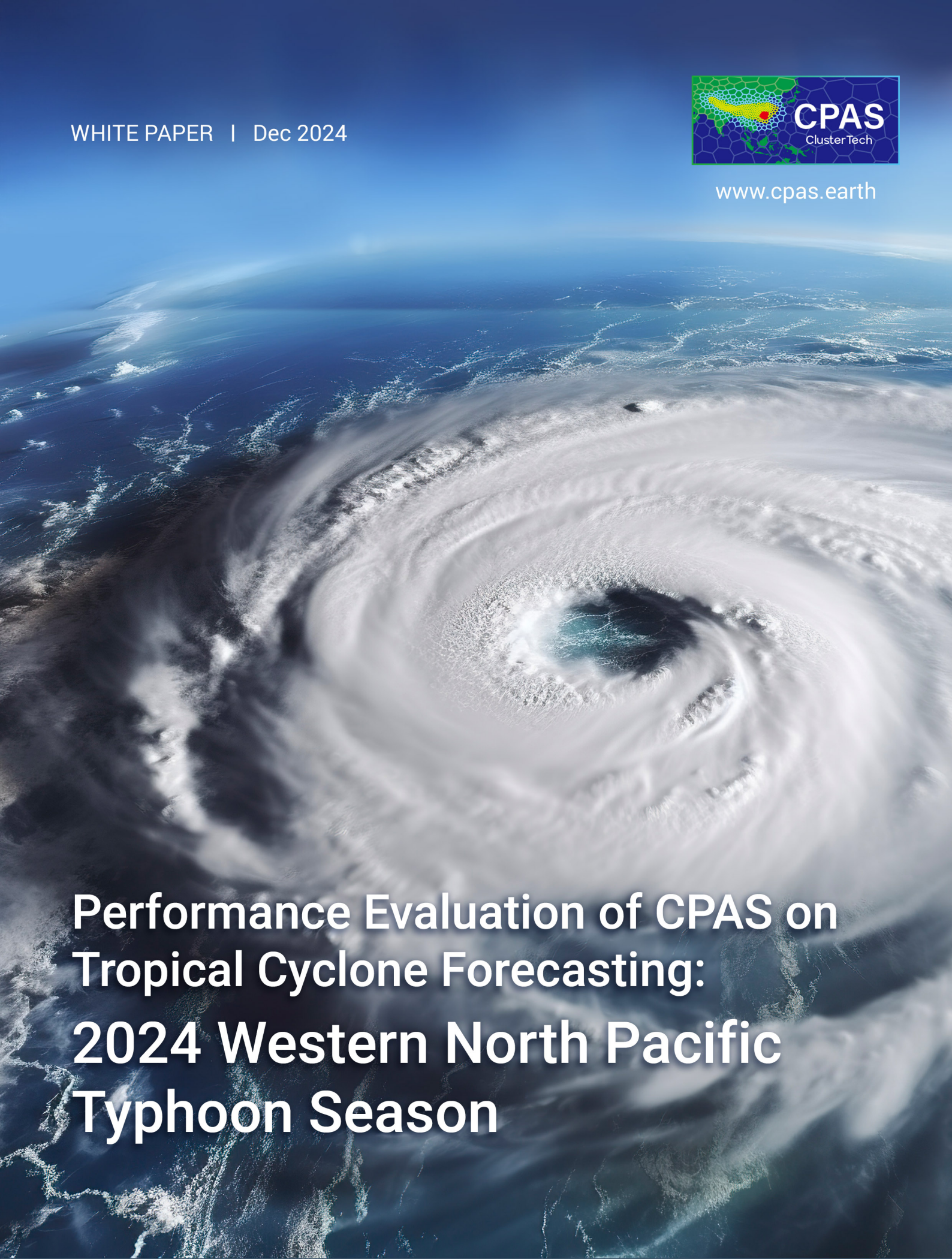
WHITE PAPER | Dec 2024

CPAS
ClusterTech

www.cpas.earth



Performance Evaluation of CPAS on Tropical Cyclone Forecasting:

2024 Western North Pacific Typhoon Season

1. Introduction

Tropical cyclones (TCs) remain one of the most destructive natural hazards, with far-reaching impacts on lives, infrastructure, and economies. In 2024, the western North Pacific experienced devastating TC events. *Super Typhoon Yagi* inflicted catastrophic damage, causing an estimated *CNY 80 billion* in direct economic losses across China in September. The storm displaced over *500,000 people*, destroyed thousands of homes, and severely impacted agricultural regions like Hainan Province. That same month, *Strong Typhoon Bebinca* struck Shanghai, resulting in *CNY 10 billion in losses*, marking it the most damaging typhoon in the region since 1949.

In the face of such escalating risks, the *ClusterTech Platform for Atmospheric Simulation (CPAS)* emerges as a transformative cloud platform for early TC prediction. Based on the Numerical Weather Prediction (NWP) MPAS-A, CPAS integrates *Hierarchical Time-Stepping (HTS)* and a *Customized Unstructured Mesh Generation (CUMG)*, allowing it to deliver decent forecasts while significantly saving computational power to enhance atmospheric forecast efficiency (Ng et al. 2019; Cheung et al. 2022).

This report evaluates the performance of the *CPAS daily run* in forecasting TCs during the 2024 season in the Western North Pacific, utilizing variable-resolution unstructured grids with resolution ranging from 4.5km to 72km (**Figure 1**). The study benchmarks CPAS against two world-leading operational forecasting models: the *NCEI Global Forecast System (GFS) with global 13km resolution* and the *ECMWF Integrated Forecast System (IFS) HRES deterministic model with global 9km resolution*, providing a comprehensive analysis of forecast accuracy, track prediction, and intensity forecasts.

2024 Tropical Cyclone Season Overview

The 2024 season consists of 25 named TCs in the Northwestern Pacific, ranging from tropical storms to super typhoons. Amongst all TCs, *9 of them made landfall in China*, aligning with the historical average of 7 - 9 landfalls. These TCs brought widespread destruction, particularly in coastal areas, emphasizing the importance of reliable forecasting systems like CPAS in mitigating their impacts. **Figures 2** and **3**, along with **Table 1**, provide a detailed timeline and classification of these TCs, highlighting their durations and intensities.

