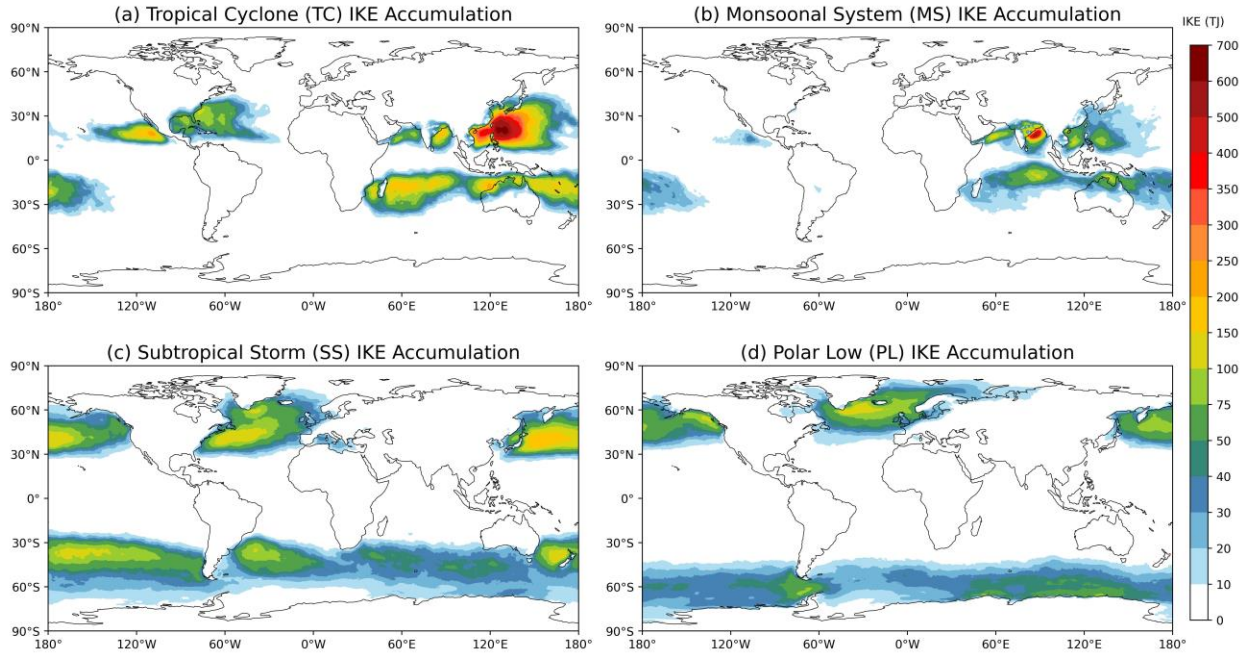



**Figure 2.11**. Integrated Kinetic Energy (IKE) of (a) tropical cyclone (TC), (b) monsoonal system (MS), (c) subtropical storm (SS), and (d) polar low (PL) accumulated over the 1979-2022 period.

Objectively tracked LPSs are often used in fractional precipitation contribution studies to tease out the contribution of each LPS type to the total precipitation (e.g., Prat & Nelson, 2013; Prein et al., 2023). The outputs from our framework could be a good source for this purpose, as precipitation blobs can be derived and labeled in a similar manner as for the size blob. Blobs (areas) that satisfy the smoothed 850 hPa cyclonic relative vorticity (CRV) threshold (CRV > $2 \times 10^{-5}$ $s^{-1}$) and a minimum 3-hourly total precipitation threshold of 0.3 mm per 3 hours (0.1 mm $hr^{-1}$) are highlighted as LPSassociated precipitation and tagged with LPS labels (see Fig. 2.S6). We consider this dynamic precipitation detection method more flexible than a fixed or uniform radius method that was often implemented in previous studies (e.g., Dare et al., 2012; Stansfield et al., 2020). We select each class of high-impact LPS nodes and their associated precipitation blobs in the same way as we do for IKE. We demonstrate the fractional contribution of precipitation from the four types of high-impact LPS nodes in Fig. 2.12. The results suggest that TCs contribute over 40% of total precipitation along the coasts of northwestern Australia and south of Baja California. MSs are responsible for a larger fraction of total precipitation than TCs

throughout South Asia and inland China. SSs make up about 5% of total precipitation along the coastal region of the Mediterranean Sea and about 6-7% near northern Japan and the coasts along the Japan Sea. PLs are responsible for several percent of total precipitation in the United Kingdom, northern Europe, and along the coast of Alaska. Since TLCs are active in the winter season, one may expect heavy snowfall as the form of their precipitation.

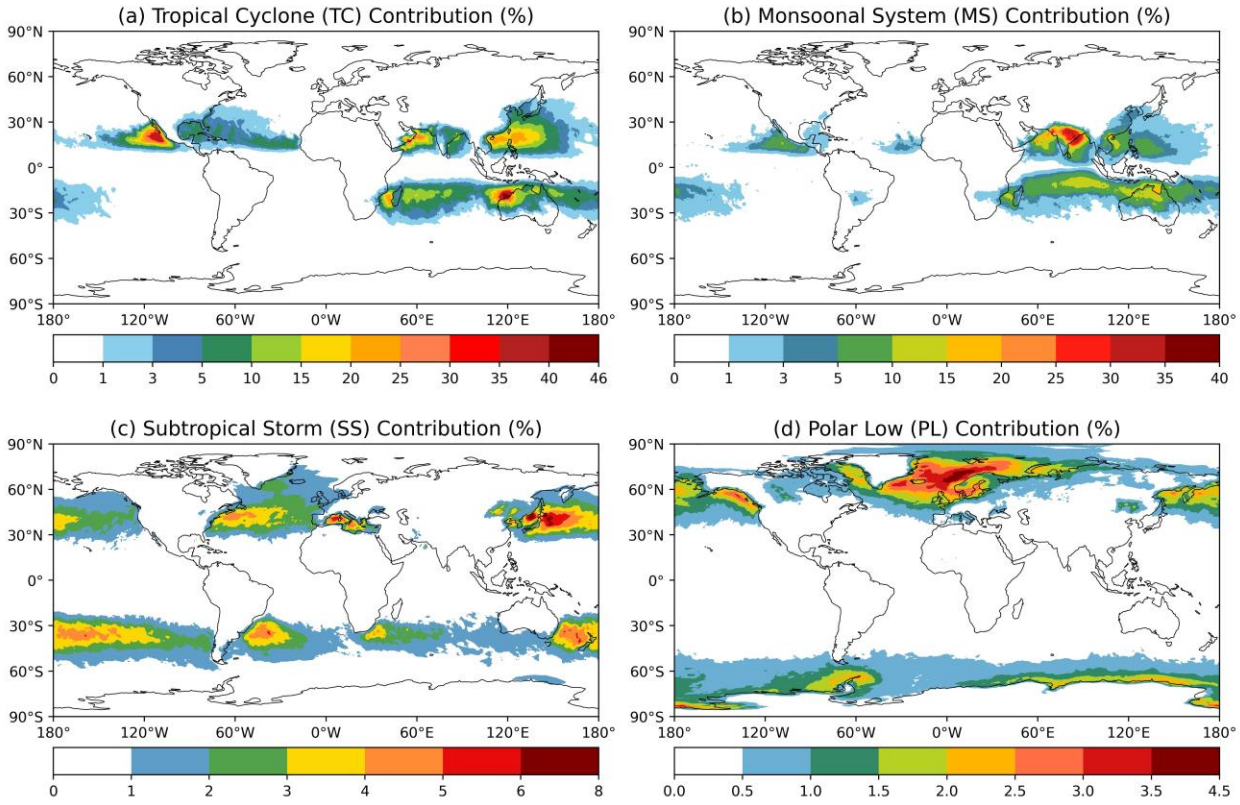



**Figure 2.12**. Fractional precipitation contributions from (a) tropical cyclone (TC), (b) monsoonal system (MS), (c) subtropical storm (SS), and (d) polar low (PL) for the 1979-2022 period.

### 2.7 Summary and Conclusions

In this study, we propose an all-in-one detection and classification framework which we refer to as the System for Classification of Low-Pressure Systems (SyCLoPS). In essence, SyCLoPS combines multiple specialized LPS detectors into a single framework, which has the effect of improving classification performance for individual features through a comprehensive treatment of boundary cases. SyCLoPS is developed atop the TempestExtremes (TE) software package (Ullrich & Zarzycki, 2017;

Ullrich et al., 2021; Ullrich, 2024). It is tuned and subsequently applied to the ERA5 reanalysis. To the authors' best knowledge, this work represents the first attempt to classify all LPSs in a single global dataset. Classification parameters and thresholds are set using conventional definitions, observations, and physical intuition, with some assistance from mathematical optimization and basic machine learning methods.

To summarize, SyCLoPS performs detection and classification following two main steps: First is to detect and track LPSs spanning all intensities globally and compute the input atmospheric parameters as shown in Appendix 2C using TE commands. Second, a Python classifier is applied to construct the input catalog and take in the input parameters for classifying (labeling) all detected LPS nodes and tracks following the workflow shown in Figure 2.1. The classifier also incorporates additional information about the nodes and tracks, such as extratropical transition nodes and whether the LPS track should be considered quasi-stationary, into the final classified LPS catalog together with the LPS node and track labels. The classification process of SyCLoPS features the following novelties:

1. Because a single intuitive workflow is employed, all LPS nodes are assigned one of the 16 LPS class labels at once. No LPS node is repeated or doubly classified. Consequently, LPS phase evolution can be traced using the labeled nodes as shown in Figure 2.10.
2. No topographical, latitudinal, or temporal restrictions need to be applied in order to use this framework, and the detection threshold is low enough to include very weak LPS nodes in the detected LPS tracks. As a result, a much more complete LPS lifecycle can be obtained for each LPS track.
3. To address the insufficiency of using warm-core criteria to disentangle tropical systems, SyCLoPS proposes a new way to objectively and physically define tropical and non-tropical systems using 100 hPa level relative humidity and large-scale deeplayer wind shear.
4. SyCLoPS features the first global tropical-like cyclone (subtropical storms and polar lows) detection scheme by detecting similar physical features to tropical cyclones among non-tropical system candidates.

Using the LPS track labels in the classified catalog, our results show that the unified framework improves upon the previously reported best TC detection skill with TE (Zarzycki & Ullrich, 2017; Bourdin et al., 2022) by both increasing HR and lowering FAR. We found that the lowered FAR is mainly contributed by successfully identifying subtropical storms in the subtropics. Detection skill for MS is comparable to previous work (Vishnu et al., 2020). Upon comparing the labels given by SyCLoPS to the corresponding IBTrACS labels, we observe that SyCLoPS can reasonably label the LPS status in different stages of a TC. We also demonstrate that the resulting classified catalog can be used to study the evolution, annual frequencies, vertical cross-section composites, IKE accumulation, and fractional precipitation contribution of each LPS class. These potential applications are useful for general LPS studies, and it could be valuable if applied to climate model outputs to investigate the trends of LPS activities under the effects of climate change. SyCLoPS may also be applicable in real-time operations and weather model outputs. Furthermore, information in the classified catalog may be used as data labels to train a machine learning model. We intend to pursue additional research tasks on those topics using SyCLoPS.

Users may also personalize the framework to meet their own needs with the data and codes provided in SyCLoPS. For example, the detection procedure for a single type of LPS in a given dataset can be easily isolated following a single path from the workflow. Users may also modify the TE specifications to optimize detection for a specific type of LPS in a specific region. More LPS sub-classes may be derived from SyCLoPS catalogs and the detection of other atmospheric features. For example, weaker tropical LPSs may be separated into different classes by matching them to distinct tropical waves in the lower troposphere, and TLCs may be divided into those that develop in a front-shear environment versus a reverse-shear environment.

There are some notable limitations in SyCLoPS. Firstly, SyCLoPS is based upon ERA5 dataset, so the thresholds and parameter choices could be biased if applied directly to another global or regional dataset. While we have proposed some suggestions that would allow SyCLoPS to adapt to different datasets, and have shown in Appendix 2B that SyCLoPS's performance is stable across different

adaptation experiments, more fine-tuning in the detection and classification processes may be required. It should also be noted that a dataset with a resolution coarser than ERA5 may be insufficient for detection and classification of smaller features such as early-stage TCs and TLCs. Second, although SyCLoPS features detection and classification of LPSs over any terrain, signals in MSLP or any lowlevel atmospheric fields can be distorted by elevated or rough topography. Detection over or near those regions are subject to greater errors, especially for weak systems. Third, ultimately, the hard cut-off threshold we impose between LPS phases is somewhat arbitrary: namely, there is always a gray zone or transition zone when it comes to the thresholds for a given LPS. Nonetheless, objective LPS detection and classification reduces biases introduced by human error and subjectivity because an objective standard can be strictly followed. However, by nature an LPS can exist in an "impure" and somewhat ambiguous state, which is contrary to fixed thresholds. This conflict is most obvious when a detected LPS persists at the edge of our defined thresholds, leading to its classification jumping between two labels. And lastly, confidence in detecting and classifying global TLC systems is still low due to a lack of global observations. The method for calculating LPS size in TE as described in Appendix 2E for classifying TLCs can be further improved to more accurately represent the size of a smaller TLC in a larger circulation or the background flow. We expect that there will be other deficiencies discovered and questions raised in the practical use of this experimental framework. Hence, we aim to address the remaining issues and evolve the algorithms for future versions of this framework.

## 2.8 Appendixes

### 2.8.1 Appendix 2A: Variables Used in SyCLoPS

Although not listed in Table 2.A1, 10-m U and V-component Wind (VAR_10U and VAR_10V) are used in SyCLoPS to calculate the maximum surface wind speed (WS) of an LPS as a reference information in the classified catalog. It does not play a role in the detection and classification process.

Although SyCLoPS is tested on 3-hourly data, as suggested in Appendix 2B below, it's possible to use 6-hourly data or even lower frequency data for parameters that change more slowly and smoothly. The latest SyCLoPS software documented in Chapter 2.10 automatically adjusts to the optimal configuration for 6-hourly frequency detection.

| Variable Name | Pressure Level (hPa) |
|---|---|
| U-component Wind (U) | 925, 850, 200[b] |
| V-component Wind (V) | 925, 850, 200[b] |
| Temperature (T) | 850 |
| Relative Humidity (R)[a] | 850, 100 |
| Mean Sea Level Pressure (MSL) | Sea Level |
| Geopotential (Z) | 925, 850, 700, 500, 300[b] |

a. Specific humidity can be converted to R with additional temperature information.
b. These default levels can be replaced by 250 hPa. See Appendix 2B for details.

**Table 2.A1**. Variables Used in SyCLoPS

### 2.8.2 Appendix 2B: Experiments for Climate Model Data Adaptations

Four adaptation experiments are performed on SyCLoPS to test how SyCLoPS can be adapted to a typical high-resolution climate model output. In these high-resolution models (e.g. those in the HighresMIP submission), RH at 100 hPa is usually only available in daily frequency, so our first experiment resamples ERA5 3-hourly RH data to daily frequency (daily-mean) and recalculates RH100MAX for classification (i.e, each grid point will have the same RHMAX100 for 8 3-hourly time steps in a day). Most climate models only output 6-hourly frequency (instead of 3) for many of the fields we use in SyCLoPS. Therefore, the second experiment resamples the 3-hourly frequency input catalog of SyCLoPS to 6-hourly frequency (similar to using a 6-hourly detection rate) before performing the same

classification. In addition, the number of nodes (time steps) needed to satisfy the TC, MS, SS and PL track conditions is halved to reflect the frequency change. For example, four TC nodes (instead of eight) in a track are required for the TC track condition under 6-hourly frequency data.

Finally, some climate model outputs only have 250 hPa data instead of the 200 hPa data (used in DEEPSHEAR and WS200PMX) or 300 hPa data (used in UPPTKCC) we use in SyCLoPS. Therefore, the third experiment opts to use 250 hPa data in ERA5 and the subsequent adjustments to the classification conditions are needed to achieve the optimal detection skills (or to match our original definition to justify the polar jet/front):

1. In the tropical condition, the DEEPSHEAR threshold is changed to: DEEPSHEAR < 13 m $s^{-1}$
2. TC Condition is changed to: MSLPCC20 > 225 Pa, LOWTKCC < 0 $m^2$ $s^{-2}$ and UPPTKCC < -147 $m^2$ $s^{-2}$.
3. TC Track Condition is changed: At least 6 3-hourly TC-labeled nodes in a track are required to satisfy the TC Track Condition.
4. WS200PMX should be renamed to WS250PMX with the following changes: In the SC condition, use WS250PMX > 35 m $s^{-1}$. And to separate SS and PL, use WS250PMX > 30 m $s^{-1}$

The results of these adaptation experiments are summarized in the table below. As shown, the detection skills and labeling accuracy of the three adaptation experiments are generally very similar to the original algorithm, suggesting the flexibility of SyCLoPS in adapting to climate model data. It's also not surprising that the TLC detection skill degrades more noticeably at a lower sampling frequency in the 6-hourly data frequency experiment given that many TLCs are transient.

| Method | TC Detection | Label Accuracy | MS Detection | TLC Detection |
|---|---|---|---|---|
| Original | HR = 78.2%<br>FAR = 14.6%<br>HRMFAR = 63.6% | TS: 96.6%<br>TC: 74%<br>EX: 78% | CSI = 0.83 | HR ≈ 80%<br>IR = 11.9% |
| Daily-mean RH100MAX | HR = 78.0%<br>FAR = 14.7%<br>HRMFAR = 63.3% | TS: 96.6%<br>TC: 74%<br>EX: 79% | CSI = 0.83 | HR ≈ 80%<br>IR = 11.7% |
| 6-hourly Frequency | HR = 78.4%<br>FAR = 15.2%<br>HRMFAR = 63.2% | TS: 96.6%<br>TC: 75%<br>EX: 78% | CSI = 0.82 | HR ≈ 70%<br>IR = 14.8% |
| 250 hPa Parameters | HR = 78.7%<br>FAR = 15.1%<br>HRMFAR = 63.6% | TS: 96.3%<br>TC: 72%<br>EX: 84% | CSI = 0.83 | HR ≈ 80%<br>IR = 11.6% |

**Table 2.B1**. Adaptation Experiment Performance

### 2.8.3 Appendix 2C: Catalog Column Table

| Column | Unit | Description |
|---|---|---|
| TID | - | LPS track ID (0-based) in both the input and classified catalog |
| ISOTIME | - | UTC timestamp (ISO time) of the LPS node in both catalogs |
| LAT | ◦ | Latitude of the LPS node in both catalogs |
| LON | ◦ | Longitude of the LPS node in both catalogs |
| MSLP | Pa | Mean sea level pressure at the LPS node in both catalogs |
| MSLPCC20 | Pa | Greatest positive closed contour delta of MSLP over a 2.0° GCD (the core of an LPS) |
| MSLPCC55 | Pa | Greatest positive closed contour delta of MSLP over a 5.5° GCD |
| DEEPSHEAR | $m\ s^{-1}$ | Average deep-layer wind speed shear between 200 hPa and 850 hPa over a 10.0° GCD |
| UPPTKCC | $m^2\ s^{-2}$ | Greatest negative closed contour delta of the upper-level thickness between 300 hPa and 500 hPa over a 6.5° GCD, referenced to the maximum value within 1.0° GCD |
| MIDTKCC | $m^2\ s^{-2}$ | Greatest negative closed contour delta of the middle-level thickness between 500 hPa and 700 hPa over a 3.5° GCD, referenced to the maximum value within 1.0° GCD |
| LOWTKCC | $m^2\ s^{-2}$ | Greatest negative closed contour delta of the lower-level thickness between 700 hPa and 925 hPa over a 3.5° GCD, referenced to the maximum value within 1.0° GCD |

| | | |
|---|---|---|
| Z500CC | $m^2\ s^{-2}$ | Greatest positive closed contour delta of geopotential at 500 hPa over a 3.5° GCD referenced to the minimum value within 1.0° GCD |
| VO500AVG | $s^{-1}$ | Average relative vorticity over a 2.5° GCD |
| RH100MAX | % | Maximum relative humidity at 100 hPa within 2.5° GCD |
| RH850AVG | % | Average relative humidity over a 2.5° GCD at 850 hPa |
| T850 | K | Air temperature at 850 hPa at the LPS node |
| Z850 | $m^2\ s^{-2}$ | Geopotential at 850 hPa at the LPS node |
| ZS | $m^2\ s^{-2}$ | Geopotential at the surface at the LPS node |
| U850DIFF | $m\ s^{-1}\ sr$ | Difference between the weighted area mean of positive and negative values of 850 hPa U-component wind over a 5.5° GCD |
| WS200PMX | $m\ s^{-1}$ | Maximum poleward value of 200 hPa wind speed within 1.0° GCD longitude |
| RAWAREA | $km^2$ | The raw defined size (see Appendix 2E) of the LPS |
| LPSAREA | $km^2$ | The adjusted defined size of the LPS in both catalogs |
| WS | $m\ s^{-1}$ | Maximum wind speed at the 10-m level within 2.0° GCD |
| Full_Name | - | The full LPS name based on the classification |
| Short_Label | - | The assigned LPS label (the abbreviation of the full name) |
| Tropical_Flag | - | 1 if the LPS is designated as a tropical system, otherwise 0 |
| Transition_Zone | - | 1 if the LPS is in the defined transition zone, otherwise 0 |
| Track_Info | - | "TC", "MS", "SS(STLC)", "PL(PTLC)", "QS" denoted for TC, MS, SS, PL and QS tracks; "EXT", "TT" denoted for extratropical and tropical transition completion nodes |
| IKE | TJ | The integrated kinetic energy computed based on the LPS size blobs that are used to define RAWAREA |

**Table 2.C1**. Column Table for the Input Catalog (Top) and Classified Catalog (Bottom)

### 2.8.4 Appendix 2D: Condition List

| Condition Name | Conditions |
|---|---|
| High-altitude Condition | Z850 > ZS |
| Dryness Condition | RH850AVG < 60% |
| Cyclonic Condition | VO500AVG ≥ (<) 0 $s^{-1}$ if LAT ≥ (<) 0º |
| Tropical Condition | RH100MAX > 20%; DEEPSHEAR < 18 m $s^{-1}$; T850 > 280 K |
| Transition Condition | Tropical Conditon = True; DEEPSHEAR > 10 m $s^{-1}$ or RH100MAX < 55% |
| TC Condition | MSLPCC20 > 215 Pa; LOWTKCC < 0 $m^2 s^{-2}$; UPPTKCC < -107.8 $m^2 s^{-2}$ |
| TD Condition | MSLPCC55 > 160 Pa; UPPTKCC < 0 $m^2 s^{-2}$ |
| MS Condition | U850DIFF > 0 m $s^{-1}$; RH850AVG > 85% |
| TLC Condition[a] | MSLPCC20 > 190 Pa; LOWTKCC and MIDTKCC < 0 $m^2 s^{-2}$; (LPSAREA < 5.5 × $10^5$ $km^2$; LPSAREA > 0 $km^2$) or (MSLPCC20 > 420 Pa; MSLPCC20 : MSLPCC55 > 0.5) |
| SC Condition | LOWTKCC < 0 $m^2 s^{-2}$; Z500CC > 0 $m^2 s^{-2}$; WS200PMX[b] > 30 m $s^{-1}$ |
| TC Track Condition | At least 8 (8+) 3-hourly TC-labeled nodes in an LPS track |
| MS Track Condition | 10+ 3-hourly "TLO(ML)" or "TD(MD)"-labeled nodes in an LPS track |
| SS Track Condition | 2+ 3-hourly TLC-labeled nodes ("SS(STLC)" or "PL(PTLC)") and 1 "SS(STLC)"-labeled node in an LPS track |
| PL Track Condition | 2+ 3-hourly TLC-labeled nodes ("SS(STLC)" or "PL(PTLC)") and 1 "PL(PTLC)"-labeled node in an LPS track |
| QS Track Condition | See Text 2.S3 for details |

a. See Chapter 2.5.3 for a potential alternative.
b. WS200PMX criteria used in this framework may be supplemented by other parameters in some regional models. See Text 2.S4 for details.

**Table 2.D1**. Classification condition

### 2.8.5 Appendix 2E: LPSAREA Computation

To calculate LPS size, we refer to the definition of TC size which is typically determined by a TC's outer surface wind radius. We first use TE to detect blobs (areas) of smoothed 850 hPa cyclonic relative vorticity (CRV) > 2 × $10^{-5}$ $s^{-1}$ and 925 hPa wind speed > 12 m $s^{-1}$. An alternative condition to this detection requirement is CRV > 4 × $10^{-5}$ $s^{-1}$ so that TC eyes and EXs' central weaker wind areas can be captured. Wind speed from 925 hPa is used for this calculation, for it is a commonly found lower

model level above the surface level. Surface level winds are not used since they can be greatly distorted by complex topography. The 12 m $s^{-1}$ threshold is obtained using a log wind profile from the 8 or 9 m $s^{-1}$ surface outer wind speed threshold often found in TC-size-related studies using ERA5 or climate models (e.g., Stansfield et al., 2020; Bian et al., 2021). The smoothed CRV field is used to control the boundary of an LPS so that the outer wind fields are less likely to connect with an unrelated system nearby. Each detected size blob is then assigned to a detected LPS node if the node is within 5° GCD of the centroid of the blob at the same timestamp, or otherwise within the region bounded by the minimum/maximum latitude/longitude (extent) of the blob. Information on the centroid, extent, and size (area) of each blob can be directly output by TE. If multiple nodes are found for one blob, the blob is assigned to the node with the lowest MSLP. Next, the areas of all the blobs paired with each node are added together as raw LPS sizes (areas).

To avoid misclassifying EXs/SCs as TLCs near shorelines with elevated topography, we adjust the raw LPS size if the LPS is close to those shorelines with its wind field largely affected by topography. Specifically, we multiply the raw LPS size by 2 if only 30 to 70% of surface geopotential within 5° GCD of an LPS is smaller than 7000 $m^2 s^{-2}$ (approximately the 925 hPa level). This adjusted LPS size is defined as LPSAREA (in $km^2$). For a quick comparison, the sizes of 2010-2021 western North Pacific TCs computed by our method and the sizes given by the Japan Meteorological Agency in IBTrACS have a reasonably high correlation coefficient of 0.63, with very high statistical significance. Furthermore, the average and median global PL size (in radius) given by LPSAREA is 231 and 206 km. which are very close to the 221 and 200 km suggested by the PL radius records in the STARS dataset. The non-adjusted (raw) LPS size computed by this method is included as the "RAWAREA" column in the input catalog. LPSAREA parameter is automatically computed for all LPSs in the SyCLoPS software.

## 2.9 Supplements

### Text 2.S1. Track Matching in the Objective ERA5 LPS dataset

TC track matching: A match between the ERA5 LPS dataset and IB-TC is realized when they lie within 2° GCD of each other at the same timestamp. 98% of the 3510 IB-TC tracks are matched, which forms the matched TC dataset. Additionally, we obtain the matched tropical LPS dataset, of which 96% of the 4495 tracks are matched, by comparing it to main-type tracks in 1979-2021 IBTrACS without the 34 knots wind speed filter. This unfiltered version of IBTrACS is also used for the LPS label POS computation described in Sec. 3 and the selection of the tropical/subtropical cluster discussed in Text 2.S2.

MS track matching: Detected ERA5 LPS tracks having at least 5 nodes lying within 5° GCD of a record in the daily-frequency Sikka dataset on the same date are considered matches. 88% of the 350 Sikka tracks are matched, which forms the matched MS dataset.

SS track matching: A match between the ERA5 LPS dataset and the ERA5 manually-tracked Mediterranean cyclone dataset is realized when they lie within 2° GCD of each other at the same timestamp. 98% of the 129 Mediterranean cyclone tracks, which consist of mainly SSs and SCs, are matched. Then, potential SSs (STLCs) are chosen by the method described in Section 2, forming the matched SS dataset.

PL track matching: We first round the hourly timestamps in the STARS archive to the nearest 3-hourly timestamps. Detected ERA5 tracks having the most nodes lying within 3° GCD of a record in the STARS dataset at the same timestamp are considered matches. 63% of the 186 STARS tracks are matched, which forms the matched PL dataset.

EW track matching: Detected ERA5 LPS tracks having the most nodes lying within 5° GCD of a record in the objectively-tracked EW dataset at the same timestamp are considered matches. 44% of the 1396 easterly wave tracks are matched, which forms the matched EW dataset.

### Text 2.S2. Selection of the Tropical/Subtropical Cluster and Sensitivity Test

To make the two clusters more alike in terms of their latitudinal locations, warm cores, and underlying SSTs, we first select a set of candidates LPS tracks that are more likely to be largely non-tropical/subtropical and have a more stable warm core for the subtropical cluster:

1. The track is within the year range of 1979-2021 and cannot be a match to any tracks in the non-filterd IBTrACS.
2. The track must begin poleward of 20° latitude from the equator to demonstrate a potential non-tropical origin.
3. The track must have at least five time steps within 25° to 50° latitudes from the equator.
4. The track must establish at least five time steps with both upper- and lower-level warm cores (UPPTKCC $< 0$ $m^2 s^{-2}$ and LOWTKCC $< 0$ $m^2 s^{-2}$).

Tracks in the non-filtered IBTrACS are considered candidates for the tropical cluster. Next, we choose LPS nodes from both candidate track sets that meet the subtropical latitudinal requirement of 20° to 45° from the equator (a range of latitudes that frequently host both TCs and SCs), the warm-core requirement of UPPTKCC $< 0$ $m^2 s^{-2}$ and LOWTKCC $< 0$ $m^2 s^{-2}$, and a potential minimum underlying SST (data from ERA5) requirement to only include subtropical cluster's LPSs over a comparable SST to that of the tropical cluster. According to observations, almost all TCs with a maximum wind speed over 35 knots in IBTrACS are over 288 K SSTs, and over 95% of those TCs are over 293 K SSTs (Stansfield & Reed, 2021). We find that the final classification could be sensitive to the SST requirement chosen. Hence, we select a range of minimum SST requirements from 288 K to 295 K with an interval of 0.5 K to perform a sensitivity test on the optimal thresholds needed for the decision trees.

As shown in Fig. 2.S2, DEEPSHEAR is not sensitive to the SST threshold we choose for the requirement, and its threshold is stable at around 10 m $s^{-1}$. Beyond 293.5 K, the decision trees with a depth of 2 only split on DEEPSHEAR (RH100MAX becomes indecisive) as the sample size of the subtropical cluster decreases more quickly, making the two clusters more imbalanced. This could mean

that a larger proportion of tropical LPS nodes are falsely included in the subtropical cluster at this point. For the SST range of 288-293.5 K, sample sizes of the two clusters are comparable, ranging between 50 and 80 thousand. RH100MAX thresholds (shown with the blue line) within this range are averaged at 20% and they are also most stable around the 20% level after being rounded off to the nearest 5% as indicated by the gray line. Hence, RH100MAX=20% and DEEPSHEAR=10 m $s^{-1}$ is considered the optimal combination to separate the two clusters. Accuracy scores for these decisions range from 77% to 80%. We use the threshold of SST ≥ 291 K to produce an example result for Fig. 2.3. The dashed line in Fig. 2.S2 shows the decision tree with a depth of 1 when only RH100MAX is considered. The threshold of RH100MAX in this situation is more stable at about the 50-55% levels, with a roughly 74% accuracy score.

**Text 2.S3. Quasi-stationary Track Condition**

We calculate three additional track parameters in the Python classifier, which are track linearity, track spread, and track inland ratio, to label QS tracks. Track spread is the standard deviation of the distance between each node in a track and the first track node; track linearity is the Pearson correlation coefficient of all the nodes within a track on the latitude-longitude coordinate; and track inland ratio is the percentage of nodes located within 1° GCD of any inland areas, which are defined as those grid points with surface geopotential exceeding 150 $m^2 s^{-2}$. One may expect a typical quasi-stationary LPS to be meandering and bouncing around some topographical features within a limited region. Hence, the linearity of those tracks should be low, the deviation of track nodes from the first detected track node should be rather small, and they reside mostly inland or close to shorelines. We select a set of quasi-permanent LPSs near Colombia in South America as being representative of quasi-stationary tracks. They exist year-round, potentially due to the positive feedback between rainforest evapotranspiration and mesoscale convection (Poveda et al., 2014). Drawing from the intuition mentioned above, we optimize the three quasi-stationary thresholds of the three parameters by maximizing the filtering of the Colombia LPSs and minimizing the overlap between the matched TC tracks and the selected quasi-stationary tracks. The final decision for the quasi-

stationary track condition is that a track needs to have a track linearity lower than 0.55, a track spread smaller than 3° GCD, and a track inland ratio greater than 65%.

**Text 2.S4. Alternatives to WS200PMX criteria**

In regional models, one may assume the region covers mostly polar regions or the subtropics, so the WS200PMX criteria are unnecessary (i.e., it will be either SS and SC or PL and EX in the extratropical branch). If the model domain does not cover very high latitudes (i.e., 65° or higher), in the event that the polar jet/front is further poleward of the domain limit, one may use the following alternatives when WS200PMX of an LPS is lower than 30 m $s^{-1}$ in the SC condition and 25 m $s^{-1}$ in separating PLs:

DEEPSHEAR and T850 are considered alternatives to WS200PMX. For the alternatives to the SC condition, the decision tree classifier selects the WS200PMX alternative criterion in the SC condition based on two clusters of 100 thousand node samples selected using the WS200PMX > 30 m $s^{-1}$ and WS200PMX ≤ 30 m $s^{-1}$ thresholds. The result shows that DEEPSHEAR > 14 m $s^{-1}$ and T850 > 273 K is the best combination, giving a 77% accuracy score. If temperature thresholds are deemed to be avoided under global warming scenarios, DEEPSHEAR > 12 m $s^{-1}$ alone is also acceptable with a 72% accuracy score.

The alternatives to the WS200PMX criterion for distinguishing PLs and SSs following the TLC condition are found in a similar way as in the SC condition. Results show that DEEPSHEAR < 11 m $s^{-1}$ and T850 < 273 K is the best combination, giving an 80% accuracy score. Similarly, DEEPSHEAR < 11 m $s^{-1}$ alone is also acceptable with a 77% accuracy score. One can also choose to treat PLs and SSs as a single type of LPS in a region, so this criterion is not needed for the TLC condition.

**Text 2.S5. Justification of Parameter Specifications**

It's nearly impossible to additionally optimize every parameter selection process (i.e., determining the optimized GCD used in some parameters), as it will require significantly larger time complexity to compute. Most of the chosen parameter specifications are derived directly or indirectly from previous studies. Here is a brief justification of some of the chosen parameters:

1. MSLPCC55, UPPTKCC: These are the same parameters used in the previous TE's TC tracking algorithms (Zarzycki & Ullrich, 2017), which have been optimized with respect to the GCD distance. The 1° GCD offset allowance in UPPTKCC to search for a thickness maximum is also used in Zarzycki & Ullrich (2017) based on observational intuition. They also found that the TC detection result is relatively insensitive to this offset value.
2. MSLPCC20: The 2.0° GCD specification in MSLPCC20 is chosen to represent the "core" of cyclones. For example, Weatherford and Gray (1988) defined the TC outer core as the 1° - 2.5° GCD region from the TC center. The region outside of the core may be regarded as the outer region/wind field (i.e., within the 5.5° GCD as specified in MSLPCC55).
3. VO500AVG, RH100MAX, RH850AVG: The 2.5° GCD specification is chosen in accordance with MSLPCC20 with a 0.5° GCD buffer given that they are upper-level parameters.
4. MIDTKCC, LOWTKCC, Z500CC: Rather than being assessed for a particular value, these three parameters are computed only to confirm if they are non-zero. Specifically, they indicate whether there are warm cores or a 500 hPa closed circulation close to the core of an LPS. Hence, a rather small GCD of 3.5° is used with a 1.0° offset allowance. Since we only need to know if a low-level warm core is present, it should be acceptable— though untested—to replace the 925 hPa geopotential needed for LOWTKCC with 850 hPa geopotential if some datasets or models use a vertical coordinate system with missing values below the surface.
5. DEEPSHEAR: The 10° GCD specification is chosen to be reasonably large to reflect the local large-scale environment.
6. U850DIFF: The 5.5° GCD specification is chosen to be the same as MSLPCC55.
7. WS200PMX: The 1.0° offset allowance is chosen to be the same as for the others.

**Figure 2.S1.** 1979-2022 mean RH100MAX at each grid point. The black solid-line contour is the 20% RH100MAX threshold we choose for the tropical condition.

