機器學習天氣預報模型

對於2023年西北太平洋颱風預報能力之探討

徐驊 劉正欽 彭順台 陳得松 張保亮 蕭玲鳳 馮欽賜 洪景山 程家平 呂國臣 陳嘉榮 郭鴻基
中央氣象署 (CWA)
University of Colorado Colorado Springs
國立台灣大學

天氣分析與預報研討會@CWA 2024/9/4

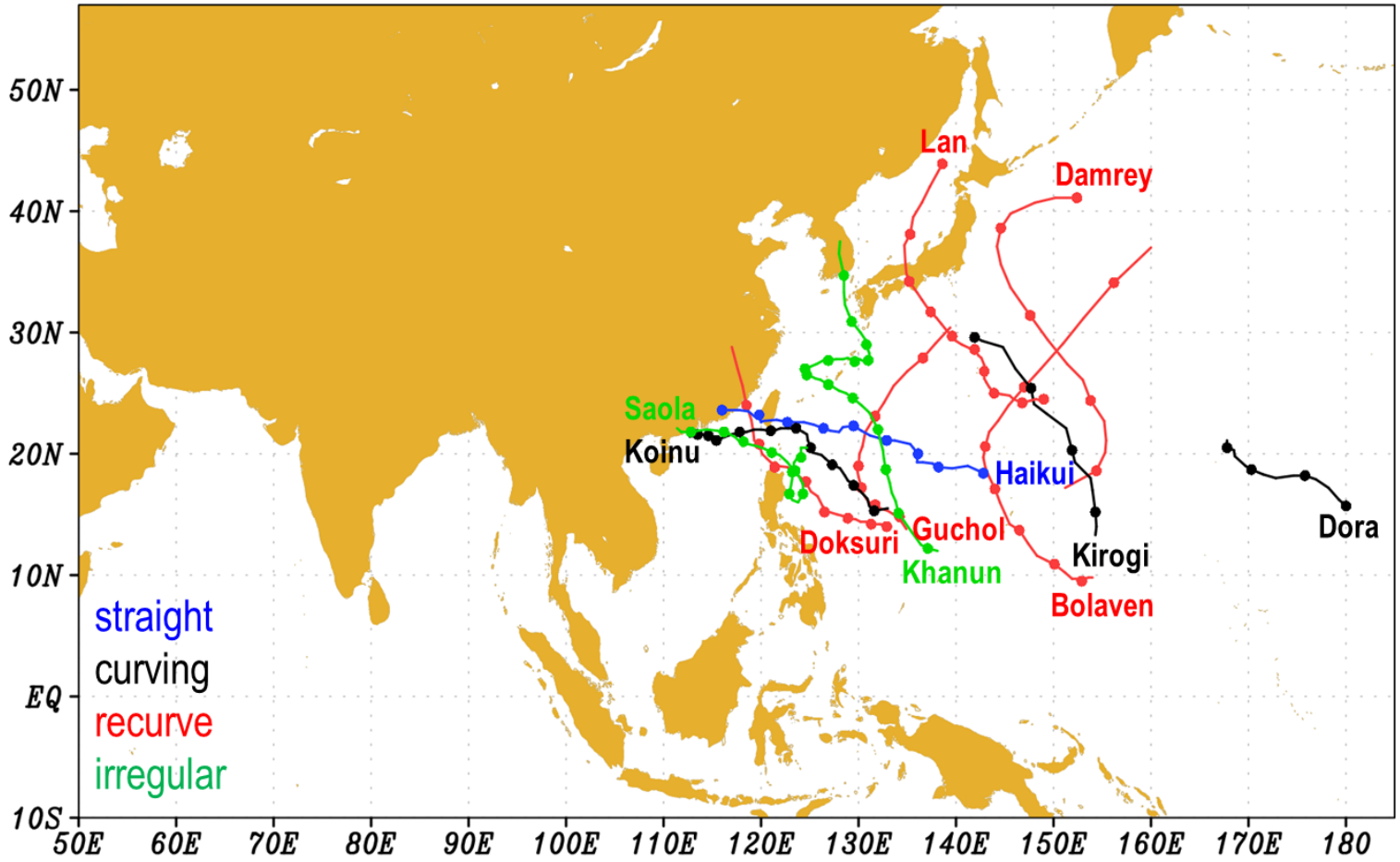
Outline

1. Purpose: Evaluate 5 MLWP (machine learning-based weather prediction) models to see if they have the potential to **provide better initial and boundary conditions for regional models**.
2. Evaluate track and intensity errors of **11 typhoons in the western North Pacific** between Jun. to Nov. 2023
3. Detailed evaluation of **Typhoon Haikui (2023)** during its passage through Taiwan

Typhoons for Evaluation

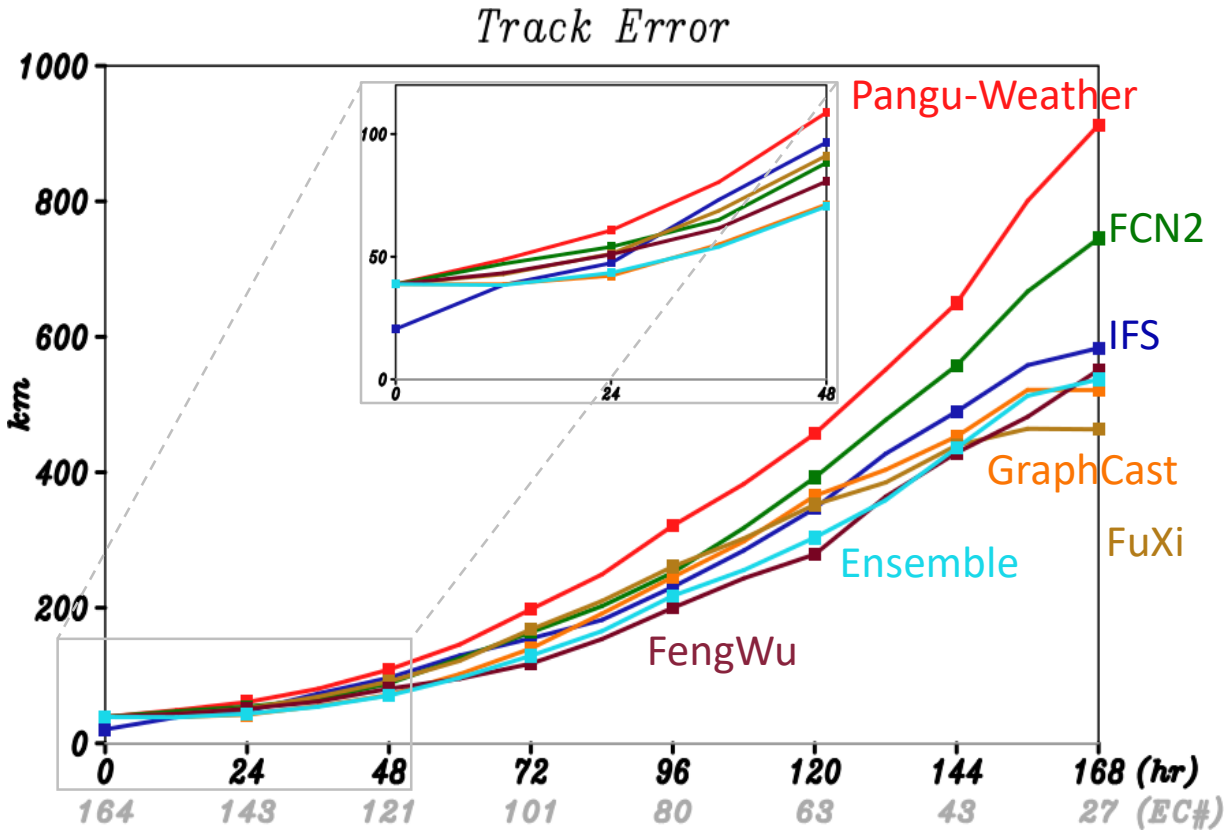
#	Typhoon	t_0 (UTC)/12h	Cases
1	GUCHOL	0606 12~0612 12	13
2	DOKSURI	0721 12~0728 00	14
3	KHANUN	0728 00~0810 12	28
4	LAN	0808 00~0817 00	19
5	DORA	0812 00~0815 00	7
6	SAOLA	0824 12~0902 12	19
7	DAMREY	0825 00~0928 12	8
8	HAIKUI	0828 12~0904 12	15
9	KIROGI	0830 12~0903 00	8
10	KOINU	0930 00~1009 00	19
11	BOLAVEN	1007 12~1014 00	14
	Total		164