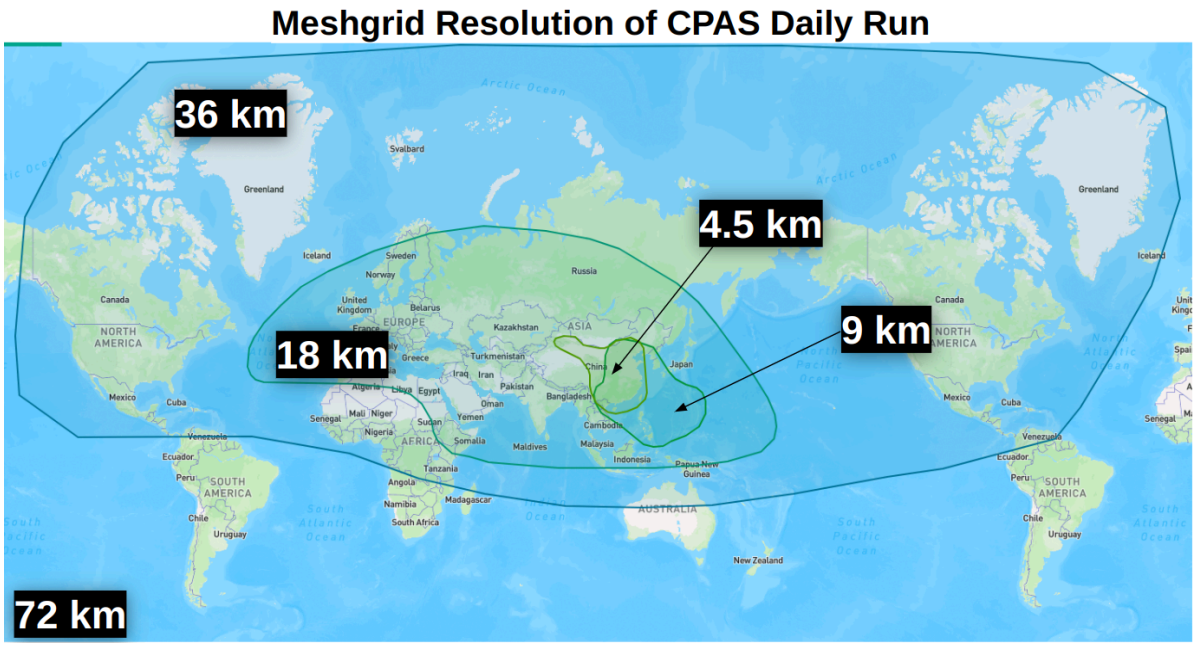


Figure 1: The mesh resolution of CPAS daily run. The resolution for TCs in the Western North Pacific and part of China is 9 km and 4.5 km respectively, and the outermost domain is 72km.

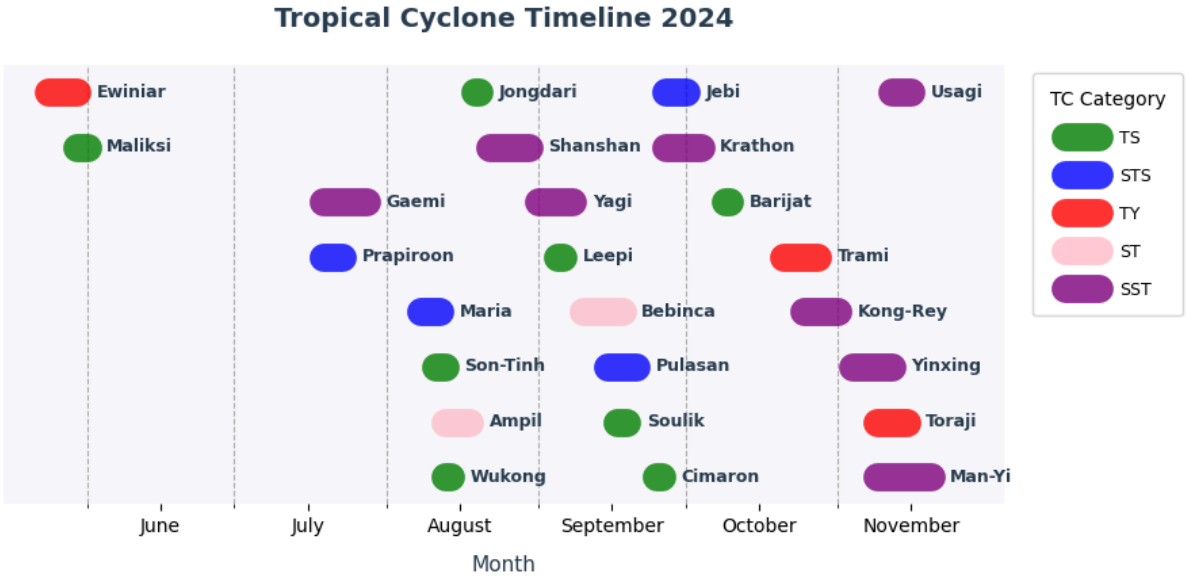
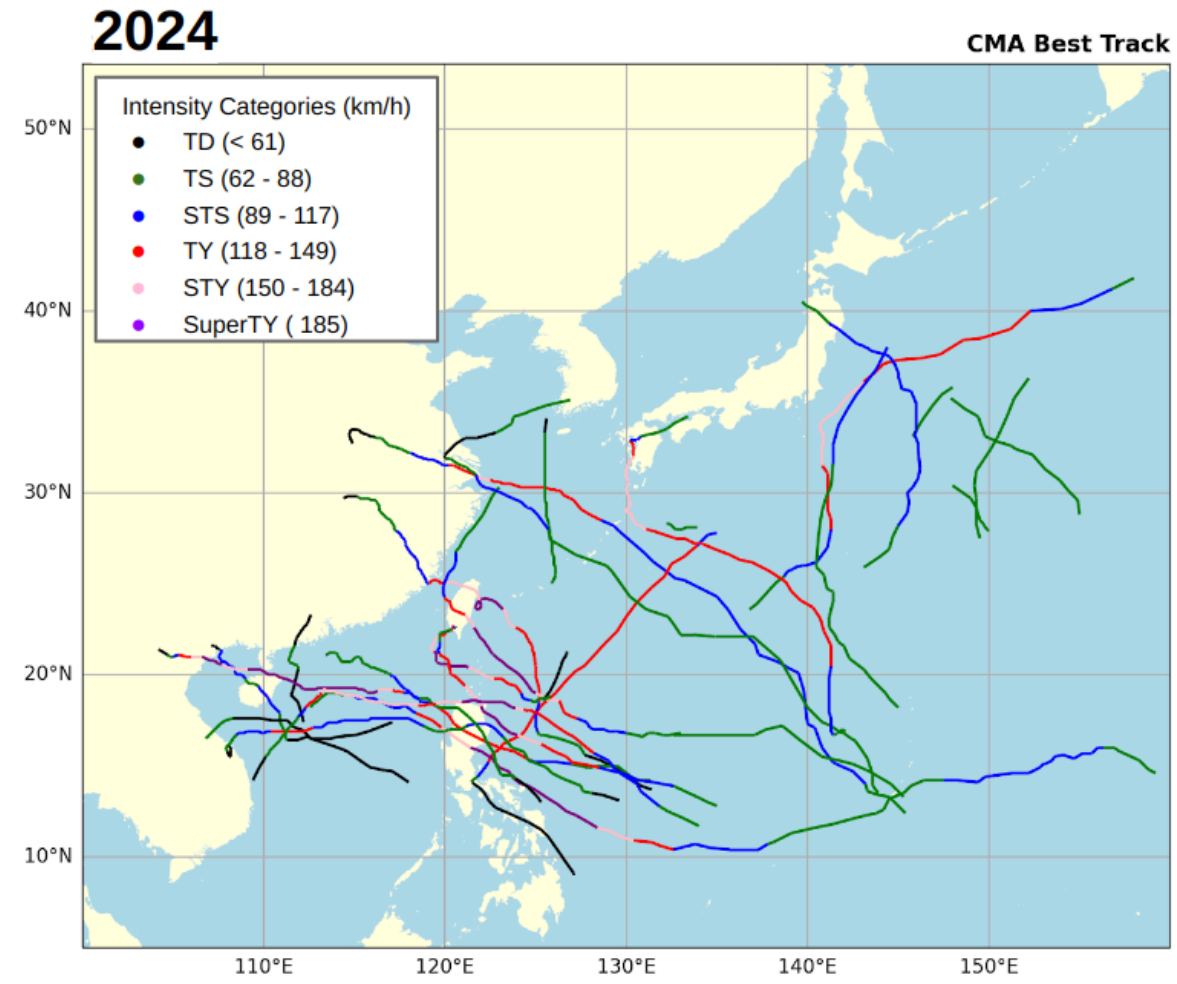
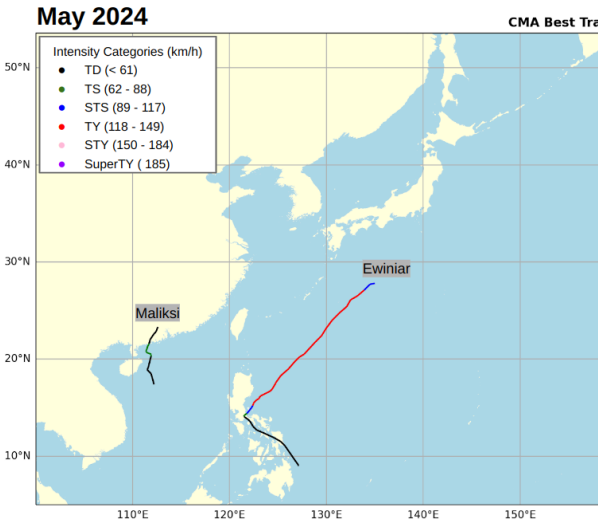