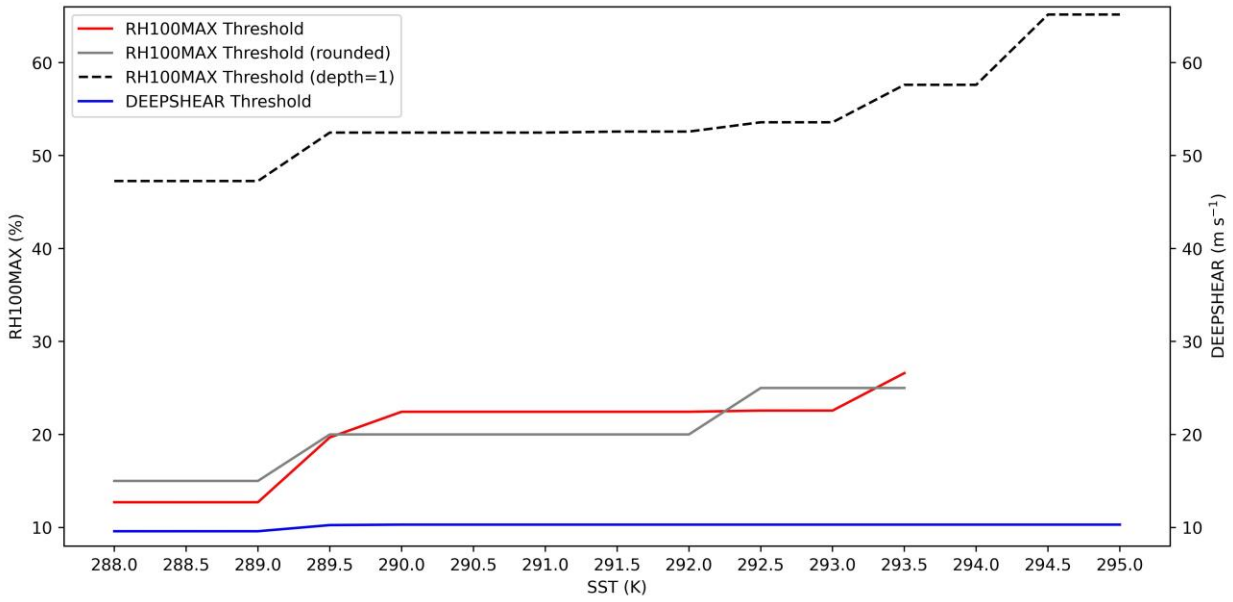
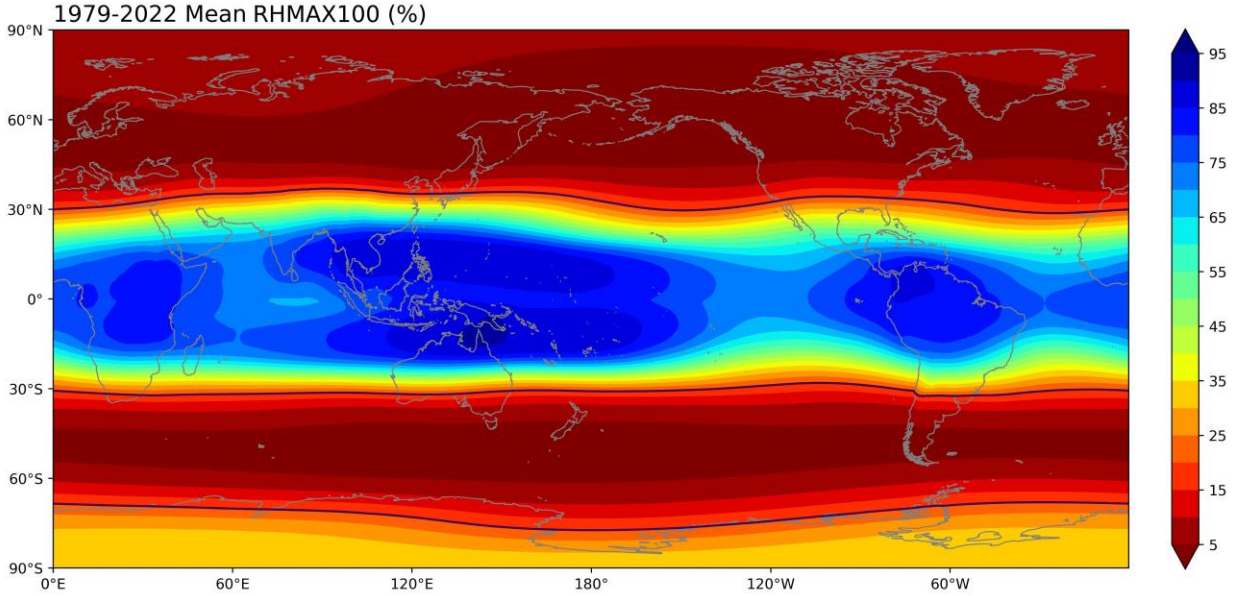



**Figure 2.S2.** Sensitivity of RH100MAX and DEEPSHEAR thresholds to the minimum SST requirement




used in the decision tree classifier. See Text 2.S2 for details.

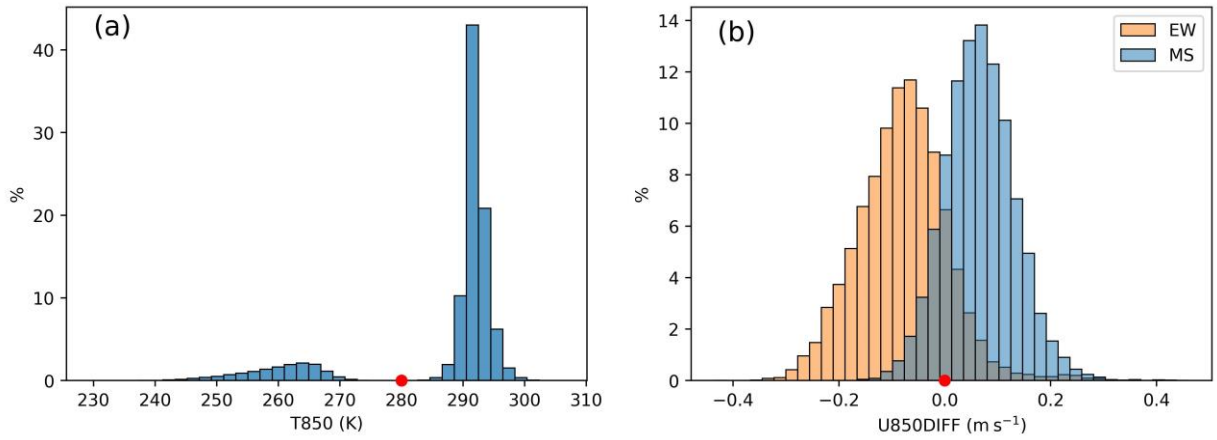



**Figure 2.S3.** (a) T850 distribution of potential tropical systems that already satisfy the RH100MAX and DEEPSHEAR thresholds of the tropical condition in the workflow, and (b) percentage distribution of U850DIFF for the matched MS dataset (blue) and EW dataset (orange). Red dots indicate the thresholds we choose for the tropical condition and the MS condition.

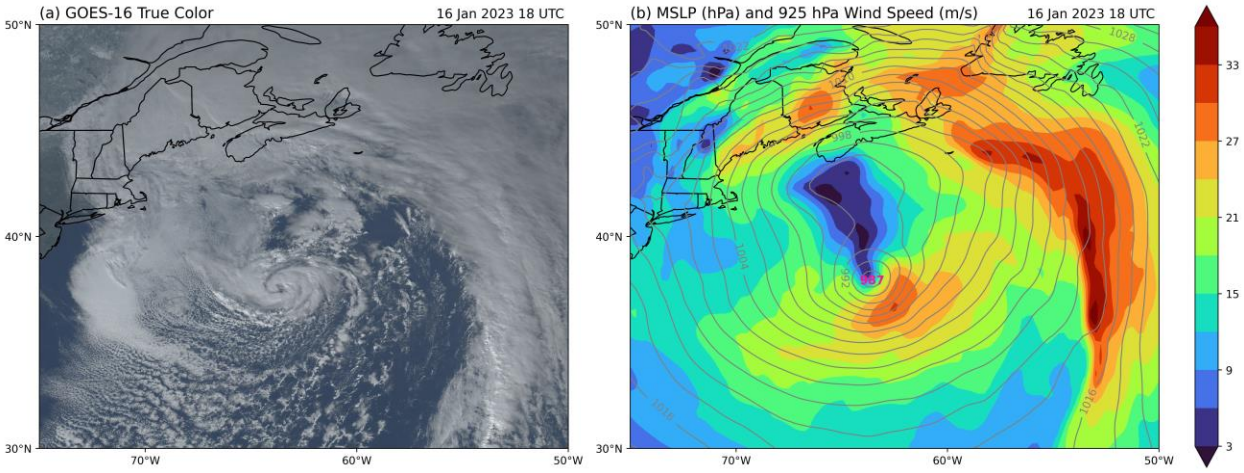



**Figure 2.S4.** An example of an "embedded" TLC on Jan 16, 2023 in the North Atlantic.

This storm was later recognized as a notable off-season subtropical storm in National Hurricane Center's 2023 best track, indicating its potential to become a "real" TC. (a) shows its visible satellite image near peak intensity. A developed eyewall structure can be clearly seen. Its 925 hPa wind and MSLP field during the same time are shown in (b). We can see that its 987 hPa core (the TLC portion) was embedded within a synoptic-scale cyclonic circulation, where winds of near-gale scale were widespread. A weaker

MSLP minima was also developing to the north of the TLC at this time, kicking off a twin-cyclone system.

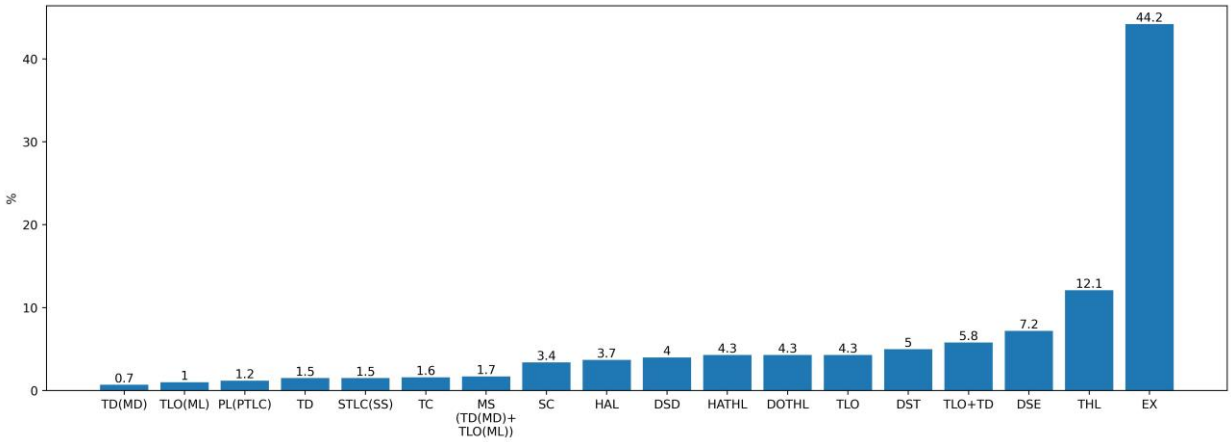



**Figure 2.S5.** Frequencies (in percentages) of each type of LPS labels in our classified LPS catalog.

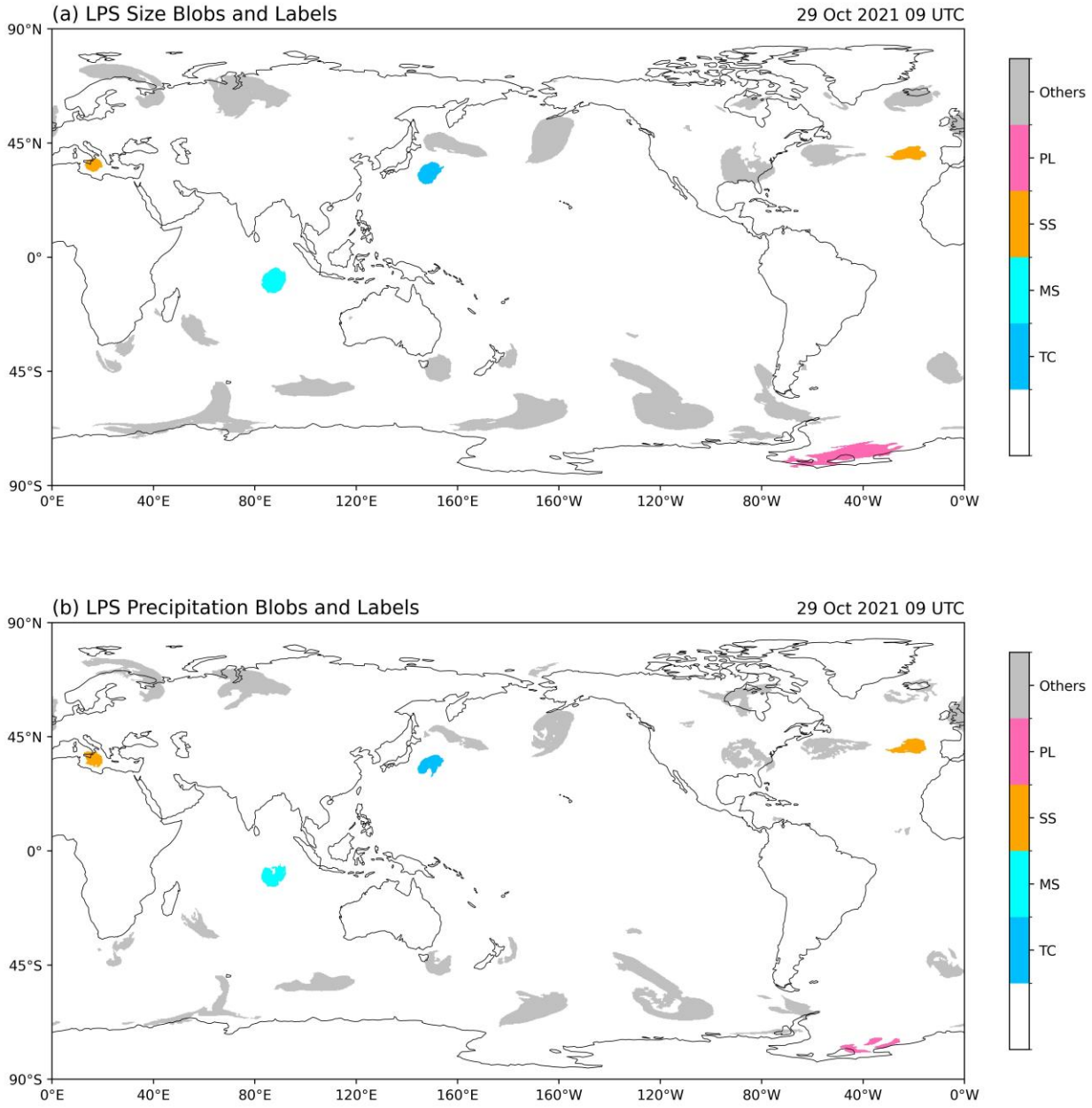



**Figure 2.S6.** (a) An illustration of LPS size blobs and (b) precipitation blobs with LPS labels when they are paired with labeled LPS nodes and tracks. We choose a time slice that contains all four types of high-impact LPS (TC, MS, SS and PL), which include 2021 Typhoon "Malou" and 2021 Cyclone “Appolo,” a.k.a. Medicane (Mediterranean hurricane) “Nearchus.” These labeled ERA5 LPS size and precipitation blob files are available in the dataset on Zenodo (see Chapter 2.10).

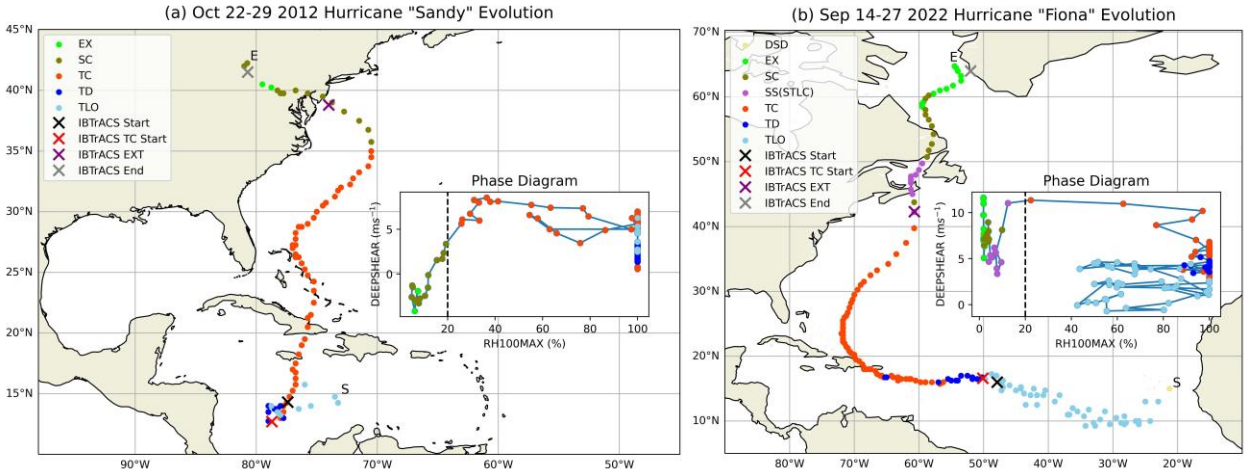



**Figure 2.S7.** As in Fig. 2.10, but for two interesting TC cases in the North Atlantic basin.

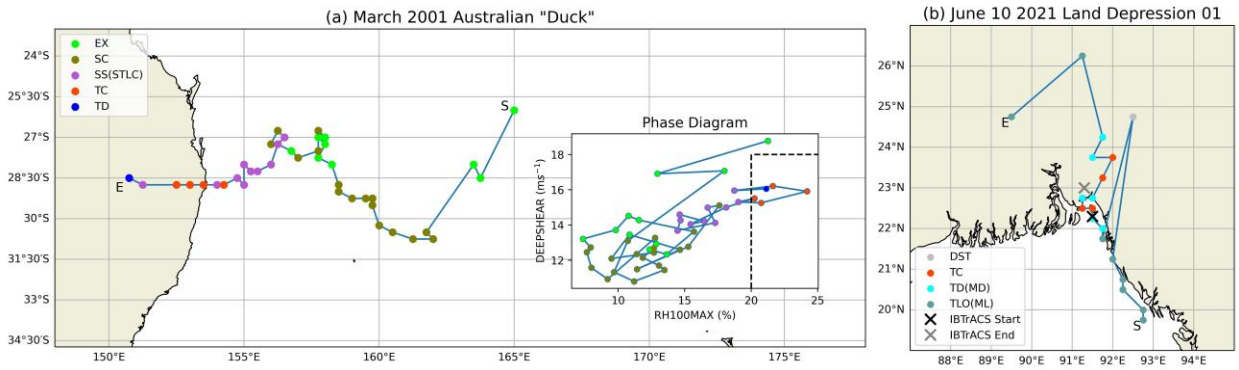



**Figure 2.S8.** As in Fig. 2.10, but for two marginal TC cases. (a) is a classic tropical transition case happened near Australia (the "Duck"), and (b) shows an example of a South Asia monsoonal system recorded as a land depression by Indian Meteorological Department in IBTrACS. They both labeled "TC" four times by our classification, but are not recorded as TCs in IBTrACS, potentially because they are too transient to be realised by the agencies and that they are close to or over land as when they form.

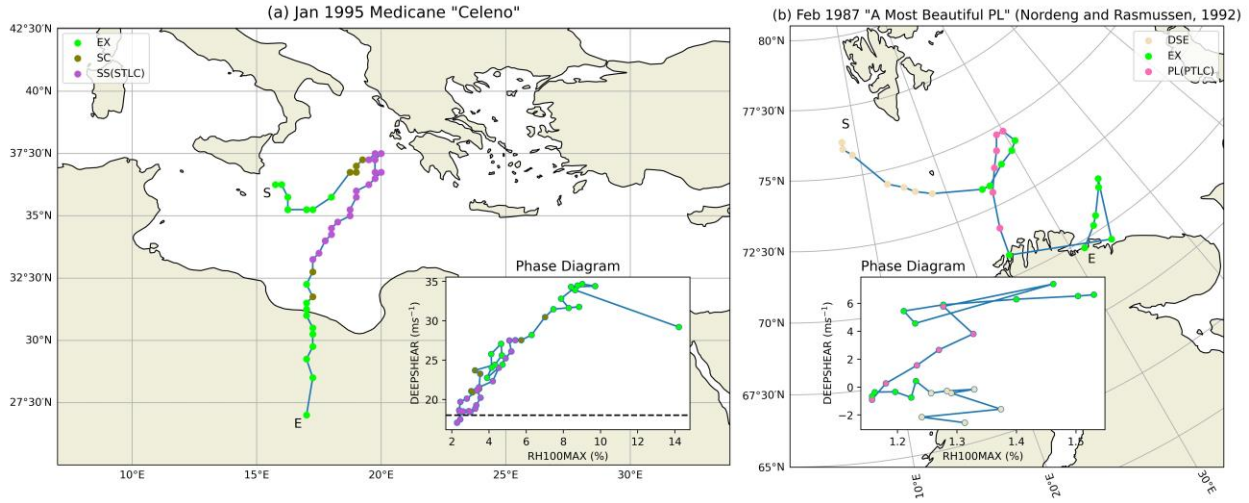



**Figure 2.S9.** As in Fig. 2.10, but for two TLC cases. (a) is a well-studied 1995 medicane, as documented in Pytharoulis et al. (2000), Emanuel (2005), and others. (b) is a polar low described as "a most beautiful polar low" in Nordeng and Rasmussen (1992), as shown by its beautiful eye on satellite images before making landfall.

## 2.10 Software and Published Datasets

The SyCLoPS software is publicly available through the GitHub repository: yepkids/SyCLoPS. The software consists of two primary components: a main driver program that wraps all TempestExtremes (TE) commands using Python subprocesses, and a classifier module that performs the low-pressure system (LPS) classification workflow. After specifying the required variable names and input files, users can execute the main program by following the prompts provided in the software. The workflow generates an intermediate LPS catalog containing all parameters required for classification, followed by a final catalog containing the classified node- and track-level labels for each detected LPS. Detailed installation instructions, implementation guidance, and documentation of all software updates are available in the GitHub repository.

Major improvements introduced since the initial release of SyCLoPS include:

1. Resolution-dependent TC criterion: A variable MSLPCC20 threshold is introduced for TC classification and is automatically adjusted according to the horizontal grid resolution of the input dataset to optimize TC detection skill.
2. Improved track stitching: The *prioritizing* option in TE's *StitchNodes* utility, together with an additional stitching refinement algorithm implemented in the classifier, minimizes "jumpy nodes" and erroneous track connections in 6-hourly LPS tracking (see Text 3.S2).
3. Adjusted labels: An *Adjusted_Label* column in the classified catalog is introduced to reduce uncertainties arising from hard threshold criteria and noise in reanalysis and model data (see Text 3.S3).
4. Refined extratropical transition detection: The extratropical transition detection scheme is improved to further reduce errors in determining the timing of extratropical transition completion (see Text 3.S4).
5. Hybrid TC classification: A new “Hybrid TC” label is introduced to recover previously missed TC detections in the North Atlantic (see Text 4.S1).
6. Simplified classification workflow: In the latest pre-release version of SyCLoPS, several ambiguous or unnecessary labels within the High-altitude and Dryness branches have been removed to produce a more physically intuitive classification workflow. The updated workflow is illustrated in Figure 2.S10.

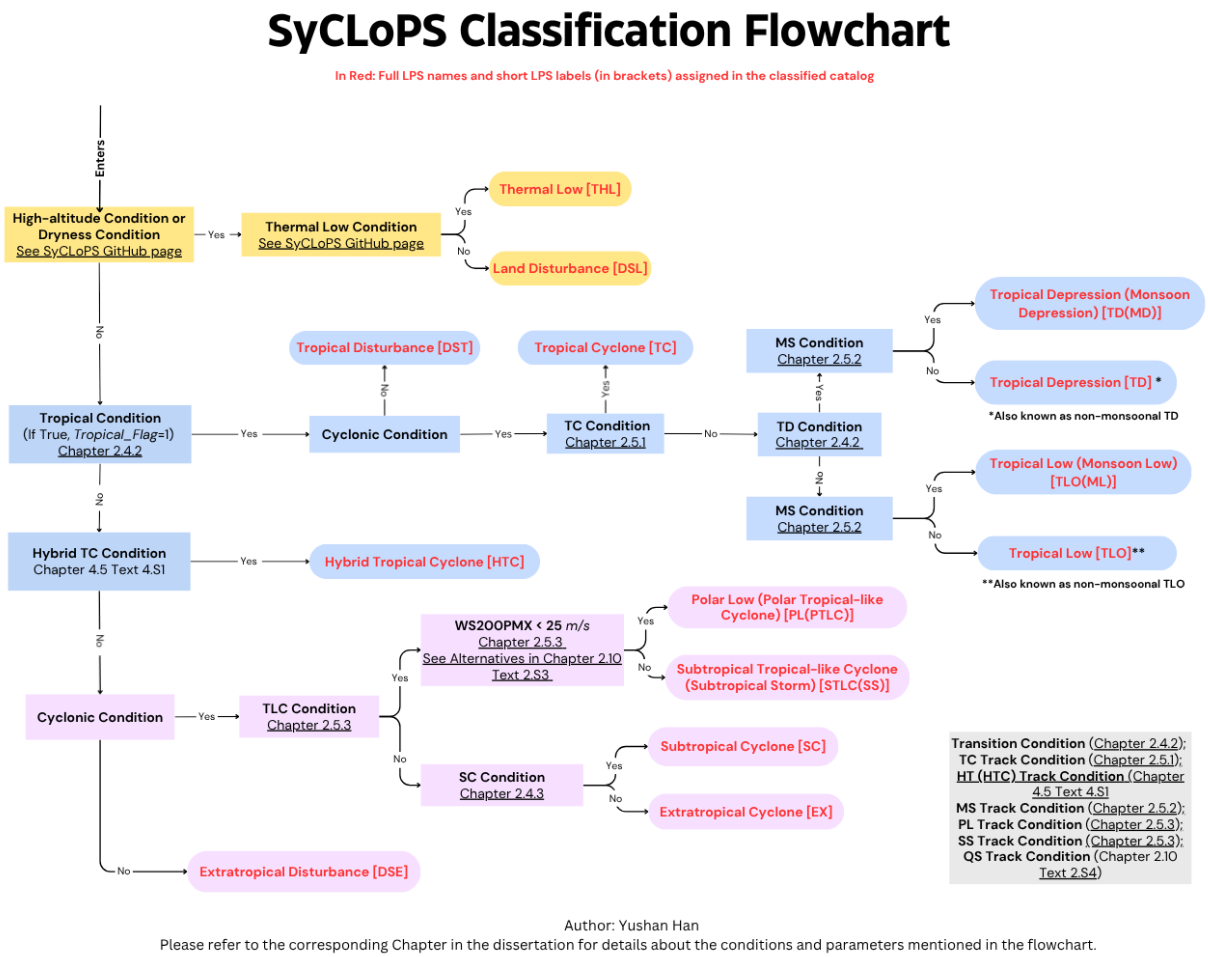



**Figure 2.S10**. As in Figure 2.1, but for a depiction of the updated LPS classification workflow in the latest pre-release version of SyCLoPS.

Published dataset relevant to Chapter 2: *SyCLoPS ERA5 Low-Pressure System (LPS) Dataset*. Zenodo. DOI: 10.5281/zenodo.10906284.

## 2.11 References


Bernhardt, J. E., & DeGaetano, A. T. (2012). Meteorological factors affecting the speed of movement and related impacts of extratropical cyclones along the US East Coast. *Natural Hazards*, *61*, 1463–1472.

Bian, G.-F., Nie, G.-Z., & Qiu, X. (2021). How well is outer tropical cyclone size represented in the ERA5 reanalysis dataset? *Atmospheric Research*, *249*, 105339.

Blackmon, M. L., Wallace, J. M., Lau, N.-C., & Mullen, S. L. (1977). An observational study of the Northern Hemisphere wintertime circulation. *Journal of the Atmospheric Sciences*, *34*(7), 1040–1053.

Bourdin, S., Fromang, S., Dulac, W., Cattiaux, J., & Chauvin, F. (2022). Intercomparison of four algorithms for detecting tropical cyclones using ERA5. *Geoscientific Model Development*, *15*(17), 6759–6786.

Cavicchia, L., Dowdy, A., & Walsh, K. (2018). Energetics and dynamics of subtropical australian east coast cyclones: Two contrasting cases. *Monthly Weather Review*, *146*(5), 1511–1525.

Chavas, D. R., & Emanuel, K. A. (2010). A QuikSCAT climatology of tropical cyclone size. *Geophysical Research Letters*, *37*(18).

Dare, R. A., Davidson, N. E., & McBride, J. L. (2012). Tropical cyclone contribution to rainfall over Australia. *Monthly Weather Review*, *140*(11), 3606–3619.

Datt, I., Camargo, S. J., Sobel, A. H., McTAGGART-COWAN, R., & Wang, Z. (2022). An Investigation of Tropical Cyclone Development Pathways as an Indicator of Extratropical Transition. *Journal of the Meteorological Society of Japan. Ser. II*, *100*(4), 707–724.

Doran, P. T., Priscu, J. C., Lyons, W. B., Walsh, J. E., Fountain, A. G., McKnight, D. M., Moorhead, D. L., Virginia, R. A., Wall, D. H., Clow, G. D., & others. (2002). Antarctic climate cooling and terrestrial ecosystem response. *Nature*, *415*(6871), 517–520.

Emanuel, K. (2005). Genesis and maintenance of" Mediterranean hurricanes". *Advances in Geosciences*, *2*, 217–220.

European Centre for Medium-Range Weather Forecasts. (2019). [Dataset]. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/BH6N-5N20

Evans, J. L., & Braun, A. (2012). A climatology of subtropical cyclones in the South Atlantic. *Journal of Climate*, *25*(21), 7328–7340.

Feng, X., Liu, C., Fan, G., Liu, X., & Feng, C. (2016). Climatology and structures of southwest vortices in the NCEP climate forecast system reanalysis. *Journal of Climate*, *29*(21), 7675–7701.

Fita, L., Romero, R., Luque, A., Emanuel, K., & Ramis, C. (2007). Analysis of the environments of seven Mediterranean tropical-like storms using an axisymmetric, nonhydrostatic, cloud resolving model. *Natural Hazards and Earth System Sciences*, *7*(1), 41–56.

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. Á., Porcù, F., Priestley, M. D. K., Reale, M., Roberts, M. J., … Ziv, B. (2023). A composite approach to produce reference datasets for extratropical cyclone tracks: Application to Mediterranean cyclones. *Weather and Climate Dynamics*, *4*(3), 639–661.

Flaounas, E., Davolio, S., Raveh-Rubin, S., Pantillon, F., Miglietta, M. M., Gaertner, M. A., Hatzaki, M., Homar, V., Khodayar, S., Korres, G., & others. (2022). Mediterranean cyclones: Current knowledge and open questions on dynamics, prediction, climatology and impacts. *Weather and Climate Dynamics*, *3*, 173–208.

Fu, S.-M., Sun, J.-H., Li, W.-L., & Zhang, Y.-C. (2018). Investigating the mechanisms associated with the evolutions of twin extratropical cyclones over the Northwest Pacific Ocean in mid-January 2011. *Journal of Geophysical Research: Atmospheres*, *123*(8), 4088–4109.

Gan, M. A., Kousky, V. E., & Ropelewski, C. F. (2004). The South America monsoon circulation and its relationship to rainfall over west-central Brazil. *Journal of Climate*, *17*(1), 47–66.
Garde, L. A., Pezza, A. B., & Bye, J. A. T. (2010). Tropical transition of the 2001 Australian Duck. *Monthly Weather Review*, *138*(6), 2038–2057.
Gozzo, L. F., da Rocha, R. P., Reboita, M. S., & Sugahara, S. (2014). Subtropical cyclones over the southwestern South Atlantic: Climatological aspects and case study. *Journal of Climate*, *27*(22), 8543–8562.
Guishard, M. P., Evans, J. L., & Hart, R. E. (2009). Atlantic subtropical storms. Part II: climatology. *Journal of Climate*, *22*(13), 3574–3594.
Han, Y., & Ullrich, P. A. (2024). *The System for Classification of Low-Pressure Systems (SyCLoPS) Dataset (Based on ERA5)* [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.10906284
Hart, R. E. (2003). A cyclone phase space derived from thermal wind and thermal asymmetry. *Monthly Weather Review*, *131*(4), 585–616.
Heo, K.-Y., & Ha, K.-J. (2008). Snowstorm over the southwestern coast of the Korean Peninsula associated with the development of mesocyclone over the Yellow Sea. *Advances in Atmospheric Sciences*, *25*, 765–777.
Hepworth, E., Messori, G., & Vichi, M. (2022). Association between extreme atmospheric anomalies over Antarctic sea ice, Southern Ocean polar cyclones and atmospheric rivers. *Journal of Geophysical Research: Atmospheres*, *127*(7), e2021JD036121.
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., & others. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.
Hodges, K., Cobb, A., & Vidale, P. L. (2017). How well are tropical cyclones represented in reanalysis datasets? *Journal of Climate*, *30*(14), 5243–5264.
Hodges, K. I. (1994). A general method for tracking analysis and its application to meteorological data. *Monthly Weather Review*, *122*(11), 2573–2586.
Hoinka, K. P., & Castro, M. D. (2003). The Iberian peninsula thermal low. *Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography*, *129*(590), 1491–1511.
Holland, G. J., Lynch, A. H., & Leslie, L. M. (1987). Australian east-coast cyclones. Part I: Synoptic overview and case study. *Monthly Weather Review*, *115*(12), 3024–3036.
Hopsch, S. B., Thorncroft, C. D., & Tyle, K. R. (2010). Analysis of African easterly wave structures and their role in influencing tropical cyclogenesis. *Monthly Weather Review*, *138*(4), 1399–1419.
Hunt, K. M., Turner, A. G., Inness, P. M., Parker, D. E., & Levine, R. C. (2016). On the structure and dynamics of Indian monsoon depressions. *Monthly Weather Review*, *144*(9), 3391–3416.
Hurley, J. V., & Boos, W. R. (2015). A global climatology of monsoon low-pressure systems. *Quarterly Journal of the Royal Meteorological Society*, *141*(689), 1049–1064.
Iwao, K., Inatsu, M., & Kimoto, M. (2012). Recent changes in explosively developing extratropical cyclones over the winter northwestern Pacific. *Journal of Climate*, *25*(20), 7282–7296.
Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., & Schreck, C. J. (2018). *International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4* [Dataset]. NOAA National Centers for Environmental Information. https://doi.org/10.25921/82ty-9e16
Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, *91*(3), 363–376.
Koch, P., Wernli, H., & Davies, H. C. (2006). An event-based jet-stream climatology and typology. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *26*(3), 283–301.
Kreussler, P., Caron, L.-P., Wild, S., Loosveldt Tomas, S., Chauvin, F., Moine, M.-P., Roberts, M. J., Ruprich-Robert, Y., Seddon, J., Valcke, S., & others. (2021). Tropical cyclone integrated kinetic

energy in an ensemble of HighResMIP simulations. *Geophysical Research Letters*, *48*(5), e2020GL090963.
Landsea, C. W., & Franklin, J. L. (2013). Atlantic hurricane database uncertainty and presentation of a new database format. *Monthly Weather Review*, *141*(10), 3576–3592.
Lau, W. K., & Kim, K.-M. (2015). Robust Hadley circulation changes and increasing global dryness due to CO2 warming from CMIP5 model projections. *Proceedings of the National Academy of Sciences*, *112*(12), 3630–3635.
Lavaysse, C., Flamant, C., Janicot, S., Parker, D. J., Lafore, J.-P., Sultan, B., & Pelon, J. (2009). Seasonal evolution of the West African heat low: A climatological perspective. *Climate Dynamics*, *33*, 313–330.
Lawton, Q. A., Majumdar, S. J., Dotterer, K., Thorncroft, C., & Schreck III, C. J. (2022). The influence of convectively coupled Kelvin waves on African easterly waves in a wave-following framework. *Monthly Weather Review*, *150*(8), 2055–2072.
Lawton, Q., & Majumdar, S. (2018). *Objective Tracking of African Easterly Waves in Reanalysis Data (Updated through 2022)* [Dataset]. OSF. https://doi.org/10.17605/OSF.IO/J4HPQ
Li, J., & Zeng, Q. (2003). A new monsoon index and the geographical distribution of the global monsoons. *Advances in Atmospheric Sciences*, *20*, 299–302.
Li, L., Zhang, R., Wen, M., Duan, J., & Qi, Y. (2019). Characteristics of the Tibetan Plateau vortices and the related large-scale circulations causing different precipitation intensity. *Theoretical and Applied Climatology*, *138*, 849–860.
Lodise, J., Merrifield, S., Collins, C., Rogowski, P., Behrens, J., & Terrill, E. (2022). Global climatology of extratropical cyclones from a new tracking approach and associated wave heights from satellite radar altimeter. *Journal of Geophysical Research: Oceans*, *127*(11), e2022JC018925.
Lu, J., & Ding, Y. (1989). Climatic study on the summer tropical easterly jet at 200 hPa. *Advances in Atmospheric Sciences*, *6*(2), 215–226.
Lu, J., Vecchi, G. A., & Reichler, T. (2007). Expansion of the Hadley cell under global warming. *Geophysical Research Letters*, *34*(6).
Masson-Delmotte, V., Zhai, P., Pirani, S., Connors, C., Péan, S., Berger, N., Caud, Y., Chen, L., Goldfarb, M., & Scheel Monteiro, P. M. (2021). *Ipcc, 2021: Summary for policymakers. In: Climate change 2021: The physical science basis. Contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change*.
Merlis, T. M., Cheng, K.-Y., Guendelman, I., Harris, L., Bretherton, C. S., Bolot, M., Zhou, L., Kaltenbaugh, A., Clark, S. K., Vecchi, G. A., & others. (2024). Climate sensitivity and relative humidity changes in global storm-resolving model simulations of climate change. *Science Advances*, *10*(26), eadn5217.
Montgomery, M. T., & Farrell, B. F. (1992). Polar low dynamics. *Journal of the Atmospheric Sciences*, *49*(24), 2484–2505.
Moreno-Ibáñez, M., Laprise, R., & Gachon, P. (2021). Recent advances in polar low research: Current knowledge, challenges and future perspectives. *Tellus A: Dynamic Meteorology and Oceanography*, *73*(1), 1–31.
National Hurricane Center. (n.d.). *Glossary of NHC Terms*. Retrieved December 2, 2023, from https://www.nhc.noaa.gov/aboutgloss.shtml
Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., Cocozza, A., Dacre, H. F., Feng, Y., Fraedrich, K., & others. (2013). IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bulletin of the American Meteorological Society*, *94*(4), 529–547.
Noer, G., Saetra, Ø., Lien, T., & Gusdal, Y. (2011). A climatological study of polar lows in the Nordic Seas. *Quarterly Journal of the Royal Meteorological Society*, *137*(660), 1762–1772.
Nordeng, T. E., & Rasmussen, E. A. (1992). A most beautiful polar low. A case study of a polar low development in the Bear Island region. *Tellus A*, *44*(2), 81–99.