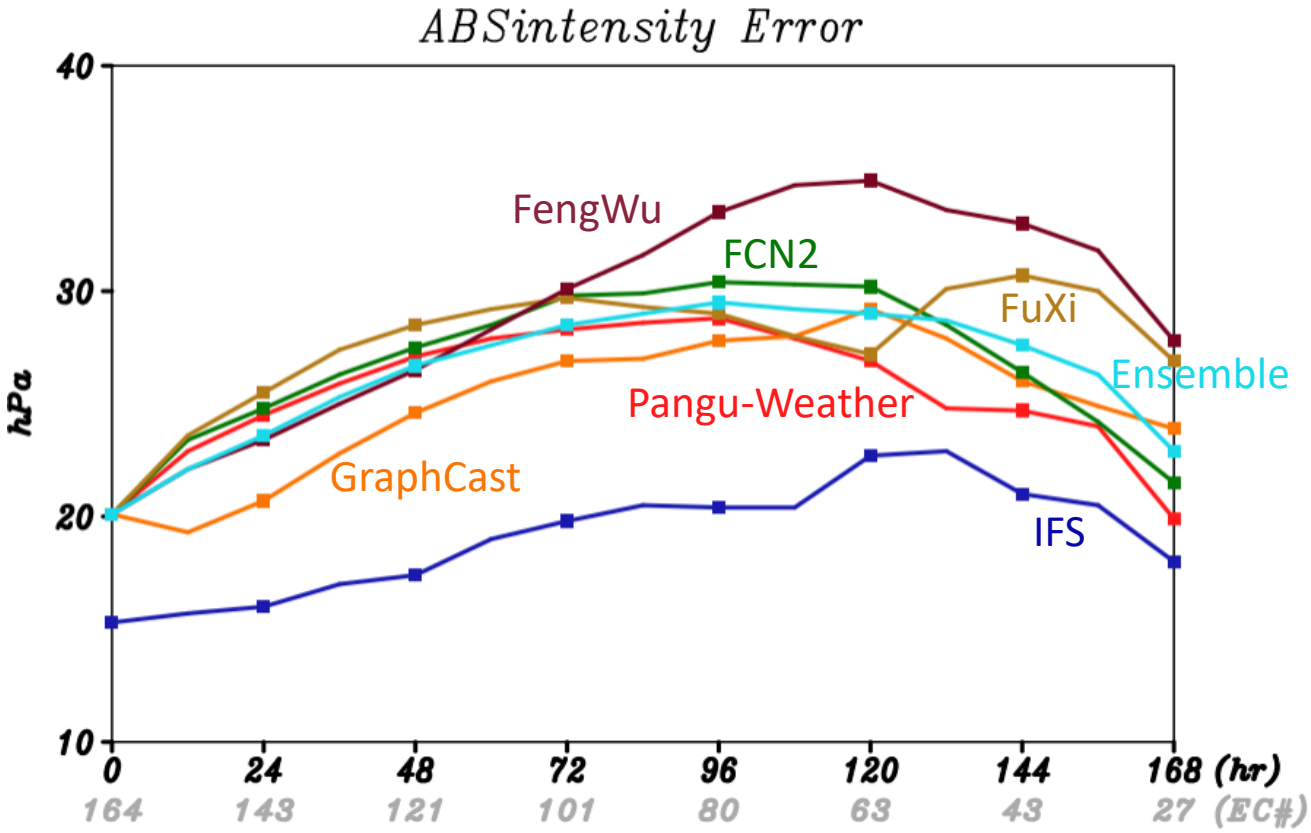
(typhoons for which Taiwan has issued warnings)



Domain for evaluating 5 MLWP models

Averaged Track and Absolute Intensity Errors

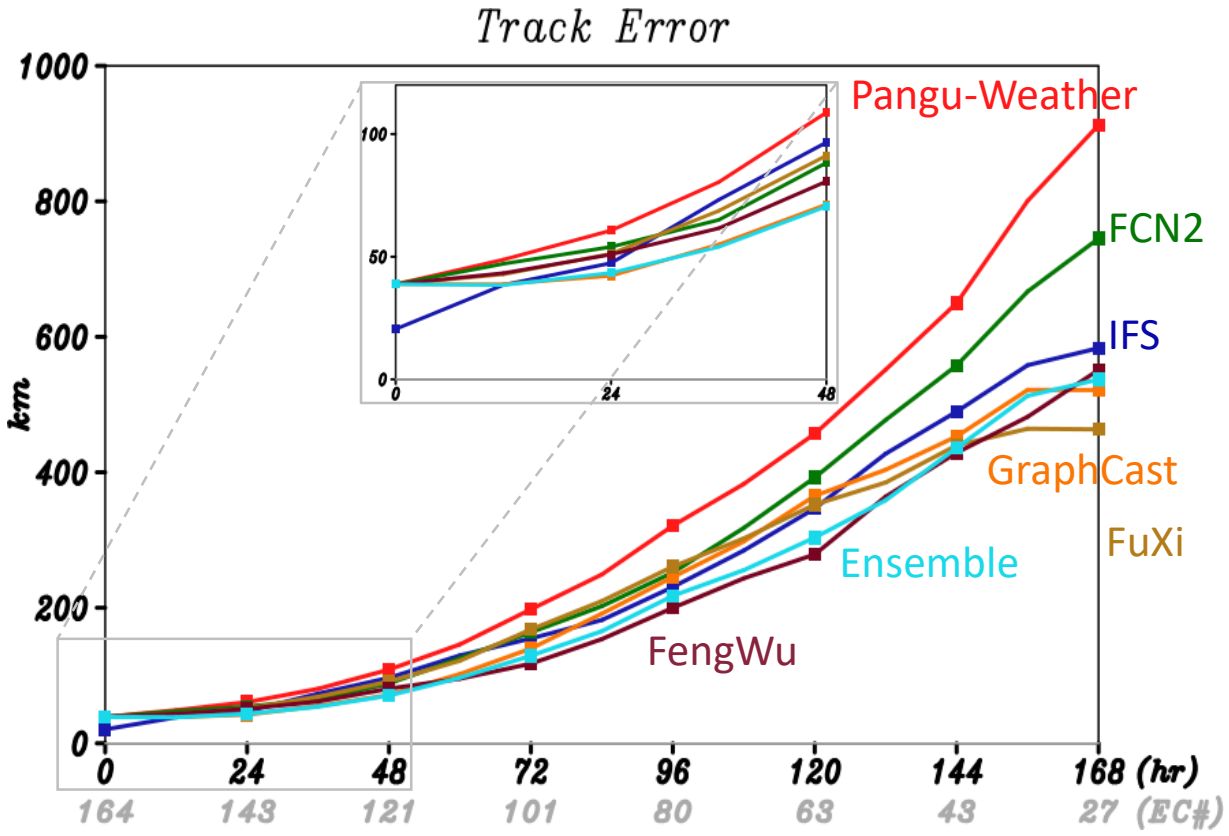


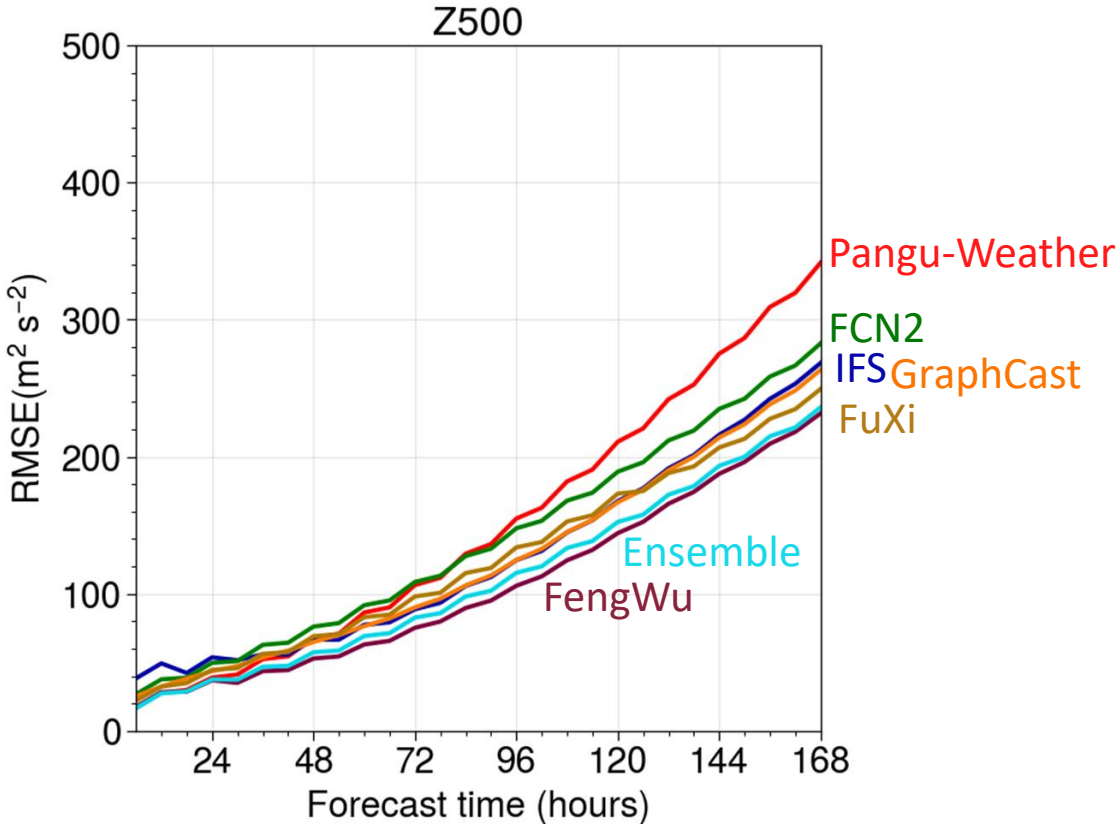


1. **Pangu-Weather** exhibited the **largest track error**
2. **FengWu** performed the **smaller track error** up to 144-h
3. IFS roughly in the middle of the group with smallest initial condition
4. The ensemble performance was very close to FengWu

1. **GraphCast** and **Pangu-Weather** tied for the **lowest intensity errors**
2. **FengWu** performed the **largest intensity error**

