Waves Workshop 2025

Santander, Spain

Combining synthetic and observed data to improve deep learning model performance for severe storm surge forecasting

Yue Qin[1], Jinghan Dong[2], Dongdong Chu[3], Jicai Zhang[2], Zhumei Che[4]

Introduction

Accurate prediction of extreme storm surges is critical for coastal disaster prevention and mitigation, yet remains challenging for data-driven models due to limited observational data of severe events in China and profound sensitivity to data uncertainty. This study reduces neural networks-based (NNs-based) severe storm surge prediction errors through three approaches: wind speed data correction, model selection, and strategic data augmentation via numerical models.

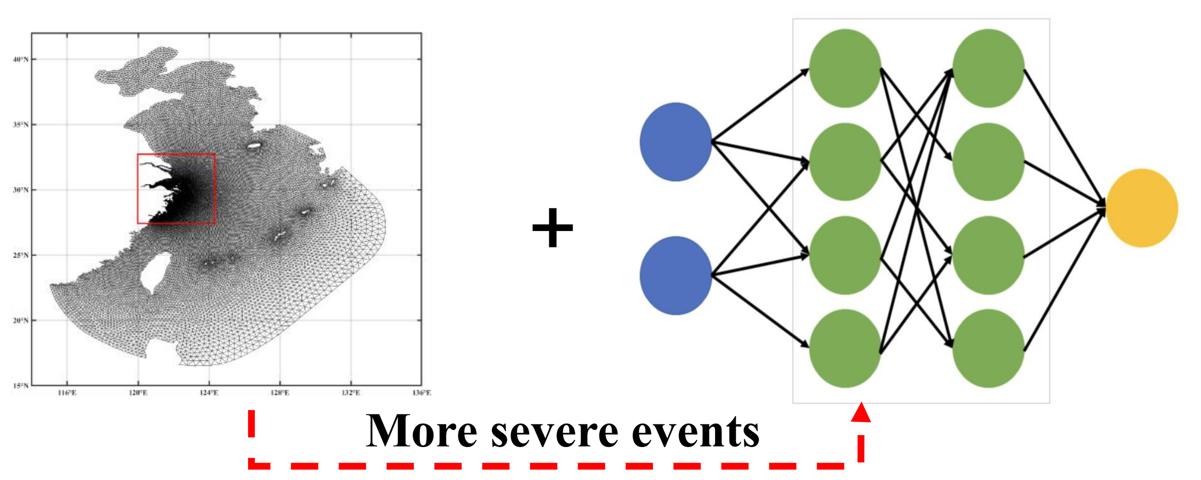


Fig. 2 Building a database for neural network training from an FVCOM model.

Data & Models

- Tidal gauges：Dajishan, Tanxvdao, Dinghai, Zhenhai (Sea), Shacheng, Sansha (Jun. to Oct., 1950~2000)
- Typhoon data：IBTrACS **(Wind speed data before 1971 were corrected)**
- Reanalysis data：ERA5 (1950~1978), NCEP-CFSR (1979~2000)