Oliva, M., Navarro, F., Hrbáček, F., Hernandéz, A., Nỳvlt, D., Pereira, P., Ruiz-Fernandéz, J., & Trigo, R. (2017). Recent regional climate cooling on the Antarctic Peninsula and associated impacts on the cryosphere. *Science of the Total Environment*, *580*, 210–223.

Poveda, G., Jaramillo, L., & Vallejo, L. F. (2014). Seasonal precipitation patterns along pathways of South American low-level jets and aerial rivers. *Water Resources Research*, *50*(1), 98–118.

Powell, M. D., & Reinhold, T. A. (2007). Tropical cyclone destructive potential by integrated kinetic energy. *Bulletin of the American Meteorological Society*, *88*(4), 513–526.

Prat, O. P., & Nelson, B. R. (2013). Mapping the world's tropical cyclone rainfall contribution over land using the TRMM Multi-satellite Precipitation Analysis. *Water Resources Research*, *49*(11), 7236–7254.

Prein, A. F., Mooney, P. A., & Done, J. M. (2023). The Multi-Scale Interactions of Atmospheric Phenomenon in Mean and Extreme Precipitation. *Earth's Future*, *11*(11), e2023EF003534.

Pytharoulis, I., Craig, G. C., & Ballard, S. P. (2000). The hurricane-like Mediterranean cyclone of January 1995. *Meteorological Applications: A Journal of Forecasting, Practical Applications, Training Techniques and Modelling*, *7*(3), 261–279.

Qian, W., & Lee, D.-K. (2000). Seasonal march of Asian summer monsoon. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *20*(11), 1371–1386.

Rasmussen, E. A., & Turner, J. (2003). *Mesoscale Weather Systems in the Polar Regions*. Cambridge University Press.

Reed, R. J., Norquist, D. C., & Recker, E. E. (1977). The structure and properties of African wave disturbances as observed during phase III of GATE. *Monthly Weather Review*, *105*(3), 317–333.

Reeder, M. J., Smith, R. K., Deslandes, R., Tapper, N. J., & Mills, G. A. (2000). Subtropical fronts observed during the 1996 central Australian fronts experiment. *Australian Meteorological Magazine*, *49*(3), 181–200.

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., & others. (2020). Impact of model resolution on tropical cyclone simulation using the HighResMIP–PRIMAVERA multimodel ensemble. *Journal of Climate*, *33*(7), 2557–2583.

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vannière, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., & others. (2020). Projected future changes in tropical cyclones using the CMIP6 HighResMIP multimodel ensemble. *Geophysical Research Letters*, *47*(14), e2020GL088662.

Romero, R., & Emanuel, K. (2017). Climate change and hurricane-like extratropical cyclones: Projections for North Atlantic polar lows and medicanes based on CMIP5 models. *Journal of Climate*, *30*(1), 279–299.

Schreck, C. J., Knapp, K. R., & Kossin, J. P. (2014). The impact of best track discrepancies on global tropical cyclone climatologies using IBTrACS. *Monthly Weather Review*, *142*(10), 3881–3899.

Schultz, D. M., & Keyser, D. (2021). Antecedents for the Shapiro–Keyser cyclone model in the Bergen School literature. *Bulletin of the American Meteorological Society*, *102*(2), E383–E398.

Sherwood, S. C., Ingram, W., Tsushima, Y., Satoh, M., Roberts, M., Vidale, P. L., & O'Gorman, P. A. (2010). Relative humidity changes in a warmer climate. *Journal of Geophysical Research: Atmospheres*, *115*(D9).

Shimada, U., Wada, A., Yamazaki, K., & Kitabatake, N. (2014). Roles of an upper-level cold vortex and low-level baroclinicity in the development of polar lows over the Sea of Japan. *Tellus A: Dynamic Meteorology and Oceanography*, *66*(1), 24694.

Sikka, D. R. (2006). *A study on the monsoon low pressure systems over the Indian region and their relationship with drought and excess monsoon seasonal rainfall*. Center for Ocean-Land-Atmosphere Studies.

Smith, E. A. (1986). The structure of the Arabian heat low. Part I: Surface energy budget. *Monthly Weather Review*, *114*(6), 1067–1083.

Spengler, T., Reeder, M. J., & Smith, R. K. (2005). The dynamics of heat lows in simple background flows. *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 3147–3165.
Stansfield, A. M., & Reed, K. A. (2021). Tropical cyclone precipitation response to surface warming in aquaplanet simulations with uniform thermal forcing. *Journal of Geophysical Research: Atmospheres*, *126*(24), e2021JD035197.
Stansfield, A. M., Reed, K. A., Zarzycki, C. M., Ullrich, P. A., & Chavas, D. R. (2020). Assessing tropical cyclones' contribution to precipitation over the eastern united states and sensitivity to the variable-resolution domain extent. *Journal of Hydrometeorology*, *21*(7), 1425–1445.
Stoll, P. J. (2022). A global climatology of polar lows investigated for local differences and wind-shear environments. *Weather and Climate Dynamics*, *3*(2), 483–504.
Stoll, P. J., Graversen, R. G., Noer, G., & Hodges, K. (2018). An objective global climatology of polar lows based on reanalysis data. *Quarterly Journal of the Royal Meteorological Society*, *144*(716), 2099–2117.
Terpstra, A., Michel, C., & Spengler, T. (2016). Forward and reverse shear environments during polar low genesis over the Northeast Atlantic. *Monthly Weather Review*, *144*(4), 1341–1354.
Toomey, T., Amores, A., Marcos, M., Orfila, A., & Romero, R. (2022). Coastal hazards of tropical-like cyclones over the Mediterranean Sea. *Journal of Geophysical Research: Oceans*, *127*(2), e2021JC017964.
Tucker, D. F. (1999). The summer plateau low pressure system of Mexico. *Journal of Climate*, *12*(4), 1002–1015.
Ullrich, P. A. (2024). *TempestExtremes v2.2.3 Repository* [Software]. GitHub. https://doi.org/10.5281/zenodo.13830377
Ullrich, P. A., & Zarzycki, C. M. (2017). TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, *10*(3), 1069–1090.
Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., & Reed, K. A. (2021). TempestExtremes v2. 1: A community framework for feature detection, tracking and analysis in large datasets. *Geoscientific Model Development Discussions*, *2021*, 1–37.
U.S. Navy. (1994). *Local area forecaster's handbook for Naval Air Station Bermuda*. Naval Atlantic Meteorology Facility, Bermuda.
Vishnu, S., Boos, W., Ullrich, P., & O'brien, T. (2020). Assessing historical variability of South Asian monsoon lows and depressions with an optimized tracking algorithm. *Journal of Geophysical Research: Atmospheres*, *125*(15), e2020JD032977.
Winckler, P., Contreras-López, M., Campos-Caba, R., Beyá, J. F., & Molina, M. (2017). El temporal del 8 de agosto de 2015 en las regiones de Valparaíso y Coquimbo, Chile Central. *Latin American Journal of Aquatic Research*, *45*(4), 622–648.
Yamamoto, M. (2012). Rapid merger and recyclogenesis of twin extratropical cyclones leading to heavy precipitation around Japan on 9–10 October 2001. *Meteorological Applications*, *19*(1), 36–53.
Yokoyama, Y., & Yamamoto, M. (2019). Influences of surface heat flux on twin cyclone structure during their explosive development over the East Asian marginal seas on 23 January 2008. *Weather and Climate Extremes*, *23*, 100198.
Zappa, G., Shaffrey, L., & Hodges, K. (2014). Can polar lows be objectively identified and tracked in the ECMWF operational analysis and the ERA-Interim reanalysis? *Monthly Weather Review*, *142*(8), 2596–2608.
Zarzycki, C. M., Thatcher, D. R., & Jablonowski, C. (2017). Objective tropical cyclone extratropical transition detection in high-resolution reanalysis and climate model data. *Journal of Advances in Modeling Earth Systems*, *9*(1), 130–148.
Zarzycki, C. M., & Ullrich, P. A. (2017). Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophysical Research Letters*, *44*(2), 1141–1149.
Zhang, W., Villarini, G., Scoccimarro, E., & Napolitano, F. (2021). Examining the precipitation associated with medicanes in the high-resolution ERA-5 reanalysis data. *International Journal of Climatology*, *41*, E126–E132.

# Chapter 3. The Evolution of Tropical Cyclone Intensity and Size in Observations and HighResMIP Models

## 3.1 Abstract

The destructive potential of a tropical cyclone (TC) depends on both its intensity and size, yet the coupled evolution of these properties remains insufficiently explored. This study investigates TC intensity–size relationships of TCs in observations and historical simulations from 12 HighResMIP models in the North Atlantic (NA) and western North Pacific (WNP). We apply the System for Classification of Low-Pressure Systems (SyCLoPS) to track TC life cycle and a novel quadrant wind radii calculation mirroring the IBTrACS format to enable consistent model–observation comparisons. SyCLoPS effectively removes weak tropical and extratropical systems misidentified as TCs in a conventional tracker, leading to improved TC tracking performance in ERA5 and models. TC track types are classified by their precursors and posterior scenarios based on SyCLoPS and IBTrACS labels. They are also grouped by intensity–size evolutionary pathways defined by peak intensity, size at peak intensity, and terminal size. Tropical depressions are more likely to intensify into intense but compact TCs, whereas monsoon depressions in the WNP and TCs undergoing extratropical transition tend to grow substantially in size as they evolve. In observations, the largest TCs are not the most intense and generally occur at higher latitudes. In contrast, most models have the strongest TCs reach peak size at lower latitudes. Only the highest-resolution (~20 km) model assessed captures intense, compact TCs that are common in observations. Despite this, the model ensemble mean performs comparably to ERA5 regarding TC track density, TC track type compositions, and differences in intensity and size characteristics between ENSO phases.

## 3.2 Introduction

Tropical cyclones (TCs) are among the most destructive weather phenomena, impacting coastal regions with strong winds, heavy rainfall, and storm surges (Hagen, 2025; Needham et al., 2015; Peduzzi et al., 2012). TC intensity determines localized wind damage, while size influences wind duration and the extent of storm surge and rainfall (Irish et al., 2008; Lin et al., 2015; Matyas, 2010; Powell & Reinhold, 2007). Examining TC evolution in intensity–size space thus provides a more complete view of their development and destructive potential.

Musgrave et al. (2012) compared the time evolution of TCs to the main sequence of stars by drawing an analogy between an intensity-size diagram and a Hertzsprung-Russell (H-R) diagram. Using the analogy, TCs can be separated into four categories: "weak dwarf (WD)," "strong dwarf (SD)," "super giant (SG)," and "weak giant (WG)." Over the 2024 Atlantic hurricane season, *Helene*, the deadliest storm of the year, was one of the largest U.S. landfalling TC according to the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al., 2010), even surpassing *Katrina* (2005). *Helene* was large and moderately intense, fitting the SG category. In contrast, *Milton*, which arrived shortly after *Helene*, was compact but extremely intense, representing a SD. These examples illustrate two destructive TC evolutionary pathways: storms that grow both large and intense (see Dunnavan & Diercks, 1980), and those that remain small but reach extreme strength (see Hawkins & Imbembo, 1976). The deadly hurricane *Noel* (2007), which traced across Hispaniola and Cuba, exemplifies WGs—weak but exceptionally large systems that still cause massive damage even as a post-TC. The different characteristics of each TC evolutionary pathway imply the need for tailored disaster mitigation strategies.

The relationship between intensity and size as TCs evolve has been explored through idealized simulations (e.g., Kilroy et al., 2017; Musgrave et al., 2012; Ooyama, 1969) and observations (Chavas et al., 2016; Chen et al., 2022; Ruan & Wu, 2022; Song et al., 2020). In model simulations, TCs often continue to grow larger and reach their lifetime maximum size (LMS) after reaching their lifetime maximum intensity (LMI). In western North Pacific (WNP) observations, Song et al. (2020) found that a

TC can reach its LMS before its LMI, or vice versa, depending on the peak intensity of the TC, the latitude, and the large-scale environment. Chen et al. (2022) classified TCs into two groups: one with a low linear regression coefficient (RC) between intensity and size, and one with a high RC. They found that high-RC TCs expand more rapidly and are more frequently found at higher latitudes, suggesting links between size expansion and the local Coriolis parameter $f$ (Chan & Chan, 2014; Chavas & Emanuel, 2014; Smith et al., 2011). Moreover, studies have shown that a larger initial size of the pre-TC vortex may favor rapid expansion but suppress rapid intensification (Carrasco et al., 2014; Cocks & Gray, 2002; Lee et al., 2010; Martinez et al., 2020). While most studies focus on the mature or developing stages of the TC life cycle, Nguyen & Schenkel (2025) investigated changes in the outer size of TCs during extratropical transition (ET) in reanalysis and found that TC size remains largely unchanged until the end of the transition. Despite previous efforts, a comprehensive analysis and detailed classification of TC intensity-size pathways based on the entire lifecycle of observed TCs is lacking.

Despite being extensively explored in idealized simulations focused on individual TCs, the general TC intensity-size relationship has received much less attention in climate models. Wu et al. (2015) investigated a 9-km regional model and found that the model failed to reproduce the TC intensity-size relationship in the WNP shown by satellite-based TC wind field observations from 2007-2013. Schenkel et al. (2018) investigated North Atlantic (NA) and WNP TC size evolution in a GFDL high-resolution (~25 km grid spacing) model. They found that the model's patterns of TC size evolution generally align with those in the reanalysis and are partially controlled by TC track patterns (i.e., track orientation) and size at genesis. Kreussler et al. (2021) examined TC integrated kinetic energy (IKE; Powell & Reinhold, 2007)—a metric primarily controlled by TC size—in three sets of high- and low-resolution CMIP6 endorsed High-Resolution Model Intercomparison Project (HighresMIP; Haarsma et al., 2016) models. They concluded that compared to the low-resolution models, high-resolution models simulate stronger, albeit smaller, TCs, and that coupled simulations exhibit similar behavior despite lower overall intensity. To the best of the authors' knowledge, no studies have drawn direct comparisons of TC size between HighResMIP models and observations of any kind.

An obstacle to analyzing the complete life cycle of TCs in models is an accurate and consistent objective detection framework. However, the recently developed System for Classification of Low-pressure Systems (SyCLoPS; Han & Ullrich, 2025), based on the state-of-the-art feature detection software TempestExtremes (TE; Ullrich et al., 2021; Ullrich & Zarzycki, 2017), provides a complete and consistent framework for labeling every detected low-pressure system (LPS) node, thereby distinguishing the pre-TC, TC, and post-TC stages of a detected TC track. This approach also substantially boosts TC detection skill scores compared to the original TE-based algorithm. The current study presents several enhancements to SyCLoPS, improving its applicability to models with varying grid resolutions. Furthermore, we develop an algorithm in TE that can conveniently output a set of four quadrant wind radii for each detected TC node in a format that matches IBTrACS. With these developments, one can track the evolution of TC intensity and size over time in model simulations and classify their evolutionary pathways according to their precursors and post-TC (posterior) scenarios. Model performance for TC track density, intensity and size can also be validated directly against IBTrACS with higher reliability.

This study evaluates the evolution of TC intensity and size in the NA and the WNP—the two basins that have very active TC activity and relatively high-quality observations—in both observations and a suite of high-resolution climate model simulations from HighResMIP. Chapter 3.3 introduces the models and observations used for evaluation and explains our TC tracking algorithms and TC size calculation method. Chapter 3.4 discusses the advantages of using SyCLoPS for tracking TCs to more accurately represent TCs in climate models, followed by an overview of TC intensity and size in observations in Chapter 3.5 and model evaluations in Chapter 3.6.

## 3.3 Data, Tracking, and TC Size Estimation

### 3.3.1 Model data

We analyze outputs from nine atmosphere-only (highresSST-present) simulations from 9 selected HighResMIP models, with each simulation representing a model, and three coupled (Coup.) simulations (hist-1950) from three of these models. All simulations run from 1950 to 2014. All models have a

nominal resolution of either 25 or 50 km. Models with coarser nominal resolution are not considered since they are likely not sufficient for resolving the TC surface wind structure as required in this study. All atmosphere-only simulations use sea surface temperature (SST) and sea ice conditions from the Hadley Centre Global Sea Ice and Sea Surface Temperature data set (HadISST.2.2.0; Kennedy et al., 2017). Table 3.1 provides a summary of model simulations considered in this study.

| Model | Run | Atmospheric mesh spacing at equator (km) | Lat × lon, level | Analysis grid (°) | Reference |
|---|---|---|---|---|---|
| CNRM-CM6-1-HR | r1i1p1f2; r1i1p1f2 (coupled) | 55 | 360 × 720, 91 | 0.5 × 0.5 | Voldoire (2019) |
| EC-Earth3P-HR | r1i1p1f1 | 39 | 512 × 1024, 91 | 0.35 × 0.35 | Consortium (EC-Earth) (2018) |
| ECMWF-IFS-HR | r1i1plf1; r1i1plf1 (coupled) | 25 | 800 × 1600, 91 | 0.5 × 0.5 | Roberts et al. (2017) |
| ECMWF-IFS-LR | r1i1plf1 | 50 | 400 × 800, 91 | 1 × 1 | Roberts et al. (2017b) |
| HadGEM3-GC31-HM | r1i1plf1; r1i3p1f1 (coupled) | 39 | 768 × 1024, 79 | 0.23 × 0.35 | Roberts (2017) |
| IPSL-CM6A-ATM-ICO-VHR | r1i1plf1 | 25 | 720 × 1440, 79 | 0.25 × 0.25 | Boucher et al. (2022) |
| IPSL-CM6A-ATM-ICO-HR | r1i1plf1 | 50 | 360 × 720, 79 | 0.5 × 0.5 | Boucher et al. (2022a) |
| MRI-AGCM3.2-S | r1i1plf1 | 20 | 960 × 1920, 64 | 0.1875 × 0.1875 | Mizuta et al. (2019) |
| MRI-AGCM3.2-H | r1i1plf1 | 60 | 320 × 640, 64 | 0.5625 × 0.5625 | Mizuta et al. (2019a) |

**Table 3.1.** Model descriptions of selected HighResMIP simulations.

### 3.3.2 Observations

IBTrACS is a well-known global TC observation dataset that integrates best track data provided by regional specialized meteorological centers (agencies) designated by the World Meteorological Organization (WMO). Due to discrepancies among agencies and changes over time in observational technologies, wind measurement standards, and operational procedures, care must be taken when comparing tropical cyclone records across basins (Schreck et al., 2014). One notable difference arises

from discrepancies in maximum wind speed measurement standards (wind averaging periods) adopted by different regional agencies. For example, U.S. agencies, including the National Hurricane Center (NHC) and the Joint Warning Center (JTWC) use a 1-min averaging period, whereas most of the WMO regional agencies use a 10-min period. Thus, the number of observed TC tracks and nodes varies slightly depending on the chosen standard for processing IBTrACS data. See Supplements (Chapter 3.10) Text 3.S1 for details on how this discrepancy affects the performance of TC tracking.

For the purposes of this study, we process IBTrACS as follows: (1) Records that are not recorded at hours of 0, 6, 12, and 18 are removed. (2) If available, wind speed data provided by WMO regional agencies are used for the analysis. However, if the data is missing, we use JTWC wind estimates after converting them to match the averaging period used in the target basin (the NA and WNP) via the conversion ratios for at-sea TCs suggested by Harper et al. (2010). (3) All tracks that never reach 34 knots (kt) are removed; remaining tracks are defined as *observed TC tracks*. (4) An *observed TC node* in a track is defined as a record with at least a 34 kt maximum wind speed that is not labeled as extratropical, subtropical, or post-tropical by both the WMO and a U.S. agency.

IBTrACS also contains consistent wind radii observations in units of nautical miles (n. mile; 1 n.mile = 1.852 km) from 2001 onward in most basins and provided by U.S. agencies. When available, these wind radii data are used in both the WNP and NA. Wind radii data from the extended best track (EBT; Demuth et al., 2006) are used to supplement missing radii data from 1988 to 2001 in the NA basin. Wind radii are normally recorded for all TCs with an intensity above 34 kt. Some scattered, small along-track gaps in the wind radii data in both basins are filled by linear interpolation. The resulting observational dataset covers the period from 1988 to 2024 in NA and from 2002 to 2024 in WNP. Only the overlapping periods are analyzed when observations are compared to model outputs. The overlapping period is 1988-2014 in the NA and 2002-2014 in the WNP.

We use the 0.25° ERA5 reanalysis (Hersbach et al., 2020) and a conservatively regridded 1° ERA5 (the lowest grid resolution among our models) to evaluate the sensitivity of the trackers to grid

resolution. TC intensity and size derived from the two ERA5 datasets are presented in figures alongside the model results for comparison.

### 3.3.3 Tracking algorithms

In this study we compare two TE-based trackers: the original TE-based TC tracking algorithm, referred to as the ZU method or ZU (Ullrich et al., 2021; Zarzycki & Ullrich, 2017) and SyCLoPS. Compared to non-TE-based trackers, ZU was shown to have the best detection skills, as measured by hit rate minus false alarm rate (HRMFAR), among all investigated trackers for the ERA5 reanalysis (Bourdin et al., 2022). SyCLoPS tracks all types of LPSs globally without topographic restrictions and further reduces false alarms among global detected TC tracks in ERA5 compared with ZU (Han & Ullrich, 2025). Therefore, SyCLoPS is used as our primary TC tracking algorithm. Links to the latest SyCLoPS and TE software repositories can be found in the Open Research section. TCs detected in the two target basins are selected based on the basin boundaries depicted in Figure 3.S1.

Both ZU and SyCLoPS require TCs to have a strong upper-level warm core and a steep mean sea-level pressure (MSLP) gradient. However, a warm-core criterion alone is likely insufficient to distinguish many warm-core or warm air-secluded extratropical systems (Bourdin et al., 2022; Han & Ullrich, 2025; Hart, 2003). SyCLoPS also verifies the tropical characteristics of the system using a 100 hPa relative humidity threshold and a large-scale deep-layer shear threshold. Additionally, we introduce a regression formula that adjusts the MSLP gradient threshold in SyCLoPS's TC detection configuration to account for varying model resolutions, thereby maximizing SyCLoPS's TC detection skill at any resolution. ZU enforces that all LPSs at the node level be TC candidates, so all final tracks are considered fully TC in nature. SyCLoPS tracks arbitrary LPSs and then labels each node, which permits TCs to be defined both at the node and aggregate track level. Here, only "TC" labeled nodes along "TC" labeled tracks are selected from the SyCLoPS dataset. Appendixes 3A and 3B provide details of the tracking algorithms implemented in this study.

SyCLoPS's use of distinct node and track labels provides us an opportunity to evaluate TC detection hit rate (HR) and false alarm rate (FAR) at both levels. Tracker performance at the track level is

measured using the HRMFAR metric for both algorithms and critical success index (CSI) when evaluating SyCLoPS at the node level. HR and miss rate (MR) of TC tracks is defined as the proportion of IBTrACS tracks that are listed by the tracking algorithm (i.e., hits) and missed by the algorithm (i.e., misses). A hit (miss) is marked when a detected TC track contains a node that is (not) found within 2.0° great-circle distance of an IBTrACS record at the same timestamp. The FAR for TC tracks is defined as the proportion of tracks listed by the tracking algorithm that don't correspond to a track in IBTrACS. At the node level, hit and false alarm rates (node HR and node FAR) are calculated similarly. A hit occurs when a TC-labeled node in SyCLoPS is found within 2.0° of an observed TC node in IBTrACS; otherwise, it's labeled as a miss. A false alarm (FA) occurs when no TCs in IBTrACS are found within 2.0° of a TC-labeled node in SyCLoPS. The CSI is defined as:

$$CSI = \frac{hits}{hits + misses + false\ alarms} \tag{3.1}$$

### 3.3.4 Track type

TC tracks may be further labeled using pre-TC and post-TC status using the SyCLoPS LPS track and node labels. Pre-TCs may be labeled as monsoon depressions (MDs), tropical depressions (TDs), or, in some rare cases, extratropical cyclones. Post-TCs may be labeled as extratropical systems if they have completed ET. Consequently, TC tracks can be placed into several categories based on TC precursor labels (precursor type) and post-TC scenario (posterior type). Precursor types include TD, MD, and tropical transition (TT), whereas posterior types include landfall (LD), ET, and Non-ET (NT). IBTrACS agencies provide enough TC status labels for the observed TC tracks to justify ET in two basins. However, observations lack specific classification before TC genesis. Therefore, we use LPS tracks identified by SyCLoPS in ERA5 and pair them with each IBTrACS TC track to complete the pre-TC classification. SyCLoPS detects all types of LPS tracks, including extremely weak features. Consequently, 98% of the observed tracks have an ERA5 counterpart.

A TT precursor type is assigned to tracks that were once identified as non-tropical systems in SyCLoPS before being labeled as TCs. From the remaining tracks, the precursor type is classified as an MD type if the track is labeled a "monsoonal system (MS)" in SyCLoPS, otherwise it is classified as a TD type. In SyCLoPS, a track is labeled as a MS track when a weak tropical LPS is embedded in a sufficiently moist, predominantly westerly wind low-level monsoonal environment for a minimum prescribed duration. See Han & Ullrich (2025) for details on the classification workflow and the physical justification for the classification. For the posterior type, a LD label is given when a TC is on land equatorward of 40° following its TC stage. If the track is not assigned the LD label, it is labeled as an ET if the LPS transitions into an extratropical system (except for extratropical disturbance) within 18 hours of the last TC label in SyCLoPS. The ET labeling condition has been optimized against IBTrACS observations in the WNP and NA using ERA5 data, which identifies 91% of ET tracks matched in IBTrACS with a comparable ET fraction. Additional details are provided in Text 3.S4 and Figure 3.S2. Observed ET tracks are identified by TC status labels in IBTrACS. The remaining tracks are designated "Non-ET (NT)" for their posterior type. These are TCs that weaken significantly over the ocean due to a deteriorating environment.

Next, tracks are divided into seven categories in the WNP and five categories in the NA, based on the precursor and posterior type mentioned above. The categories are MD-LD, TD-LD, MD-ET, LD-ET, MD-NT, TD-NT, and TT in the WNP, and MD, TD-LD, TD-ET, TD-NT, and TT in the NA. TT is a relatively rare precursor type in both basins, and MD is relatively rare in the NA, so they are not further divided into three sub-categories. Figure 3.1 illustrates the tracks and their categories in both basins.

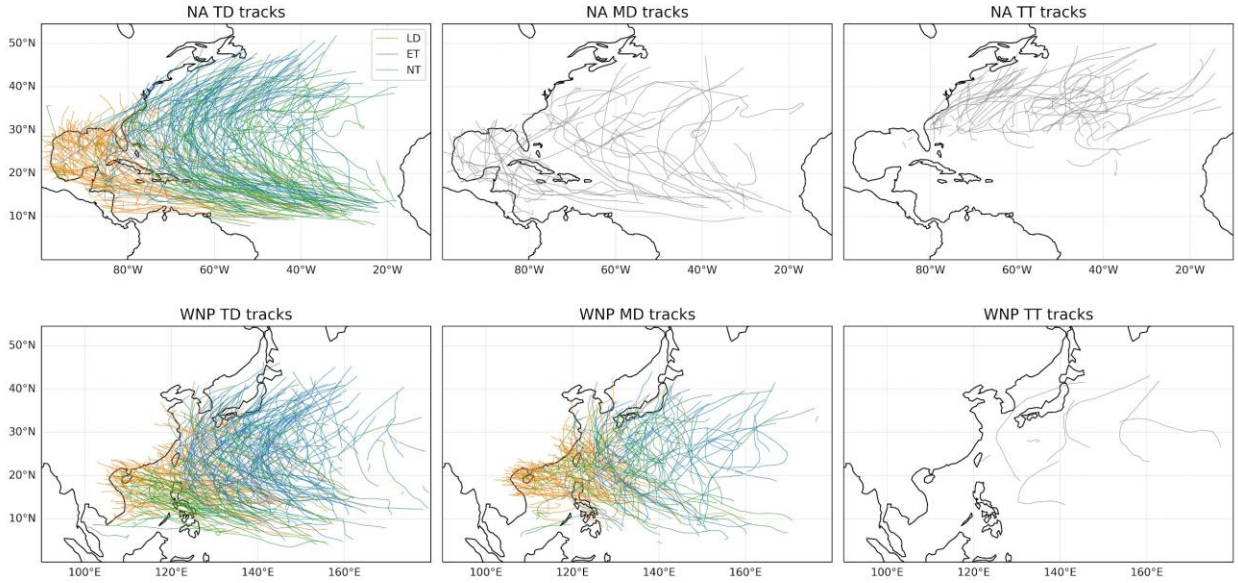



**Figure 3.1**. Tracks of TC in different track categories based on precursor types (TD, MD, TT) and posterior type (LD, ET, NT).

### 3.3.5 TC size estimation

A threshold of 34 kt, the minimum near-surface (10 m) wind speed required for damaging gale-force winds and a named TC, is often used to evaluate TC size. In IBTrACS, this threshold is applied in each quadrant (i.e., northeast, southeast, southwest, and northwest) to give TC size in terms of wind radii (R34). R34 has also been used to quantify TC size in many previous studies focusing on observations and operational forecasts (e.g., Cangialosi & Landsea, 2016; Chan & Chan, 2015; Quiring et al., 2011). IBTrACS wind radii are classically represented as four uniform R34 radii in quarter circles (Figure 3.2a), and TC size sometimes calculated by averaging over the four quadrants (Cangialosi & Landsea, 2016; Quiring et al., 2011). However, this metric underestimates the size of asymmetric TCs (i.e., TCs with large variations between quadrants) because area scales with the square of the mean radius and the variance of the four quadrants, so the mean radius does not conserve TC size under different magnitudes of asymmetry. Many studies also use the average of the nonzero quadrantal radius to represent TC size (e.g., Avenas et al., 2023; Chavas & Knaff, 2022; Sampson et al., 2017). However, this method can overestimate TC size when the asymmetry is large, especially when a zero radius (34 kt threshold not

reached) occurs in some quadrants. Instead, we propose to linearly interpolate 360 radius values between the center of each quadrant to best represent the smoothly varying wind field in real TCs (Figure 3.2a). This imitates the operational wind radii representation graphics of the Advanced Dvorak Technique (ADT; Olander & Velden, 2007, 2019) that are available in real-time online (Figure 3.2b). Herein, TC size ($R_{size}$) is defined as the radius of the circle whose area is equal to that of the feature demarcated by the R34 curve, defined in polar coordinates as $R_c(\theta)$, and discretized using 1° azimuthal wedges as follows:

$$R_{size} = \sqrt{\frac{1}{2\pi}\int_0^{2\pi} R_c(\theta)^2 d\theta} \approx \left[\frac{1}{360}\sum_{i=1}^{360} R_c(\theta_i)^2\right]^{1/2} \quad (3.2)$$

Using equation (3.2), the example TC shown in Figure 3.2b has a $R_{size}$ of 55 n.mile, which is greater than the 50 n.mile radius calculated by the arithmetic mean of the four quadrants (see dashed circles in Figure 3.2a). If the zero-radii southwest quadrant is discarded when calculating the mean, its size will inflate to 67 n.mile.

Wind speeds in models and reanalysis are typically much lower than reality because of their relatively coarse model grids, necessitating lower wind speed thresholds to produce realistic estimates of storm size. A storm radius computed with a much lower 10 m wind speed of 6 m/s – 9 m/s (r6 - r9) is often used for representing TC size in reanalysis and models (e.g., Schenkel et al., 2018; Stansfield et al., 2020). These size metrics have been shown to best match a short period of satellite observations for TCs in reanalysis products (Bian et al., 2021; Schenkel et al., 2017). However, these metrics average wind speeds over all azimuths, again leading to possible underestimation of the size of weaker and more asymmetric TCs, and cannot be directly compared to the quadrant format in IBTrACS.

Following the insights above, our proposed size algorithm is implemented in TE as follows: We first bin grid points within a fixed radius of the storm by quadrant and distance from the center of the storm to produce a radial profile of wind speeds within each quadrant. Then, we find the radius

corresponding to a particular wind threshold, which is optimized for the model grid spacing. Finally, equation (3.2) is then used to compute storm size. See Text 3.S5 for additional implementation details.

Note that there is a discrepancy between the definition of the "maximum extent" of the IBTrACS quadrantal wind radii and the azimuthally averaged quadrantal wind radii in our size estimate algorithm. Namely, we do not adopt the maximum extent approach because it is susceptible to large errors from noise in the model's near-surface wind field and from nearby systems in the vicinity. Chavas & Knaff (2022) and DeMaria et al. (2009) suggest a conversion ratio of 0.85 between maximum and central or mean quadrantal radii. Here, we also apply the 0.85 coefficient to convert IBTrACS data to align with the output of our size estimation algorithm.