Averaged Track and Absolute Intensity Errors

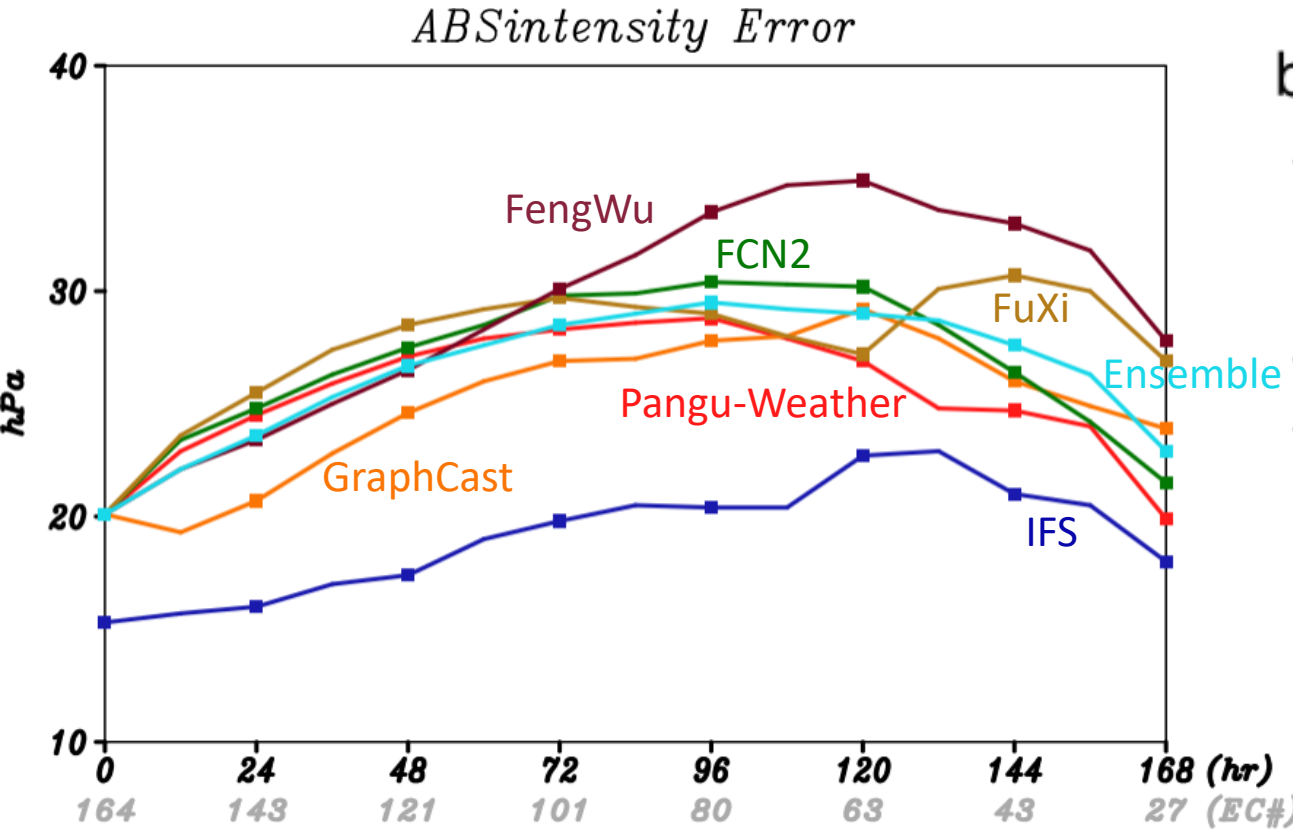


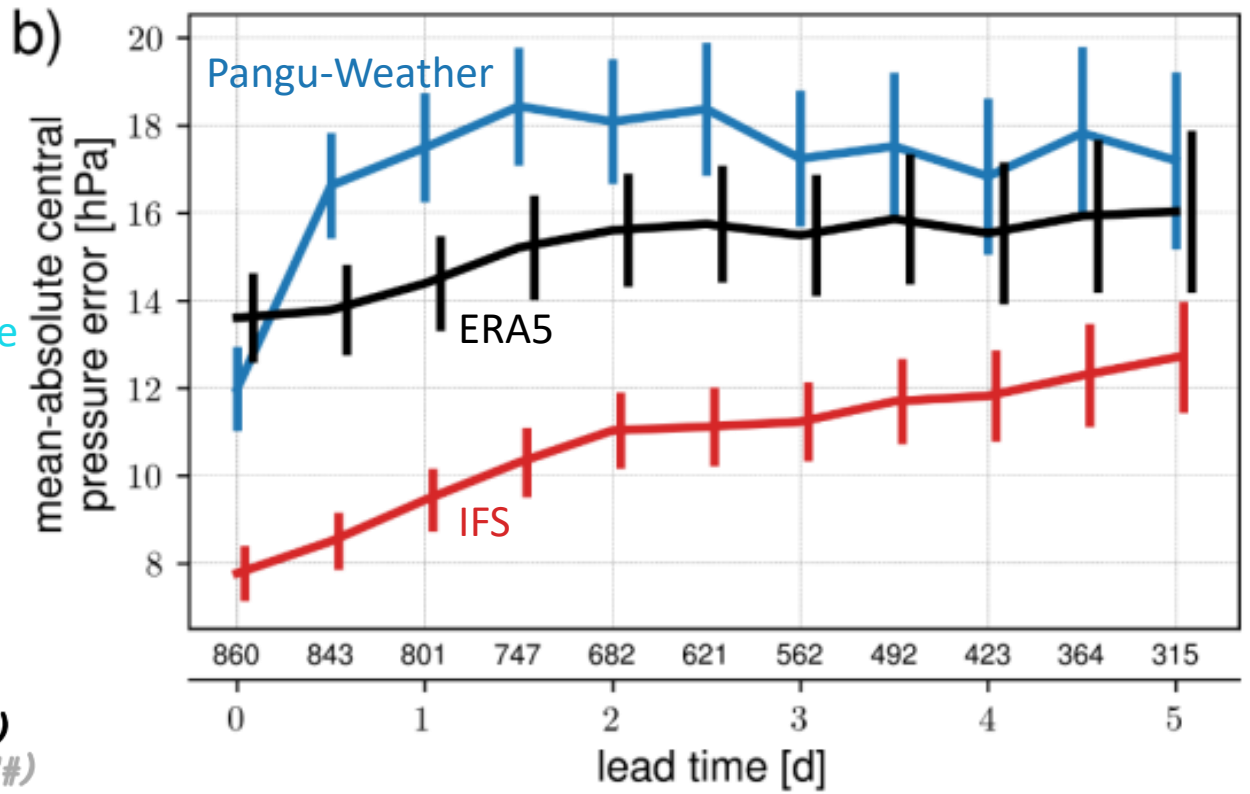


Track prediction performance aligns with the ranking by RMSE

First order of steering typhoon is the large-scale flow and mid-tropospheric levels