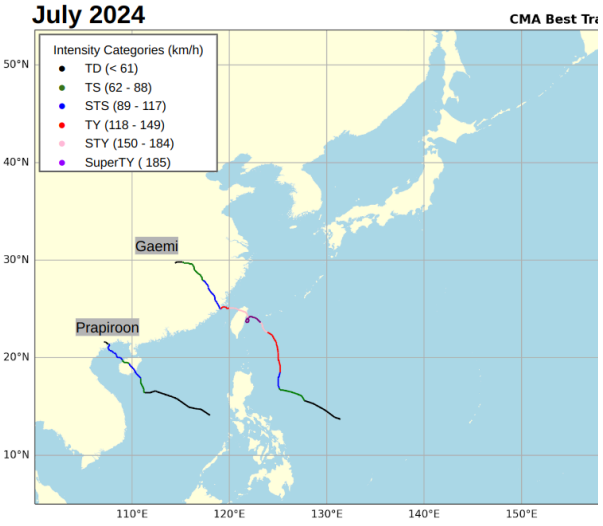
Figure 2: 2024 Western North Pacific Typhoon Season Timeline. TC categories: TS for tropical storm, STS for strong tropical storm, TY for typhoon, ST for strong typhoon, and SST for super strong typhoon.


2024
CMA Best Track
Intensity Categories (km/h)
TD (< 61)
TS (62 - 88)
STS (89 - 117)
TY (118 - 149)
STY (150 - 184)
SuperTY (185)
50°N
40°N
30°N
20°N
10°N
110°E
120°E
130°E
140°E
150°E

(a)


May 2024
CMA Best Tra
Intensity Categories (km/h)
TD (< 61)
TS (62 - 88)
STS (89 - 117)
TY (118 - 149)
STY (150 - 184)
SuperTY (185)
Ewiniar
Maliksi
50°N
40°N
30°N
20°N
10°N
110°E
120°E
130°E
140°E
150°E

(b)


July 2024
CMA Best Tra
Intensity Categories (km/h)
TD (< 61)
TS (62 - 88)
STS (89 - 117)
TY (118 - 149)
STY (150 - 184)
SuperTY (185)
Gaemi
Prapiroon
50°N
40°N
30°N
20°N
10°N
110°E
120°E
130°E
140°E
150°E

(c)

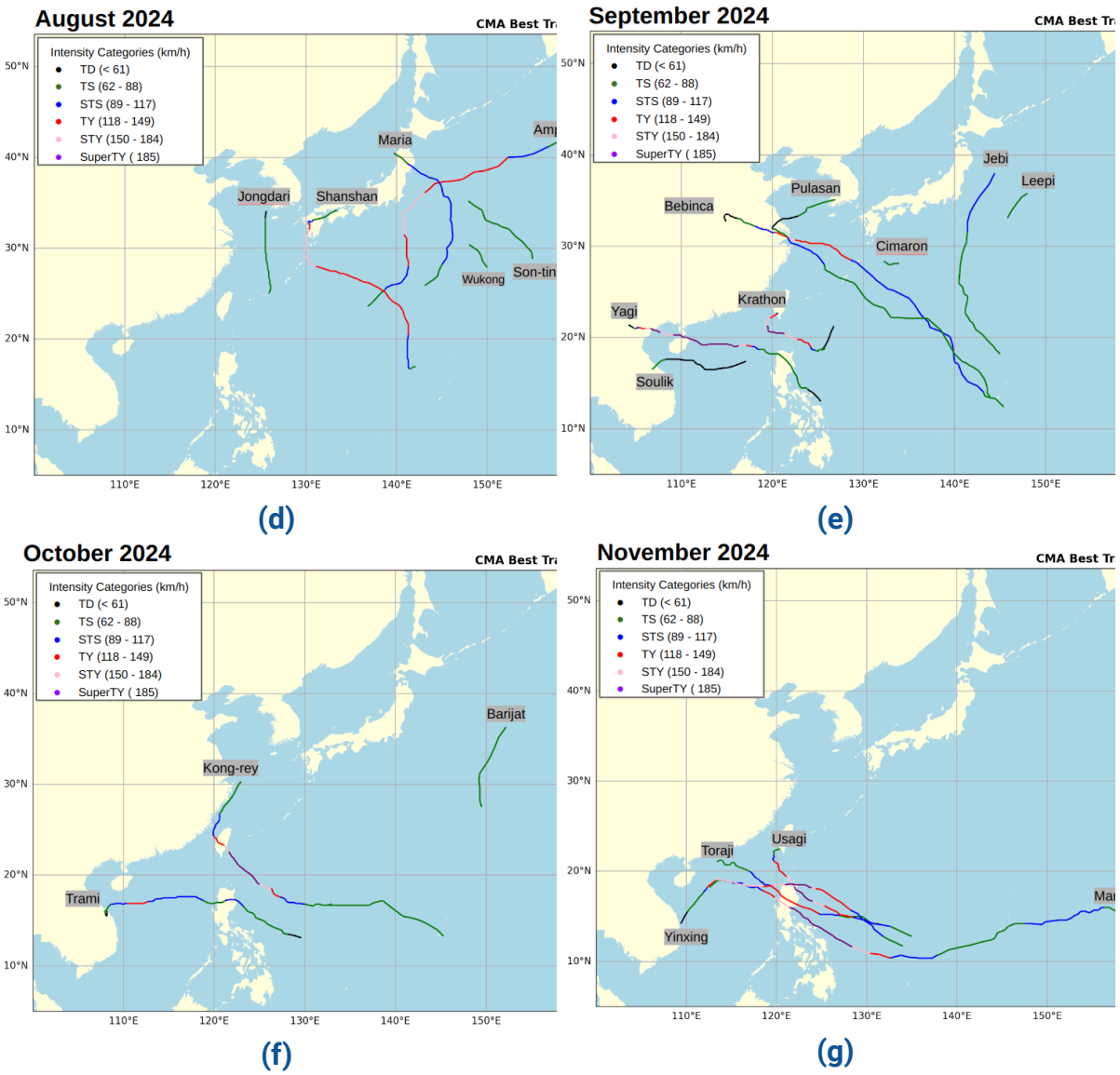


Figure 3: China Meteorological Administration (CMA) Tropical Cyclone Best Track Dataset of 2024 Tropical Cyclones: (a) all months (b) May, (c) July, (d) August, (e) September, (f) October; and (g) November.