Reanalysis and high-resolution climate models are expected to more accurately simulate mean sea-level pressure (MSLP) gradients near TC boundaries than near TC centers. However, wind speed biases can still occur due to the limitations in model grid resolution. For model grid spacing between 0.25º to 1º, the "instantaneous" wind speed is effectively an averaged wind speed over approximately 27 to 108 minutes. This is significantly longer than the 1-minute and 10-minute averaging periods used by most operational agencies. Therefore, the model-derived R34 may not directly correspond to the R34 metric employed in IBTrACS. To address this discrepancy, we compare the TC sizes of 2500 NA and 2500 WNP TCs in the most recent IBTrACS records with their corresponding objectively detected TCs in different reanalysis products. A range of $R_c$ from 20 to 40 kt are used to calculate the size of all detected TCs. The results in Figure 3.2c indicate that, excluding the small number of underrepresented TCs with zero size, the $R_c$ of TCs in reanalysis data sets, regardless of grid resolution, align best with IBTrACS at 30-31 kt with a mean absolute error (MAE) of ~25 n.mile. The bias suggests a ~28% uncertainty (relative to the average size of selected cases) in the size estimation in reanalysis, which is comparable to the uncertainties (~30%) in the NA R34 observations in IBTrACS (Landsea et al., 2012). Note that a 30-31 kt wind is equivalent to a 1-min 34 kt wind speed converted to an average wind speed over one hour using the 0.90 conversion ratio as suggested by Harper et al (2010). Therefore, the 30 kt wind radius (R30) is considered the $R_c$ and an equivalent measurement to observational R34 for all the models investigated in

this study. R30 is then used to calculate TC size ($R_{size}$) in the models. Finally, consistent with our approach for IBTrACS, we apply linear interpolation to fill minor data gaps in $R_{size}$ for modeled TC nodes that exceed 30 kt intensity but do not have a detectable R30 at any quadrant.

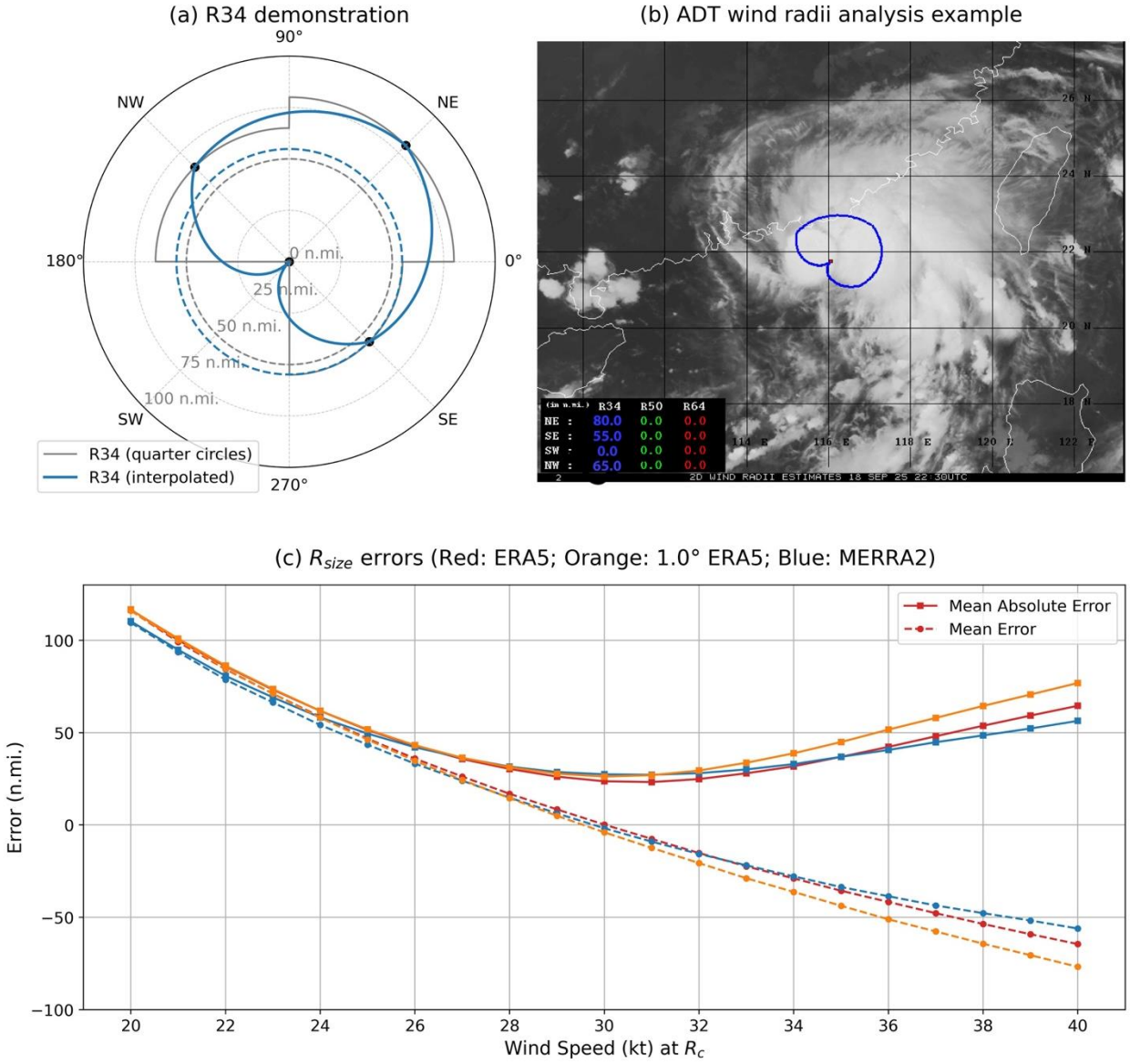



**Figure 3.2**. (a) Illustration of the quarter circle and interpolated R34 using the quadrantal wind radii information given in the lower left of (b), which reads 80, 55, 0, and 65 in the four quadrants. (b) is adapted from the ADT real-time website, showing the ADT wind radii analysis on Typhoon *Mitag* (2025) at 20:30 UTC, September 18, 2025. (c) Mean absolute error and mean error between the average TC size

($R_{size}$) of selected IBTrACS records and corresponding TC nodes within 2º of IBTrACS locations at the same timestamp in ERA5 (red), 1º ERA5 (orange), and MERRA2 (blue; Gelaro et al., 2017).

### 3.4 Comparing ZU and SyCLoPS TC Trackers

To highlight SyCLoPS's strengths in more fairly representing TC performance in earth-system models, we first evaluate the performance of ZU and SyCLoPS TC trackers in ERA5 against IBTrACS over 1988-2014 (the longest overlapping period). The results are summarized in Figure 3.3. Figures 3.3a and b show, consistent with the main improvement of SyCLoPS found in Han & Ullrich (2025), SyCLoPS improves the HRMFAR detection skill score by ~22% in the NA and ~8% in the WNP mainly by successful removing most of the false alarm tracks. HRs in SyCLoPS are also slightly higher than those in ZU. This is likely because SyCLoPS does not apply ZU's stricter minimum time step and 10-m wind speed requirements for filtering tracks. SyCLoPS yields detection skill improvements in other basins similar to those in the WNP (not shown). Nevertheless, both ZU and SyCLoPS miss ~40% of tracks in the NA, which is a striking difference from the WNP. We briefly discuss this NA behavior in Chapter 3.7.1. The detection results of the regridded 1º ERA5 also indicate that SyCLoPS's detection skills are much less sensitive to grid resolution changes than ZU. As shown in Figure 3.3c, ZU detects a much larger number of total TC nodes than SyCLoPS and observations. However, after removing all FAs from ZU, ZU matches IBTrACS very well in the mid-latitudes, indicating that ZU is incorrectly tagging extratropical cyclones as TCs at these latitudes. At lower latitudes, FA-free ZU detects a smaller number of TC nodes than observations and SyCLoPS, especially in the Southern Hemisphere. Upon further investigation, we find that the low-latitude FA nodes removed from ZU are mostly very weak LPS nodes (e.g., MSLP > 1005 hPa). These results are consistent with Bourdin et al., (2022) who performed a detailed evaluation of FAs in multiple trackers and found common failure modes were strong extratropical cyclones and weak tropical storms.

Figure 3.3d provides the global node-level detection skill for SyCLoPS. SyCLoPS labels ~70% of TC nodes in IBTrACS correctly but includes ~20% FAs, producing a CSI of 60%. In some cases, an

observed TC can be paired with an LPS node detected by SyCLoPS that is not labeled as a TC node. This could be attributed to reanalysis errors from an overly weak LPS in ERA5. An ERA5 LPS node is deemed too weak if its MSLP exceeds 1005 hPa (a typical standard to distinguish tropical storms and TDs) and is at least 10 hPa higher than its paired observed TC node, it lacks a 500 hPa cyclonic circulation or upper-level warm core per SyCLoPS requirements, or is otherwise absent from ERA5. If these nodes are treated as true negatives, the MR drops to about 15%. Similarly, some FAs may occur due to underestimation of TC intensity from IBTrACS agencies or appearing overly strong in ERA5. By discarding FAs that meet SyCLoPS's strict warm-core and MSLP gradient criteria and have an MSLP below 1005 hPa, but are classified only as a TD or disturbance in IBTrACS, the FAR decreases to ~13%. With these modifications, the adjusted CSI reaches 73% for ERA5 and 72% for 1º ERA5.

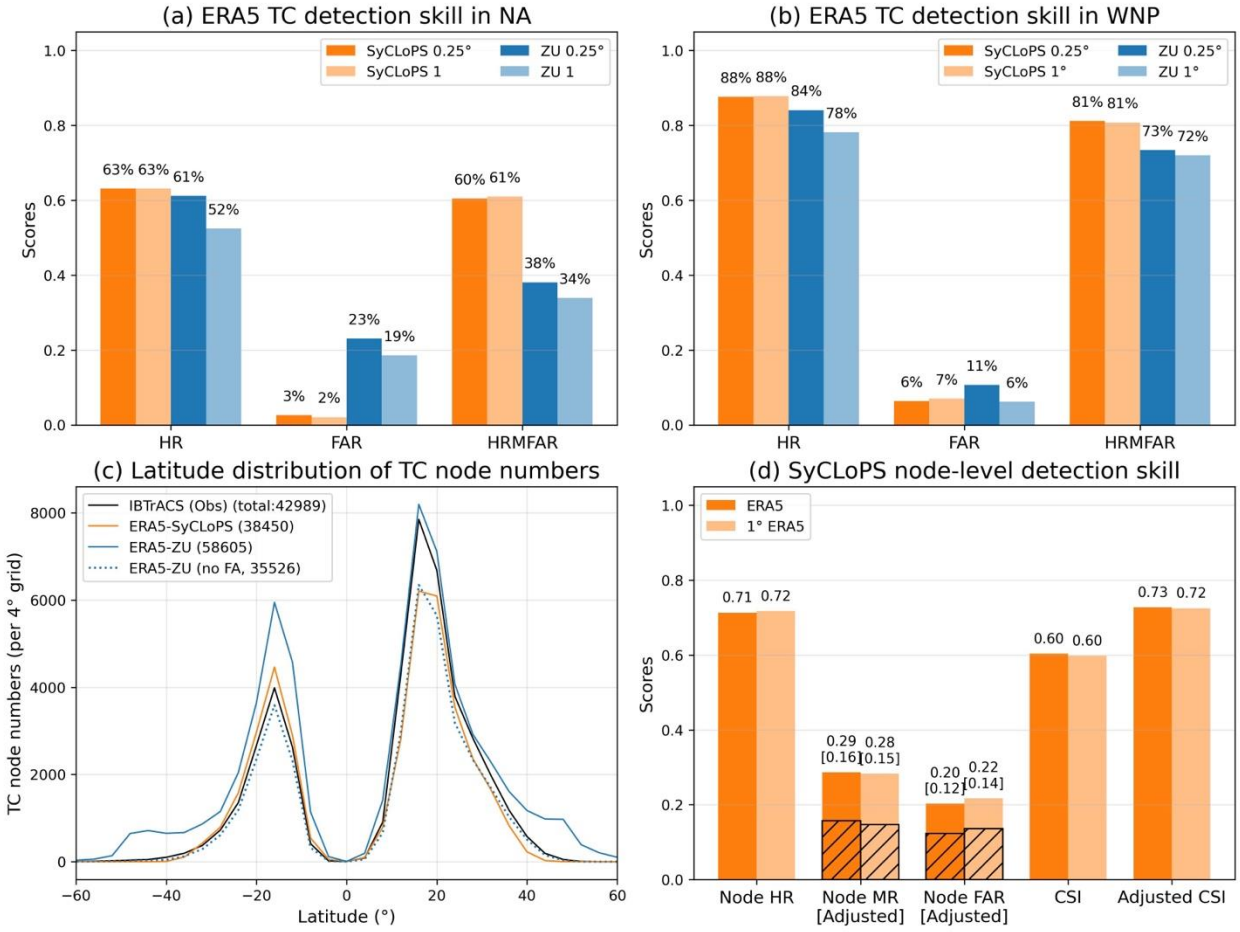


**Figure 3.3**. Summary of TC detection performance of the SyCLoPS (in oranges) and ZU (in blues) tracker based on evaluations of 1988-2014 ERA5 against IBTrACS.

Figure 3.4 presents models' performance for replicating observed global TC node density, as assessed by both trackers. For most models and ERA5, SyCLoPS provides greater accuracy than ZU, primarily by reducing errors in standard deviation magnitude. However, SyCLoPS performs more poorly for the EC-Earth3P model. Closer examination reveals that ZU selects a significantly larger number of weak tropical LPS nodes than SyCLoPS (not shown), many of which are likely FAs. The two models with the highest resolution (MRI-S and IPSL-VHR) are not clearly better than the other models when it comes to TC performance under either tracker, mainly due to their overestimations of TC frequency. In general, we find that the patterns of TC node density differences between the two trackers in coupled simulations do not differ from those in their uncoupled counterparts (not shown explicitly). Performance of CNRM improves most notably in its coupled run in both trackers. The ocean coupling likely reduces TC frequency and greatly improves its TC density spatial correlation. Noticeably, SyCLoPS leads to comparable improvements in both the 12-model ensemble mean and ERA5. In SyCLoPS, the multi-model ensemble mean outperforms every individual model in terms of the centered root-mean-square error of TC density, with performance comparable to that of 1º ERA5. Geographically, after SyCLoPS removes the many likely FAs detected in ZU, both ERA5 and the ensemble mean produce fewer TCs than IBTrACS in the NA and WNP. Noticeably, SyCLoPS produces much better TC spatial distributions in the Southern Hemisphere. SyCLoPS slightly (but consistently) underestimates TCs north of 30º N in the NA, indicating it is more conservative in identifying high-latitude TCs. TCs at these latitudes can display both subtropical and tropical features, making many of them hybrid systems (see Chapter 3.7.1 for further discussion).

In summary, we show that using SyCLoPS significantly reduces the number of false alarm TC tracks while maintaining HRs under different grid resolutions. It also reduces false alarm nodes along TC tracks, including post-TC extratropical or subtropical cyclones and LPS nodes that are likely too weak to be considered TCs. Together, these improvements help reduce the loss of detected TCs due to

discrepancies in model grid resolution and prevent non-TCs from polluting the TC evolution datasets used for this study. The results shown in Figure 3.4 further demonstrate that SyCLoPS is an effective tracking algorithm for more accurately representing and evaluating TCs simulated by climate models.

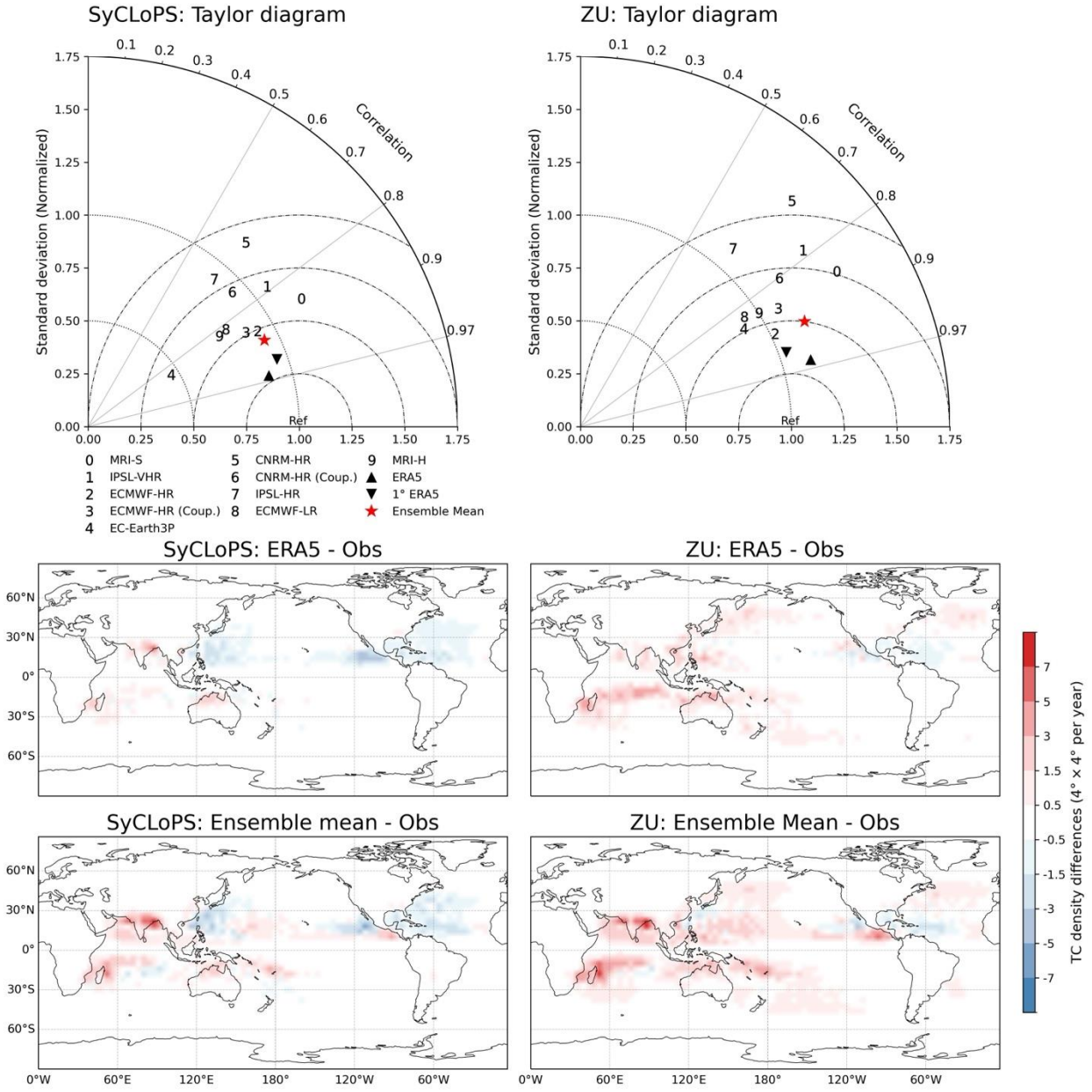



**Figure 3.4.** 1988-2014 4º × 4º TC annual mean density differences between models and IBrACS (the reference (Ref) and observations (Obs)) by both trackers. The two Taylor diagrams at the top combine information about the normalized variance (along radial distance) and spatial correlation (along azimuthal angle) of TC density. Points closer to the "Ref" location have a lower root-mean-square error (dashed

concentric isoline), indicating more skillful performance. Model identifiers in the Taylor diagrams are ordered by decreasing resolution measured by the model atmospheric mesh spacing at the equator. "(Coup.)" in model names indicate that they are coupled simulations. The two HadGEM models are not shown in the Taylor diagrams because their estimates of TC frequencies are too high and would be placed outside the display range in both trackers.

### 3.5 TC Intensity and Size in Observations

The time evolution of an average NA and WNP TC in intensity-size space is depicted in Figure 3.5. We choose MSLP over maximum wind speed as the intensity indicator to avoid interagency inconsistencies in wind measurement standards and because models tend to better represent MSLP minima than wind speed maxima (Davis, 2018; Manganello et al., 2012; Zarzycki et al., 2021). To be consistent with the wind radii and NA data sources, we use MSLP data from the JTWC in the WNP. The best-fit curve in Figure 5a suggests that the evolution of an average TC is analogous to the evolution of "main sequence" stars on a H-R diagram. The point of lifetime maximum intensity (LMI) for an average TC can be used to divide this diagram into four subspaces, by using the lowest MSLP (~970hPa) and associated TC size at this point (105 n.mile). Under this classification, the four subspaces are named "weak dwarf (WD)," "strong dwarf (SD)," "super giant (SG)," and "weak giant (WG)" based on their intensity and size. The evolutionary pathways of each TC are termed as follows: (1) the WD path is assigned to TCs that never move beyond WD; (2) the WG path includes TCs that achieve the WG category but never intensify below 970 hPa; (3) the SD-expanding (SD-E) and SD-shrinking (SD-S) paths include TCs that reach the SD category at LMI and maximizing or shrinking in size, respectively, at the end of their TC stage; and (4) the SG-E and SG-S paths include TCs that reach the SG category at LMI, maximizing or shrinking in size, respectively, at the end of their TC stage.

We mark the average initial TC intensity and size, LMI and size at LMI, and the final TC intensity and size for each pathway in Figure 3.5a. All paths begin in the lower right WD space and generally exhibit a similar initial size, except for WG-path TCs which begin with a larger size. The

intensity and size of a WD-path TC changes little, while a WG path TC expands more rapidly with intensification but does not change substantially after reaching its LMI. An SD-path TC expands slowly but quickly intensifies to the SD stage. Depending on the posterior scenario, these storms can then rapidly weaken and shrink (SD-S) or moderately expand towards a WG (SD-E). SG paths tend to start at a slightly larger size, expanding moderately and intensifying quickly. This makes them the largest and strongest on average at their LMIs. SG-E paths grow even larger at the end of their lifecycle, highlighting the threat they pose at each stage, while SG-S storms dissipate into the larger end of the WD quadrant.

In Figures 3.5b-c, we present examples from each of the evolutionary pathways in the NA and WNP, overlaid with the joint probability density function (pdf) showing the normalized frequencies of TCs in intensity-size space. The largest TC size in both basins is around 350 n.mile during their observation period. Notably, the WNP has a larger number of intense TCs, particularly in the SG category, than the NA. Consequently, the WNP has a slightly larger average TC size (89 n.mile) than NA (79 n.mile) over 2002-2024. Some examples from the NA are shown in Figure 3.5b and were mentioned in Chapter 3.2. *Katrina* (2015) is classified as an SG-S, while *Milton* (2024), which had a peak intensity similar to *Katrina*'s but was considerably smaller, is classified as an SD-E. The “expanding” annotation indicates that *Milton* expanded as it underwent ET. The costly and deadly Hurricane *Sandy* (2012) is an SG-E, but is on the weaker side. Its unusual expansion at the end of ET makes it one of the largest TCs in the NA during the analysis period. In the WNP, the highly impactful Typhoon *Mankghut* (2018), which follows a path through the upper left portion of the SG space, represents a type of TC not found in the NA. In contrast, the very intense *Haiyan* (2013) was relatively small, placing it on the larger end of the SD-S category.

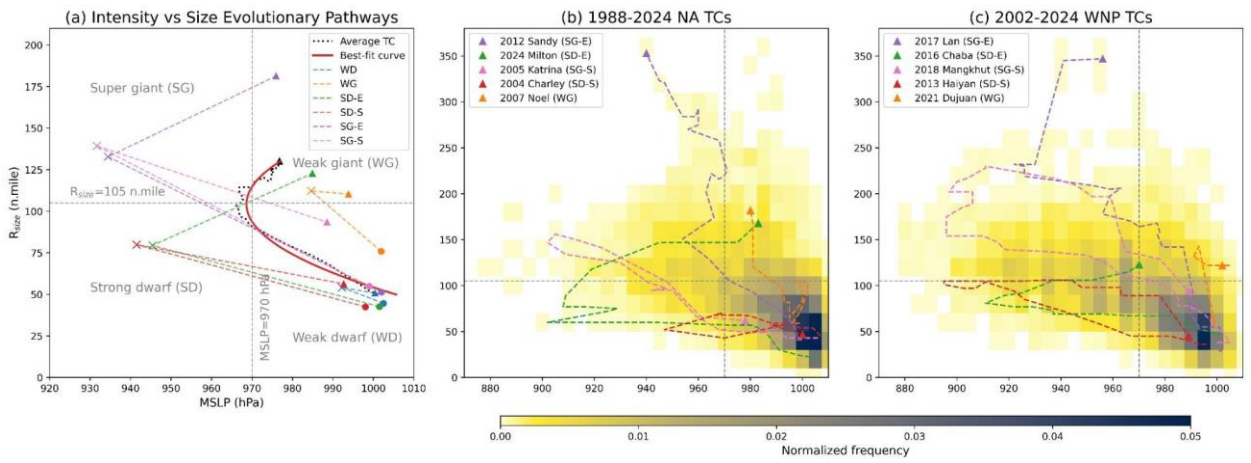



**Figure 3.5**. (a) The 4 subspaces of the intensity (hPa)-size (n.mile) space and summaries of different TC evolutionary pathways. (b) The Observed TC frequencies (joint pdf) and classic examples in the intensity-size space in the NA during 1988-2024, and (c) in the WNP during 2002-2024. Circle, cross, and triangle in the figure marks the track at its start, LMI, and end, respectively.

Next, we analyze the relationship between the track type as defined by the precursor and posterior scenarios and the evolutionary pathway defined above. Figure 3.6 shows that in both basins, ET tracks are more likely to follow the more dangerous SG-E path, while LD, non-ET, and TT type TCs are more likely to stay weaker as a WD. In the NA, not surprisingly, there are many more TT tracks and far fewer MD tracks compared to the WNP. WGs are mostly found in TD-ET and TT tracks, while SG-E paths are dominated by the TD-ET track type featuring longer track lengths and recurvature over ocean as depicted in Figure 3.1. TD-LD tracks favor shrinking SD and SG paths as TC winds are suppressed by overland friction. In the WNP, WG paths are most frequently found with MD-ET and TT tracks. A WNP MD track is more likely to develop into an SG than an SD. Notably, these MDs usually form closer to the continent (see Figure 3.1) and are more likely to landfall than other precursor types; therefore, they tend to shrink in their later stage. However, this result also highlights their ability to grow large in a shorter period of time close to land. Conversely, TD tracks, especially those ending in LD or NT, are more likely to start shrinking into a smaller SD than a SG. Lastly, we compare the differences between NA and WNP TD-type TCs. The results show that, although the WNP favors stronger TD-type TCs, they don't necessarily

grow larger than those in the NA. This implies that the greater SG numbers in the WNP mainly arise from MD-type TCs.

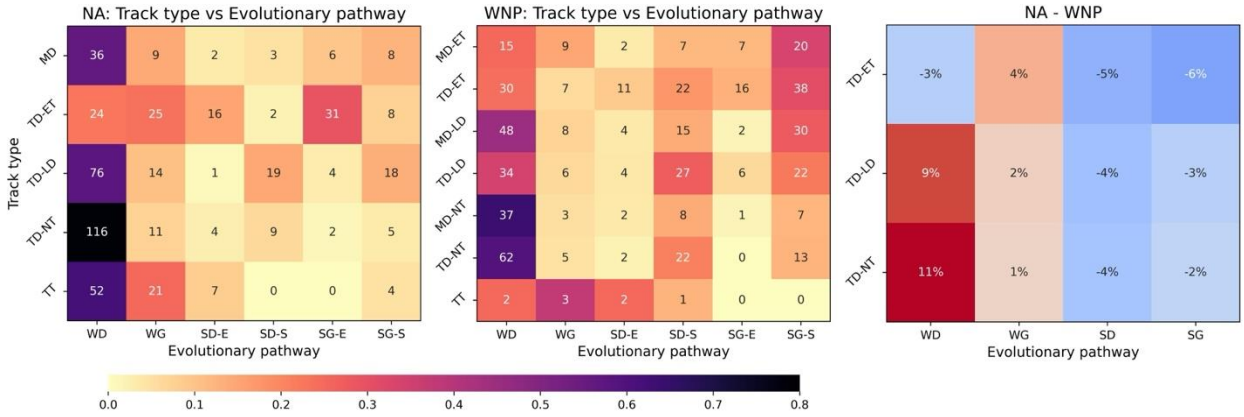



**Figure 3.6**. Track type versus evolutionary pathway heatmaps for observations in both basins for the NA during 1988-2024 and the WNP during 2002-2024. The heatmaps for each basin are normalized horizontally for each track type, and the total TC track counts are shown on top. The difference between the two basins is based on normalization across all the 12 categories shown in the subfigure.

### 3.6 TC Intensity and Size in HighResMIP Models

The joint pdfs of TC frequencies in the intensity-size space for all models during the overlapping period (1988-2014 for the NA and 2002-2014 for the WNP) in the two basins are depicted in Figure 3.7. Unsurprisingly, many models fail to produce TCs as intense as those in observations. MRI-S (the highest resolution model in this study) and CNRM-HR simulate more intense TCs than the other models, with high-end intensities comparable to observations. In alignment with observations, models produce the most intense TCs around or below 20° N, with the exceptions of ECMWF-LR and the two IPSL models. Unlike intensity, the largest TC sizes simulated by the models (~300-400 n.mile) are mostly comparable to those in IBTrACS. This suggests that modeled TCs tend to increase in size much more quickly with intensity than observations, rendering most of the models, including ERA5, incapable of producing the substantial number of observed SDs. Although models such as CNRM-HR simulate more realistic TC

intensities, they still fail to capture the observed intensity–size relationship. In contrast to IBTrACS, models' most intense TCs are also their largest ones.

In observations, one group of more intense TCs tends to be smaller and reside in the lower tropics, indicative of the constraint on size by the Rhines scale in the low latitudes (Chavas & Reed, 2019; Lu & Chavas, 2022). These very intense TCs are most commonly observed to be close to the threshold that separates "giants" from "dwarfs" (105 n. miles). Another group of weak, larger TCs are generally found between 30° and 40° N, which should contain TCs undergoing ET. Overall, there is a monotonically increasing relationship between size and latitude, which can be explained by the $\beta$-plane of the Rhines scale and the transport of larger planetary angular momentum (Chan & Chan, 2014; Chavas & Lin, 2016). The increase in average latitude with size is less pronounced in the models, likely due to the blending of large, intense TCs that primarily reside in low latitudes in the models. This observation is most evident in CNRM-HR, where the average latitude of TCs at the higher end of the size dimension steadily decreases to around 20° N, matching that of the most intense TCs in the model. ERA5 exhibits an intensity–size relationship similar to that of ECMWF-HR, the climate model that shares the same underlying atmospheric model. However, the largest TCs in ERA5 occur, on average, at more realistic latitudes than most models, likely owing to the benefits of data assimilation.

The coupled simulation counterpart indicated by "(Coup.)" in the model names generally reduces the number of large, intense TCs in HadGEM-HR and CNRM-HR, and it only reduces the number of very intense TCs in ECMWF-HR. Nevertheless, the relationships between latitude and size as well as intensity and size remain largely unchanged across the three models.

On a positive note, we find that, in general, size expands more moderately and realistically with intensity as model resolution improves. The MRI-S model (~20 km resolution) successfully reproduces many SDs, as well as the flattening of size with increasing intensity as seen in observations. However, it is not directly clear why there exist two inflection points in the intensity-size relationship from MRI-S. We attempt to discuss this in Chapter 3.7.2. Resolution does not appear to be the primary factor affecting TC performance in terms of intensity and size: MRI-H (~60 km resolution) somewhat captures the

flattening of the intensity-size relationship at medium intensity. A similar signal is present in CNRM, though much less pronounced. This suggests that it may be some feature of the MRI model design that is better capturing this relationship.

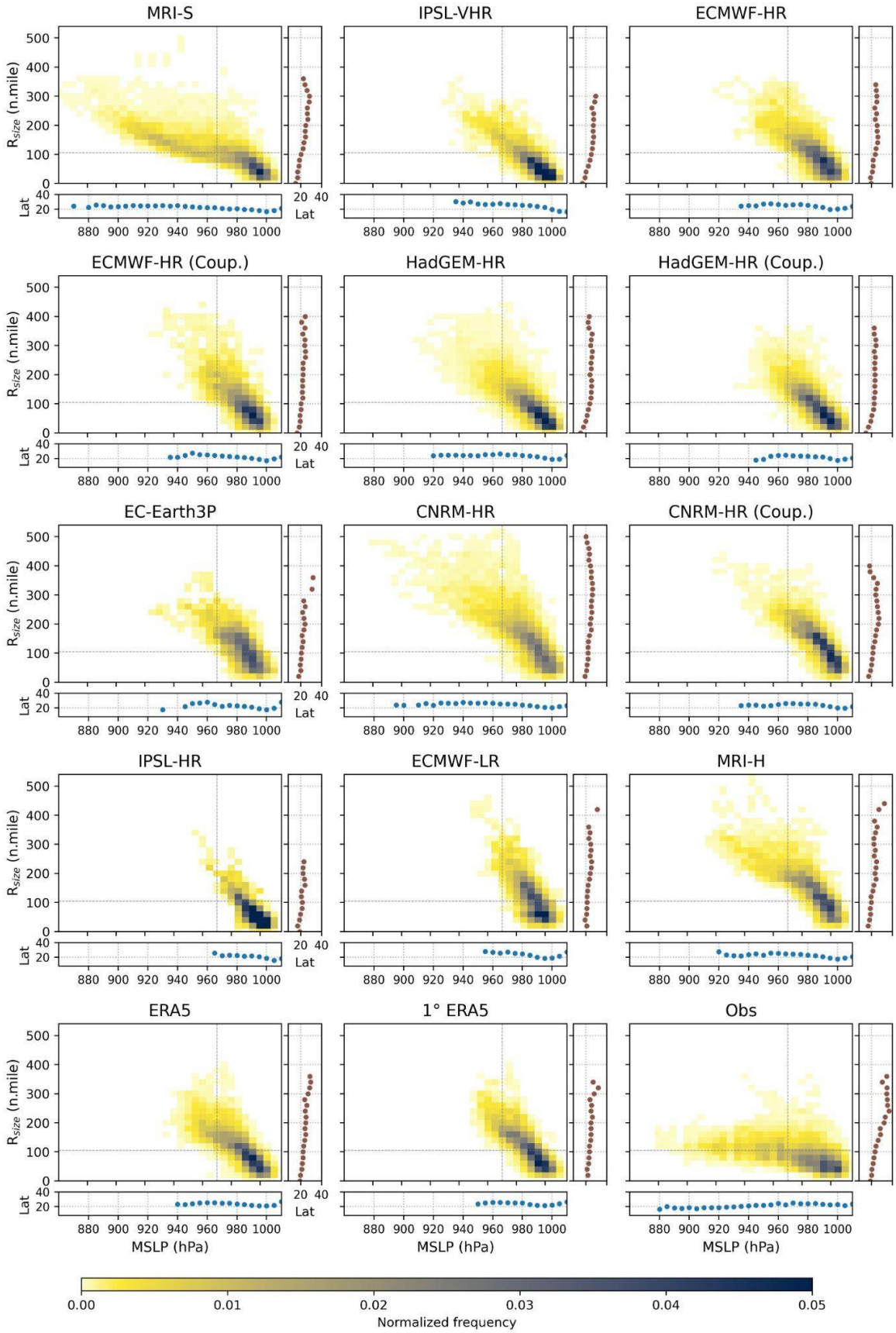

MRI-S
IPSL-VHR
ECMWF-HR
ECMWF-HR (Coup.)
HadGEM-HR
HadGEM-HR (Coup.)
EC-Earth3P
CNRM-HR
CNRM-HR (Coup.)
IPSL-HR
ECMWF-LR
MRI-H
ERA5
1° ERA5
Obs
R_size (n.mile)
Lat
MSLP (hPa)
Normalized frequency
0.00
0.01
0.02
0.03
0.04
0.05

**Figure 3.7**. Joint pdf of the normalized frequency of MSLP (hPa) and TC size represented by $R_{size}$ (n.mile) of each model, reanalysis, and IBTrACS observations over the overlapping period in the two basins. The average latitude at which TCs occur are shown for each bin in the intensity and size dimensions (values at bins with less than five samples are not shown). Subfigures of models (the first three rows) are ordered from highest to lowest model horizontal resolution.

Heatmaps of track type and evolutionary pathway for all the models are depicted for the NA (Figure 3.8) and WNP (Figure 3.9). Consistent with the pdfs in Figure 3.7, most models (including ERA5) produce few SD-path TCs in either basin. However, models that simulate more realistic intensities do produce several SG-path TCs. These larger TCs are mostly classified as the TD-ET track type, which aligns with observations.