Averaged Track and Absolute Intensity Errors

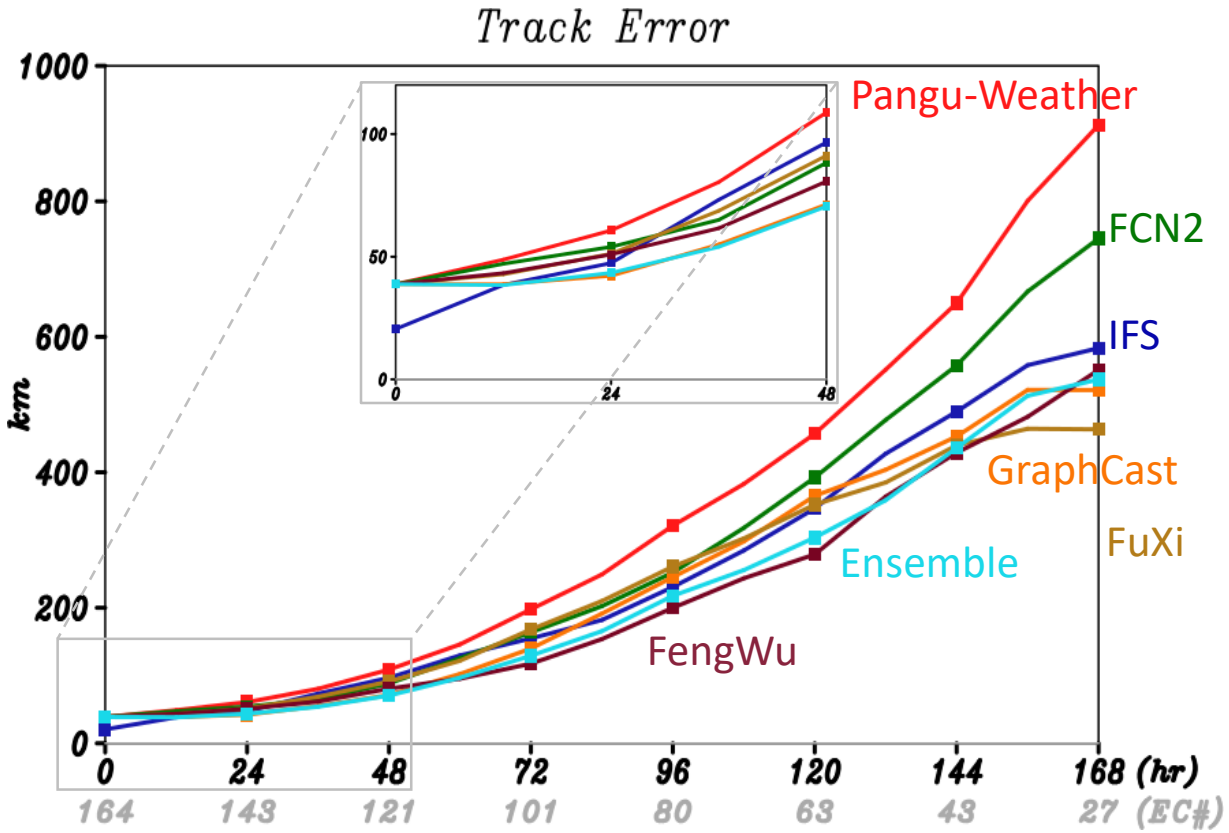


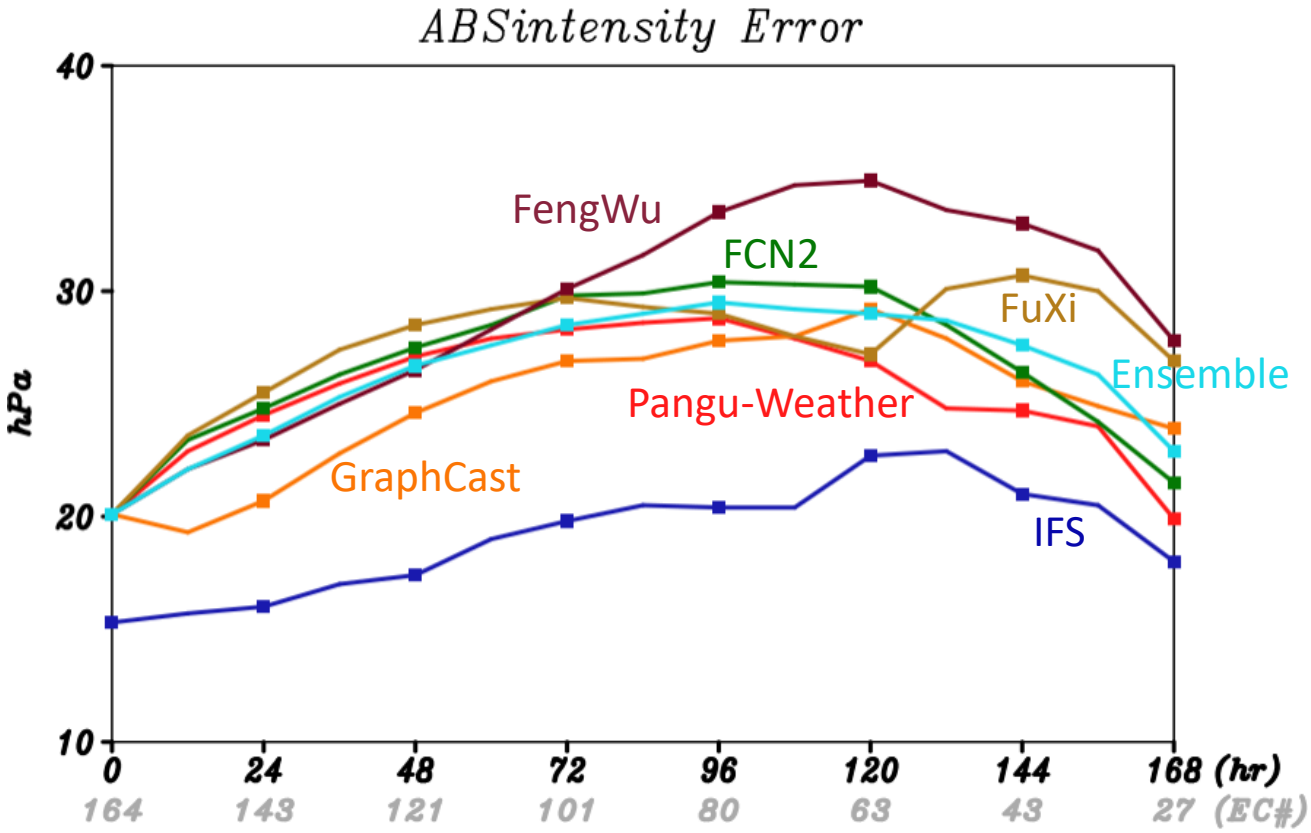


2018 global TC (Bouallègue et al. 2023, Fig. 8)

The better performance of IFS compared with ERA5 is mainly explained by its higher resolution (9 km vs 28 km)

Averaged Track and Absolute Intensity Errors





1. **Pangu-Weather** exhibited the **largest track error**
2. **FengWu** performed the **smaller track error** up to 144-h
3. IFS roughly in the middle of the group with smallest initial condition
4. The ensemble performance was very close to FengWu

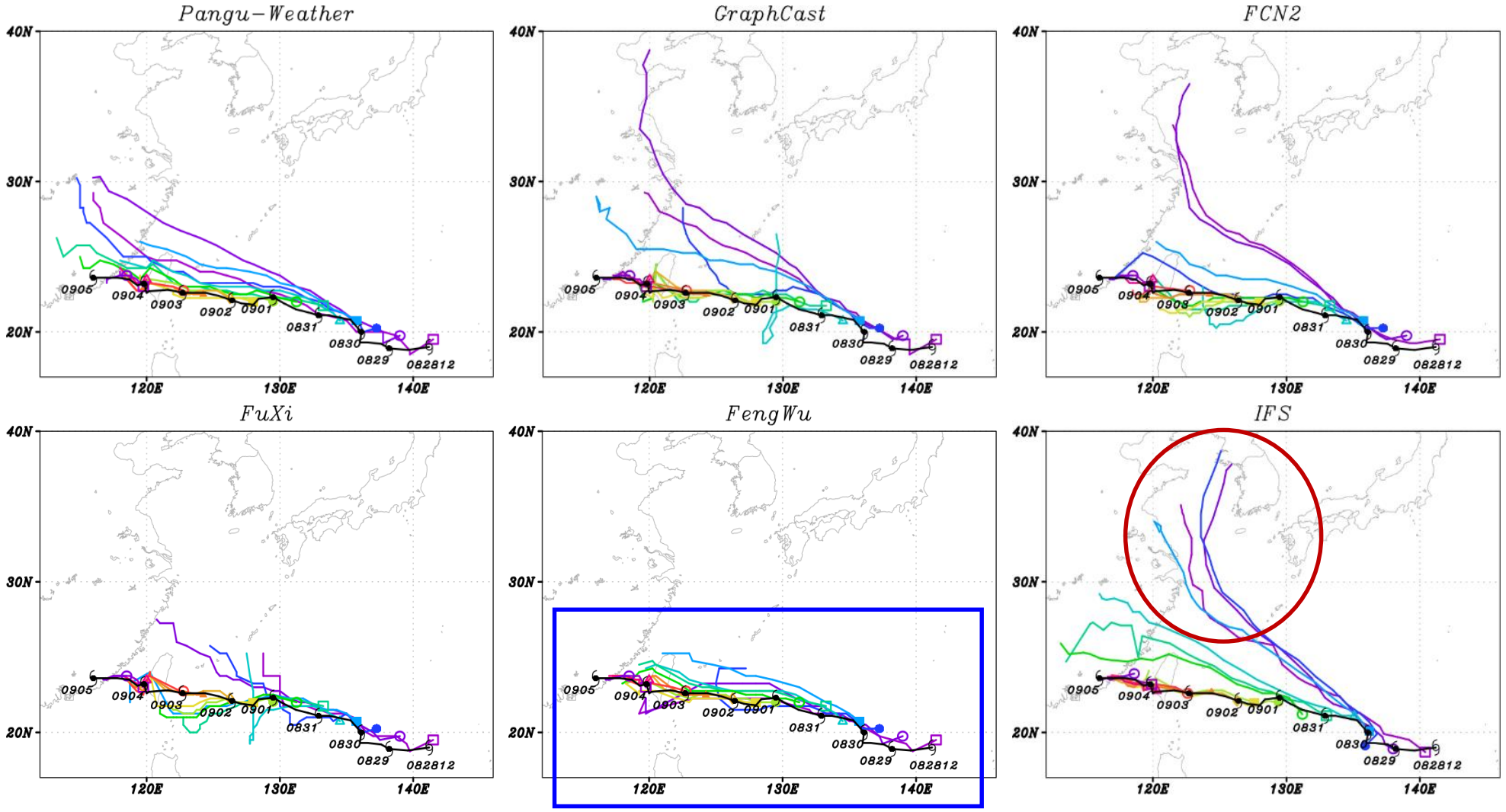
1. **GraphCast** and **Pangu-Weather** tied for the **lowest intensity errors**
2. **FengWu** performed the **largest intensity error**