Table 1: 2024 named tropical cyclone list (Tropical Storm or stronger) and the classification standard can be found on Hong Kong Observatory[1]. TCs marked with bold font are those making landfall in China.

NAME (中文)	Classification	Start Date	End Date
EWINIAR (艾雲尼)	TY	2024-05-24	2024-05-30
MALIKSI（馬利斯）	TS	2024-05-30	2024-06-01
GAEMI（格萊美）	SuperTY	2024-07-19	2024-07-28
PRAPIROON（派比安）	STS	2024-07-19	2024-07-23
MARIA（瑪莉亞）	STS	2024-08-08	2024-08-12
SON-TINH（山神）	TS	2024-08-11	2024-08-13

[1] HKO Tropical Cyclone Classification

AMPIL (安比)	STY	2024-08-13	2024-08-18
WUKONG (悟空)	TS	2024-08-13	2024-08-14
JONGDARI (雲雀)	TS	2024-08-19	2024-08-20
SHANSHAN (珊珊)	SuperTY	2024-08-22	2024-08-30
YAGI (摩羯)	SuperTY	2024-09-01	2024-09-08
LEEPI (麗琵)	TS	2024-09-05	2024-09-06
BEBINCA (貝碧嘉)	STY	2024-09-10	2024-09-18
PULASAN (普拉桑)	STS	2024-09-15	2024-09-21
SOULIK (蘇力)	TS	2024-09-17	2024-09-19
CIMARON (西馬崙)	TS	2024-09-25	2024-09-26
JEBI (飛燕)	STS	2024-09-27	2024-10-01
KRATHON (山陀兒)	SuperTY	2024-09-27	2024-10-04
BARIJAT (百里嘉)	TS	2024-10-09	2024-10-10
TRAMI (潭美)	TY	2024-10-21	2024-10-28
KONG-REY (康妮)	SuperTY	2024-10-25	2024-11-01
YINXING (銀杏)	SuperTY	2024-11-04	2024-11-12
TORAJI (桃芝)	TY	2024-11-09	2024-11-15
MAN-YI (萬宜)	SuperTY	2024-11-09	2024-11-19
USAGI (天兔)	SuperTY	2024-11-12	2024-11-16

2. Data

The CPAS daily run uses GFS 00UTC data (0.25°x0.25°) as its initial condition, producing forecasts up to 4 days and 16 hours (Total 112 hours). CPAS output is interpolated to a grid resolution of approximately 10 km (0.1°x0.1°) to support the TC tracking process, enabling detailed and consistent analysis across all forecast models.

GFS/ECMWF TC Forecast Track dataset

TCs forecast track data from GFS and ECMWF are extracted at NCAR Research Data Archive Tigge Model Tropical Cyclone Track Data[2].

CMA Tropical Cyclone Best Track Dataset

The China Meteorological Administration (CMA) Tropical Cyclone Best Track Dataset[3] provides detailed information on TCs originating in the western North Pacific. This basin lies north of the equator and west of 180°E, encompassing the

[2] NCAR Research Data Archive Tigge Model Tropical Cyclone Track Data
[3] CMA Tropical Cyclone Best Track Data was obtained from tcdata.typhoon.org.cn

South China Sea. The dataset includes 6-hourly analyses of TC tracks and intensities.

3. Methodology

3.1. Vortex Tracker

TCs from CPAS results are located using GFDL vortex tracker 3.9[4] developed by the National Center for Atmospheric Research (NCAR).

3.2. Track Analysis

Track error analysis for TC forecasts employs **direct positional error (DPE)**, which calculates the distance between forecast and best-tracked data of TC position.

3.3. Intensity Analysis

The intensity performance assessment focuses on two key parameters of TCs: maximum wind speed and central pressure (mean sea level pressure). The evaluation involves two metrics:

Bias: This represents the difference between the model-predicted intensity and the observed intensity (take CMA best track as ground true), calculated as:

$$\text{Bias} = \text{Model} - \text{Observation}$$

Absolute Error: This measures the magnitude of the bias, indicating the overall discrepancy without considering the sign of the deviation:

$$\text{Error} = |\ \text{Bias}\ |$$

These metrics provide insight into the model's ability to accurately capture both the strength and central pressure characteristics of TCs. The absolute error offers a straightforward measure of forecast accuracy, regardless of over- or under-prediction.

3.4. Hypothesis Testing

Hypothesis testing will be used to compare the error of CPAS to GFS and ECMWF. Taken with absolute values, the samples exhibit skewness which does not satisfy the assumption of data normalization as required in the t-test, the *Wilcoxon signed rank test* instead, is used as a non-parametric test which does not require specified

[4] Biswas, M. K., D. Stark, and L. Carson, 2018: GFDL Vortex Tracker Users Guide V3.9a, 35 pp.

sample distribution. The computation steps are briefly described as follows, details of the implementation and theory can be found at Wilcoxon (1965):

1. For each pair of observations, calculate the difference between the two related groups (e.g., Error of $Group_{CPAS}$ - $Group_{GFS}$)
1. Take the absolute values of the non-zero differences and rank them from smallest to largest. If there are ties, assign the average rank to the tied values.
2. Assign the sign of the original difference (positive or negative) to the corresponding ranks.
3. Calculate the sum of the ranks for the positive differences (W^+) and negative differences (W^-). The test statistic W is defined as the smaller of these two sums: $W = \min(W^+, W^-)$.
4. For small sample sizes (< 20), refer to Wilcoxon signed-rank test tables to find the critical value of W. For larger sample sizes (typically n>= 20), the distribution of W can usually be approximated by a normal distribution.
5. Compare the calculated critical value from the table. If using a normal approximation for larger samples, calculate the z-score of W, i.e.

$$z = \frac{(W - \mu_W)}{\sigma_w},$$

where μ_w and σ_w are the mean and standard deviation of the distribution of W. If the calculated p-value is less than the *significance level (0.05)*, or if W is less than the critical value, reject the null hypothesis.

3.5. Rapid Intensification Analysis

Rapid intensification is a rapid strengthening of a TC over a short time. In convention, it is defined as a TC increase in wind speed of at least 30 kt within 24 hours, as proposed by Kaplan and DeMaria (2003). Although several studies have suggested different definitions, CMA adopts the definition of 15m/s (~29.2kt) (Xiang, 2022), which is close to Kaplan and Demaria's definition. We will follow the CMA convention for 15m/s in 24 hours as the threshold.