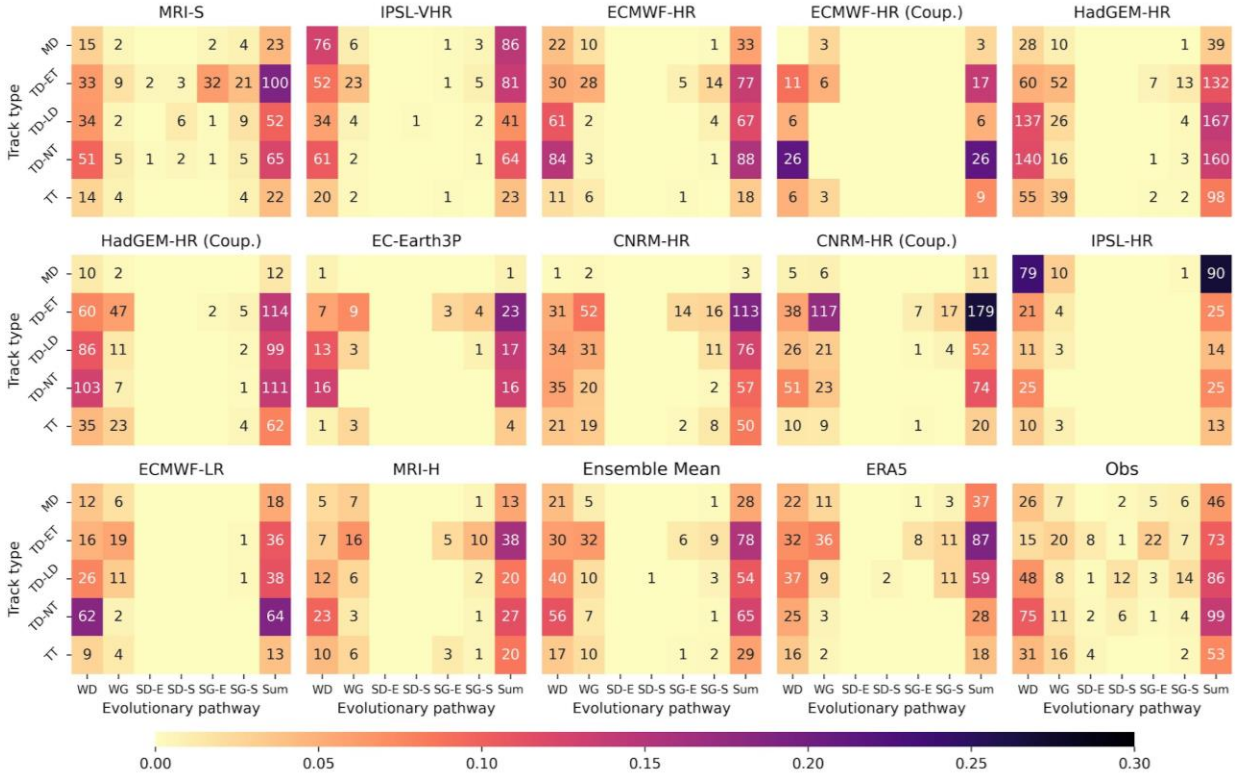



**Figure 3.8**. Heatmaps of normalized TC frequency of different TC track type and evolutionary pathway, overlaid with TC track counts in the NA basin for the overlapping period (1988-2014). The last column sums up the counts in all pathways for each track type.

In the NA, models that produce SGs generate too few SG-path TCs for the TD-LD track type compared to observations. MRI-S, the best-performing model, produces generally more realistic numbers for SD and SG paths. However, MRI-S tends to produce a larger number of ET tracks and too few LD tracks, resulting in a significantly higher ET fraction compared to IBTrACS—a trait also observed in ERA5 and some of the other models. On the other hand, models (including ERA5) tend to significantly underestimate TT track numbers in the NA. These discrepancies could be partially due to biases introduced by the tracking algorithm because these hybrid TCs are more likely to be missed by the tracker (see Chapter 3.7.1 for discussion). Additionally, the HadGEM models simulate too many TCs, and the IPSL models significantly overestimate MD tracks. ECMWF-HR produces reasonable numbers for each TC track type, except for the TT type. Interestingly, despite an overall underestimation of total TC counts, the proportion of TC tracks of each track type in the ensemble mean matches the observed proportions better than ERA5 in the NA basin. Coupled simulations generally produce fewer NA TCs in each category than their uncoupled counterparts. While this improves the overestimation in HadGEM-HR, it reduces too many TCs in ECMWF-HR, and the results are mixed for different track types in CNRM-HR.

In the WNP, MRI-S again outperforms other models in simulating SD and SG paths for each track type. However, it does not capture the pattern of more landfalling SG-path TCs with an MD precursor and more landfalling SD-path TCs with a TD precursor. Interestingly, the coarser MRI model (MRI-H) captures this pattern more accurately than its high-resolution counterpart. CNRM-HR reasonably reproduce the composition and count of each track type for the SG paths, while the three ECMWF models simulate total TC numbers that closely match observations. However, the two HadGEM models generate too many TD-type TCs and too few MD-type TCs. Overall, as shown in the ensemble mean, the models produce a reasonable total number of WNP TCs but simulate too few MD tracks and too many TD-NT tracks compared to observations. The effects of coupling in the WNP are mixed. Total TC frequency is substantially reduced in CNRM-HR, resulting in an underestimation relative to observations, whereas it is further elevated in HadGEM-HR, which already severely overestimates TC activity.

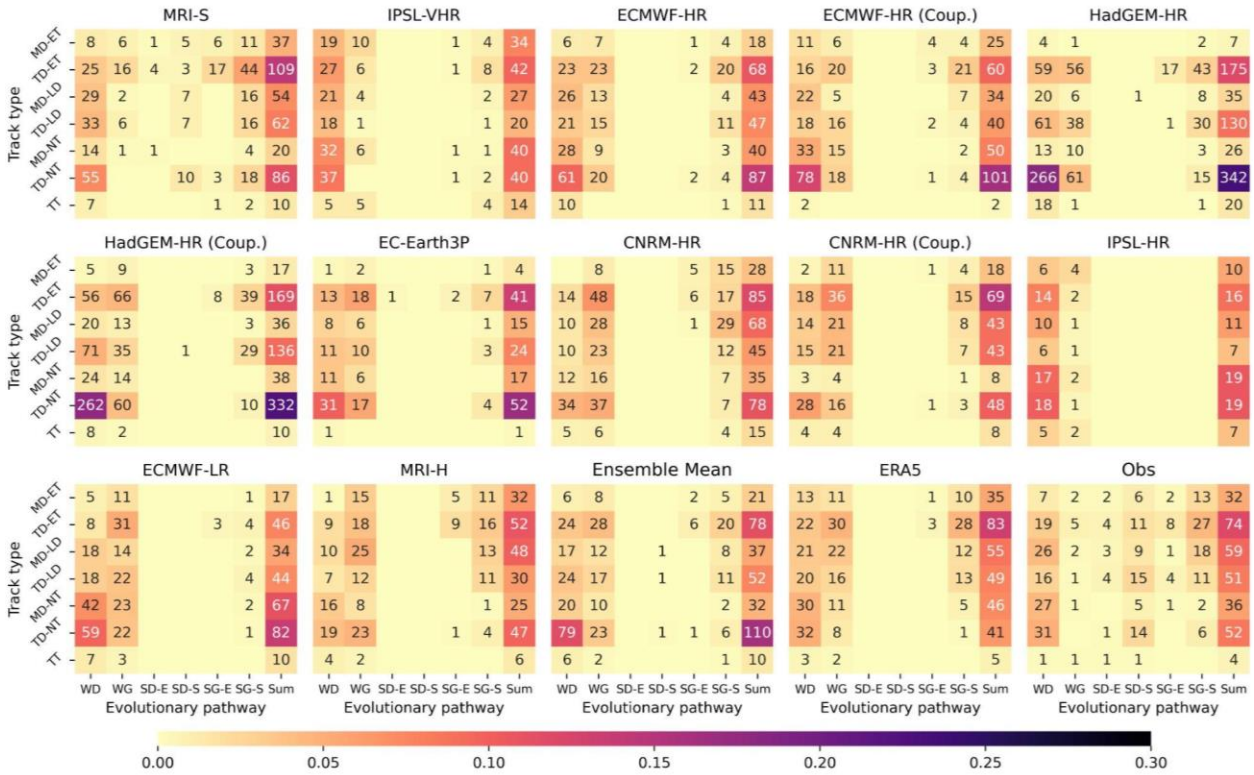



**Figure 3.9**. The same as in Figure 3.8, but for the WNP during the overlapping period (2002-2014).

Figure 3.10 compares TC counts across track types and evolutionary pathways between El Niño and La Niña phases of ENSO for atmosphere-only models, ERA5, and observations in both basins. Statistical significance (p-value < 0.1) is assessed for each individual model and for ERA5 using a two-tailed t-test. Ensemble-mean significance is evaluated with a t-test applied to the mean differences across the nine models. In addition, the ensemble-mean result is considered statistically significant only if at least 7 of the 9 models show differences with the same sign. Only when the category is statistically significant are annotations of differences in percentage shown.

In the NA, the most robust signals in observations are the smaller total number of super giants and the accompanied elevated numbers of SDs in some track types in El Niño years compared to La Niña years. Additionally, El Niño years clearly favor more TT cases. Although model results vary, most models and the 9-model ensemble mean successfully reproduce the significantly larger number of TT tracks and smaller number of SGs in some categories. ERA5 and its corresponding ECMWF-HR model

show similar patterns and do not outperform the ensemble mean compared to observations. In the WNP, the most robust signal in observations is the larger number of SGs accompanied by the reduction in WDs in in El Niño years compared to La Niña years. This pattern is very well captured by the model ensemble mean.

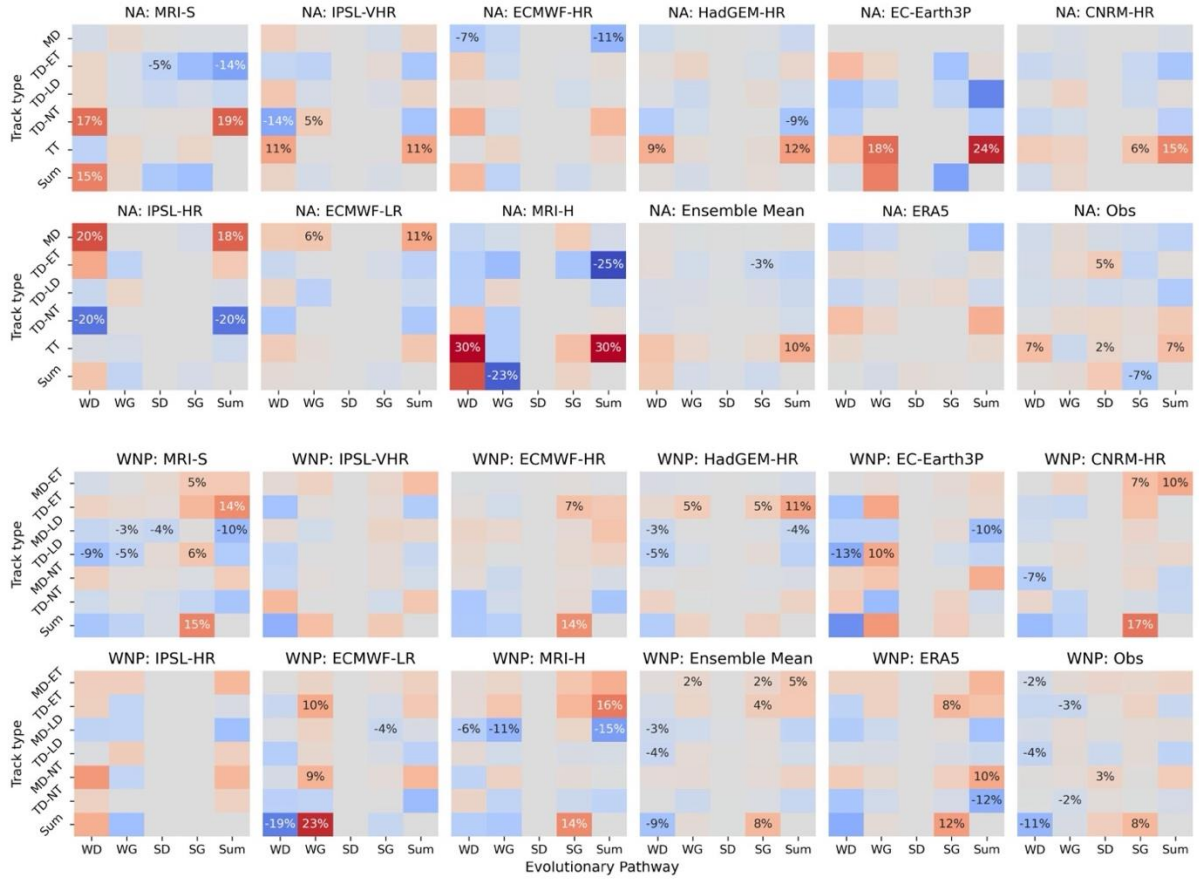



**Figure 3.10**. Heatmaps showing the percentage differences in the number of TCs in each track type and evolutionary pathway category between El Niño and La Niña years in both basins for each atmosphere-only (uncoupled) model, ERA5, and IBTrACS. El Niño (La Niña) years in the models and in ERA5 are defined as those during the overlapping period with an average Niño 3.4 index from May through November greater than 0.4 K (less than -0.4 K), as calculated using the HadISST dataset. For IBTrACS, the ending year of the overlapping period is extended from 2014 to 2024 to account for more samples for producing more robust observed differences. Percentage notations are only shown when the underlying category passes the significance test with a p-value of less than 0.1.

## 3.7 Discussion

### 3.7.1 Missing TC tracks in the NA

The remarkably lower HR in the NA may contribute to biases in model TC performance in this study. ERA5 NA TC numbers are underestimated in both SyCLoPS and ZU trackers. The ERA5 and IBTrACS heatmaps in Figure 3.8 further suggest that the misses in SyCLoPS are primarily landfalling and weak Non-ET and TT tracks. As shown in Figure 3.1, TT tracks are also much more prevalent in the NA than in the WNP. Some small, weak TCs are likely underrepresented in ERA5 due to biases in the reanalysis and its ~31 km resolution, but this issue is likely not unique to the NA basin. We further inspect tracks not labeled "TC" but detected in SyCLoPS and find that many of these tracks are classified as subtropical storms, a type of subtropical hybrid LPS that shares similarities with TCs but are usually weaker and smaller with shallower convection. These systems may also be responsible for the majority of the large number of FAs detected in ZU. This misclassification might be attributed to the deficiencies in the tracking criteria used by both trackers and in ERA5 itself. However, we do not observe this phenomenon in the WNP, suggesting the distinct hybrid nature of many NA TCs as identified by the NHC. Guishard et al. (2009) found that 27% of North Atlantic subtropical hybrid (shallow warm-cored) storms detected in the reanalysis overlap with IBTrACS records, highlighting a "gray zone" in how agencies define TCs. Since these storms are prevalent in the South Atlantic (Evans & Braun, 2012), this could imply that some of these South Atlantic systems would be classified as TCs by the agency if they occurred in the NA. In the latest pre-release version of SyCLoPS, we identify a new LPS category, "Hybrid TC," which primarily resides in the subtropical NA. Preliminary results indicate that incorporating this new category into TC detection in the NA recovers a portion of the missed hits while maintaining low FARs in both ERA5 and JRA-3Q.

### 3.7.2 Factors modulating TC intensity-size relationship in models

Moon et al. (2020) and Roberts et al. (2015) found that overestimation of the TC radius of maximum wind (RMW) and size is greater in models with coarser resolutions, and further that the bias

remains with the finer (0.25°) resolution models when more intense wind speeds are simulated. Similarly, according to Figure 3.7, a connection may be drawn between model performance in reproducing the observed TC intensity-size relationship and model resolution. Nevertheless, improvements in model resolution appear to contribute only partially to capturing the relationship: Flattening is only apparent in the MRI and CNRM family of models, with the most obvious flattening occurring in the 20 km MRI-S model.

Curiously, the MRI model exhibits two clear inflection points. The initially faster-than-observed size expansion in all models likely stems from their relatively coarse horizontal grid spacing, which is insufficient to resolve the rapid contraction of RMW during early-stage intensification. Next, in alignment with observations (but distinct from other models), TC size in MRI models largely stops growing as the intensity increases further, indicating that RMW stabilizes. This implies that the TC grows large enough to feel the constraints of the Rhines scale, which acts to shrink TCs to an upper limit (Lu & Chavas, 2022). The ability to maintain a stable TC size at this stage likely depends in part on the model formulation, so different configurations in other models may contribute to their overestimation of size at this stage. As Nardi et al. (2022) demonstrated with a ~25 km model, the inward movement of the eyewall outer wind radii, as well as the profiles of the secondary circulation—which reflects the radially inward transport of angular momentum—are sensitive to perturbations in the parameters of the planetary boundary layer turbulence scheme. Similar conclusions were reached in Gopalakrishnan et al. (2013) using a model with 3 km resolution. Previous studies have also emphasized the role of the diabatic heating generated by the TC outer region's convection activity and rainbands in the expansion of TC size (Fudeyasu & Wang, 2011; Hill & Lackmann, 2009; Wang, 2009). Therefore, cumulus convection scheme parametrizations can also be critical to modeling realistic TC sizes. The second inflection point in the MRI models appears to be because the deepening MSLP has exceeded the MSLP gradient permitted for the given model resolution, leading the wind field to progressively expand as the system intensifies. While outside the scope of this study, it could be beneficial to analyze the full wind structure evolution of

very intense TCs simulated in different HighResMIP models, and to compare this with finer wind field observations of analogous TCs.

Lastly, as implied by Figures 3.8-10, the TC intensity-size relationship, as indicated by the evolutionary pathways, can also be modulated by TC track types (i.e., the geographic locations and orientations of tracks) and ENSO. These factors are largely controlled by SST patterns and the large-scale environment. Theoretically, air-sea interactions in the coupled HighResMIP model runs should hold an advantage over uncoupled simulations in more realistically representing the underlying environment. However, our analysis finds no improvement in the representation of the intensity-size relationship in the coupled simulations compared to their uncoupled counterparts.

### 3.8 Summary and Conclusions

This study examines the relationship between the intensity and size of TCs and their time evolution in the intensity-size domain using observations and high-resolution HighResMIP models. Leveraging a new tracking algorithm and a new method to estimate TC size in gridded data that is better aligned with IBTrACS's wind radii format, we can detect TCs more robustly in models and compare modeled TC size with IBTrACS. Our main findings are summarized below.

Chapter 3.4 compares two TC-tracking algorithms in ERA5 and demonstrates that SyCLoPS improves upon existing TC trackers. Its ability to remove false alarm TC tracks and nodes along each track prevents extratropical LPSs and weak tropical systems from polluting an objective TC dataset. As a result, TC performance is improved with SyCLoPS for many models and ERA5. With this new tracker, we find that the 12-model HighResMIP ensemble mean performs comparable to the regridded 1º ERA5. A new adjustable MSLP threshold in SyCLoPS has been implemented to optimize TC detection performance for a range of grid resolutions, which allows for a fairer comparison between lower and higher resolution models. This feature could also prove valuable for detecting TCs in increasingly popular AI-driven climate models that are trained on 0.25º ERA5 data but operate at much lower grid resolutions (e.g., Brenowitz et al., 2025; Kochkov et al., 2024).

Inspired by Musgrave et al. (2012), Chapter 3.5 compares the distribution and evolution of observed TC in intensity–size space and draws an analogy with stars in a H–R diagram. We identify four categories of TCs (subspaces)—weak dwarfs (WD), strong dwarfs (SD), weak giants (WG), and super giants (SG)—and classify TC evolution into 6 evolutionary pathways based on the subspaces where TCs reach its LMI and terminate. The main takeaways are:

- WG storms start with the largest initial size, show the fastest size growth, but reach low peak intensities. SG storms begin with a moderate initial size yet reach the highest LMI and the largest size at LMI and the final stage. The extremely impactful *Katrina* (2005) and *Sandy* (2012), as well as typhoon *Mangkhut* (2018) all belong to this group.
- SDs and WGs can also be highly destructive—SDs via extremely intense winds, WGs via extensive precipitation. The evolution-path clustering framework thus provides useful guidance for stakeholders to prepare for different disasters brought by different types of TCs.

By leveraging LPS class/status labels provided in SyCLoPS and IBTrACS, we can classify TC track types based on their precursors and whether they make landfall or complete ET at the end. We show that in observations:

- TC tracks from a MD precursor in the WNP favors "giant" TC paths over "dwarf" paths. This aligns with Liu & Chan (2002) and Teng et al. (2019), who suggest that larger TCs in the WNP are frequently associated with a monsoonal flow surge and the monsoon trough. Because a larger portion of MDs make landfall than other precursor types, these giant, landfalling MDs are likely among the most destructive.
- A significant portion of ET tracks continue to grow and maintain relatively high strength at the end of their TC stage, as represented by the SG-E path. These TCs could pose a significant threat to extratropical regions as post-TCs (Sainsbury et al., 2020). The expansion of some TCs during ET does not contradict previous studies that found TC sizes generally remain unchanged during ET (Nguyen & Schenkel, 2025; Schenkel et al., 2018). However, we note that ET tracks are more likely than other track types to follow the SG-E path, especially in the NA.

In Chapter 3.6, we analyze and validate TC intensity-size relationship and the relationship between TC track types and evolutionary pathways in HighResMIP models. The main findings are:

- All models, except for the highest resolution model MRI-S, and ERA5, fail to capture the TC intensity-size relationship and produce barely any "strong dwarfs." We suggest that MRI-S performs better not only because of its higher resolution, but also because of the model formulation, such as its planetary boundary layer and cumulus/convection scheme.
- While the models vary in their representation of the relationship between track types and evolutionary pathways, the ensemble mean reasonably reproduce the observed TC track type composition in the NA and even outperforms ERA5.
- The differences in the number of TCs among track types and evolutionary pathways between El Niño and La Niña years are broadly captured by most atmosphere-only models. The ensemble mean in both basins closely matches ERA5 and observations.

This study demonstrates the value of using the ensemble mean to minimize model errors, as using the HighResMIP model ensemble mean produces superior results in the distribution of TC density, TC track type composition, and differences between ENSO phases. The study also suggests that a model resolution finer than ~0.25° is required to fully reproduce the observed TC intensity-size relationship. Although, we also show that atmosphere-only models with the highest resolutions tend to overestimate TC frequencies. This overestimate could be partly neutralized with a coupled ocean, as the ocean provides essentially a limitless source of energy in atmosphere-only models with prescribed SST. However, coupling opens the door to SST biases that are potentially even more problematic for TC distribution, and we find no evidence that it improves the intensity–size relationship in the three HighResMIP models. Therefore, the most successful Earth system model should balance performance between correct TC frequencies and its ability to capture the intensity-size relationship. The upcoming HighResMIP2 model outputs, which promise higher atmosphere and ocean resolutions (Roberts et al., 2025), the recent emergence of many kilometer-scale dynamical models (e.g., in Judt et al., 2021), and the combination of

generative AI and dynamical downscaling of climate model ensembles, as proposed by Google (Lopez-Gomez et al., 2025), may provide useful insights towards solving this compound problem.

## 3.9 Appendixes

### Appendix 3A: Details of Tracking algorithms

Both the ZU method and SyCLoPS define an LPS candidate as a nodal point whose MSLP must increase by at least a threshold amount within a great-circle distance (GCD) in all directions. This requirement is referred to as the MSLP closed contour criteria (MSLPCC; CC stands for "closed contour"). MSLP minima are merged to the lowest MSLP nodal point within 6º GCD for both algorithms. Here, we use a 6-hourly temporal frequency for both algorithms. We also use 250 hPa level data instead of the default 300 and 200 hPa level data originally used in the two algorithms to accommodate data availability in the climate models we use. Han and Ullrich (2025) demonstrated that resampling the resulting LPS catalogs to 6 hours or replacing 300 and 200 hPa data with 250 hPa data does not negatively impact SyCLoPS's detection skill.

ZU finds LPS candidate nodes that satisfy:

1. An increase in MSLP of 200 Pa within a radius of 5.5° GCD (MLPSCC55 > 200 Pa).
2. A warm core indicated by a maximum upper level (250-500 hPa) thickness within 1° of the LPS center that decreases by at least 6 m within 6.5º (UPPTKCC < -6 m).

Then, consecutive candidates are stitched in time when they are found within 8º GCD of the current candidate node (8º stitching distance) to form LPS tracks. Tracks that meet the additional criteria below is defined as TC tracks:

1. The underlying surface elevation is below 15 m for at least 10 6-hourly time steps.
2. The maximum near-surface wind speed exceeds 10 m/s within 2.0º for at least 10 6-hourly time steps.
3. The track remains within the latitude band of 50°S–50°N for at least 10 6-hourly time steps.

SyCLoPS finds all detectable LPS candidates that are at least 10 Pa lower than their surroundings within 5.5º (MSLPCC55 > 10 Pa). The detected LPS candidates are stitched in time when they are within 6° of each other. Note that the stitching distance is reduced from 8º in ZU. An explanation of this change and an alternative method to keep the 8º distance is provided in Text 3.S2. Tracks with an MSLPCC55 greater than 100 Pa for at least three 6-hourly time steps are retained. The first two steps are done in TE. Next, a Python program classifies TC nodes and TC tracks using the following criteria:

1. The LPS candidate's mean relative humidity (RH) at 850 hPa within 2.5º is at least 60%.
2. The LPS candidate has a 100 hPa maximum RH within 2.5º that exceeds 20%, a mean 250-850 hPa deep layer wind shear within 10.0º below 13 m/s, and a central 850 hPa temperature over 280 K (the tropical LPS criteria). Note that all the model 100 hPa RH data is provided at daily mean frequency. This inconsistency does not affect SyCLoPS performance (see Han and Ullrich, 2025, appendix B).
3. LPS candidates that satisfy the TC criteria are labeled as TC nodes. The TC criteria states that, first, a TC node must have a low-level (700-925 hPa) warm core indicated by a local thickness maximum within 1.0º (LOWTKCC < 0 m) and a UPPTKCC < -117.6 $m^2\ s^{-2}$ (12 m). The MSLPCC20 threshold varies by grid resolution (see Appendix B for details on how these thresholds are determined).
4. LPS tracks that have at least three 6-hourly LPS time steps that are classified as TC nodes are identified as TC tracks.
5. LPS labels in the TC tracks are adjusted to reduce uncertainties arising from hard cutoff TC criteria thresholds and noise in the reanalysis and model data. This adjustment improves TC node CSI by ~1%. Please see Text 3.S4 for details on how we make the adjustment.
6. Detected TC nodes are defined as the TC-labeled nodes in the TC tracks.

SyCLoPS tracking and classification can be performed via the SyCLoPS Python program (Han & Ullrich, 2026) upon configuring a few user settings.

**Appendix 3B: Varying detection thresholds with grid resolution in SyCLoPS**

SyCLoPS, as well as many other trackers, are trained on a specific reanalysis with fixed horizontal resolution. For example, the TRACK algorithm (Hodges et al., 2017) was calibrated using the European Center for Medium-Range Weather Forecast operational analyses from 2003 to 2005 (Bengtsson et al., 2007), the ZU method (Zarzycki & Ullrich, 2017) is initially optimized on JRA55 and CFSR (Saha et al., 2010), and SyCLoPS is trained on ERA5. Although the latter two algorithms are based on TE's closed contour criteria, which by its design, minimizes the sensitivity of the closed contour delta value to grid resolution, resolution may still be a factor in determining the optimal thresholds of the TC detection scheme's closed contour criteria, as gradients are smoothed out at lower resolution.

To verify this theory, we test the optimal thresholds for the 0.25° ERA5 on the coarsened 1° ERA5. We find that the detection skill, as measured by HRMFAR, decreased by ~4% from its maximum possible skill of ~71%, and the number of selected TC nodes was far lower than that of observations and 0.25° ERA5, indicating clear underrepresentation. The results show that, given the minimum LOWTKCC requirement, the UPPTKCC threshold of -117.6 $m^2\ s^{-2}$ and the time step requirement of three 6-hourly intervals are rather stable and insensitive to changes in grid resolution. However, the optimal MSLPCC20 threshold decreases by 100 Pa when the grid resolution is reduced 16-fold. This suggests that the strength of a TC's upper-level warm core is less sensitive to model resolution than the pressure gradient within the inner region, which is consistent with the finding in Moon et al. (2020). Because the MSLP field generally varies smoothly, a linear relationship between MSLPCC20 (y) and grid spacing (x in $deg^2$) can be developed using linear regression:

$$y = -107x + 247 \tag{3.B1}$$

$y$ levels off at ~240 Pa as $x$ surpass ERA5 grid resolution of 0.25° × 0.25° and approaches 0 $deg^2$. Therefore, a threshold of MSLPCC20 > 240 Pa is used for any models with a grid resolution equal to or finer than 0.25°. The threshold at any other resolutions is determined by Eq. B1 and then rounded to the nearest 5 Pa. The threshold reaches 0 when the resolution exceeds 1.5°, indicating that this detection

scheme should not be used with a resolution coarser than 1.5º. However, tracking TCs in a dataset with a grid resolution coarser than this level is generally not recommended. Note that the chosen MSLPCC20 and UPPTKCC thresholds differ from those suggested in Han & Ullrich (2025) for 250 hPa data. The updated thresholds proposed here are expected to be less sensitive to grid resolution.

We evaluate the detection scheme using the default SyCLoPS configuration and the 250 hPa configuration, optimizing thresholds to maximize HRMFAR. The default setup uses 200 hPa winds for the deep-layer shear parameter and 300 hPa geopotential for UPPTKCC. The resulting thresholds and detection performance are summarized below:

| ERA5 resolution | 250 hPa / default | Best configuration, given LOWTKCC < 0 $m^2/s^2$ | Detection skills (HR; FAR; HRMFAR; Node number; Adjusted CSI) |
|---|---|---|---|
| 0.25º | 250 hPa | UPPTKCC < -117.6 $m^2/s^2$ (-12 m); MSLPCC20 > 240 Pa; time step >= 3 (6-hourly) | 83.2%; 12.9%; 70.3%; 38048; 72.7% |
| | default | UPPTKCC < -107.8 $m^2/s^2$ (-11 m); MSLPCC20 > 215 Pa; time step >= 4 (6-hourly) | 81.9%; 12.0%; 69.8%; 40416; 73.5% |
| 1.0º | 250 hPa | UPPTKCC < -117.6 $m^2/s^2$; MSLPCC20 > 140 Pa; time step >= 3 (6-hourly) | 84.1%; 13.3%; 70.8%; 37857; 72.5% |
| | default | UPPTKCC < -107.8 $m^2/s^2$; MSLPCC20 > 135 Pa; time step >= 4 (6-hourly) | 83.0%; 13.2%; 69.8%; 37548; 72.8% |

**Table 3.B1**. Recommendations for TC detection configurations and their optimal detection skills based on evaluations of 6-hourly ERA5 against the 1988-2014 IBTrACS.

For the default setting, a linear regression between grid resolution and MSLPCC20 can be similarly constructed as $y = -85x + 220$. As shown in Table 3.B1, the detection skills, as measured by HRMFAR and the adjusted CSI, saturate at around 70% and 73%, respectively, for all scenarios tested. TC node numbers are generally slightly lower in the 1° data than in 0.25°, implying slightly higher rates of missing TC nodes when compared to 42989 observed TC nodes in IBTrACS. Similarly, using 250 hPa data slightly improves the HRMFAR but slightly compromises TC node frequencies. All SyCLoPS

updates mentioned in this work are included in the latest SyCLoPS software version (Han & Ullrich, 2026; see Chapter 2.10 for details).

## 3.10 Supplements

### Text 3.S1. Sensitivity of TC subjective data processing to TC detection skill

For instance, the Indian Meteorological Department (IMD) did not assume its WMO regional responsibilities until 1990; therefore, TC records are missing from the WMO records in IBTrACS in the North Indian Ocean before 1990. Wind estimates by the Joint Typhoon Warning Center (JTWC), a United States agency responsible for issuing TC reports outside NA and Eastern North Pacific (ENP), supplement much of the missing data in the basin. However, JTWC mainly uses the Dvorak model (Dvorak, 1975, 1984) for wind estimates in earlier years, which compromises estimation quality (Chu & Sampson, 2002).

Most global agencies use 10-minute sustained wind speed to measure TC near-center 10-meter maximum wind speed, following the guidance of the WMO, whereas U.S. agencies, such as the National Hurricane Center (NHC) and JTWC, use 1-minute measurements. WMO (Saha et al., 2010) has suggested a conversion ratio of 0.93 between the 1-minute and 10-minute measurement standards for tropical cyclones at sea, whereas a ratio of 0.88 was traditionally considered. Additionally, IMD uses 3-minute sustained winds. Due to these differences in measurement standards, the definition of a named TC—a tropical system that possesses a maximum wind speed of at least 34 knots near its center—will vary across basins. Since most contemporary TC tracking algorithms are optimized based on global named TCs or the 10-minute standard (Tory et al., 2013; Zarzycki & Ullrich, 2017; Han & Ullrich, 2025) and more TCs (~60%) are named based on the 10-minute standard, objective trackers likely stand at a middle ground, albeit slightly biased toward the 10-minute standard. This could contribute to underestimating TC frequency in the North Atlantic (NA) and ENP in models. Nevertheless, when comparing model results using TC trackers to best track, we recommend using the definition of named TCs in each basin to best match the past and future observations and to fairly evaluate trackers' performance. Below, we compare

four different best track data sets to demonstrate the impact of differently processed IBTrACS data on the performance of the two tracking algorithms applied to ERA5 for the period of 1988–2014.

The following four 6-hourly data sets from 1988-2014 are compared here:

1. The original IBTrACS data set (IB-ORI) used in this study, which has 42964 TC records.
2. IB-WMO that uses wind estimates from WMO agencies and does not include records with missing WMO wind estimates. It has 41619 TC nodes.
3. IB-1MIN that only uses global JTWC's 1-min data. It contains 43574 TC nodes
4. IB-10MIN that converts all wind estimates in IB-RAD to the 10-min standard. It contains 40629 TC nodes.

In each dataset, TC records with maximum wind speeds of at least 34 knots in tracks with at least one instance of maximum wind speed reaching at least 34 knots are selected. See Table 3.S1 for a comparison of TC detection skills when SyCLoPS is applied to ERA5 against the four subjective data sets. The number of TC records (frequency) in each dataset is also shown in Table 3.S1. As a reference, SyCLoPS selects 38,048 TC instances in ERA5. As shown in the table, TC detection skill remains relatively stable across the different IBTrACS datasets. The lowest hit rate minus false alarm rate (HRMFAR_ is obtained using IB-WMO, primarily because of its higher false alarm rate (FAR). We speculate that this elevated FAR is partly attributable to inconsistencies among TC records reported by different WMO agencies.

**Text 3.S2. Justification of the stitching distance for tracking TCs**

The stitching distance is reduced from 8º in the ZU method to 6º for the SyCLoPS algorithm to minimize the chance of the end of one track being stitched to another LPS track. This adjustment is implemented because the designated merging distance at 6º is short compared to a 8º stitching distance. Under this condition, the very low detectable LPS threshold (MSLPCC55 > 10 Pa) can make it much more likely for a weaker LPS to falsely connect to another LPS track rather than terminate itself. TCs almost never travel at a speed greater than 6º per 6 hours (or 111 km/hr)—99.99% of the TC records in

the original IBTrACS dataset have a translation speed below this threshold. Hence, the 6º GCD stitching distance should be considered safe for tracking the TC stage of TC tracks using SyCLoPS.

We also do an experiment using 8º stitching distance with a newly developed "prioritize MSLP" function in TE's SitchNodes to minimize false track connections. This option orders all candidates by their intensity (indicated by MSLPCC55) instead of the earliest candidate to occur in the data set. It prioritizes the most intense LPS candidates across all time steps to connect to a possible future candidate within the prescribed stitching distance, effectively identifying the strongest connections between sequential candidates while jumping around in time. We find that this method removes almost all false connections at the post-TC stage. Although some track differences are observed in the weak pre-TC stage compared with the 6º method, the TC stage of the TC tracks remains unchanged. Therefore, it can serve as an alternative method to detect TC and other LPS tracks at a 6-hourly interval, and we recommend choosing this option when post-TCs and extratropical cyclones are considered as they generally have a higher translation speed. We also found that using the "prioritize MSLPCC55" function slightly improves extratropical transition detection skills as defined in Text 3.S4 at the same stitching distance. In the SyCLoPS software, "prioritize MSLPCC55" is automatically turned on if the input stitching distance is larger than the merging distance.