96-h Best/Worst Track and Intensity Forecast

Storm name	No of cases	Track types	Storm maximum intensity (hPa)	Best track performer @ 96 h (km)	Worst track performer @ 96 h (km)	Best intensity performer @ 96 h (hPa)	Worst intensity performer @ 96 h (hPa)	Average track error from all models @ 96 h (km/No of cases)	Average intensity error from all models @ 96 h (hPa/No of cases)
Haikui	15	straight	945	FW (41.3)	IFS (1012.5)	FuXi (5.2)	P-W (46.6)	442.4/48	26.0/42
Dora @ 72 h	7	curving	975	FW (63.7)	P-W (195.7)	P-W (16.0)	FW (16.8)	111.0/3	14.9/3
Kirogi @ 72 h	8	curving	985	GC (19.3)	P-W (458.9)	P-W (6.5)	FuXi (9.7)	287.7/8	7.2/8
Koinu	19	curving	930	P-W (49.9)	FW (441.8)	FuXi (27.7)	GC (63.3)	179.1/61	44.6/55
Guchol	13	recurve	960	FW (22.3)	IFS (533.8)	P-W (0.4)	FW (23.0)	162.4/30	8.7/30
Doksuri	14	recurve	935	FuXi (0.0)	FW (642.4)	GC (0.2)	FuXi (50.8)	158.9/42	32.2/42
Lan	19	recurve	940	IFS (24.3)	FuXi (599.3)	P-W (0.1)	FCN2 (41.6)	268.4/66	14.9/66
Damrey	8	recurve	985	IFS (102.5)	FuXi (447.5)	GC (10.9)	FW (20.0)	303.6/6	14.1/6
Bolaven	14	recurve	900	P-W (38.0)	IFS (575.8)	GC (1.1)	FuXi (93.7)	222.8/36	53.6/36
Khanun	28	irregular	930	FW (19.4)	P-W (760.1)	FuXi (0.2)	FW (60.5)	256.2/120	15.5/120
Saola	19	irregular	915	P-W (43.2)	GC (618.8)	FCN2 (0.4)	FW (79.3)	263.6/66	46.8/66
Averages	------	----------	945.5	(38.5)	(571.5)	(2.9)	(45.9)	251.1/486	27.9/474

1. **FengWu** has the highest number of **best track** performances (4), but **none for best intensity** performance
2. **Pengu-Weather** with the **worst average track** error, ranks **second in best track** performances and has **4 best intensity** performances

Typhoon Case Study: Haikui (0828 12Z~0904 12Z)
168-h Track Forecasts



1. **IFS** exhibited the **largest average track error** mainly from the early stage

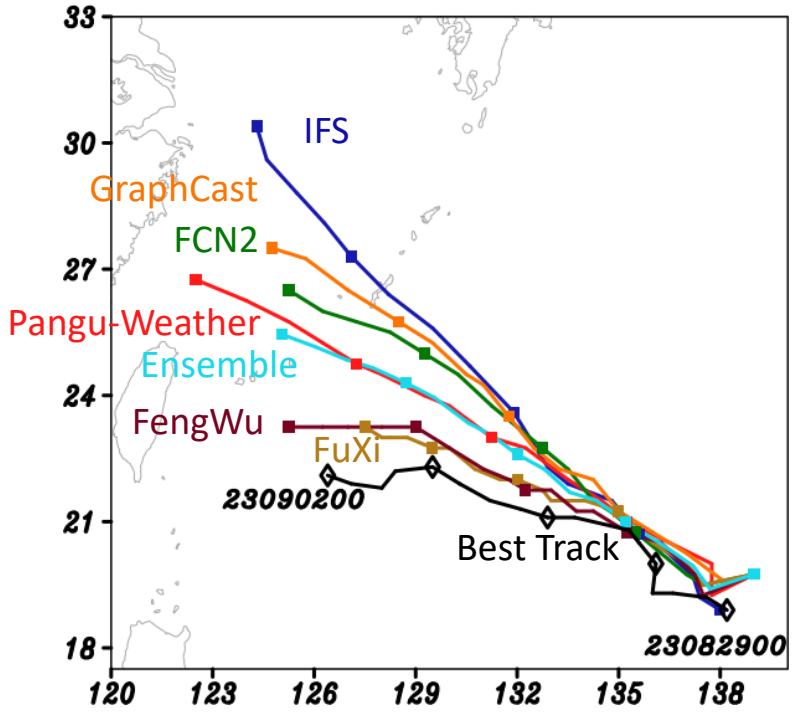
2. **FengWu** predicted tracks that **closely aligned with the best track**

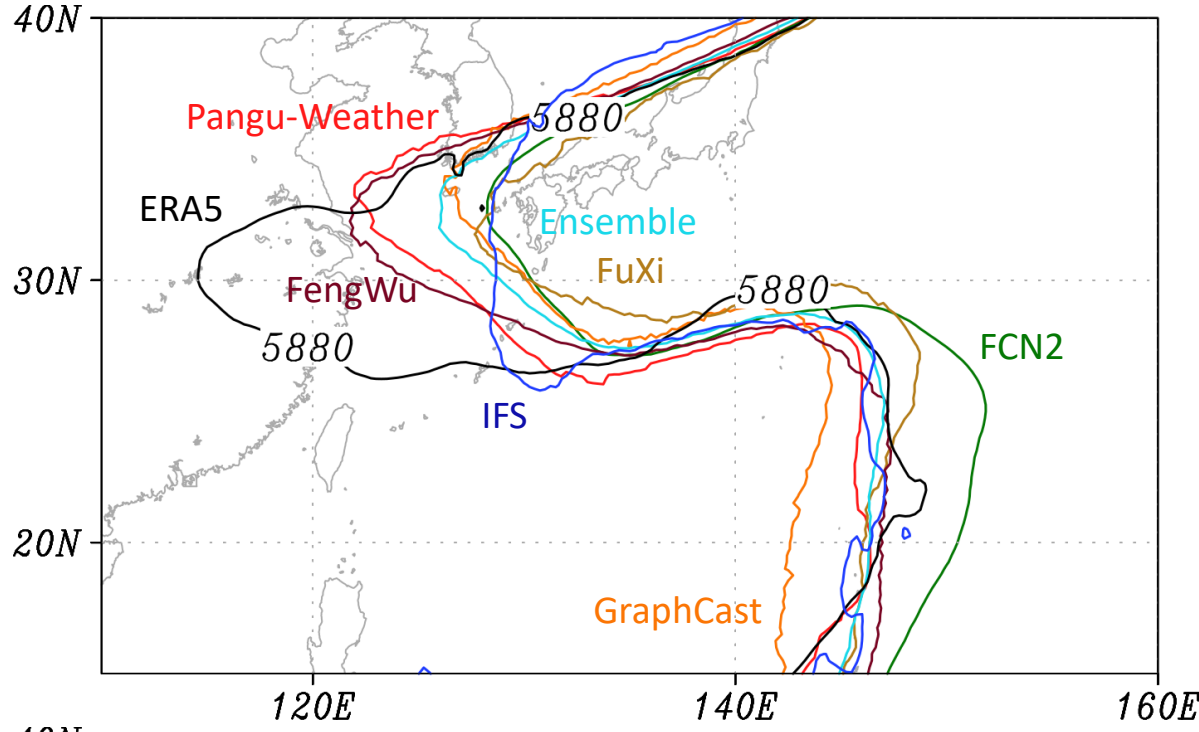
Typhoon Haikui
96-h Track Forecasts

1. The distribution of the **western edges** of the western Pacific subtropical high (WPSH) generally **aligns with their individual tracks**

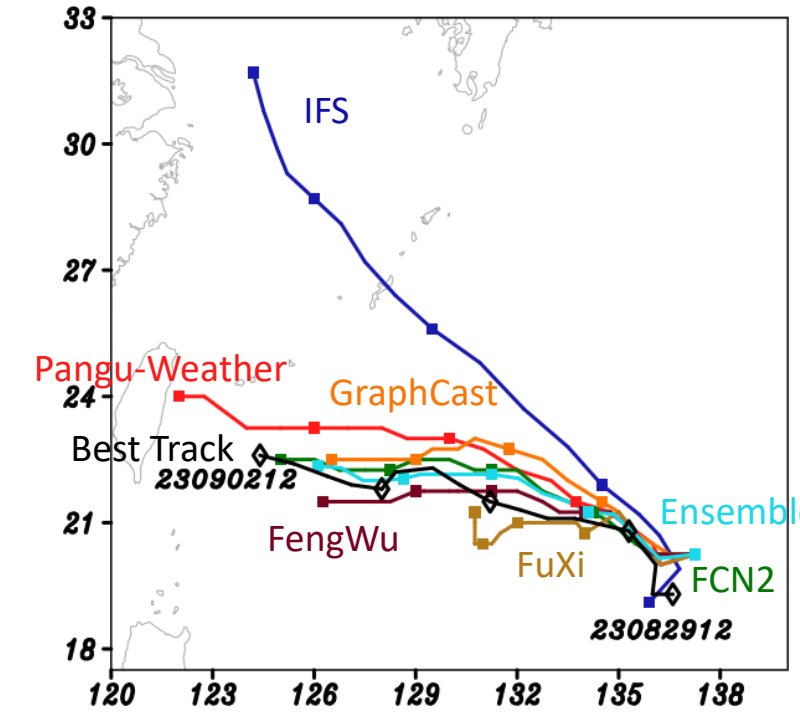
2. The changes in MLWP forecast **TC tracks corresponding to** the changes in **WPSH** demonstrate the **reasonable forecast result** of MLWP track forecast