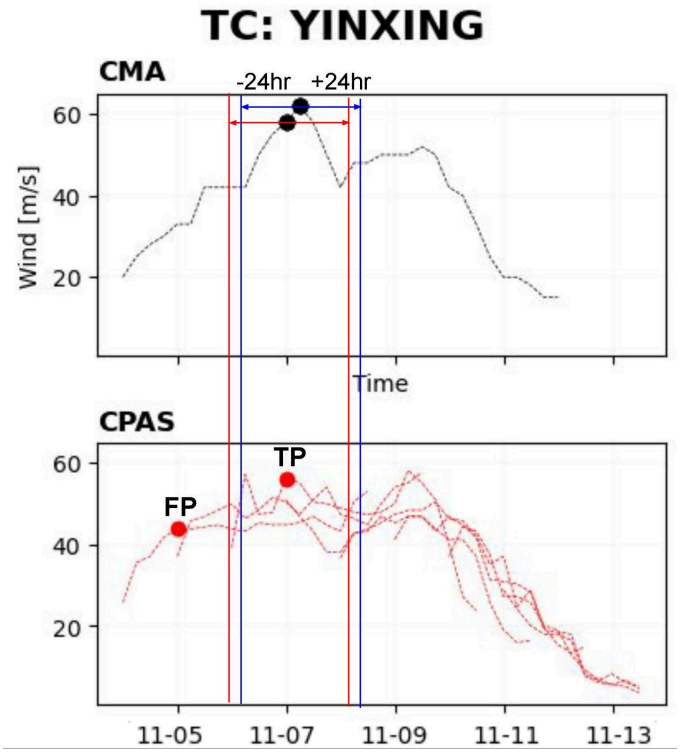


Figure 4: Demonstration of valid RI forecast, setting a maximum 24-hour deviation from the observation record. TP and FP labels represent True Positive and False Positive respectively.

Three metrics are used to evaluate the models' effectiveness for forecasting TC RI, where the TP, TN, FP and FN represent True Positve, True Negative, False Positive and False Negative respectively. Their relationships are shown in **Figure 5**.

- The *probability of detection (POD)*: evaluates the effectiveness of a detection system. It quantifies the likelihood that a true positive will be correctly identified.

$$POD = \frac{TP}{TP + FN}$$

- The *False Alarm Ratio (FAR):* measures the frequency at which a detection system incorrectly identifies a negative case as positive. It is calculated using the formula:

$$FAR = \frac{FP}{FP + TN}$$

- The *Critical Success Index (CSI):* evaluates how effectively a system predicts events compared to actual occurrences. It is calculated as:

$$CSI = \frac{TP}{TP+FP+FN}$$

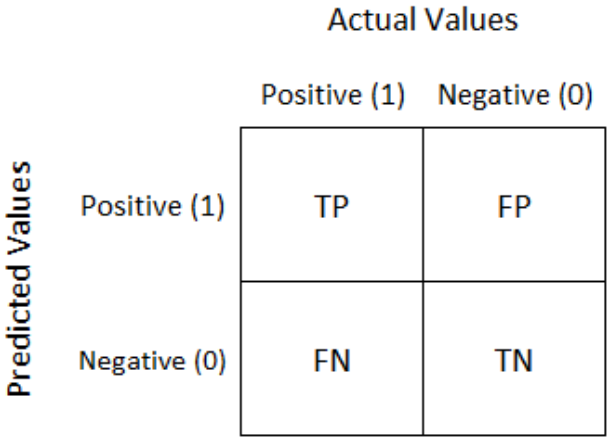


Figure 5: Demonstration of confusion matrix table, extracted from Towards Data Science[5].

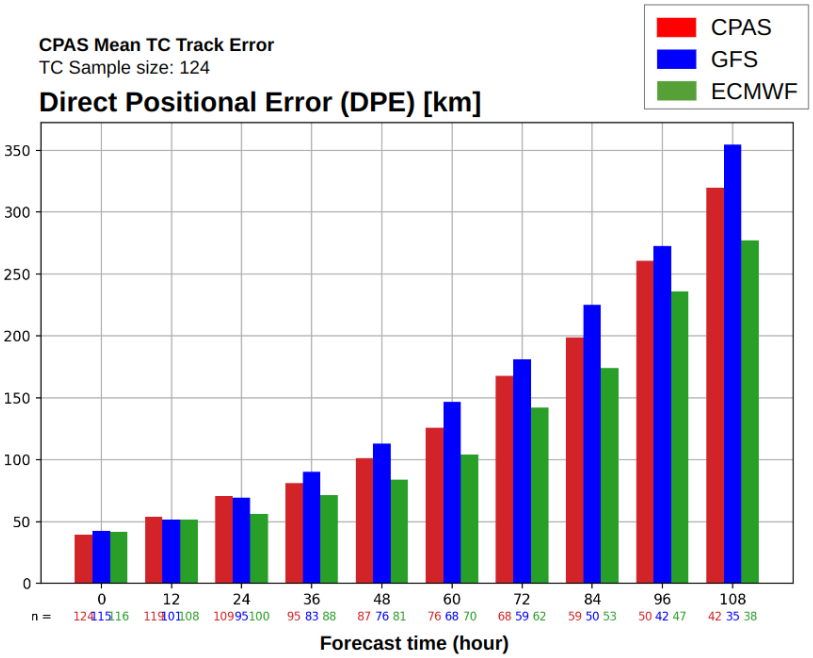
4. Performance

4.1. Track Analysis

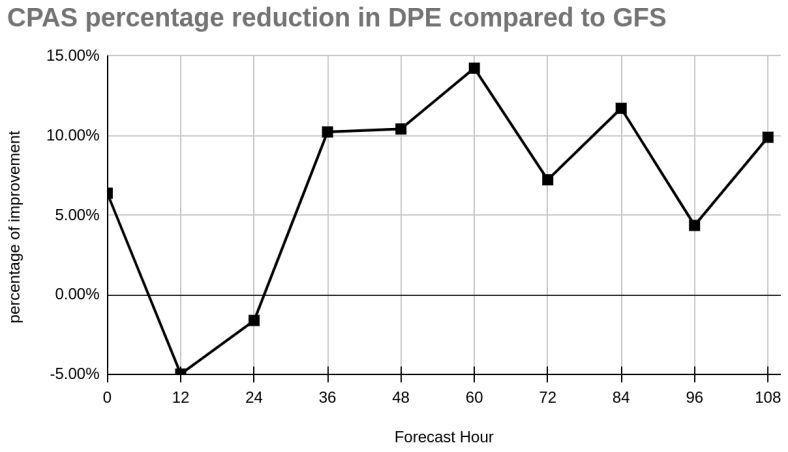
Overall, CPAS performance in the TC track forecast is comparable to the other two leading-edge operating forecast models.

The DPE of CPAS lies between that of ECMWF and GFS (**Figure 6a**), with GFS exhibiting the highest DPE and ECMWF the lowest. **Figure 6b** shows that CPAS demonstrates a 5 - 15% DPE improvement for one day or longer forecast time compared to GFS by using its dataset.

[5] A., Ragan (2018). Taking the Confusion Out of Confusion Matrices, *Towards Data Science*.



(a)



(b)

Figure 6: **(a)** Bar plot comparing CPAS (red), GFS (blue), and ECMWF (green) of TC *direct positional error (DPE)* versus forecast hour, and **(b)** CPAS percentage reduction in DPE compared to GFS. $Percentage\ reduction\ = \frac{DPE_{GFS} - DPE_{CPAS}}{DPE_{GFS}}$.