**Text 3.S3. Adjusting LPS labels in SyCLoPS**

As mentioned in the SyCLoPS paper (Han & Ullrich, 2025), using a hard cutoff threshold to determine a class of LPS can lead to bias when LPSs are in a transitional or ambiguous state. In the experiment, several combinations of UPPTKCC and MSLPCC20 thresholds close to the optimal configuration produce detection skills not statistically different from those of the optimal configuration, indicating the flexibility of an optimal threshold. Additionally, noise and glitches in reanalysis and model data (e.g., data assimilation and topographical noise) may fluctuate the values of the MSLPCC20, UPPTKCC, and LOWTKCC parameters used in the TC criteria. Considering the factors above, we introduce an "Adjusted Label" column in the classified LPS catalog produced by the new version of

SyCLoPS Python software (Han, 2026). The labels in this column are adjusted from the original labels in the following manners:

1. Only TD nodes (including nodes with TD and TD(MD) labels) and TD/TLO nodes in monsoonal system (MS) tracks are under consideration for adjustment.
2. For TCs, if no more than 5 consecutive 6-hourly TD nodes occur within a stable TC-labeled period and do not show evidence of notable short-term weakening, these nodes within the short TC label break period are reclassified as TC nodes. A short break within a stable TC-labeled period is identified when the break is sandwiched by two uninterrupted TC-labeled periods that are longer than the duration of the TC label break. A notable short-term weakening is recognized when labels other than TD are found during the break, or when the MSLP at TD nodes during the break exceeds the MSLP from linear regression between the start and end of the break by at least 5 hPa. 12% of TD nodes in the 1988-2014 ERA5 TC tracks are reclassified under this adjustment.
3. Similarly, TD and TLO nodes in MS tracks are adjusted to “TD(MD)” and “TLO(ML)” during the monsoonal system label (“TD(MD)” and “TLO(ML)”) breaks. 12% of TD or TLO-labeled nodes in the 1988-2014 ERA5 MS tracks are reclassified under this adjustment.

**Text 3.S4. Extratropical transition condition optimization**

In SyCLoPS, the ET condition is redefined to optimize its labeling accuracy compared to the IBTrACS data used in this study. ET detection skills are evaluated using the hit rate minus false alarm rate (HRMFAR). The hit rate is defined as the number of successfully selected ERA5 ET tracks in SyCLoPS divided by the total number of ET tracks successfully detected by SyCLoPS's TC detection algorithm. The false alarm rate is defined as the proportion of selected ERA5 ET tracks that have a match with the observed TC track, but not with any of the observed ET tracks. An ET track in observations is defined as a TC that transform directly from the TC stage (i.e. a tropical system that maintains at least 34 kt wind speed) to the extratropical stage (i.e., labeled as an extratropical cyclone or subtropical storm) in 6 hours without first weakening into a tropical depression (TD). In SyCLoPS, the boundary between TC and TD labels are not defined by LPS’s near-surface wind speed and is subjective to some degrees of

uncertainty. We test various maximum allowable time windows, ranging from 6 to 36 hours after the last TC node and before an LPS is classified as extratropical, to define an ET track. The result shows that the HRMFAR is maximized at 65% (HR: 91%) when the time window is chosen to be 18 hours. Under this condition, the ET fraction in ERA5 is at 40%, comparable to the observed value of 36%. We then define the ET completion time to be at the first timestep during the 18 hours window. After removing a few outliers, the mean and median ET completion time errors compared to IBTrACS are -3 and -6 hours, respectively. And the mean and median absolute errors in ET completion time are 15 and 12 hours, respectively. See the error distribution in Figure 3.S2 for details. We also test this ET condition using the 200 hPa and 300 hPa ERA5 data as the default setting in SyCLoPS and achieve similar results despite a slightly higher false alarm rate.

**Text 3.S5. Implementation of TC size algorithms in TE**

The new function implemented in TempesetExtremes (TE) is the *radial_quadrant_profile* and *firstwhere* operator in NodeFileEditor. NodeFileEditor takes TC node locations and 10 m wind data as input. Next, *radial_quadrant_profile* divides the TC wind field surrounding the TC center into four geographic quadrants and a certain number of bins of a certain bin width (in GCD), both of which are customized by the user. For all models except the 1.0º ECMWF-LR, which uses a 0.5º bin width, the bin width is chosen to be 0.25º in this study. Then, TE computes the radial quadrant wind profile of a TC by averaging wind speed data points within each bin for each quadrant, generating four distinct wind profiles within 13º GCD. The *firstwhere* operator is used to search for the bin index of the first data point along each of the four quadrant wind profiles that drops below the 30 kt (15.4 m/s) cutoff radius wind threshold (30 kt or 15.4 m/s). The search only begins after the bin where the maximum wind speed is found within 3° of the TC centers. The software finds the bin indices and multiplies them by the bin width to convert them to wind radii in GCD in the four quadrants ordered by northeast, southeast, southwest, and northwest. Finally, before calculating the final TC size, the TC size Python calculator subtracts half of the bin width from the input wind radii and converts them into quadrantal wind radii in nautical miles. This Python calculator is available in the GitHub repository mentioned in Chapter 3.10.

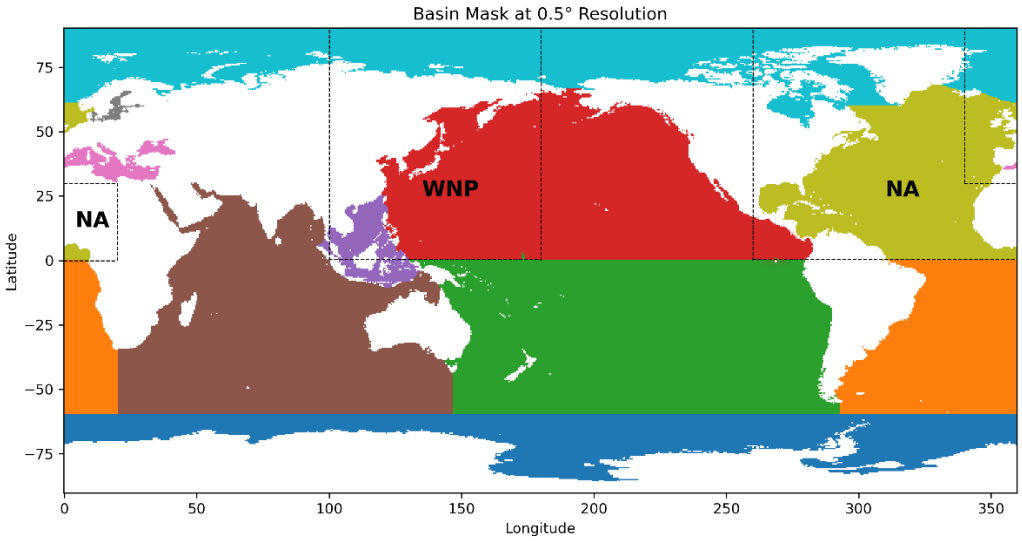



**Figure S3.1**. The boundaries of the NA and WNP basins defined in this study. The shapes (basin masks at 0.5º resolution created by Englander et al. in 2025) with different colors show the major global oceans. TCs detected between 100° and 180° E north of the equator are considered WNP TCs. TCs detected between 100°W and 180°W north of the equator, as well as between 0° and 20°E and 0° and 30°N, are classified as NA TCs, except for those located within the lower-left portion of the defined boundary that overlaps with the northern Pacific polygon (red).

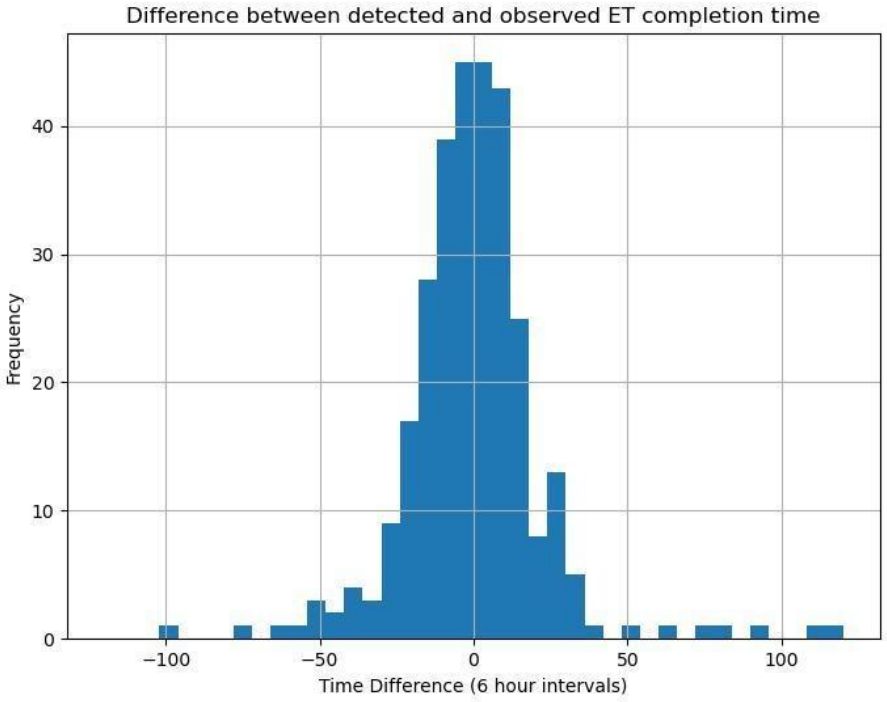



**Figure S3.2**. ET completion timing errors of detected ET tracks compared to IBTrACS records. Five outliers are not displayed in this distribution and are not included when calculating the error statistics.

| Data set | HR | FAR | HRMFAR | Node Adjusted CSI |
|---|---|---|---|---|
| IB-ORI | 83.2% | 12.9% | 70.3% | 72.7% |
| IB-WMO | 83.9% | 14.1% | 69.8% | 71.6% |
| IB-1MIN | 82.2% | 12.0% | 70.3% | 73.5% |
| IB-10MIN | 84.7% | 13.8% | 70.9% | 73.4% |

**Table S3.1.** Comparison of TC detection skills among differently processed IBTrACS data.

### 3.11 Software and Published Datasets

The Python program for calculating modeled TC size is archived on the GitHub repository: yepkids/TC-Size. This program takes near-surface (10-m) U-component and V-component wind fields from model or reanalysis data and output the wind radii at the four quadrants and the final size measured in radius (n. mile). TempestExtremes commands for extracting quadrantal radial wind profiles ran in the

subprocesses of the Python TC size calculator. Please see the GitHub page for more information on implementing the TC size calculator.

Published dataset relevant to Chapter 3:

**1**. *Objectively Detected HighResMIP Tropical Cyclone Tracks and Tropical Cyclone Size (1988–2014) for Model Evaluation.* Zenodo. DOI: https://doi.org/10.5281/zenodo.18178615

**2**. *Objectively Detected and Classified HighResMIP Surface Low-pressure System Tracks (1988-2014).* Zenodo. DOI: https://doi.org/10.5281/zenodo.18226201

## 3.12 References


Avenas, A., Mouche, A., Tandeo, P., Piolle, J.-F., Chavas, D., Fablet, R., Knaff, J., & Chapron, B. (2023). Reexamining the Estimation of Tropical Cyclone Radius of Maximum Wind from Outer Size with an Extensive Synthetic Aperture Radar Dataset. *Monthly Weather Review*, *151*(12), 3169–3189. https://doi.org/10.1175/MWR-D-23-0119.1

Bengtsson, L., Hodges, K. I., & Esch, M. (2007). Tropical cyclones in a T159 resolution global climate model: Comparison with observations and re-analyses. *Tellus A: Dynamic Meteorology and Oceanography*, *59*(4), 396. https://doi.org/10.1111/j.1600-0870.2007.00236.x

Bian, G.-F., Nie, G.-Z., & Qiu, X. (2021). How well is outer tropical cyclone size represented in the ERA5 reanalysis dataset? *Atmospheric Research*, *249*, 105339. https://doi.org/10.1016/j.atmosres.2020.105339

Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., Meurdesoif, Y., & Ghattas, J. (2022a). IPSL IPSL-CM6A-ATM-ICO-HR model output prepared for CMIP6 HighResMIP [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.16361

Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., Meurdesoif, Y., & Ghattas, J. (2022b). IPSL IPSL-CM6A-ATM-ICO-VHR model output prepared for CMIP6 HighResMIP [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.16367

Bourdin, S., Fromang, S., Dulac, W., Cattiaux, J., & Chauvin, F. (2022). Intercomparison of four algorithms for detecting tropical cyclones using ERA5. *Geoscientific Model Development, 15*(17), 6759–6786. https://doi.org/10.5194/gmd-15-6759-2022

Brenowitz, N. D., Ge, T., Subramaniam, A., Manshausen, P., Gupta, A., Hall, D. M., Mardani, M., Vahdat, A., Kashinath, K., & Pritchard, M. S. (2025). Climate in a Bottle: Towards a Generative Foundation Model for the Kilometer-Scale Global Atmosphere (No. arXiv:2505.06474) [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2505.06474

Cangialosi, J. P., & Landsea, C. W. (2016). An Examination of Model and Official National Hurricane Center Tropical Cyclone Size Forecasts. *Weather and Forecasting, 31*(4), 1293–1300. https://doi.org/10.1175/WAF-D-15-0158.1

Carrasco, C. A., Landsea, C. W., & Lin, Y.-L. (2014). The Influence of Tropical Cyclone Size on Its Intensification. *Weather and Forecasting, 29(3),* 582–590. https://doi.org/10.1175/WAF-D-13-00092.1

Chan, J. C. L., & Yip, C. K. M. (2003). Interannual variations of tropical cyclone size over the western North Pacific. *Geophysical Research Letters*, *30*(24), 2003GL018522. https://doi.org/10.1029/2003GL018522

Chan, K. T. F., & Chan, J. C. L. (2012). Size and Strength of Tropical Cyclones as Inferred from QuikSCAT Data. *Monthly Weather Review*, *140*(3), 811–824. https://doi.org/10.1175/MWR-D-10-05062.1

Chan, K. T. F., & Chan, J. C. L. (2014). Impacts of initial vortex size and planetary vorticity on tropical cyclone size. *Quarterly Journal of the Royal Meteorological Society*, *140*(684), 2235–2248. https://doi.org/10.1002/qj.2292

Chan, K. T. F., & Chan, J. C. L. (2015). Global climatology of tropical cyclone size as inferred from QuikSCAT data. *International Journal of Climatology*, *35*(15), 4843–4848. https://doi.org/10.1002/joc.4307

Chavas, D. R., & Emanuel, K. (2014). Equilibrium Tropical Cyclone Size in an Idealized State of Axisymmetric Radiative–Convective Equilibrium*. *Journal of the Atmospheric Sciences*, *71*(5), 1663–1680. https://doi.org/10.1175/JAS-D-13-0155.1

Chavas, D. R., & Lin, N. (2016). A Model for the Complete Radial Structure of the Tropical Cyclone Wind Field. Part II: Wind Field Variability. *Journal of the Atmospheric Sciences*, *73*(8), 3093–3113. https://doi.org/10.1175/JAS-D-15-0185.1

Chavas, D. R., Lin, N., Dong, W., & Lin, Y. (2016). Observed Tropical Cyclone Size Revisited. *Journal of Climate*, *29*(8), 2923–2939. https://doi.org/10.1175/JCLI-D-15-0731.1

Chavas, D. R., & Reed, K. A. (2019). Dynamical Aquaplanet Experiments with Uniform Thermal Forcing: System Dynamics and Implications for Tropical Cyclone Genesis and Size. *Journal of the Atmospheric Sciences*, *76*(8), 2257–2274. https://doi.org/10.1175/JAS-D-19-0001.1

Chen, K., Chen, G., & Shi, D. (2022). Reexamination of the Relationship between Tropical Cyclone Size and Intensity over the Western North Pacific. *Advances in Atmospheric Sciences*, *39*(11), 1956–1968. https://doi.org/10.1007/s00376-022-1450-6

Chu, J.-H., Sampson, C. R., & Fukada, E. (2002). *ATCR report plan*. The Joint Typhoon Warning Center Tropical Cyclone Best-Tracks, 1945-2000. NRL/MR/7540-02-16. https://www.metoc.navy.mil/jtwc/products/best-tracks/tc-bt-report.html

Cocks, S. B., & Gray, W. M. (2002). Variability of the Outer Wind Profiles of Western North Pacific Typhoons: Classifications and Techniques for Analysis and Forecasting. *Monthly Weather Review*, *130*(8), 1989–2005. https://doi.org/10.1175/1520-0493(2002)130%253C1989:VOTOWP%253E2.0.CO;2

Consortium (EC-Earth), E.-E. (2018). EC-Earth-Consortium EC-Earth3P-HR model output prepared for CMIP6 HighResMIP [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.2323

Davis, C. A. (2018). Resolving Tropical Cyclone Intensity in Models. *Geophysical Research Letters*, *45*(4), 2082–2087. https://doi.org/10.1002/2017GL076966

DeMaria, M., Knaff, J. A., Knabb, R., Lauer, C., Sampson, C. R., & DeMaria, R. T. (2009). A New Method for Estimating Tropical Cyclone Wind Speed Probabilities. *Weather and Forecasting*, *24*(6), 1573–1591. https://doi.org/10.1175/2009WAF2222286.1

Demuth, J. L., DeMaria, M., & Knaff, J. A. (2006). Improvement of Advanced Microwave Sounding Unit Tropical Cyclone Intensity and Size Estimation Algorithms. *Journal of Applied Meteorology and Climatology*, *45*(11), 1573–1581. https://doi.org/10.1175/JAM2429.1

Dunnavan, G. M., & Diercks, J. W. (1980). An Analysis of Super Typhoon Tip (October 1979). *Monthly Weather Review*, *108*(11), 1915–1923. https://doi.org/10.1175/1520-0493(1980)108%253C1915:AAOSTT%253E2.0.CO;2

Dvorak, V. F. (1975). Tropical Cyclone Intensity Analysis and Forecasting from Satellite Imagery. Monthly Weather Review, 103(5), 420–430. https://doi.org/10.1175/1520-0493(1975)103%253C0420:TCIAAF%253E2.0.CO

Dvorak, V. F. (1984). Tropical cyclone intensity analysis using satellite data. *NOAA Tech. Rep. NESDIS 11, NOAA*, 47 pp., https://satepsanone.nesdis.noaa.gov/pub/Publications/Tropical/Dvorak_1984.pdf.

Englander, G., Bone, J., Costello, C., McDonald, G., & Raynor, J. (2025). Reproducibility package for Global Expansion of Marine Protected Areas and the Redistribution of Fishing Effort [Dataset]. *World Bank*. https://doi.org/10.60572/G1K8-FX28

European Centre for Medium-Range Weather Forecasts. (2019). ERA5 Reanalysis (0.25 Degree Latitude-Longitude Grid) [Dataset]. *NSF National Center for Atmospheric Research*. https://doi.org/10.5065/BH6N-5N20

Evans, J. L., & Braun, A. (2012). A Climatology of Subtropical Cyclones in the South Atlantic. *Journal of Climate*, *25*(21), 7328–7340. https://doi.org/10.1175/JCLI-D-11-00212.1

Fudeyasu, H., & Wang, Y. (2011). Balanced Contribution to the Intensification of a Tropical Cyclone Simulated in TCM4: Outer-Core Spinup Process*. *Journal of the Atmospheric Sciences*, *68*(3), 430–449. https://doi.org/10.1175/2010JAS3523.1

Gahtan, J., K. R. Knapp, C. J. Schreck, H. J. Diamond, J. P. Kossin, M. C. Kruk, 2024: International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4r01. [Dataset]. NOAA National Centers for Environmental Information. https://doi:10.25921/82ty-9e16 [Accessed on 2025-10-03].

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L.B., et al. (2017). *The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)*. https://doi.org/10.1175/JCLI-D-16-0758.1

Gopalakrishnan, S. G., Marks, F., Zhang, J. A., Zhang, X., Bao, J.-W., & Tallapragada, V. (2013). A Study of the Impacts of Vertical Diffusion on the Structure and Intensity of the Tropical Cyclones Using the High-Resolution HWRF System. *Journal of the Atmospheric Sciences*, *70*(2), 524–541. https://doi.org/10.1175/JAS-D-11-0340.1

Guishard, M. P., Evans, J. L., & Hart, R. E. (2009). Atlantic Subtropical Storms. Part II: Climatology. *Journal of Climate*, *22*(13), 3574–3594. https://doi.org/10.1175/2008JCLI2346.1

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., et al. (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, *9*(11), 4185–4208. https://doi.org/10.5194/gmd-9-4185-2016

Hagen, A. B. (2025). The 2024 Atlantic Hurricane Season: Helene and Milton Highlight Five U.S. Hurricane Landfalls. *Weatherwise*, *78*(4), 19–28. https://doi.org/10.1080/00431672.2025.2502305

Han, Y. (2026a). Objectively Detected and Classified HighResMIP Surface Low-pressure System Tracks (1988-2014) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.18226202

Han, Y. (2026b). yepkids/SyCLoPS: SyCLoPS v1.1 [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.18287269

Han, Y. (2026c). yepkids/TC-Size: Initial Release (Version 1.0.0) [Software]. *Zenodo*. https://doi.org/10.5281/ZENODO.18287288

Han, Y., & Ullrich, P. (2026). Objectively Detected HighResMIP Tropical Cyclone Tracks and Tropical Cyclone Size (1988–2014) for Model Evaluation [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.18178616

Han, Y., & Ullrich, P. A. (2025). The System for Classification of Low-Pressure Systems (SyCLoPS): An All-In-One Objective Framework for Large-Scale Data Sets. *Journal of Geophysical Research: Atmospheres*, *130*(1), e2024JD041287. https://doi.org/10.1029/2024JD041287

Harper, B. A., Kepert, J. D., & Ginger, J. D. (2010). Guidelines for converting between various wind averaging periods in tropical cyclone conditions. *World Meteorological Organization,* WMO/TDNo. 1555. https://library.wmo.int/viewer/48652

Hart, R. E. (2003). A Cyclone Phase Space Derived from Thermal Wind and Thermal Asymmetry. *Monthly Weather Review*, *131*(4), 585–616. https://doi.org/10.1175/1520-0493(2003)131%253C0585:ACPSDF%253E2.0.CO;2

Hawkins, H. F., & Imbembo, S. M. (1976). The Structure of a Small, Intense Hurricane—Inez 1966. *Monthly Weather Review*, *104*(4), 418–442. https://doi.org/10.1175/1520-0493(1976)104%253C0418:TSOASI%253E2.0.CO;2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hill, K. A., & Lackmann, G. M. (2009). Influence of Environmental Humidity on Tropical Cyclone Size. *Monthly Weather Review*, *137*(10), 3294–3315. https://doi.org/10.1175/2009MWR2679.1

Hodges, K., Cobb, A., & Vidale, P. L. (2017). How Well Are Tropical Cyclones Represented in Reanalysis Datasets? *Journal of Climate*, *30*(14), 5243–5264. https://doi.org/10.1175/JCLI-D-16-0557.1

Irish, J. L., Resio, D. T., & Ratcliff, J. J. (2008). The Influence of Storm Size on Hurricane Surge. *Journal of Physical Oceanography*, *38*(9), 2003–2013. https://doi.org/10.1175/2008JPO3727.1

Judt, F., Klocke, D., Rios-Berrios, R., Vanniere, B., Ziemen, F., Auger, L., et al. (2021). Tropical Cyclones in Global Storm-Resolving Models. *Journal of the Meteorological Society of Japan. Ser. II*, *99*(3), 579–602. https://doi.org/10.2151/jmsj.2021-029

Kennedy, J., Titchner, H., Rayner, N., & Roberts, M. (2017). input4MIPs.MOHC.SSTsAndSeaIce.HighResMIP.MOHC-HadISST-2-2-0-0-0 [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/input4MIPs.1221

Kilroy, G., Smith, R. K., & Montgomery, M. T. (2017). A unified view of tropical cyclogenesis and intensification. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 450–462. https://doi.org/10.1002/qj.2934

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data. *Bulletin of the American Meteorological Society*, *91*(3), 363–376. https://doi.org/10.1175/2009BAMS2755.1

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., et al. (2024). Neural general circulation models for weather and climate. *Nature*, *632*(8027), 1060–1066. https://doi.org/10.1038/s41586-024-07744-y

Kosaka, Y., Kobayashi, S., Harada, Y., Kobayashi, C., Naoe, H., Yoshimoto, K., et al. (2024). The JRA-3Q Reanalysis. *Journal of the Meteorological Society of Japan. Ser. II*, *102*(1), 49–109. https://doi.org/10.2151/jmsj.2024-004

Kreussler, P., Caron, L.-P., Wild, S., Loosveldt Tomas, S., Chauvin, F., Moine, M.-P., Roberts, M. J., Ruprich-Robert, Y., Seddon, J., Valcke, S., Vannière, B., & Vidale, P. L. (2021). Tropical Cyclone Integrated Kinetic Energy in an Ensemble of HighResMIP Simulations. *Geophysical Research Letters*, *48*(5), e2020GL090963. https://doi.org/10.1029/2020GL090963

Lee, C.-S., Cheung, K. K. W., Fang, W.-T., & Elsberry, R. L. (2010). Initial Maintenance of Tropical Cyclone Size in the Western North Pacific. *Monthly Weather Review*, *138*(8), 3207–3223. https://doi.org/10.1175/2010MWR3023.1

Lin, Y., Zhao, M., & Zhang, M. (2015). Tropical cyclone rainfall area controlled by relative sea surface temperature. *Nature Communications*, *6*(1), 6591. https://doi.org/10.1038/ncomms7591

Liu, K. S., & Chan, J. C. L. (2002). Synoptic Flow Patterns Associated with Small and Large Tropical Cyclones over the Western North Pacific. *Monthly Weather Review*, *130*(8), 2134–2142. https://doi.org/10.1175/1520-0493(2002)130%253C2134:SFPAWS%253E2.0.CO;2

Lopez-Gomez, I., Wan, Z. Y., Zepeda-Núñez, L., Schneider, T., Anderson, J., & Sha, F. (2025). Dynamical-generative downscaling of climate model ensembles. *Proceedings of the National Academy of Sciences*, *122*(17), e2420288122. https://doi.org/10.1073/pnas.2420288122

Lu, K.-Y., & Chavas, D. R. (2022). Tropical Cyclone Size Is Strongly Limited by the Rhines Scale: Experiments with a Barotropic Model. *Journal of the Atmospheric Sciences*, *79*(8), 2109–2124. https://doi.org/10.1175/JAS-D-21-0224.1

Manganello, J. V., Hodges, K. I., Kinter, J. L., Cash, B. A., Marx, L., Jung, T., et al. (2012). Tropical Cyclone Climatology in a 10-km Global Atmospheric GCM: Toward Weather-Resolving Climate Modeling. *Journal of Climate*, *25*(11), 3867–3893. https://doi.org/10.1175/JCLI-D-11-00346.1

Martinez, J., Nam, C. C., & Bell, M. M. (2020). On the Contributions of Incipient Vortex Circulation and Environmental Moisture to Tropical Cyclone Expansion. *Journal of Geophysical Research: Atmospheres*, *125*(21), e2020JD033324. https://doi.org/10.1029/2020JD033324

Matyas, C. J. (2010). Associations between the size of hurricane rain fields at landfall and their surrounding environments. *Meteorology and Atmospheric Physics*, *106*(3–4), 135–148. https://doi.org/10.1007/s00703-009-0056-1

Mizuta, R., Yoshimura, H., Ose, T., Hosaka, M., & Yukimoto, S. (2019a). MRI MRI-AGCM3-2-H model output prepared for CMIP6 HighResMIP [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.10942

Mizuta, R., Yoshimura, H., Ose, T., Hosaka, M., & Yukimoto, S. (2019b). MRI MRI-AGCM3-2-S model output prepared for CMIP6 HighResMIP [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.1625

Moon, Y., Kim, D., Camargo, S. J., Wing, A. A., Sobel, A. H., Murakami, H., Reed, K. A., Scoccimarro, E., Vecchi, G. A., Wehner, M. F., Zarzycki, C. M., & Zhao, M. (2020). Azimuthally Averaged Wind and Thermodynamic Structures of Tropical Cyclones in Global Climate Models and Their Sensitivity to Horizontal Resolution. *Journal of Climate*, *33*(4), 1575–1595. https://doi.org/10.1175/JCLI-D-19-0172.1

Musgrave, K. D., Taft, R. K., Vigh, J. L., McNoldy, B. D., & Schubert, W. H. (2012). Time evolution of the intensity and size of tropical cyclones. *Journal of Advances in Modeling Earth Systems*, *4*(3). https://doi.org/10.1029/2011MS000104

Nardi, K. M., Zarzycki, C. M., Larson, V. E., & Bryan, G. H. (2022). Assessing the Sensitivity of the Tropical Cyclone Boundary Layer to the Parameterization of Momentum Flux in the Community Earth System Model. *Monthly Weather Review*, *150*(4), 883–906. https://doi.org/10.1175/MWR-D-21-0186.1

Needham, H. F., Keim, B. D., & Sathiaraj, D. (2015). A review of tropical cyclone-generated storm surges: Global data sources, observations, and impacts. *Reviews of Geophysics*, *53*(2), 545–591. https://doi.org/10.1002/2014RG000477

Nguyen, D., & Schenkel, B. A. (2025). Characteristics of Tropical Cyclone Outer Size and Structure Associated With Extratropical Transition. *Geophysical Research Letters*, *52*(5), e2024GL111703. https://doi.org/10.1029/2024GL111703

Olander, T. L., & Velden, C. S. (2007). The Advanced Dvorak Technique: Continued Development of an Objective Scheme to Estimate Tropical Cyclone Intensity Using Geostationary Infrared Satellite Imagery. *Weather and Forecasting*, *22*(2), 287–298. https://doi.org/10.1175/waf975.1

Olander, T. L., & Velden, C. S. (2019). The Advanced Dvorak Technique (ADT) for Estimating Tropical Cyclone Intensity: Update and New Capabilities. *Weather and Forecasting*, *34*(4), 905–922. https://doi.org/10.1175/WAF-D-19-0007.1

Ooyama, K. (1969). Numerical Simulation of the Life Cycle of Tropical Cyclones. *Journal of the Atmospheric Sciences*, *26*(1), 3–40. https://doi.org/10.1175/1520-0469(1969)026%253C0003:NSOTLC%253E2.0.CO;2

Peduzzi, P., Chatenoux, B., Dao, H., De Bono, A., Herold, C., Kossin, J., Mouton, F., & Nordbeck, O. (2012). Global trends in tropical cyclone risk. *Nature Climate Change*, *2*(4), 289–294. https://doi.org/10.1038/nclimate1410

Powell, M. D., & Reinhold, T. A. (2007). Tropical Cyclone Destructive Potential by Integrated Kinetic Energy. *Bulletin of the American Meteorological Society*, *88*(4), 513–526. https://doi.org/10.1175/BAMS-88-4-513

Quiring, S., Schumacher, A., Labosier, C., & Zhu, L. (2011). Variations in mean annual tropical cyclone size in the Atlantic. *Journal of Geophysical Research: Atmospheres*, *116*(D9). https://doi.org/10.1029/2010JD015011

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., & Keeley, S. (2017a). ECMWF ECMWF-IFS-HR model output prepared for CMIP6 HighResMIP [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.2461

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., & Keeley, S. (2017b). ECMWF ECMWF-IFS-LR model output prepared for CMIP6 HighResMIP [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.2463

Roberts, M. (2017). MOHC HadGEM3-GC31-HM model output prepared for CMIP6 HighResMIP [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.446

Roberts, M. J., Reed, K. A., Bao, Q., Barsugli, J. J., Camargo, S. J., Caron, L.-P., Chang, P., et al. (2025). High-Resolution Model Intercomparison Project phase 2 (HighResMIP2) towards CMIP7. *Geoscientific Model Development*, *18*(4), 1307–1332. https://doi.org/10.5194/gmd-18-1307-2025

Roberts, M. J., Vidale, P. L., Mizielinski, M. S., Demory, M.-E., Schiemann, R., Strachan, J., Hodges, K., Bell, R., & Camp, J. (2015). Tropical Cyclones in the UPSCALE Ensemble of High-Resolution Global Climate Models*. *Journal of Climate*, *28*(2), 574–596. https://doi.org/10.1175/JCLI-D-14-00131.1

Ruan, Z., & Wu, Q. (2022). Relationship Between Size and Intensity in North Atlantic Tropical Cyclones With Steady Radii of Maximum Wind. *Geophysical Research Letters*, *49*(3), e2021GL095632. https://doi.org/10.1029/2021GL095632

Sainsbury, E. M., Schiemann, R. K. H., Hodges, K. I., Shaffrey, L. C., Baker, A. J., & Bhatia, K. T. (2020). How Important Are Post-Tropical Cyclones for European Windstorm Risk? *Geophysical Research Letters*, *47*(18), e2020GL089853. https://doi.org/10.1029/2020GL089853

Sampson, C. R., Fukada, E. M., Knaff, J. A., Strahl, B. R., Brennan, M. J., & Marchok, T. (2017). Tropical Cyclone Gale Wind Radii Estimates for the Western North Pacific. *Weather and Forecasting*, *32*(3), 1029–1040. https://doi.org/10.1175/WAF-D-16-0196.1

Schenkel, B. A., Lin, N., Chavas, D., Oppenheimer, M., & Brammer, A. (2017). Evaluating Outer Tropical Cyclone Size in Reanalysis Datasets Using QuikSCAT Data. *Journal of Climate*, *30*(21), 8745–8762. https://doi.org/10.1175/JCLI-D-17-0122.1

Schenkel, B. A., Lin, N., Chavas, D., Vecchi, G. A., Oppenheimer, M., & Brammer, A. (2018). Lifetime Evolution of Outer Tropical Cyclone Size and Structure as Diagnosed from Reanalysis and Climate Model Data. *Journal of Climate*, *31*(19), 7985–8004. https://doi.org/10.1175/JCLI-D-17-0630.1

Schreck, C. J., Knapp, K. R., & Kossin, J. P. (2014). The Impact of Best Track Discrepancies on Global Tropical Cyclone Climatologies using IBTrACS. *Monthly Weather Review*, *142*(10), 3881–3899. https://doi.org/10.1175/MWR-D-14-00021.1

Smith, R. K., Schmidt, C. W., & Montgomery, M. T. (2011). An investigation of rotational influences on tropical-cyclone size and intensity. *Quarterly Journal of the Royal Meteorological Society*, *137*(660), 1841–1855. https://doi.org/10.1002/qj.862

Song, J., Duan, Y., & Klotzbach, P. J. (2020). Revisiting the Relationship Between Tropical Cyclone Size and Intensity Over the Western North Pacific. *Geophysical Research Letters*, *47*(13), e2020GL088217. https://doi.org/10.1029/2020GL088217

Stansfield, A. M., Reed, K. A., & Zarzycki, C. M. (2020). Changes in Precipitation From North Atlantic Tropical Cyclones Under RCP Scenarios in the Variable-Resolution Community Atmosphere Model. *Geophysical Research Letters*, *47*(12), e2019GL086930. https://doi.org/10.1029/2019GL086930

Teng, H.-F., Lee, C.-S., Hsu, H.-H., Done, J. M., & Holland, G. J. (2019). Tropical Cloud Cluster Environments and Their Importance for Tropical Cyclone Formation. *Journal of Climate*, *32*(13), 4069–4088. https://doi.org/10.1175/JCLI-D-18-0679.1

Tory, K. J., Chand, S. S., Dare, R. A., & McBride, J. L. (2013). The Development and Assessment of a Model-, Grid-, and Basin-Independent Tropical Cyclone Detection Scheme. *Journal of Climate*, *26*(15), 5493–5507. https://doi.org/10.1175/JCLI-D-12-00510.1

Ullrich, P. A., & Zarzycki, C. M. (2017). TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, *10*(3), 1069–1090. https://doi.org/10.5194/gmd-10-1069-2017

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., & Reed, K. A. (2021). TempestExtremes v2.1: A community framework for feature detection, tracking, and analysis in large datasets. *Geoscientific Model Development*, *14*(8), 5023–5048. https://doi.org/10.5194/gmd-14-5023-2021

Ullrich, P. A., Chen, H., Stachowicz, K., Zarzycki, C. M., Han, Y., Huang, K., & Hoyer, S. (2026). ClimateGlobalChange/tempestextremes: V2.4.2 [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.18274602

Voldoire, A. (2019). CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 HighResMIP [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.1387

Wang, Y. (2009). How Do Outer Spiral Rainbands Affect Tropical Cyclone Structure and Intensity?*. *Journal of the Atmospheric Sciences*, *66*(5), 1250–1273. https://doi.org/10.1175/2008JAS2737.1

Wu, L., Tian, W., Liu, Q., Cao, J., & Knaff, J. A. (2015). Implications of the Observed Relationship between Tropical Cyclone Size and Intensity over the Western North Pacific. *Journal of Climate*, *28*(24), 9501–9506. https://doi.org/10.1175/JCLI-D-15-0628.1

Zarzycki, C. M., & Ullrich, P. A. (2017). Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophysical Research Letters*, *44*(2), 1141–1149. https://doi.org/10.1002/2016GL071606

Zarzycki, C. M., Ullrich, P. A., & Reed, K. A. (2021). Metrics for evaluating tropical cyclones in climate data. *Journal of Applied Meteorology and Climatology*. 60(*5*), 643–660. https://doi.org/10.1175/JAMC-D-20-0149.1

# Chapter 4. Improvements in Tropical Cyclone Representation in JRA-3Q over ERA5

## 4.1 Abstract

Reanalysis products are widely used for long-term tropical cyclone (TC) studies, yet the fidelity of their TC representation remains uncertain. We evaluate TC representation in the recently released JRA-3Q reanalysis against the higher-resolution ERA5 using the SyCLoPS tracker. JRA-3Q achieves substantially higher detection skill, with a hit rate of 90% and a false alarm rate of 6% during 1979–2024. Unlike ERA5, which exhibits markedly degraded detection skill during the pre-satellite era, JRA-3Q maintains stable performance over time. Analyses of false alarms and missed TCs in the North Indian basin, TC vortex evolution following TC bogussing, and the spatial variance of column-integrated moist static energy suggest that JRA-3Q's improvements arises from both reduced model biases and a more effective, consistently applied TC bogus scheme. TC representation errors are qualitatively decomposed into contributions from the tracking algorithm, the reanalysis, and uncertainties in IBTrACS, highlighting the relative strength of both reanalyses.