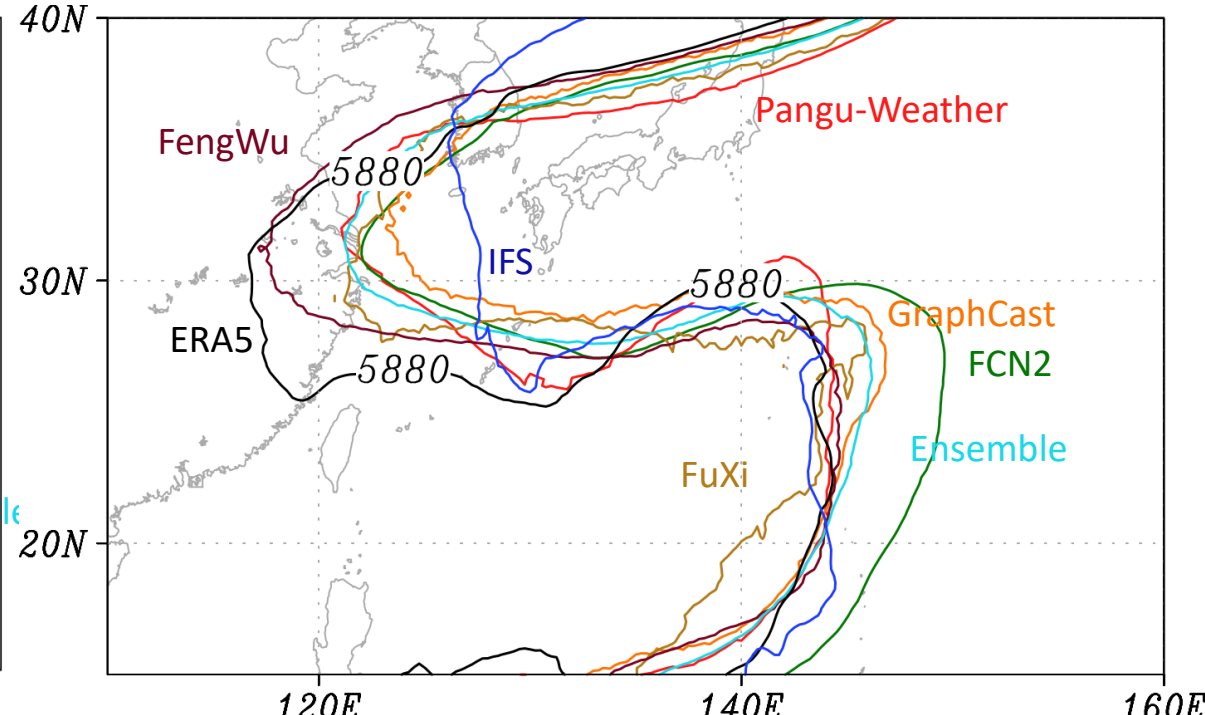
Initial Time: 23082900 UTC





Initial Time: 23082912 UTC





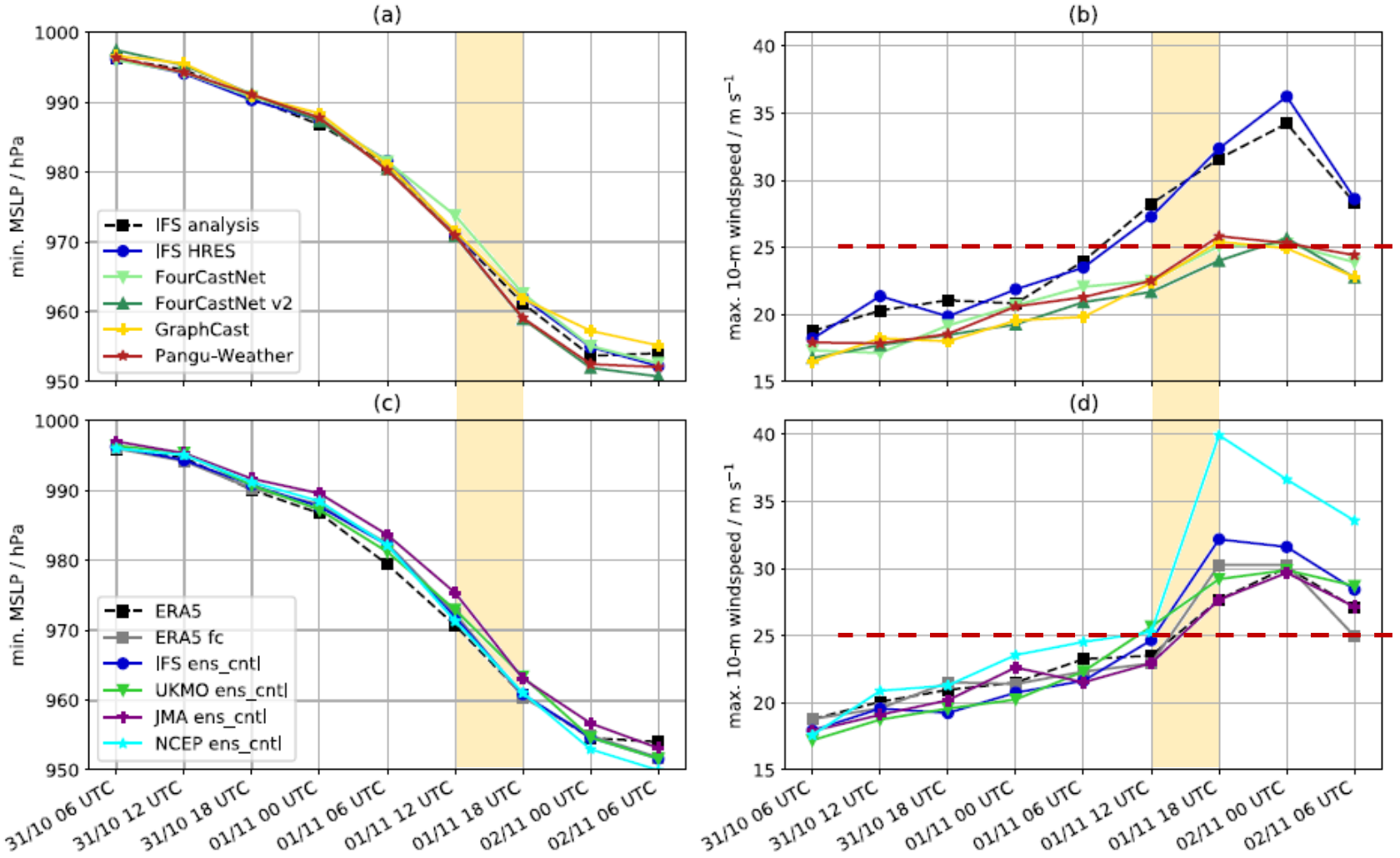
Min Sea Level Pressure v.s. Max Wind Speed

1. The forecasted **min sea level pressure are all closely follow analysis**, capture the rapid deepening phase.

2. In contrast, **the spread in the max wind speed** evolution is **far greater.**

3. **MLWP models failed** to capture the **rapid intensification of the wind** (too weak).

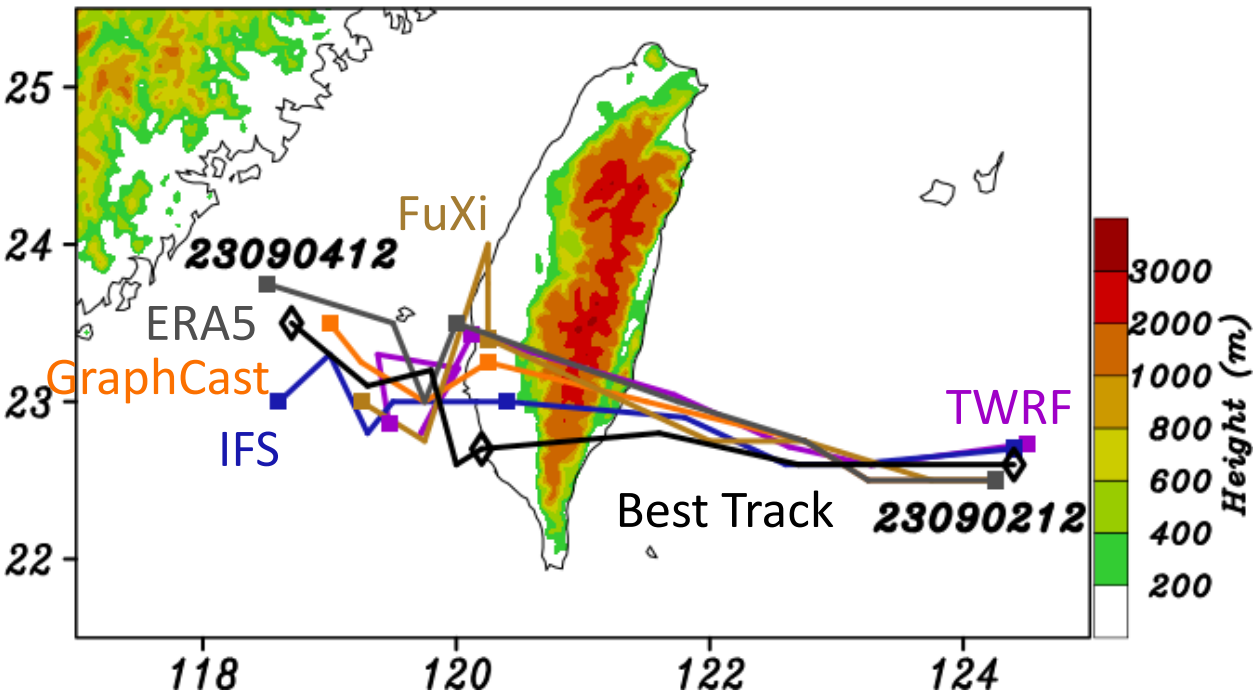
Storm Ciarán (2023)