4.2. Intensity Analysis

In **Figure 7** (upper left), CPAS performs better than ECMWF and is comparable to GFS in wind speed prediction. At shorter forecast times (T+0h to T+36h), CPAS has lower wind errors than ECMWF, though errors increase steadily over time, reaching approximately 9 m/s at T+108h, below ECMWF and marginally above GFS. For Wind speed bias (**Figure 7** lower left), CPAS overestimates wind speeds to a slight degree,

showing a mean positive bias of around 0 - 2.5m/s, while GFS and ECMWF exhibit a stronger negative bias, varying from -2 to -4m/s and -6 to -9m/s respectively.

In pressure error (**Figure 7** upper right), CPAS outperforms ECMWF and is slightly better than GFS. It maintains lower errors than ECMWF from about 3 - 5 hPa in the first 94 hours. In the first 72 forecast hours, CPAS's pressure error is slightly lower than GFS to a small extent. By T+108h, CPAS's pressure error reaches 15 hPa, remaining lower than ECMWF but higher than GFS. All models show positive pressure bias in the lower right of **Figure 7**, but CPAS demonstrates its competitively low mean bias. ECMWF considerably overestimates pressure, while GFS maintains a partly positive bias at shorter lead times. Notably, both CPAS and GFS indicate decreasing bias as the forecast proceeds.

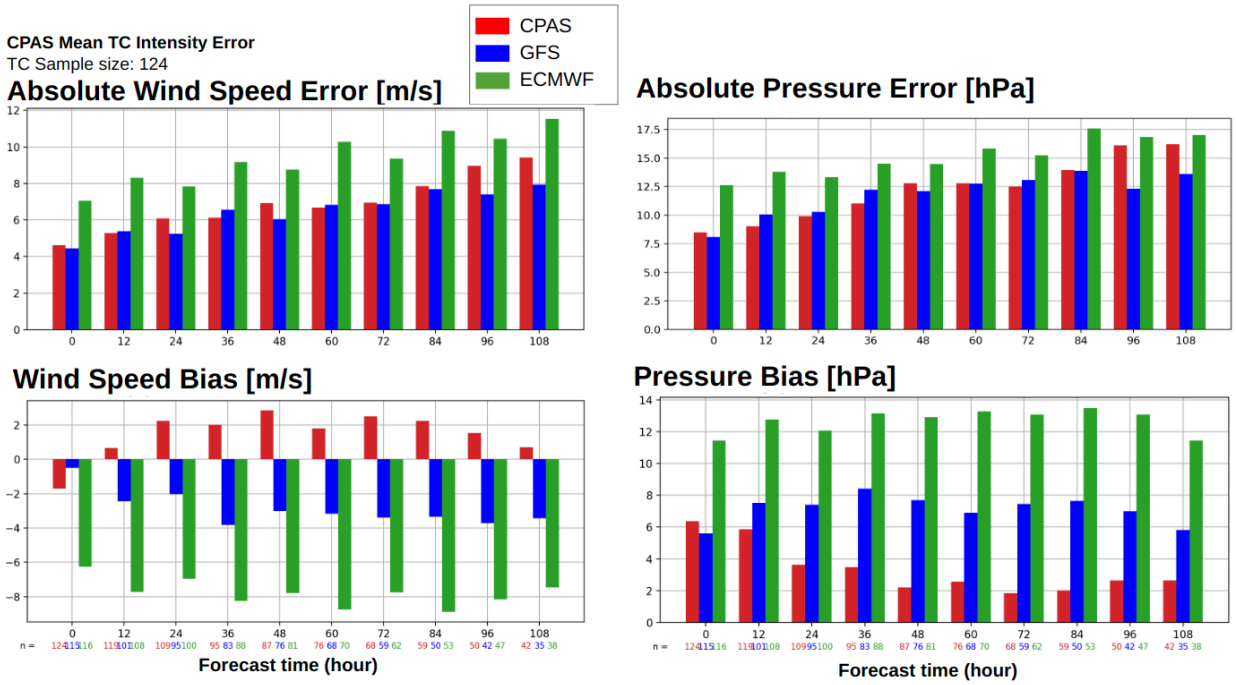


Figure 7: Bar plot comparing CPAS (red), GFS (blue) and ECMWF (green) of TC wind speed error (top-left) and bias (bottom-left); and central pressure error (top-right) and bias (bottom-right) versus forecast hour

For both wind speed and central pressure (**Figures 8a** and **8b**), CPAS demonstrates strong predictability in the first 72 hours, with regression line slopes exceeding 0.8, outperforming GFS and ECMWF. CPAS moderately overestimates TC strength for TCs in weaker intensities. The decline in CPAS performance in T+96h could be attributed to fewer sample points and increased model error uncertainty.