## 4.2 Introduction

Tropical cyclones (TCs) are a central focus of climate and Earth system modeling studies because they can generate catastrophic weather events worldwide. Recent advances in reanalysis have enabled a wide range of climatological and physical studies of TCs in continuous space and time. Given the uncertainties in the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al., 2010), particularly during the earlier decades due to limitations in observing systems and operational practices (Kossin et al., 2013; Landsea & Franklin, 2013; Vecchi & Knutson, 2011), previous studies have highlighted the value of objectively identifying TCs in reanalyses to augment confidence in best-track datasets by uncovering undetected historical TCs (Hodges et al., 2017; Truchelut et al., 2013).

The fifth-generation European Centre for Medium-Range Weather Forecasts reanalysis (ERA5; Hersbach et al., 2020; Soci et al., 2024) is often preferred for TC studies because it provides relatively high horizontal and vertical resolution and a comprehensive set of physical variables over an extended period (e.g., Bhatia et al., 2022; Slocum et al., 2022; Zhang & Chan, 2023). For similar reasons, ERA5 is also widely used to train modern AI weather and climate models that have shown value for TC forecasts and outlooks across multiple time scales (e.g., Alet et al., 2025; Kochkov et al., 2024; Lopez-Gomez et al., 2025; Watt-Meyer et al., 2025). Studies have shown that, compared with previous generations of reanalyses, ERA5 reproduces more TC and monsoonal low-pressure system (LPS) tracks than other reanalyses (Vishnu et al., 2020; Zarzycki et al., 2021) and better represents TC size and structure (Bian et al., 2021; Carstens et al., 2024; Malakar et al., 2020). However, ERA5 biases are also well documented, particularly its underestimation of TC intensity (Dulac et al., 2024; Xu et al., 2024) and biases in environmental fields (Slocum et al., 2022). Its TC intensity underestimation can be larger than that of lower-resolution reanalyses, such as the Climate Forecast System Reanalysis (CFSR) and Japanese 55-year Reanalysis (JRA-55; Kobayashi et al., 2015), which incorporate vortex reconstruction or TC bogus data (i.e., pseudo-observations from best-track datasets) in the assimilation cycle (Zarzycki et al., 2021; Carstens et al., 2024). Dirkes et al. (2023) further showed that ERA5 does not necessarily improve TC moist static energy (MSE) variance budget profiles relative to other reanalyses. Previous studies have also pointed to ERA5's overestimated TC frequency in the Indian Ocean and a low bias in the North Atlantic (Bourdin et al., 2022; Han & Ullrich, 2026; M. J. Roberts et al., 2020), though it is unclear if the error can be fully attributed to reanalysis.

The recently introduced Japanese Reanalysis for Three Quarters of a Century (JRA-3Q; Kosaka et al., 2024) provides new opportunities for long-term TC analysis at a higher horizontal resolution than its predecessor, JRA-55 (0.375° versus 0.563°), although it has coarser spatial grid spacing than ERA5 (0.25°). JRA-55 performs competitively in representing TCs (Hodges et al., 2017; Murakami, 2014; M. J. Roberts et al., 2020; Zarzycki et al., 2021) but suffers from an artificial long-term decline in TC intensity caused by assimilating wind-profile retrievals around TCs (Kobayashi et al., 2015). JRA-3Q addresses

this bias with a new TC bogus scheme based on IBTrACS, yielding a higher and more stable TC detection rate (i.e., hit rate) throughout the reanalysis period (Kosaka et al., 2024). ERA5 also assimilates TC bogus during its back-extension period (1940–1978), but with a substantially different implementation.

In this study, we show that JRA-3Q not only represents TC intensity more accurately than ERA5 and achieves a substantially higher HR, likely owing to its TC bogus scheme, but also markedly reduces the false alarm rate (FAR) across all basins—an improvement that cannot be explained by TC bogussing alone. Using the physics-based, machine learning-aided LPS tracker SyCLoPS (Han & Ullrich, 2025), JRA-3Q attains a hit rate (HR) of 90% and a FAR of 6% during the satellite era (1979–2024), representing, to our knowledge, the best TC detection skill scores reported for a global reanalysis. The corresponding values for ERA5 are 78% and 11%. ERA5 also performs substantially worse than JRA-3Q during the pre-satellite era when both reanalyses employ TC bogus observations. These results motivate a detailed comparison of TC representation in the two reanalyses in Chapter 4.4, quantified by the HR and FAR, and an investigation of the factors contributing to JRA-3Q's improved performance. In Chapter 4.5, we further decompose the sources of error contributing to TC representation biases.

### 4.3 Data and Methods

This study uses three reanalysis products: ERA5 from 1940 to 2024 at 0.25º × 0.25º resolution; JRA-3Q from 1948 to 2024 at 0.375º × 0.375º resolution; and JRA-55 from 1958 to 2024 at 0.5º × 0.625º resolution. We focus on ERA5 and JRA-3Q reanalysis for this study, so the common analysis period is 1948–2024, when the two reanalyses overlap. The six major TC basins—North Atlantic (NA), eastern Pacific (EP), western Pacific (WP), North Indian (NI), South Indian (SI), and South Pacific (SP)—are analyzed separately for each reanalysis. Basin boundaries are shown in Supplements (Chapter 4.6) Figure 4.S1. TC observations are obtained from the IBTrACS. We retain only "main" tracks and remove tracks that never reach a 10-m wind speed of 34 kt. To minimize interagency heterogeneities and maximize

coverage across eras and basins, track filtering uses only wind speeds from U.S. sources. Shifting data sources yields similar results but slightly higher FAR, especially in the pre-satellite era (1948-1978).

The System for Classification of Low-pressure Systems (SyCLoPS; Han & Ullrich, 2025) is the primary TC tracker used for all three reanalyses at 6-hourly intervals because of its superior detection skill and comprehensive LPS classification ability across system types and intensities. In this study, this framework facilitates diagnosis of the LPS characteristics associated with missed TC tracks and TC precursors in reanalyses. The conventional Zarzycki and Ullrich (ZU; Zarzycki & Ullrich, 2017) tracker is included for comparison. Both trackers are built on TempestExtremes (Ullrich et al., 2021), with the ZU algorithm modified by removing its surface elevation constraint to ensure consistency with SyCLoPS. Details of these two tracking algorithms can be found in the appendixes of Han & Ullrich (2026).

The HR and miss rate (MR) are defined as the proportion of IB tracks detected and missed by the tracking algorithm, respectively (i.e., hits and misses). A hit occurs when a detected TC track contains a node within 2.0° great-circle distance of an IBTrACS record at the same timestamp. FAR is defined as the proportion of tracks identified by the tracking algorithm that are not found in IB; that is, tracks with no detected TC nodes within 2.0° of any IBTrACS record. TC detection skill is summarized by HR minus FAR (HRMFAR).

Previous studies using different trackers have shown notably lower HRs in the NA basin when evaluated against IBTrACS observations (Bourdin et al., 2022; Han & Ullrich, 2026; Ye et al., 2026). Han & Ullrich (2026) found that many of these missing tracks are detected as subtropical storms in SyCLoPS. To address this limitation, we update SyCLoPS to recover as many missing tracks as possible while avoiding introducing FAs. Specifically, we do so by identifying a class of "hybrid TCs" that fail SyCLoPS's tropical 100-hPa humidity criterion but satisfy the remaining TC requirements—primarily the upper-level warm-core and mean sea-level pressure (MSLP) gradient criteria. These systems occur predominantly in the NA and are counted as detected TC tracks in the NA for ERA5 and JRA-3Q. Details on hybrid TC detection are provided in Text 4.S1 and Figure 4.S2.

The column-integrated Moist static energy (MSE; $\hat{h}$) analyzed in this study is calculated as:

$$\hat{h} = \int_{p_t}^{p_b} (C_p T + gZ + L_v q)\, dp \tag{4.1}$$

where $p_b$ and $p_t$is the bottom and top pressure levels, here chosen as 925 and 1 hPa, as suggested by Wing et al. (2019). $q$ is the water vapor mixing ratio, and the other terms have their conventional definitions. The spatial variance of $\hat{h}$ is defined as the square of $\hat{h}'$. The prime indicates the deviation from a 10º box centered at TC locations.

### 4.4 Results

The first row of Figure 4.1 compares TC intensity distributions and wind–pressure relationships among the three reanalyses for TCs since 2000. JRA-3Q most realistically reproduces observed TC intensity, likely owing to its more aggressive TC bogus scheme. Despite its finer horizontal and vertical resolution, ERA5 only marginally improves upon JRA-55 in the lifetime minimum MSLP distribution, and this advantage largely disappears in the wind–pressure relationship because ERA5 more severely underestimates near-surface wind speeds. All three reanalyses reproduce the observed peak near 990–1000 hPa but fail to capture the low-MSLP tail associated with the strongest TCs.

The remaining rows of Figure 4.1 summarize the nine-year moving average time series of detection skill (HRMFAR) of SyCLoPS and ZU in each reanalysis. Overall, SyCLoPS demonstrates superior TC detection skill to ZU across most basins in the three reanalyses, primarily through reducing FAs, especially subtropical/extratropical cyclones and weak tropical systems, as indicated in the SP and EP, respectively. SyCLoPS also removes dozens of FAs in the South Atlantic basin (not shown). In the NA, SyCLoPS further improves both HR and FAR by including hybrid TC tracks. The benefit is substantial for JRA-3Q, modest for ERA5, and absent for JRA-55, where the additional FAs outweigh the recovered hits. This highlights the inherent difficulty of objectively identifying hybrid TCs, whose characteristics lie near the classification boundaries of both operational agencies and objective tracking algorithms.

As shown in Figure 4.1, JRA-3Q consistently outperforms ERA5 and JRA-55 across all basins under both tracking algorithms through a combination of higher HR and lower FAR. In the well-observed NA, EP, and WP basins, JRA-3Q maintains a consistently low FAR, with HRMFAR approaching or exceeding 90%. The temporal evolution of detection skill further distinguishes the reanalyses: In the Northern Hemisphere, JRA-3Q remains largely stable throughout the record, except for a modest increase in FAR over the EP before 1970. In contrast, ERA5 exhibits a pronounced improvement from the pre-satellite (1948–1978) to the satellite era (1979–2024), reflecting both higher HR and lower FAR. This suggests that the TC bogussing scheme used in ERA5 is less effective than that of JRA-3Q in compensating for sparse conventional observations, resulting in degraded TC representation. The unphysical upward trend in HRMFAR extends into the satellite era, likely due to assimilation of more comprehensive satellite observations. Similar behavior is also documented in the older generation ERA40 reanalysis (Murakami, 2014). JRA-55 shows the opposite behavior, with relatively high initial detection skill that gradually deteriorates over time, which is consistent with the downward trend of TC detection rate reported in Kosaka et al. (2024). In the Southern Hemisphere, all three reanalyses exhibit substantially lower detection skill during the pre-satellite era, suggesting that much of the degradation arises from limitations in historical best-track observations rather than reanalysis biases.

The NI basin exhibits the largest advantage in FAR for JRA-3Q relative to ERA5, yet its peak HRMFAR remains lower than in the other basins. Unlike other basins, the NI contains numerous monsoonal systems that can intensify near coastlines even inland, while hybrid cyclones are rare because of its low latitude. These characteristics make the NI an ideal and interesting basin for investigating differences in TC representation over tropical land and ocean.

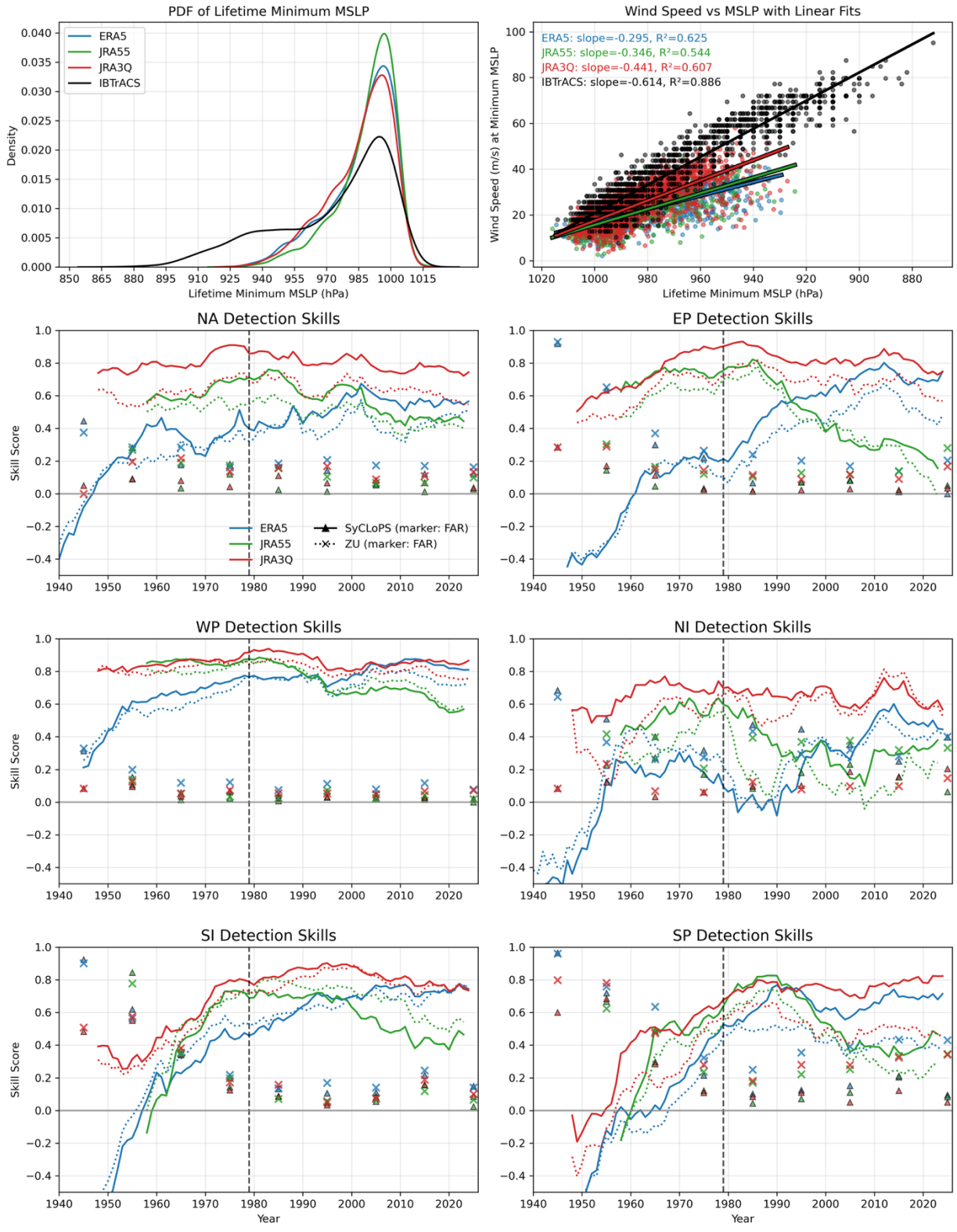



**Figure 4.1.** First row: TC intensity representation in ERA5 (blue), JRA-55 (green), and JRA-3Q (red).

Second to fourth rows: Nine-year moving average time series of TC detection skill (HRMFAR; solid lines

for SyCLoPS and dotted lines for ZU) and decadal-mean FAR (markers). The vertical dotted line separates the pre-satellite and satellite era.

Figure 4.2 shows that, even during the satellite era, ERA5 produces substantially more unique FAs than JRA-3Q, with most occurring over the Indian subcontinent. Although JRA-3Q also produces inland FAs, they are less concentrated west of 80°E and over the western Ganges Delta, where monsoon depressions frequently propagate (Vishnu et al., 2020). Inland ERA5 FAs span a wide intensity range (980–1005 hPa) and exhibit a mature cyclone structure, with strong asymmetric ascent and elevated near-surface equivalent potential temperature indicative of a warm, moist boundary layer near the center (Fig. 4.2c). Their asymmetric and westward-tilting circulation resembles that described by Hunt et al., (2016), likely reflecting the influence of the nearby Himalayas that modifies the circulation and suppresses convection on the eastern flank. In contrast, the corresponding systems in JRA-3Q are substantially weaker and less organized (Fig. 4.2d). Unique FAs in JRA-3Q are generally weaker but more symmetric (Fig. 4.2e), although part of this difference may reflect its coarser vertical resolution. The composite of the corresponding systems in ERA5 yields a less organized structure but similarly weak wind speeds (Fig. 4.2f). These findings suggest that the underlying forecast model in ERA5 is more prone to generating overly strong circulations in this region compared to JRA-3Q, despite assimilating broadly similar observations.

IBTrACS tracks uniquely reproduced in ERA5 are considerably fewer than those uniquely reproduced in JRA-3Q, and the two reanalyses capture distinct subsets of tracks (Fig. 4.2g). ERA5 unique hits are dominated by northwestward-moving monsoon depressions from the northern Bay of Bengal, whereas JRA-3Q unique hits originate farther south, reach higher intensities, and include substantially more Arabian Sea TCs. Their vertical structures further differ: ERA5 unique hits closely resemble its FAs but with slightly stronger circulation (Fig. 4.2i), while the corresponding systems in JRA-3Q display slightly weaker wind speeds and a similar asymmetric structure (Fig. 4.2j). In contrast, JRA-3Q unique hits exhibit a mature, nearly symmetric vortex with near-surface winds exceeding 18 m $s^{-1}$, whereas the same systems appear much weaker and less organized in ERA5 (Figs. 4.2k–l).

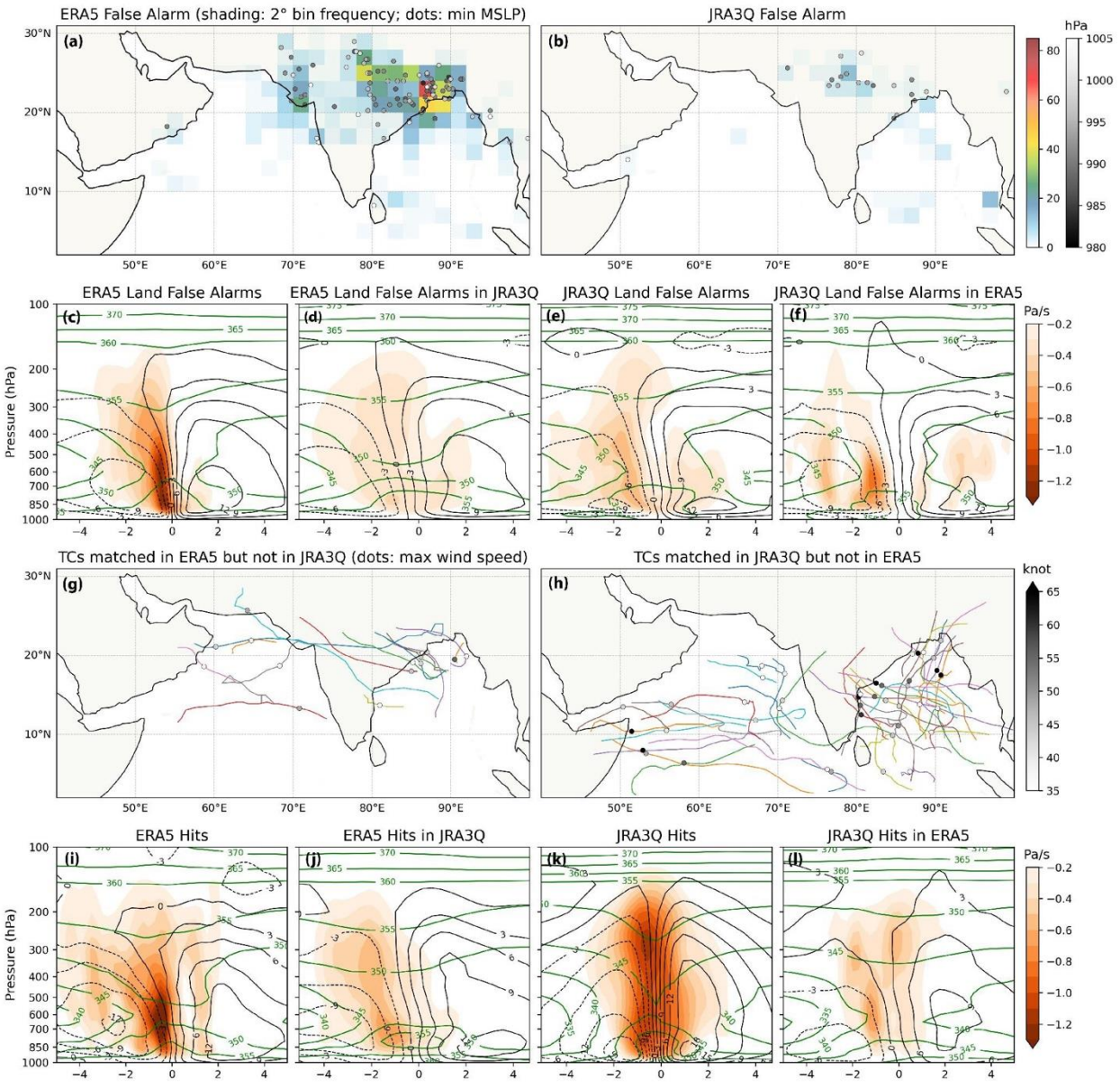



**Figure 4.2.** NI unique FAs during the satellite era in (a) ERA5 and (b) JRA-3Q (2° × 2° shading). Dots mark the locations of peak intensity for inland FA tracks, colored by minimum sea-level pressure (MSLP; grayscale). The second row shows 10° west–east vertical composites centered on these occurrences, with shading indicating vertical velocity, black contours the meridional wind (m $s^{-1}$), and green contours equivalent potential temperature (K). Panels (g) and (h) show unique hits, with dots marking peak

intensity based on 10-m wind speed (grayscale). The final row shows the vertical composites centered on the unique hit occurrences in both reanalyses.

Previous studies have shown that reanalyses employing TC vortex bogussing outperform those without it (Schenkel & Hart, 2012; Zick & Matyas, 2015). Therefore, it is reasonable to hypothesize that the TC bogus scheme in JRA-3Q contributes to its superior reproduction of observed TCs. Although ERA5 also assimilates TC bogus data derived from IBTrACS MSLP observations, the two systems differ substantially in implementation. ERA5 assimilates each IBTrACS MSLP record as a single point observation, whereas JRA-3Q constructs a synthetic three-dimensional TC vortex using the reported MSLP, an empirically estimated gale-force wind radius, gradient wind balance, and prescribed wind and pressure profiles. Observations sampled from the vortex center and surrounding locations are then assimilated as TC bogus data (Kosaka et al., 2024).

The first two rows of Figure 4.3 compare the impacts of TC bogussing in the WP during 1951–1978, when both reanalyses assimilate TC bogus observations and relatively consistent MSLP observations are available. Direct-bogus cases are defined as tracks lacking a detectable LPS record in SyCLoPS within 12 h before the first MSLP record in the corresponding IBTrACS track but become immediately detectable at the first record; all other tracks are classified as "indirect-bogus" cases, representing cases with pre-existing disturbances that likely benefited more indirectly from TC bogus. In JRA-3Q direct-bogus cases, neither a closed MSLP contour (white contours in Figure 4.3) nor a coherent elevated 850-hPa relative vorticity region exists before bogussing, yet within six hours the vortex develops dramatically into a well-organized cyclone with a distinct inner core. In contrast, ERA5 direct-bogus cases show only modest vortex intensification, with slightly stronger low-level vorticity and limited MSLP deepening. The effect is even weaker for ERA5 no-bogus cases, whose vortices are already well developed before the first IBTrACS record. Although JRA-3Q indirect-bogus cases also contain preexisting weak vortices, they undergo substantially greater structural development following bogussing than their ERA5 counterparts.

The last two rows of Figure 4.3 show the spatial variance of column-integrated MSE (h') for 100 randomly selected weaker (985–995 hPa) and stronger (935–945 hPa) Northern Hemisphere TCs across the two reanalyses and eras. Under weak temperature gradients, h' reflects moisture variability associated with organized convection (Carstens & Wing, 2022). ERA5 consistently exhibits smaller inner-core h' than JRA-3Q within both intensity bins, suggesting weaker convection associated with the central dense overcast region, and weaker diabatic activities related to surface enthalpy flux and radiative processes. However, as noted by Dirkes et al. (2023), differences in h' among reanalyses may also arise from MSE advection, data assimilation tendencies, and other unresolved physical processes, warranting further investigation through process-oriented diagnostics. ERA5 also shows substantially larger h' in the satellite era than in the pre-satellite era, whereas JRA-3Q remains relatively consistent, suggesting reduced sensitivity to observing-system changes likely associated with its consistent use of TC bogussing.

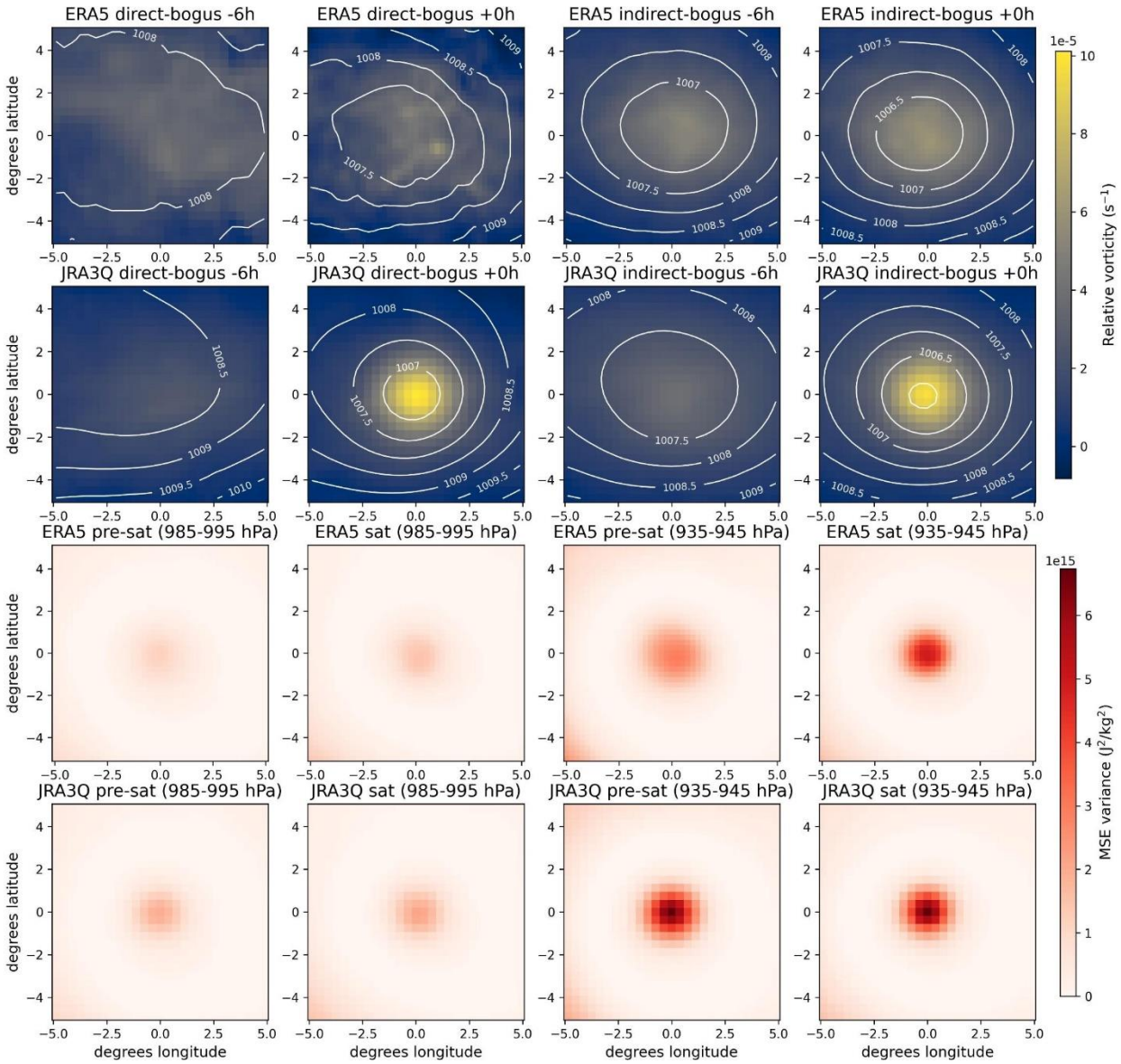



**Figure 4.3**. See the text for details.

### 4.5 Discussion and Conclusions

In this study, we identify substantial improvements in TC representation in JRA-3Q relative to ERA5 and JRA-55. Figure 4.1 shows that JRA-3Q achieves higher detection skill across all basins and maintains stable performance in Northern Hemisphere basins where better observational coverage is anticipated. On the other hand, ERA5 shows an unnatural upward trend in the HRMFAR scores across both the pre- and satellite eras. This suggests that caution should be exercised when using ERA5 for long-

term TC climatological analysis. In the Southern Hemisphere basins, the performance of both ERA5 and JRA-3Q degrades substantially, suggesting inherent limitations in TC representation in reanalysis under conditions of poorer-quality best-track data. We further investigate the sources of JRA-3Q's improvement. Over the NI, where JRA-3Q shows the largest HRMFAR gain, it produces far fewer inland FAs than ERA5, suggesting smaller biases in its underlying forecast model (Figure 4.2). Figure 4.3 indicates that JRA-3Q' TC bogus scheme plays an important role in establishing TC-like vortices, whereas the bogussing impact is diffuse in ERA5. JRA-3Q also exhibits larger spatial variance of inner-core column-integrated MSE for TCs of comparable intensities and maintains consistency between the pre-satellite and satellite eras. This further suggests that its bogussing approach contributes to a more stable TC representation over different periods.

Ultimately, misses and FAs mentioned in this study arise from three broad sources: (1) tracking algorithm biases, (2) reanalysis deficiencies including biases in the underlying model, TC bogussing, and other assimilation processes, and (3) uncertainties (subjectivities) in the IBTrACS best-track dataset. Figure 4.4 attempts to qualitatively decompose these sources of error for JRA-3Q and ERA5 by leveraging the unfiltered SyCLoPS LPS and IBTrACS datasets. While this decomposition cannot fully isolate the contribution of each source, it provides insight into their relative roles.

Missed TCs attributed to tracking errors are those classified by SyCLoPS as non-tropical systems, or, as tropical depressions for at least one day. These boundary cases are most sensitive to classification thresholds. Cases failing only the one-day TC duration criterion are classified as mixed errors because it is inherently subjective to determine the minimum duration for which a TC is considered formed. Remaining misses are attributed to reanalysis deficiencies because they are classified as weak tropical systems (i.e., without a detectable upper-level warm core). Black hatching indicates cases without even a matchable LPS in SyCLoPS. Because FAR is nearly zero in the well-observed NA, EP, and WP basins in both reanalyses, we assume negligible tracking-related FAR except for hybrid TC–related FAs in the NA. FAs matched to IBTrACS tracks that fail the 34-kt wind speed requirement are classified as potential IBTrACS errors. Such cases are more likely to reflect IBTrACS errors when the same FA is reproduced

by both reanalyses (brown hatching), or when they occur during the pre-satellite era. Finally, green hatching in the HR column indicates cases lacking detectable precursors in both SyCLoPS and a low-level relative vorticity tracker (Hodges et al., 2017; Bourdin, 2026) before the first observation of the corresponding IBTrACS track. These cases are expected to benefit more directly from TC bogussing because pre-genesis disturbances are generally detectable as weak LPSs or low-level waves before operational monitoring begins.

Figure 4.4 shows that JRA-3Q generally exhibits smaller "pure" reanalysis errors in both misses and FAs than ERA5, contributing to its lower MR and FAR across most basins. Improvements in MR and FAR over ERA5 can be roughly attributed, respectively, to differences in JRA-3Q's TC bogus scheme and its underlying model. JRA-3Q's HR also remains relatively stable between the pre-satellite and satellite eras, whereas ERA5 experiences a marked degradation in the pre-satellite era, particularly over the EP, where both MR and FAR increase substantially. This degradation likely reflects the combined effects of sparser observations and the relatively weak influence of TC bogussing, allowing model biases in ERA5 to "show through." Though, neither reanalysis recovers the artificially low observed TC frequency in the EP during the pre-satellite era (~12 annually versus ~18 in the satellite era), even with FAs included.

Despite JRA-3Q's superior overall performance, some aspects warrant caution. First, its FAR in the NI, particularly during the pre-satellite era, is unusually low (~0.3 annually) given the larger uncertainties in IBTrACS during this period. This may indicate that JRA-3Q misses some TCs that were also absent from IBTrACS in the earlier decades, while some ERA5 FAs may correspond to undocumented TCs in both eras. That is, whereas ERA5 tends to overestimate the intensity of these NI systems, JRA-3Q may instead underestimate them similarly due to its model design. Fig. 4.2g supports this interpretation, showing several inland-moving monsoonal systems captured by ERA5 but missed by JRA-3Q. This is also evidenced by the relatively larger number of NI TC tracks missed by JRA-3Q with no detectable match and by Malakar et al. (2020), who found that ERA5 represents TC intensity and structure evolution in the NI better than other reanalyses, including JRA-55. These inland systems are

also susceptible to long-term IBTrACS uncertainties because, although they may establish TC-like vertical structures, they may not meet the threshold for sustained surface winds implemented by operational agencies due to land friction. Furthermore, JRA-3Q contains a larger fraction of tracks without detectable pre-record disturbances, suggesting greater reliance on TC bogussing for vortex initiation. Consistent with this, Figure 4.3 shows that mature vortices in JRA-3Q emerge unphysically from largely nonexistent precursors after bogussing. The much weaker impact of TC bogus observations in ERA5 might result from the procedure to reduce their assimilation weight from the preliminary ERA5 extension (Bell et al., 2021) to avoid generating unrealistically strong TCs during the back-extension period in the final product (Soci et al., 2024). These underscore the challenges of reconciling TC bogus with other parts of the reanalysis.