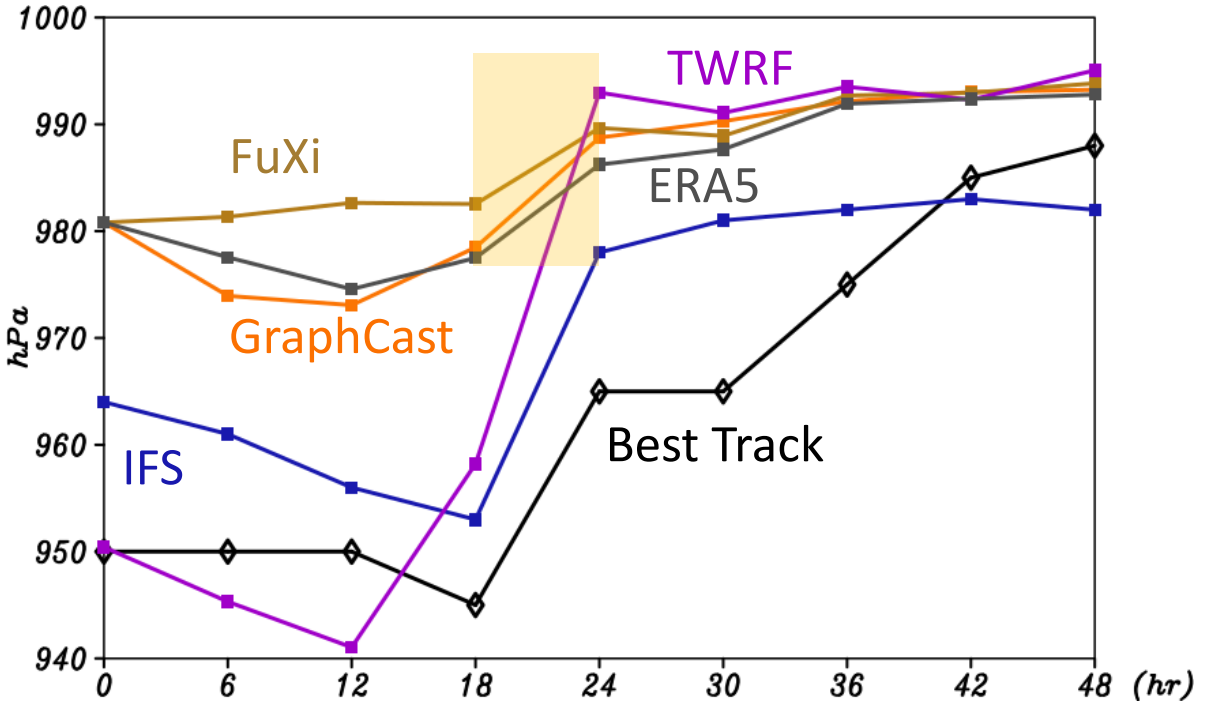


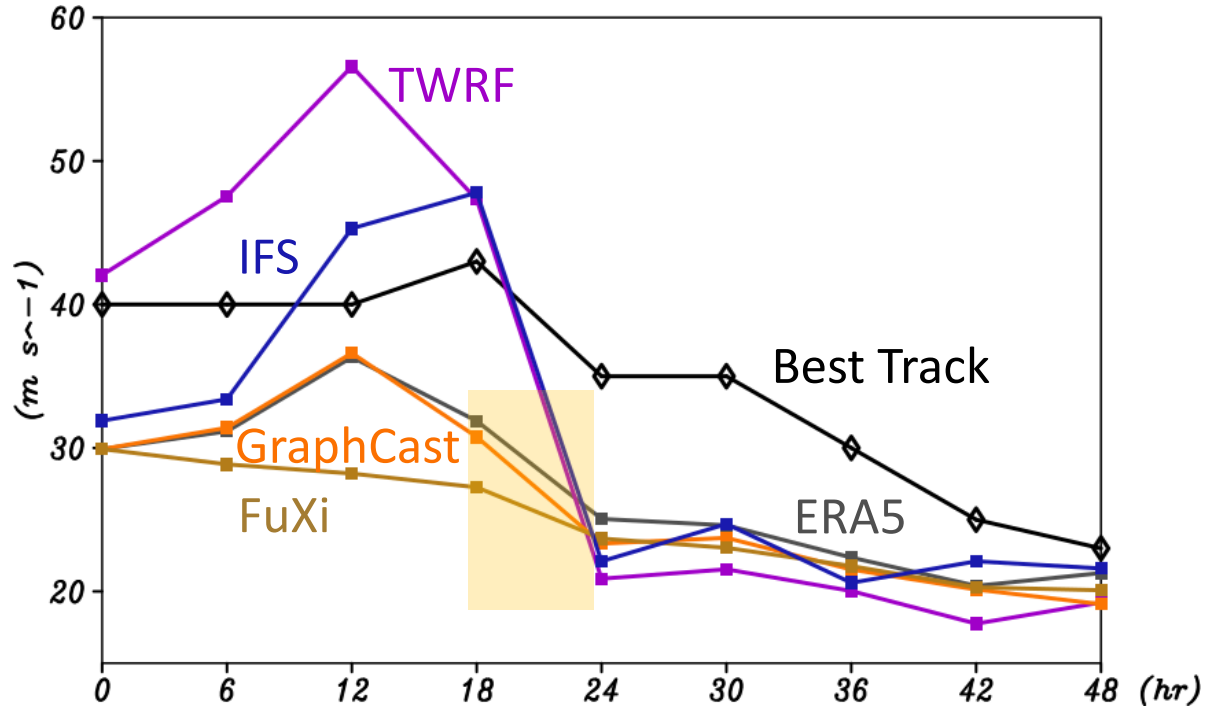
(Charlton-Perez, 2024)

Typhoon Haikui
48-h Forecasts Track & Intensity

There is strong relationship between the wind and pressure profile in this case

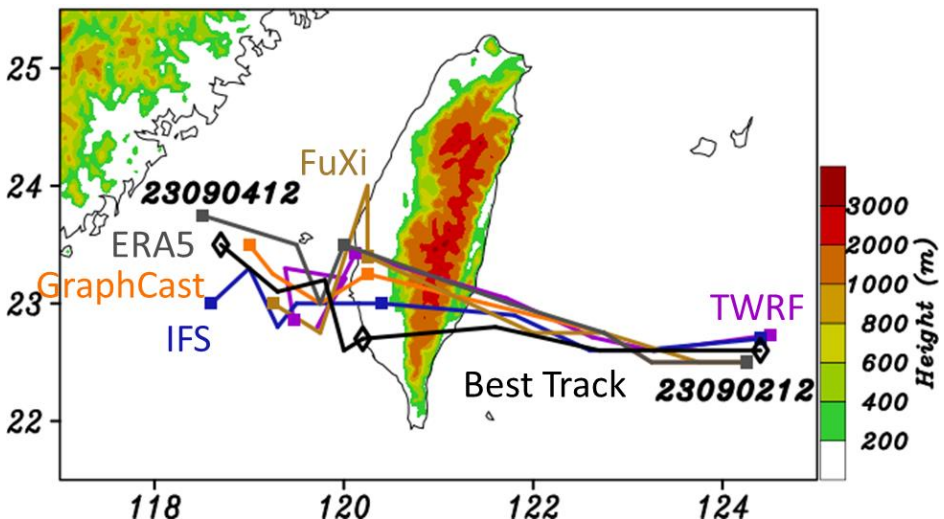


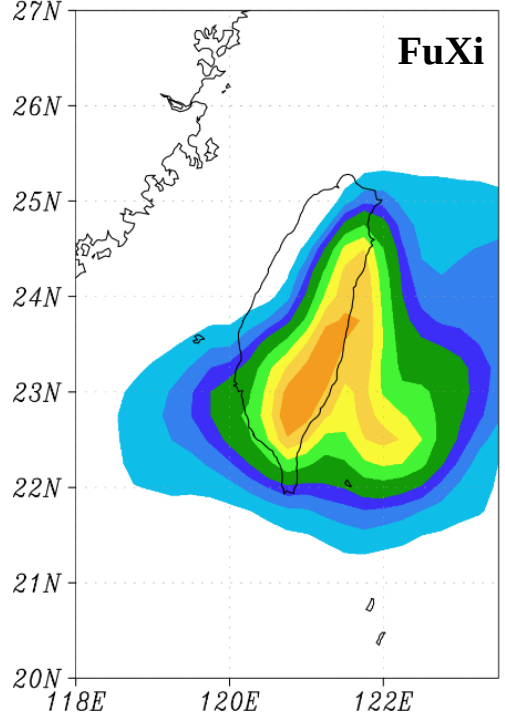


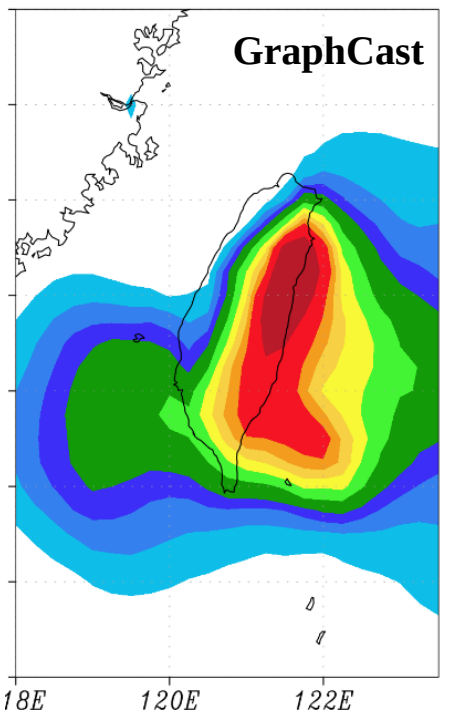


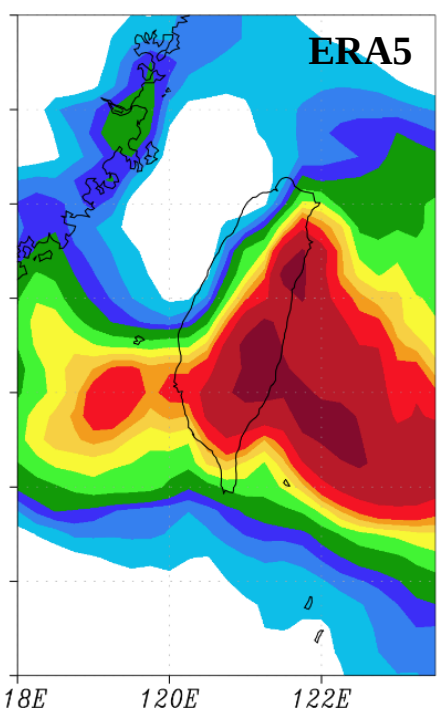
Typhoon Haikui
48-h Accumulated Rainfall