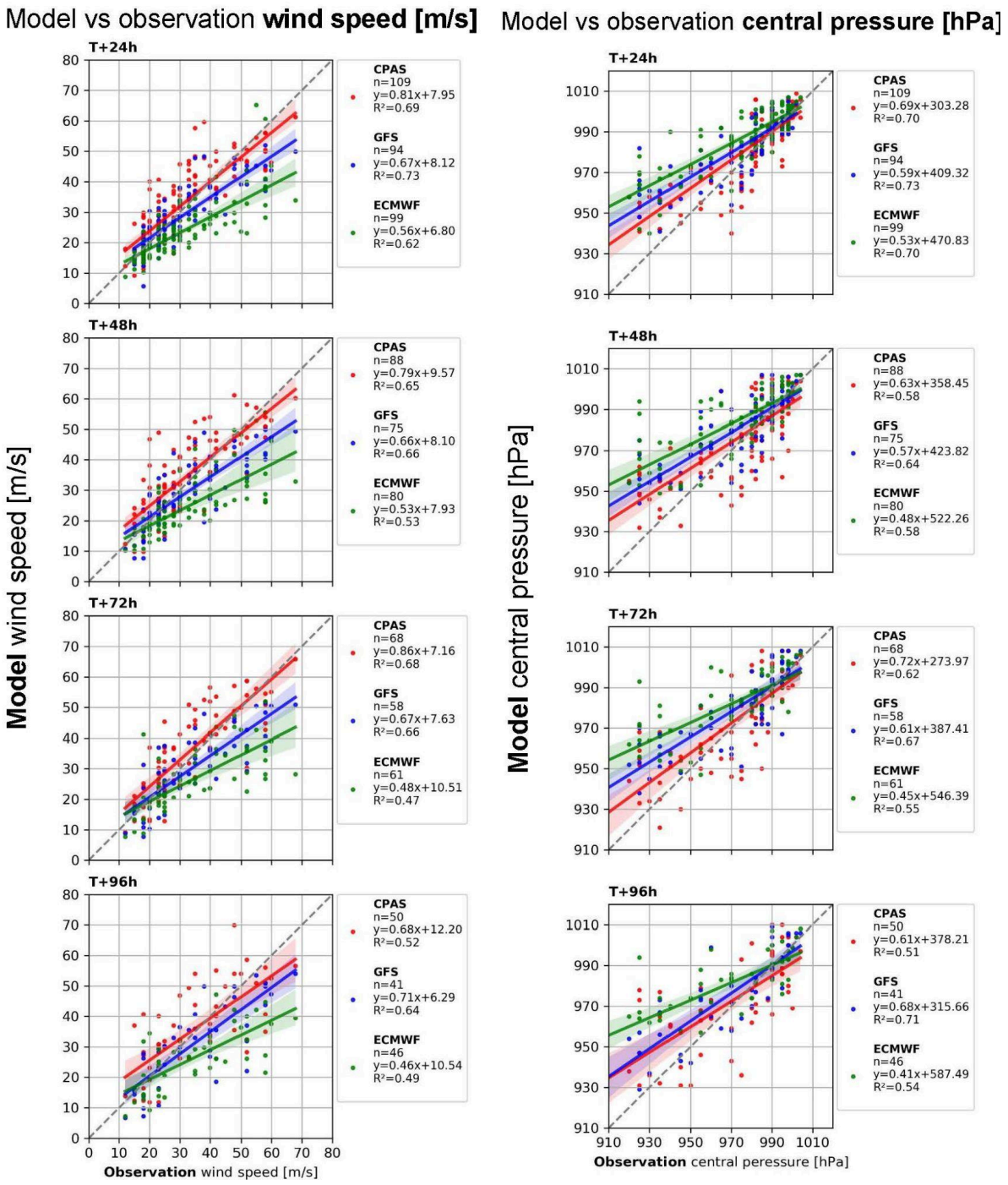


Figure 8: 3 models versus observation **(a)** wind speed and **(b)** central pressure at T+24h, T+48h, T+72h, and T+96h respectively. n, y=mx+c, and R^2 shown in legends represent the number of samples, the equation of the fitted line, and the coefficient of determination respectively.

4.3. Hypothesis Test

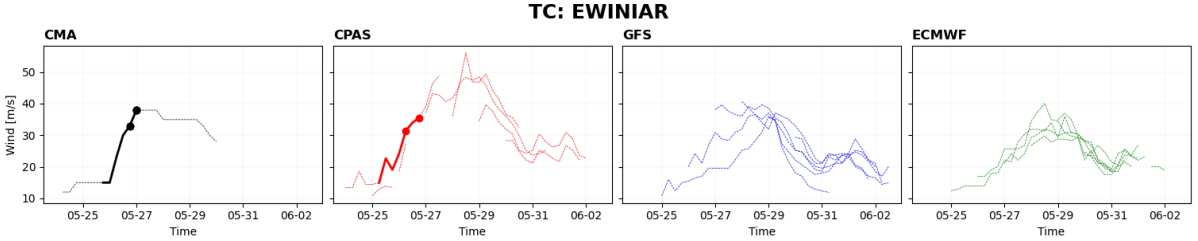
Table 2 presents the p-values from the Wilcoxon signed-rank test. CPAS provides slightly better track forecasts than GFS, with significant differences observed between T+24h and T+72h. Regarding wind and pressure absolute error, CPAS significantly outperforms ECMWF until T+96h, while its performance in wind prediction is comparable to GFS, and a significantly lower pressure error from T+24h to T+72h.

Table 2: Wilcoxon signed-rank test absolute error p-values for $P_{Direct\_Positional\_Error}$, P_{Wind_Error} , and $P_{Pressure\_Error}$. P-values that are less than 0.05 are highlighted, signifying that CPAS demonstrates a statistically better result in comparison.

Forecast time (hour)	Wilcoxon signed-rank test					
	$P_{Direct\_Positional\_Error}$		P_{Wind_Error}		$P_{Pressure\_Error}$	
	GFS	ECMWF	GFS	ECMWF	GFS	ECMWF
0 - 24	0.719	1.000	0.940	**0.000**	0.056	**0.000**
24 - 48	**0.010**	1.000	0.712	**0.000**	**0.039**	**0.014**
48 - 72	**0.007**	1.000	0.702	**0.000**	**0.001**	**0.036**
72 - 96	0.285	0.994	0.545	**0.000**	0.385	**0.026**
96 - 108	0.468	0.873	0.818	**0.003**	0.907	0.313

4.4. Rapid Intensification (RI)

8 TCs underwent RI: Ewiniar, Gaemi, Yagi, Krathon, Kong-rey, Yinxing, Usagi, and Manyi as shown in **Figure 9**. CPAS generally exhibits better alignment with the CMA best track during the RI phase of TCs, as evident from its slopes being closer to the CMA reference compared to GFS and ECMWF, which is particularly noticeable in cases like Ewiniar, Yagi, and Krathon. CPAS effectively captures the trend and intensity changes during the critical first 24 hours of RI. In contrast, GFS and ECMWF show significant deviations, with GFS often producing steeper or flatter predictions and ECMWF displaying a smoother but less responsive pattern.