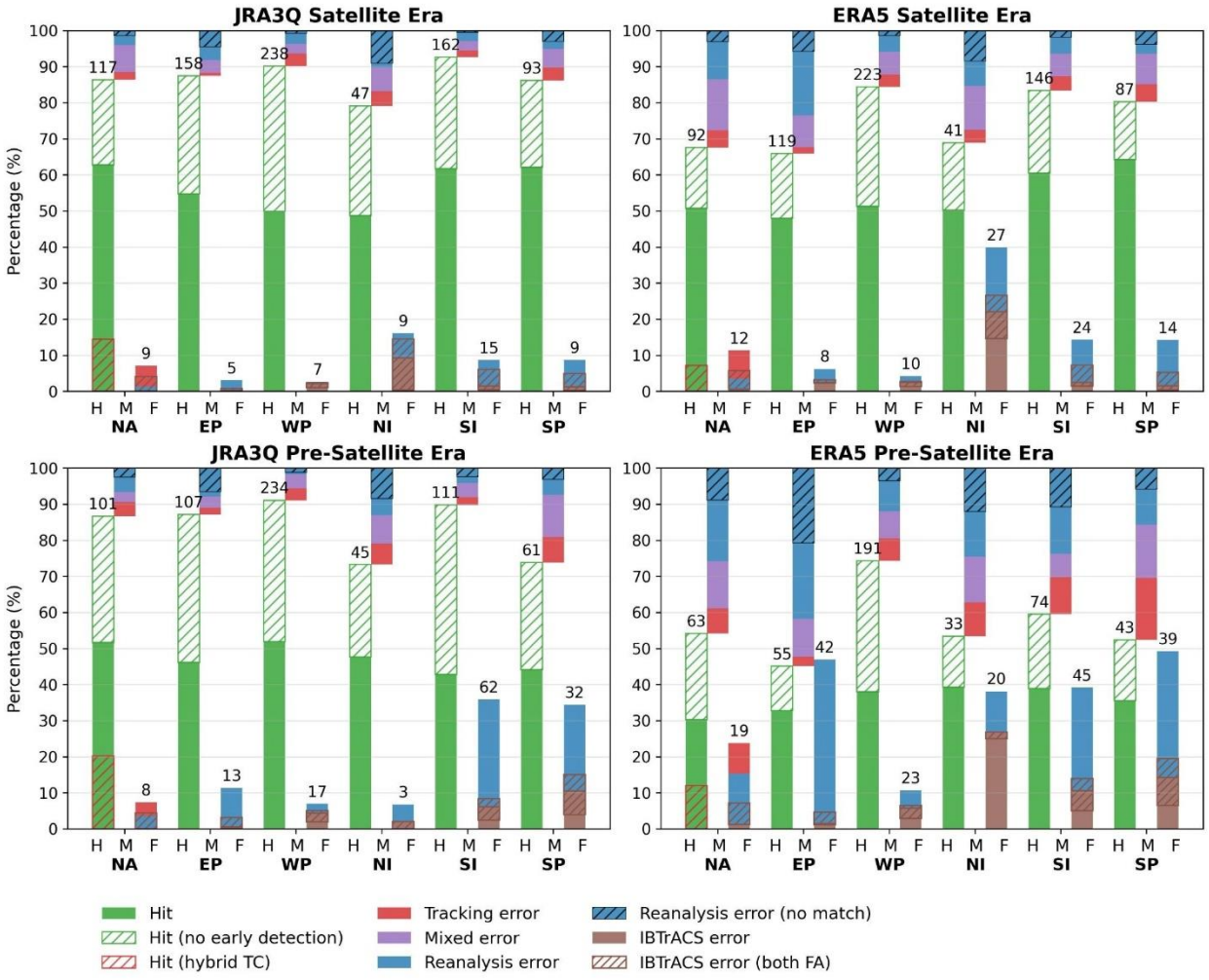



**Figure 4.4**. Decomposition of error sources and uncertainties in the hit rate (H), missing rate (M) and false alarms (F) in ERA5 and JRA3Q over satellite and pre-satellite era across the six basins. Numbers over the bars are decadal-mean TC counts. See the text for more details.

We acknowledge that further understanding and decomposition of these TC representation deficiencies require process-oriented diagnosis of the reanalyses and more importantly, their underlying general circulation models. Our preliminary analysis shows that ECMWF-IFS-HR (C. D. Roberts et al., 2017), a HighResMIP model that mirrors the underlying model of ERA5, exhibits similar biases to ERA5 in its uncoupled historical run over a 27-year period. These biases include an underrepresentation of EP and NA TCs and an overrepresentation of inland NI TCs (Figure 4.S3). Similar behavior occurs in some, but not all, analyzed HighResMIP models. Interestingly, ECMWF-IFS-HR also overproduces TCs near

the tropical West African coast and adjacent inland regions, although whether this shares the same mechanism as NI FAs remains unclear. We hope this study provides guidance for the appropriate use and selection of reanalysis products in TC research and motivates future improvements in reanalysis systems.

## 4.6 Supplements

### Text 4.S1. Introduction of Hybrid TC Label in SyCLoPS

Many of the missed North Atlantic TC detections in ERA5 are classified as subtropical systems because they fail to satisfy the RH100MAX criterion (maximum relative humidity at 100 hPa within 2.0°) in SyCLoPS's Tropical Condition. Since the thresholds defining the tropical condition were trained using the global TC population, the RH100MAX threshold may be overly restrictive for the North Atlantic, leading to a very low FAR but also a comparatively low HR. To evaluate this possibility, we perform a sensitivity analysis by lowering the original RH100MAX threshold (RH100MAX > 20%) in 5% increments while keeping all other TC classification criteria unchanged. The results indicate that both ERA5 and JRA-3Q achieve their best overall performance with a relaxed threshold of RH100MAX > 5%. Under this criterion, the HR for North Atlantic TCs increases in both reanalyses, with only a comparatively modest increase in FAR. In contrast, JRA-55 exhibits only marginal improvements in HR across all tested thresholds while incurring a substantially larger increase in FAR. Based on these results, SyCLoPS assigns the “Hybrid TC” label to low-pressure systems that satisfy the TC classification criteria under this modified tropical condition.

Figure 4.S2 shows the spatial distribution of global hybrid TCs in hybrid TCs tracks, defined as tracks containing at least four 6-hourly Hybrid TC nodes. In both ERA5 and JRA-3Q, Hybrid TCs occur most frequently over the subtropical North Atlantic, corresponding closely to many observed TCs that were previously missed by the standard detection algorithm. Because hybrid TC tracks are rare outside the North Atlantic, including these systems produces little change in the HR and FAR of either reanalysis in the other basins. On the other hand, hybrid TC in JRA-55 are much more spatially dispersed, likely explaining its poor performance when hybrid TCs are included in the North Atlantic TC evaluation.

These results suggest that future users of SyCLoPS should carefully consider whether to include hybrid TCs when evaluating North Atlantic TCs in other datasets. Before doing so, users should verify that the dataset reproduces a realistic spatial distribution of hybrid TCs comparable to that found in ERA5 and JRA-3Q. Otherwise, including hybrid TCs may lead to worse TC detection performance.

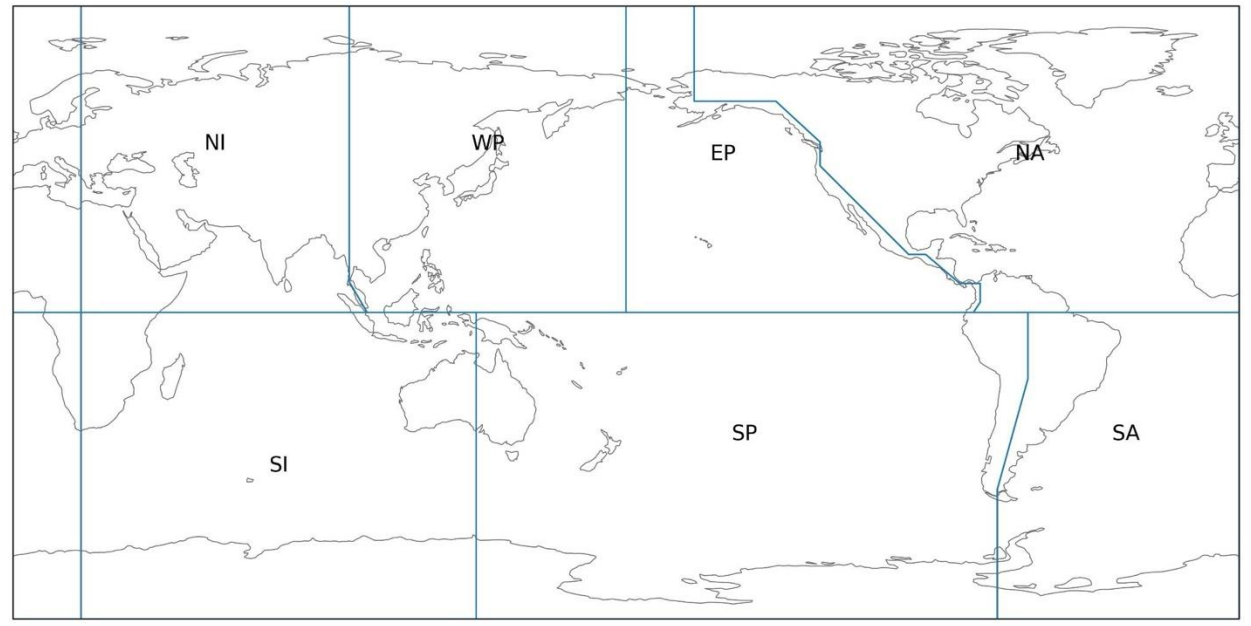



**Figure 4.S1**. Basin boundaries used in this study.

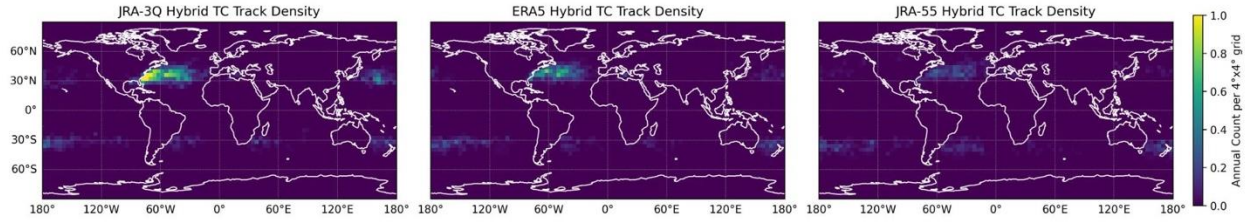



**Figure 4.S2**. Hybrid TC 4° × 4° 1958-2023 annual-mean track density.

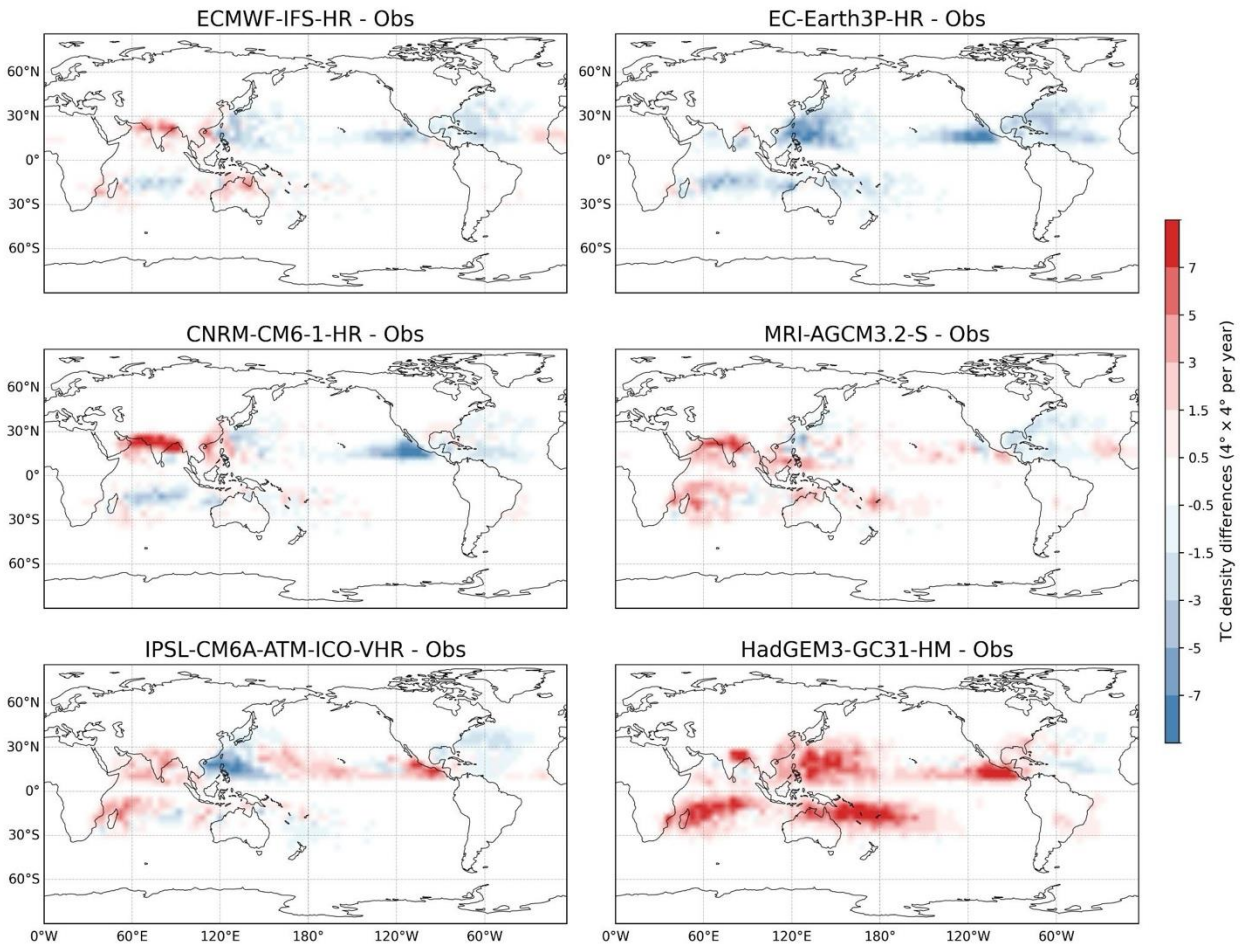



**Figure 4.S3**. TC 4° × 4° annual-mean density differences between HighResMIP models and IBrACS over 1988-2014.

## 4.7 References


Alet, F., Price, I., El-Kadi, A., Masters, D., Markou, S., Andersson, T. R., Stott, J., Lam, R., Willson, M., Sanchez-Gonzalez, A., & Battaglia, P. (2025). Skillful joint probabilistic weather forecasting from marginals (arXiv:2506.10772; Version 1). arXiv. https://doi.org/10.48550/arXiv.2506.10772

Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Soci, C., Villaume, S., Bidlot, J.-R., Haimberger, L., Woollen, J., Buontempo, C., & Thépaut, J.-N. (2021). The ERA5 global reanalysis: Preliminary extension to 1950. Quarterly Journal of the Royal Meteorological Society, 147(741), 4186–4227. https://doi.org/10.1002/qj.4174

Bhatia, K., Baker, A., Yang, W., Vecchi, G., Knutson, T., Murakami, H., Kossin, J., Hodges, K., Dixon, K., Bronselaer, B., & Whitlock, C. (2022). A potential explanation for the global increase in

tropical cyclone rapid intensification. Nature Communications, 13(1), 6626. https://doi.org/10.1038/s41467-022-34321-6
Bian, G.-F., Nie, G.-Z., & Qiu, X. (2021). How well is outer tropical cyclone size represented in the ERA5 reanalysis dataset? Atmospheric Research, 249, 105339. https://doi.org/10.1016/j.atmosres.2020.105339
Bourdin, S. (2026). RABTracks V1 (Version V1) [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.20426510
Bourdin, S., Fromang, S., Dulac, W., Cattiaux, J., & Chauvin, F. (2022). Intercomparison of four algorithms for detecting tropical cyclones using ERA5. Geoscientific Model Development, 15(17), 6759–6786. https://doi.org/10.5194/gmd-15-6759-2022
Carstens, J. D., Didlake, A. C., & Zarzycki, C. M. (2024). Tropical Cyclone Wind Shear-Relative Asymmetry in Reanalyses. Journal of Climate, 37(22), 5793–5861. https://doi.org/10.1175/JCLI-D-23-0628.1
Carstens, J. D., & Wing, A. A. (2022). Simulating Dropsondes to Assess Moist Static Energy Variability in Tropical Cyclones. Geophysical Research Letters, 49(15), e2022GL099101. https://doi.org/10.1029/2022GL099101
Dirkes, C. A., Wing, A. A., Camargo, S. J., & Kim, D. (2023). Process-Oriented Diagnosis of Tropical Cyclones in Reanalyses Using a Moist Static Energy Variance Budget. https://doi.org/10.1175/JCLI-D-22-0384.1
Dulac, W., Cattiaux, J., Chauvin, F., Bourdin, S., & Fromang, S. (2024). Assessing the representation of tropical cyclones in ERA5 with the CNRM tracker. Climate Dynamics, 62(1), 223–238. https://doi.org/10.1007/s00382-023-06902-8
Han, Y., & Ullrich, P. (2026). The Evolution of Tropical Cyclone Intensity and Size in Observations and HighResMIP Models. ESS Open Archive, 2026(0206). https://doi.org/10.22541/essoar.177038847.73410174/v1
Han, Y., & Ullrich, P. A. (2025). The System for Classification of Low-Pressure Systems (SyCLoPS): An All-In-One Objective Framework for Large-Scale Data Sets. Journal of Geophysical Research: Atmospheres, 130(1), e2024JD041287. https://doi.org/10.1029/2024JD041287
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. https://doi.org/10.1002/qj.3803
Hodges, K., Cobb, A., & Vidale, P. L. (2017). How Well Are Tropical Cyclones Represented in Reanalysis Datasets? Journal of Climate, 30(14), 5243–5264. https://doi.org/10.1175/JCLI-D-16-0557.1
Hunt, K. M. R., Turner, A. G., Inness, P. M., Parker, D. E., & Levine, R. C. (2016). On the Structure and Dynamics of Indian Monsoon Depressions. Monthly Weather Review, 144(9), 3391–3416. https://doi.org/10.1175/MWR-D-15-0138.1
Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., & Schreck, C. J. (2018). International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4 [Dataset]. NOAA National Centers for Environmental Information. https://doi.org/10.25921/82ty-9e16
Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data. Bulletin of the American Meteorological Society, 91(3), 363–376.
Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., & Takahashi, K. (2015). The JRA-55 Reanalysis: General Specifications and Basic Characteristics. 気象集誌. 第 2 輯, 93(1), 5–48. https://doi.org/10.2151/jmsj.2015-001
Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., &

Hoyer, S. (2024). Neural general circulation models for weather and climate. Nature, 632(8027), 1060–1066. https://doi.org/10.1038/s41586-024-07744-y

Kosaka, Y., Kobayashi, S., Harada, Y., Kobayashi, C., Naoe, H., Yoshimoto, K., Harada, M., Goto, N., Chiba, J., Miyaoka, K., Sekiguchi, R., Deushi, M., Kamahori, H., Nakaegawa, T., Tanaka, T. Y., Tokuhiro, T., Sato, Y., Matsushita, Y., & Onogi, K. (2024). The JRA-3Q Reanalysis. Journal of the Meteorological Society of Japan. Ser. II, 102(1), 49–109. https://doi.org/10.2151/jmsj.2024-004

Kossin, J. P., Olander, T. L., & Knapp, K. R. (2013). Trend Analysis with a New Global Record of Tropical Cyclone Intensity. Journal of Climate, 26(24), 9960-9976 https://doi.org/10.1175/JCLI-D-13-00262.1

Lopez-Gomez, I., Wan, Z. Y., Zepeda-Núñez, L., Schneider, T., Anderson, J., & Sha, F. (2025). Dynamical-generative downscaling of climate model ensembles. Proceedings of the National Academy of Sciences, 122(17), e2420288122. https://doi.org/10.1073/pnas.2420288122

Malakar, P., Kesarkar, A. p., Bhate, J. n., Singh, V., & Deshamukhya, A. (2020). Comparison of Reanalysis Data Sets to Comprehend the Evolution of Tropical Cyclones Over North Indian Ocean. Earth and Space Science, 7(2), e2019EA000978. https://doi.org/10.1029/2019EA000978

Murakami, H. (2014). Tropical cyclones in reanalysis data sets. Geophysical Research Letters, 41(6), 2133–2141. https://doi.org/10.1002/2014GL059519

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., & Keeley, S. (2017). ECMWF ECMWF-IFS-HR model output prepared for CMIP6 HighResMIP [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2461

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., Caron, L.-P., Chauvin, F., Terray, L., Valcke, S., Moine, M.-P., Putrasahan, D., Roberts, C., Senan, R., Zarzycki, C., & Ullrich, P. (2020). Impact of Model Resolution on Tropical Cyclone Simulation Using the HighResMIP–PRIMAVERA Multimodel Ensemble. Journal of Climate, 33(7), 2557–2583. https://doi.org/10.1175/JCLI-D-19-0639.1

Schenkel, B. A., & Hart, R. E. (2012). An Examination of Tropical Cyclone Position, Intensity, and Intensity Life Cycle within Atmospheric Reanalysis Datasets. Journal of Climate, 25(10), 3453–3475. https://doi.org/10.1175/2011JCLI4208.1

Slocum, C. J., Razin, M. N., Knaff, J. A., & Stow, J. P. (2022). Does ERA5 Mark a New Era for Resolving the Tropical Cyclone Environment? Journal of Climate, 35(21), 7147–7164. https://doi.org/10.1175/JCLI-D-22-0127.1

Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Villaume, S., Haimberger, L., Woollen, J., Buontempo, C., & Thépaut, J.-N. (2024). The ERA5 global reanalysis from 1940 to 2022. Quarterly Journal of the Royal Meteorological Society, 150(764), 4014–4048. https://doi.org/10.1002/qj.4803

Truchelut, R. E., Hart, R. E., & Luthman, B. (2013). Global Identification of Previously Undetected Pre-Satellite-Era Tropical Cyclone Candidates in NOAA/CIRES Twentieth-Century Reanalysis Data. Journal of Applied Meteorology and Climatology, 52(10), 2243–2259. https://doi.org/10.1175/JAMC-D-12-0276.1

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., & Reed, K. A. (2021). TempestExtremes v2.1: A community framework for feature detection, tracking, and analysis in large datasets. Geoscientific Model Development, 14(8), 5023–5048. https://doi.org/10.5194/gmd-14-5023-2021

Vecchi, G. A., & Knutson, T. R. (2011). Estimating Annual Numbers of Atlantic Hurricanes Missing from the HURDAT Database (1878–1965) Using Ship Track Density. Journal of Climate, 24(16), 1736–1746. https://doi.org/10.1175/2010JCLI3810.1

Vishnu, S., Boos, W. R., Ullrich, P. A., & O'Brien, T. A. (2020). Assessing Historical Variability of South Asian Monsoon Lows and Depressions With an Optimized Tracking Algorithm. Journal of Geophysical Research: Atmospheres, 125(15), e2020JD032977. https://doi.org/10.1029/2020JD032977

Watt-Meyer, O., Henn, B., McGibbon, J., Clark, S. K., Kwa, A., Perkins, W. A., Wu, E., Harris, L., & Bretherton, C. S. (2025). ACE2: Accurately learning subseasonal to decadal atmospheric variability and forced responses. Npj Climate and Atmospheric Science, 8(1), 205. https://doi.org/10.1038/s41612-025-01090-0

Xu, Z., Guo, J., Zhang, G., Ye, Y., Zhao, H., & Chen, H. (2024). Global tropical cyclone size and intensity reconstruction dataset for 1959–2022 based on IBTrACS and ERA5 data. Earth System Science Data, 16(12), 5753–5766. https://doi.org/10.5194/essd-16-5753-2024

Ye, G., Leung, J. C.-H., Dong, W., Toumi, R., Xu, J., Li, W., Qian, W., Kong, H., & Zhang, B. (2026). A Reanalysis-Based Global Tropical Cyclone Tracks Dataset for the Twentieth Century (RGTracks-20C). Earth System Science Data, 18(5), 3587–3607. https://doi.org/10.5194/essd-18-3587-2026

Zarzycki, C. M., & Ullrich, P. A. (2017). Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. Geophysical Research Letters, 44(2), 1141–1149. https://doi.org/10.1002/2016GL071606

Zarzycki, C. M., Ullrich, P. A., & Reed, K. A. (2021). Metrics for evaluating tropical cyclones in climate data. Journal of Applied Meteorology and Climatology, 60(5), 643-660. https://doi.org/10.1175/JAMC-D-20-0149.1

Zhang, K., & Chan, K. T. F. (2023). An ERA5 global climatology of tropical cyclone size asymmetry. International Journal of Climatology, 43(2), 950–963. https://doi.org/10.1002/joc.7846

Zick, S. E., & Matyas, C. J. (2015). Tropical cyclones in the North American Regional Reanalysis: An assessment of spatial biases in location, intensity, and structure. Journal of Geophysical Research: Atmospheres, 120(5), 1651–1669. https://doi.org/10.1002/2014JD022417

# Chapter 5. Conclusions

## 5.1 Other Ongoing Applications of SyCLoPS

This dissertation has established the foundation for the development of SyCLoPS and presented two application studies that use SyCLoPS to evaluate tropical cyclone (TC) representation in climate models and atmospheric reanalyses, highlighting their respective strengths and limitations. The potential applications of SyCLoPS extend well beyond the studies presented in this dissertation, as the framework is capable of detecting and classifying 14 types of low-pressure systems (LPSs) in a wide range of large-scale atmospheric datasets. Beyond the studies included in the previous chapters, I have been involved in several additional projects that leverage SyCLoPS, including the detection of North Atlantic TCs in the fully coupled high-resolution E3SM model (Balaguru et al., 2026), the association of LPSs with mesoscale convective systems over the continental United States, the evaluation of tropical and extratropical cyclones in kilometer-scale Earth system model simulations, and the development of Reanalysis-Augmented Best Tracks for Tropical Cyclones (RABTracks; Bourdin et al., 2026).

Recent published studies and ongoing work have further demonstrated the potential applications of SyCLoPS to extratropical transition (ET) pathways of TCs (Jin et al., 2026) and extratropical predictability in AI-hybrid models (Zhang, 2026). Jin et al. (2026) used the SyCLoPS ERA5 dataset to investigate the pathways of TC extratropical transition in the relatively understudied Southern Hemisphere and identified four clusters of distinct synoptic environments associated with ET, each leading to different downstream weather regimes. Zhang (2026) used SyCLoPS to evaluate the Neural General Circulation Model (NGCM) developed by Google Research by initializing model forecasts at the TC genesis times identified in the SyCLoPS ERA5 dataset. The study then evaluated the forecasted extratropical atmospheric fields, represented by 500-hPa geopotential height, against ERA5 across TC track clusters defined based on the SyCLoPS dataset.

Applications of SyCLoPS across diverse datasets and scientific problems have revealed limitations in the original framework and motivated continuous improvements to enhance its adaptability

and reliability. Many of the deficiencies discovered have been addressed in the later releases of SyCLoPS, as descripted in Chapter 2.10. As SyCLoPS continues to be applied to new datasets and research questions, we anticipate continued maintenance and further development of the framework to support broader applications in atmospheric science.

### 5.2 Future Open Science Questions

SyCLoPS was developed as a tool to facilitate the exploration of open scientific questions in atmospheric science. As demonstrated in the previous sections, SyCLoPS has already been applied to evaluate TC and extratropical cyclone performance in reanalysis products and physical and AI models, investigate tropical–extratropical interactions associated with TC propagation, and to be linked with other weather features of different scales. Beyond these applications, this subsection highlights several scientific questions for which SyCLoPS, in combination with other feature-detection algorithms, may provide a useful framework for further investigation.

Within the broader field of TC research, numerous unresolved questions remain regarding TC genesis pathways and the roles of different types of TC seeds. Although SyCLoPS does not track all TC seeds back to their initial atmospheric-wave stages, it identifies surface LPS signals throughout their lifecycles, including initial tropical systems that subsequently dissipate or develop into TCs in reanalyses and model outputs. This provides a straightforward framework for comprehensively quantifying TC seed survival rates and investigating the mechanisms that govern the transition from precursor disturbances to developed TCs. When combined with an atmospheric-wave tracker and classifier or a TC tracker based on low-level vorticity signals (e.g., TRACK; Hodges et al., 2017), this framework could enable a more complete characterization of TC genesis pathways, from initial LPSs embedded in synoptic-scale precursors (i.e., seeds)—including monsoon troughs, easterly waves, equatorial waves, the subtropical jet, and ITCZ-related disturbances—to mesoscale development pathways. These development pathways may include bottom-up development (e.g., Halverson et al., 2007; Dunkerton et al., 2009), top-down development (e.g., Bister & Emanuel, 1997; Ritchie & Holland, 1997), and genesis associated with

processes originating near the tropopause, including tropical transition and formation induced by tropical upper-tropospheric troughs (e.g., McTaggart-Cowan, 2013). Such a combined framework could address several fundamental questions regarding TC genesis: Is there a preferred or dominant pathway of TC formation across different geographic regions and seasons? Which combinations of TC seed type and mesoscale development pathway are most favorable for TC genesis and subsequent intensification, and which are most likely to lead to the most destructive SG-E intensity-size evolution pathway as discussed in Chapter 3.5? What environmental conditions separate the development trajectory between developing and non-developing TC seeds? How do the composition and characteristics of different TC genesis pathways change under climate change?

As discussed in Chapter 2.6, SyCLoPS can be used to evaluate the compound hazards (e.g., winds and precipitation) associated with all detectable LPSs. Our results show that not only TCs, but also other systems—including global monsoonal lows, subtropical storms, and polar lows—can produce compound hazards that are more severe and frequent than those associated with TCs in many regions. Therefore, we advocate taking an "LPS class-blinded" and "hazard-centered" approach to characterizing LPS activity, emphasizing that all LPSs should be considered important and meaningful when assessing their potential to produce extreme hazards, including precipitation (rainfall and snowfall), winds, storm surges, and heat waves, regardless of their classification or intensity. Indicators of these hazards can be evaluated over the lifecycle of individual LPS tracks or aggregated into two-dimensional spatial fields that provide stakeholders with information on the accumulated magnitude, probability (risk), or return period of hazardous events over a designated period. Applied to future climate projections, this hazard-centered framework can provide a more comprehensive perspective on how LPS activity and its associated hazards may change under a warming climate.

Another intriguing future direction for SyCLoPS applications is a global, comprehensive investigation of the much less studied population of tropical-like cyclones, including subtropical storms (e.g., Mediterranean cyclones) and polar lows. These systems have traditionally been documented individually across different basins (e.g., Dias Pinto et al., 2013; Emanuel, 2005; Nordeng & Rasmussen,

1992). More recently, Emanuel et al., (2025) proposed grouping these systems together as "surface flux-driven cyclones outside the tropics," or CYCLOPs. In that study, several case studies were examined individually to investigate the potential role of surface fluxes in the genesis and evolution of these systems. However, substantial uncertainties remain, as surface fluxes alone may not fully explain the development and maintenance of all analyzed cases. SyCLoPS provides a potential framework for addressing these uncertainties by enabling systematic detection and tracking of tropical-like cyclones across diverse atmospheric datasets. A global climatology derived from SyCLoPS could facilitate investigations of tropical-like cyclone seeds, genesis pathways, and life cycle evolution using approaches analogous to those widely applied in TC research. When combined with environmental analysis, process-oriented diagnosis, and other feature detection algorithms in reanalyses and idealized simulations, SyCLoPS could help clarify the relative roles of surface fluxes, atmospheric dynamics, and environmental conditions, and other large-scale features (e.g., blocking, atmospheric rivers, and cutoff lows) in controlling the formation and evolution of these understudied cyclone populations.

Finally, SyCLoPS produces Lagrangian, label-rich datasets together with environmental parameters and spatial information, providing a useful foundation for training AI emulators and predictive models for a range of applications. For example, SyCLoPS could provide training data for AI emulators and weather models, including the latest Google weather forecast model WeatherNext2 (Alet et al., 2026), to predict TC genesis, track characteristics, intensity, and lifecycle evolution from the early stages of TC development, at substantially lower computational cost than full-scale physical forecast models. SyCLoPS may also provide a useful foundation for developing AI-based hazard prediction systems. For individual storms, an AI model could use the Lagrangian evolution of individual LPSs as a physically meaningful representation of the weather systems responsible for hazards, enabling predictions of the probability and magnitude of precipitation, wind, storm surge, and other hazards conditioned on the characteristics and environmental evolution of an LPS. At subseasonal-to-seasonal timescales, spatial maps of the aggregated magnitude or risk of specific LPS-related hazards could be generated directly from LPS activity and lower-frequency environmental indicators, providing a computationally efficient

framework for predicting the spatial distribution of weather hazards and their potential changes under current and future climate conditions.

Taken together, these potential applications demonstrate that SyCLoPS is more than just a tool for identifying and classifying LPSs. It also provides a common framework for connecting LPSs with atmospheric features across spatial and temporal scales, as well as relating them to relevant hazards of concern to different stakeholders. Using SyCLoPS and other feature-detection algorithms alongside climate datasets, numerical simulations, and emerging AI models can bridge the gap between traditional, process-based understanding and data-driven approaches to atmospheric predictability. We envision SyCLoPS as an evolving framework that will facilitate new investigations into the formation, evolution, predictability, and hazards of the full spectrum of low-pressure systems, thereby contributing to a more comprehensive understanding of their roles in weather and climate.

### 5.3 References


Alet, F., Andersson, T. R., Price, I., Markou, S., El-Kadi, A., Masters, D., Li, A., Merchant, S., Williams, N., Thornton, G., MacKay, K., Graham, O., Uddin, A., Gaiarin, B., Shah, D., Kruse, E., Hogsett, W., Zelinsky, D., Cangialosi, J., … Battaglia, P. (2026). Operational Tropical Cyclone Forecasting with AI. *Nature*, 1–3. https://doi.org/10.1038/s41586-026-10953-2

Bister, M., & Emanuel, K. A. (1997). The Genesis of Hurricane Guillermo: TEXMEX Analyses and a Modeling Study. *Monthly Weather Review*, 125(10), 2662–2682.

Dias Pinto, J. R., Reboita, M. S., & da Rocha, R. P. (2013). Synoptic and dynamical analysis of subtropical cyclone Anita (2010) and its potential for tropical transition over the South Atlantic Ocean. *Journal of Geophysical Research: Atmospheres*, 118(19), 10,870-10,883. https://doi.org/10.1002/jgrd.50830

Dunkerton, T. J., Montgomery, M. T., & Wang, Z. (2009). Tropical cyclogenesis in a tropical wave critical layer: Easterly waves. *Atmospheric Chemistry and Physics*, 9(15), 5587–5646. https://doi.org/10.5194/acp-9-5587-2009

Emanuel, K. (2005). Genesis and maintenance of" Mediterranean hurricanes". *Advances in Geosciences*, 2, 217–220.

Emanuel, K., Alberti, T., Bourdin, S., Camargo, S. J., Faranda, D., Flaounas, E., Gonzalez-Aleman, J. J., Lee, C.-Y., Miglietta, M. M., Pasquero, C., Portal, A., Ramsay, H., Reale, M., & Romero, R. (2025). CYCLOPs: A Unified Framework for Surface Flux-Driven Cyclones Outside the Tropics. *Weather and Climate Dynamics*, 6(3), 901–926. https://doi.org/10.5194/wcd-6-901-2025

Halverson, J., Black, M., Braun, S., Cecil, D., Goodman, M., Heymsfield, A., Heymsfield, G., Hood, R., Krishnamurti, T., McFarquhar, G., Mahoney, M. J., Molinari, J., Rogers, R., Turk, J., Velden, C., Zhang, D.-L., Zipser, E., & Kakar, R. (2007). Nasa's Tropical Cloud Systems and Processes Experiment. *Bulletin of the American Meteorological Society*, 88(6), 867–882. https://doi.org/10.1175/BAMS-88-6-867

Hodges, K., Cobb, A., & Vidale, P. L. (2017). How Well Are Tropical Cyclones Represented in Reanalysis Datasets? *Journal of Climate*, 30(14), 5243–5264. https://doi.org/10.1175/JCLI-D-16-0557.1

Nordeng, T. E., & Rasmussen, E. A. (1992). A most beautiful polar low. A case study of a polar low development in the Bear Island region. *Tellus* A, 44(2), 81–99. https://doi.org/10.1034/j.1600-0870.1992.00001.x

Ritchie, E. A., & Holland, G. J. (1999). Large-Scale Patterns Associated with Tropical Cyclogenesis in the Western Pacific. *Monthly Weather Review*, 127(9), 2027–2043.