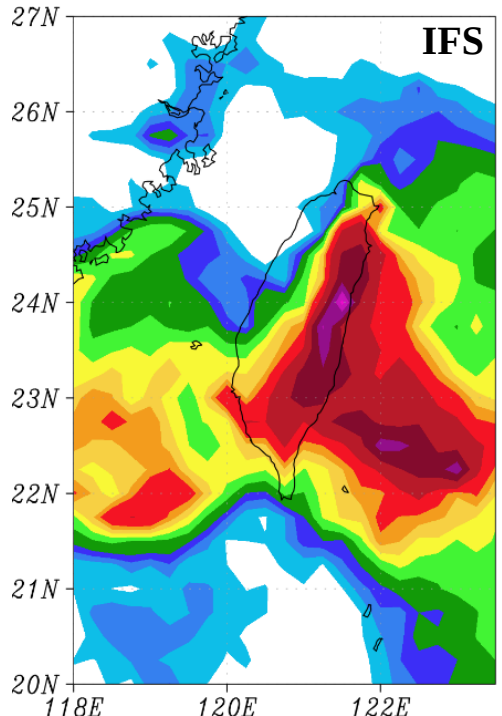
1. Only **TWRF (3 km) reaches** rainfall over **700 mm**
2. **ERA5 and IFS shows** good rainfall **pattern with less detail**
3. **FuXi and GraphCast** show **smaller** amount.
4. Despite a small TC track difference, **FuXi's weaker rainfall** may **be from weaker TC intensity** during the first 24 hours

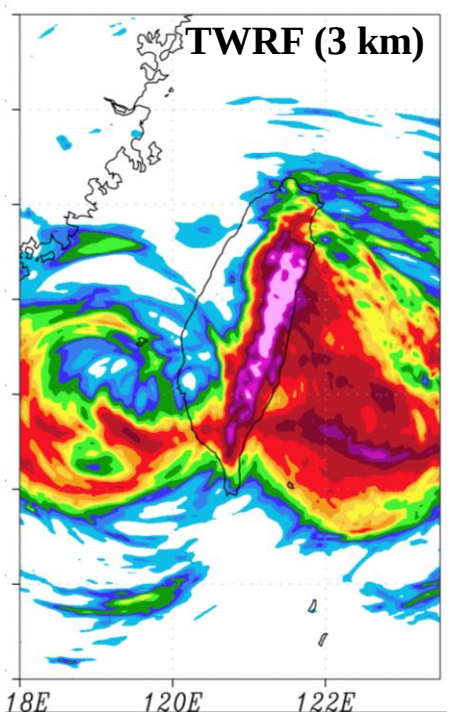


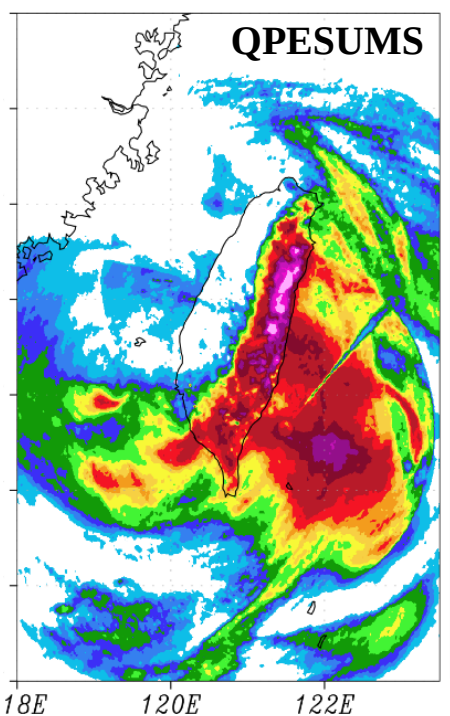












Summary

1. Evaluate track and intensity errors of **11 typhoons** in the western North Pacific between Jun. to Nov. 2023.

   a) **FengWu** demonstrated the **best averaged track** prediction, led in **4 individual typhoons**, and **poorest intensity prediction**. While **Pangu-Weather** has the **largest averaged track error**, it performed best for **3 individual typhoons**.

   b) A **multi-model ensemble** is the reduction in the error range. It consistently stays within the error range, preventing outliers from individual models. Our result shows the importance of ensemble for MLWP models predicting individual typhoon.

Summary

2. CWA uses the advantages of the **MLWP ensemble** in **the ETQPF system** to provide **improved rainfall amount** and distribution
3. Detailed evaluation of Typhoon Haikui (2023) during its passage through Taiwan

   a) The 168-hour prediction from 5 MLWP models shows a **reasonable relationship between WPSH variation and track** prediction

   b) MLWP models **successfully demonstrate the impact of terrain on rainfall** prediction
4. We will test using the **MLWP** model predictions as **initial and boundary conditions** for the regional model, incorporating the initialization of the TC structure

TC Formation
HAIKUI

Model	The earliest Initial time of the forecast predicting TC formation (UTC)	Predicted days before TC formation (8/28 00 UTC) (days)
IFS	8/23 12	4.5
Pangu-Weather	8/23 00	5
FCN2	8/22 00	6
GraphCast	8/23 12	4.5
FuXi	8/24 00	4
FengWu	8/24 00	4

Since the typhoon track and intensity errors do not follow a normal distribution, we adopted the Mann-Whitney U test (Mann and Whitney 1947; Wilcoxon 1945) to assess statistical significance.

The results indicate a 95% confidence level in the comparison of track errors between the selected model (PW or FW) and other MLWP models. Additionally, there is a 90% confidence level in the difference in TC intensity errors between FW and the other MLWP models.

PW v.s. other MLWP **Intensity** Error @96 h

Model	P Value
FCN2	4.10E-01
GraphCast	4.39E-01
FuXi	2.85E-01
FengWu	4.54E-02 <= 0.1
Ensemble	3.85E-02 <= 0.1

PW v.s. other MLWP **Track** Error @96 h

Model	P Value
FCN2	5.86E-03
GraphCast	1.11E-03
FuXi	1.40E-02
FengWu	5.37E-08
Ensemble	2.74E-04

FW v.s. other MLWP **Intensity** Error @96 h

Model	P Value
FCN2	9.01E-02 <= 0.1
GraphCast	4.66E-02 <= 0.1
FuXi	8.19E-02 <= 0.1
Ensemble	4.45E-01

FW v.s. other MLWP **Track** Error @96 h

Model	P Value
FCN2	2.84E-03
GraphCast	1.46E-03
FuXi	4.00E-05
Ensemble	2.07E-03

Mann-Whitney U Test

A Mann-Whitney U test (Wilcoxon rank-sum test) is used to compare the differences between two samples when the sample **distributions are not normally distributed** and the sample sizes are small (n <30).

It is considered to be the **nonparametric equivalent to the two sample t-test**.

$$U_1 = n_1\, n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1\, n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$U = \min(U_1, U_2)$$

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Then put z to normal distribution to get the p-value

P Value

IFS v.s. MLWP **Intensity** Error @72 h

Model	P Value
Pangu-Weather	2.59E-04
FCN2	2.95E-05
GraphCast	1.05E-04
FuXi	5.48E-06
FengWu	3.02E-06
Ensemble	1.83E-07

IFS v.s. MLWP **Intensity** Error @96 h

Model	P Value
Pangu-Weather	4.32E-04
FCN2	1.92E-04
GraphCast	4.69E-04
FuXi	1.02E-04
FengWu	6.36E-07
Ensemble	3.22E-07

All intensity errors pass 0.05 significance level

IFS v.s. MLWP **Track** Error @72 h

Model	P Value
Pangu-Weather	1.89E-05
FCN2	3.68E-02
GraphCast	**3.80E-01**
FuXi	3.49E-03
FengWu	**1.16E-01**
Ensemble	**3.09E-01**

IFS v.s. MLWP **Track** Error @96 h

Model	P Value
Pangu-Weather	9.35E-05 <= 0.05
FCN2	9.52E-02
GraphCast	8.89E-02
FuXi	1.12E-02 <= 0.05
FengWu	6.15E-02
Ensemble	**1.24E-01**

Some track errors are less than 90% confidence

Typhoon Haikui
48-h Forecasts Track & Intensity

There is strong relationship between the wind and pressure profile in this case