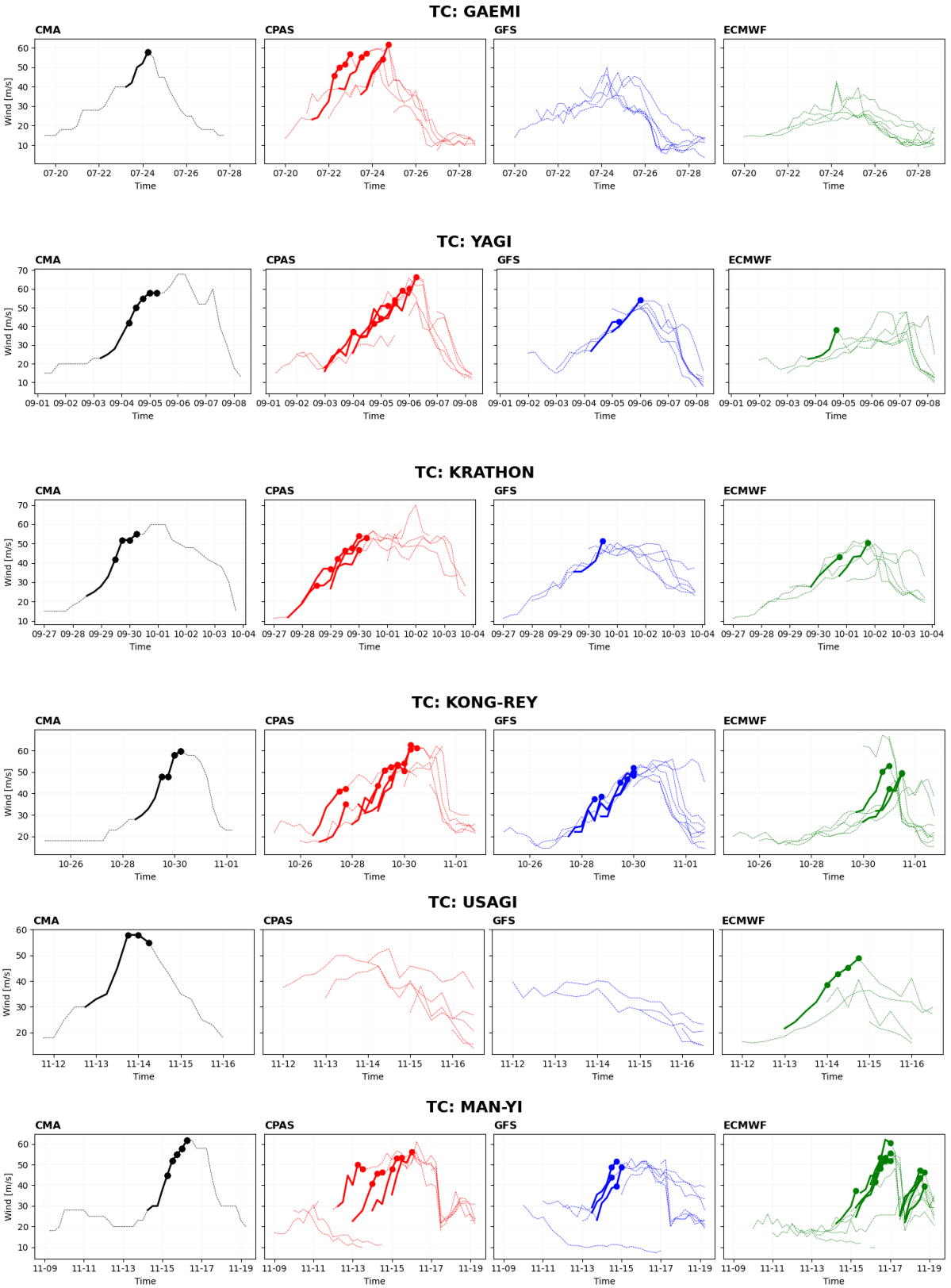


Figure 9: Time series of the intensity of seven TCs that underwent rapid intensification (RI). The CMA best track is shown in black, while CPAS (red), GFS (blue), and ECMWF (green) represent model predictions. Multiple lines correspond to

different model initialization times. Dots indicate instances of RI over the past 24 hours, and solid lines depict the entire RI process.

Table 3 summarizes the model performances regarding RI events as a contingency table with skill scores, including POD, FAR, and CSI. The computation methods of each metric can be found in Section 3.4. The results show that CPAS outperforms both GFS and ECMWF, achieving a POD of 0.85, indicating a strong ability to correctly identify RI events, while GFS and ECMWF have lower PODs of 0.44 and 0.43 respectively. However, CPAS has a higher FAR of 0.14 compared to 0.03 for both GFS and ECMWF, suggesting that it may generate more false alarms. Despite this, CPAS excels in overall performance with a CSI of 0.54, significantly better than the 0.39 scores of GFS and ECMWF. These metrics highlight CPAS's strengths in predicting RI, emphasizing its value for stakeholders requiring timely and accurate forecasting in TC scenarios.

Table 3 Contingency Table and verification metrics of the models regarding their capability in forecasting TC RI. **Bold** values highlight the best performance score among models for each metric.

		CPAS		GFS		ECMWF	
		True	False	True	False	True	False
CMA Best Track	True	22	4	11	14	11	17
	False	16	99	3	113	3	111
Probability of Detection		**0.85**		0.44		0.43	
False Alarm Ratio		0.14		**0.03**		0.03	
Critical Success Index		**0.54**		0.39		0.39	

5. Conclusion

The performance evaluation of the ClusterTech Platform for Atmospheric Simulation (CPAS) highlights its capability as a reliable and efficient tool for TC forecasting. Compared to the world-leading operational models GFS and ECMWF, CPAS demonstrates competitive performance in predicting TC track and intensity, addressing several key challenges in TC forecasting.

Key Findings

1. **Track Prediction:** CPAS produces statistically smaller DPE (Direct Position Error) than GFS, particularly within the T+72h forecast window, showcasing its strength in track accuracy during critical days of prediction periods. While ECMWF maintains better overall track accuracy, CPAS consistently narrows the performance gap and provides robust predictions.
2. **Intensity Prediction:** CPAS statistically outperforms ECMWF in wind speed and central pressure errors, particularly in the first 96 hours, and shows comparable performance to GFS. This is crucial for forecasting TC intensity, as precise predictions of wind speed and pressure are critical for risk assessment and mitigation planning.

Addressing TC Forecasting Challenges

CPAS tackles several persistent challenges in TC forecasting, including:

1. **Reliable Prediction of TCs' RI**: CPAS's steady ability to reliably predict TCs' RI provides critical lead time for stakeholders. This helps reduce the risk of catastrophic losses by enabling proactive decision-making, resource allocation, and evacuation planning.
2. **Reduce Track Forecasting Uncertainty**: By narrowing track errors and delivering consistent performance across different forecast periods, CPAS reduces uncertainty in TC track predictions, a common issue in operational forecasting.
3. **Better Intensity Forecasting Accuracy**: Accurate wind speed and pressure forecasts allow stakeholders to better prepare for the severity of TCs, particularly in vulnerable regions.

References

Cheung, C. C., Tam, C. Y., Leung, W. N., Ng, K. K., & Sze, W. P. (2022, June). Applications of flexible spatial and temporal discretization techniques to a numerical weather prediction model. In Proceedings of the platform for advanced scientific computing conference (pp. 1-8).

Heming, J. T., 2016: Tropical cyclone tracking and verification techniques for Met Office numerical weather prediction models. Meteorological Applications, 24, 1-12. doi: 10.1002/met.1599

Kaplan, J., & DeMaria, M. (2003). Large-Scale Characteristics of Rapidly Intensifying Tropical Cyclones in the North Atlantic Basin. Weather and Forecasting, 18(6), 1093–1108. doi: 10.1175/1520-0434(2003)018<1093:LCORIT>2.0.CO;2

Lu, X. Q., H. Yu, M. Ying, B. K. Zhao, S. Zhang, L. M. Lin, L. N. Bai, and R. J. Wan, 2021: Western North Pacific tropical cyclone database created by the China Meteorological Administration. Adv. Atmos. Sci., 38(4), 690−699. doi: 10.1007/s00376-020-0211-7

Ng, K. K., Tse, K. S., Lui, Y. S., Leung, W. N., Cheung, C. C., & Suen, S. C. (2019). Using hierarchical time-stepping to utilize mpas-a computational resources for customized extreme variable-resolution meshes. In Joint wrf/mpas users' workshop 2019.

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, and Coauthors, 2008: THORPEX Interactive Grand Global Ensemble (TIGGE) Model Tropical Cyclone Track Data. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, doi: 10.5065/D6GH9GSZ.

Sridevi, Ch., et al. (2022). Tropical Cyclone Track and Intensity Prediction Skill of GFS Model over NIO during 2019 & 2020. Atmosphere, 13(4), 586. doi: 10.1016/j.tcrr.2022.04.002

Xiang, C.Y., Xu, Y.L., Gao, S.Z., Wang, Q., Wang H.P. (2022). 2021年西北太平洋台风活动特征和预报难点分析 [Analysis of the Characteristics and Forecast Difficulties of Typhoon over the Western North Pacific in 2021]. Meteorological Monthly, 48(9), 195-1508. doi: 10.7519/j.issn.1000-0526.2022.053001

Ying, M., W. Zhang, H. Yu, X. Lu, J. Feng, Y. Fan, Y. Zhu, and D. Chen, 2014: An overview of the China Meteorological Administration tropical cyclone database. J. Atmos. Oceanic Technol., 31, 287-301. doi: 10.1175/JTECH-D-12-00119.1

Wilcoxon, F. (1965). Individual comparisons by ranking methods. In A. A. van der Vaart (Ed.), Nonparametric statistics (pp. 1-10). New York: Wiley. doi: 10.1093/biomet/52.1-2.203

For more information, please contact:

Clustertech Limited

Units 211 - 213, Lakeside 1, No. 8 Science Park West Avenue, Hong Kong Science Park, N.T., Hong Kong

Tel: +852 2655 6100

Email: enquiry@cpas.earth

https://cpas.earth/