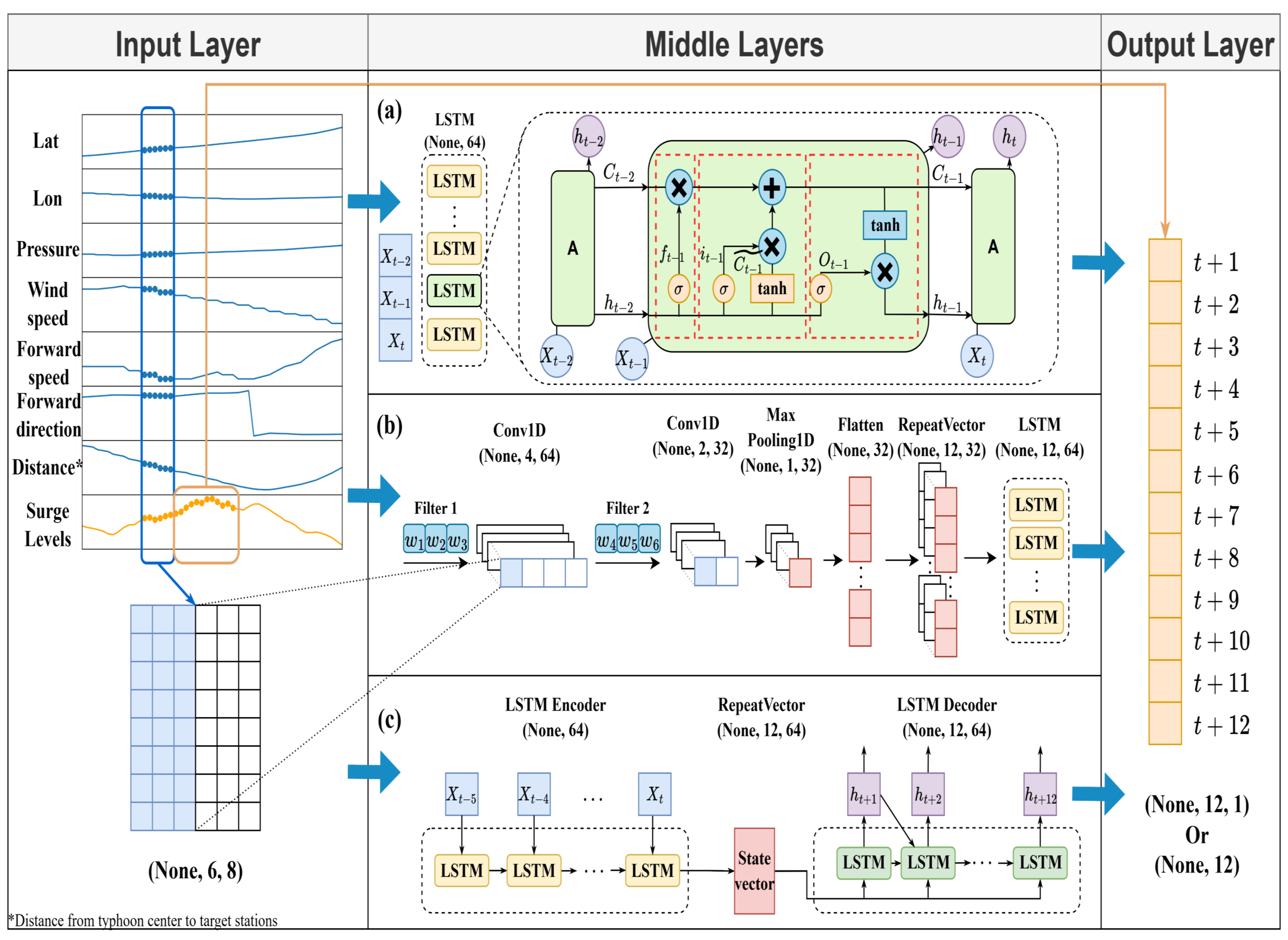


Fig. 3 Multivariate multi-step storm surge prediction framework based on deep learning models. (a) LSTM Dense model; (b) CNN-LSTM model; (c) Encoder-Decoder LSTM (E-D LSTM) model.

Fig. 1 Typhoon 9711 (Winnie) is approaching the Zhejiang coast. The stations, from top to bottom, are Dajishan, Tanxvdao, Dinghai, Zhenhai (Sea), Shacheng, and Sansha.

Model Comparisons based on Observational Data

- **Dataset:** Dinghai (22 typhoons)
- **Evaluation:** 5-fold cross validation
- **Metrics:** Correlation coefficient (CC), Root mean square error (RMSE), Peak error (PE)
- **Conclusions:**

(1) Multi-input Multi-output (MIMO) ANN performs best from $t+1$ to $t+3$

(2) E-D LSTM performs best at longer lead times ($t+6$ to $t+12$)

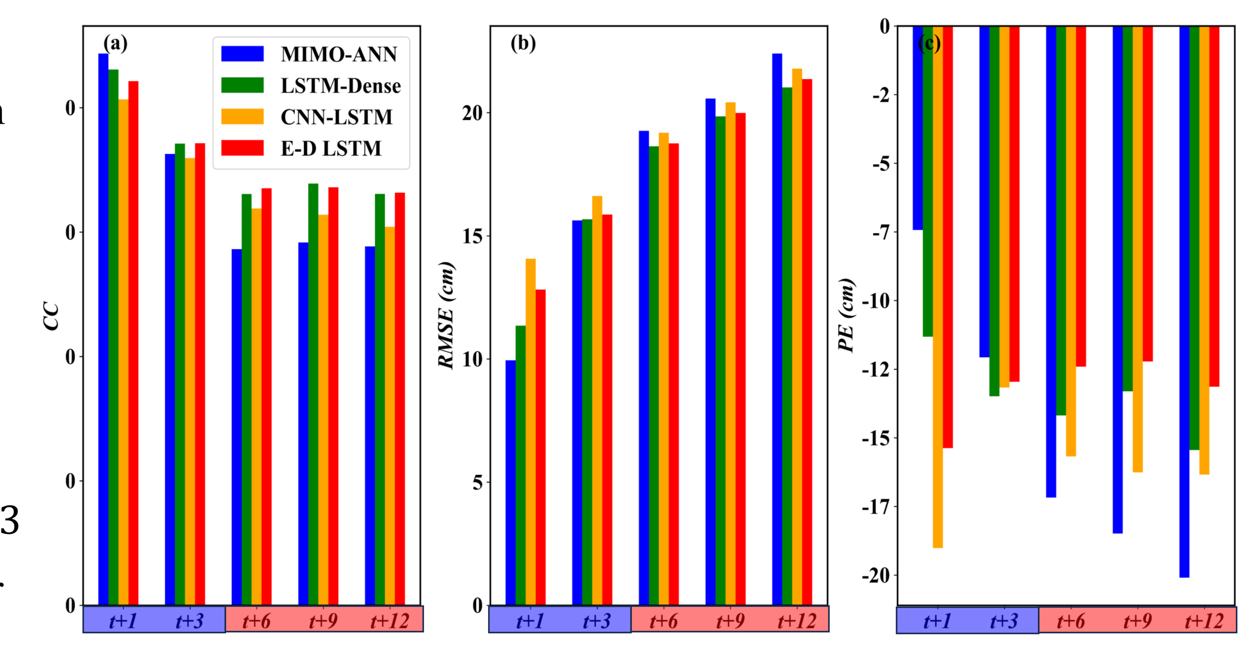


Fig. 7 Comparisons of model performance trained with observations at Dinghai station.

Incorporating FVCOM Synthetic Data

- **Model:** E-D LSTM
- **Experimental setup:** The training sets are divided into five groups: Obs. data only, Obs. data combined with any one synthetic dataset, and Obs. data combined with all synthetic typhoon datasets.
- **Testing dataset:** Landfalling typhoon (9711) and Non-landfalling typhoon (8114)

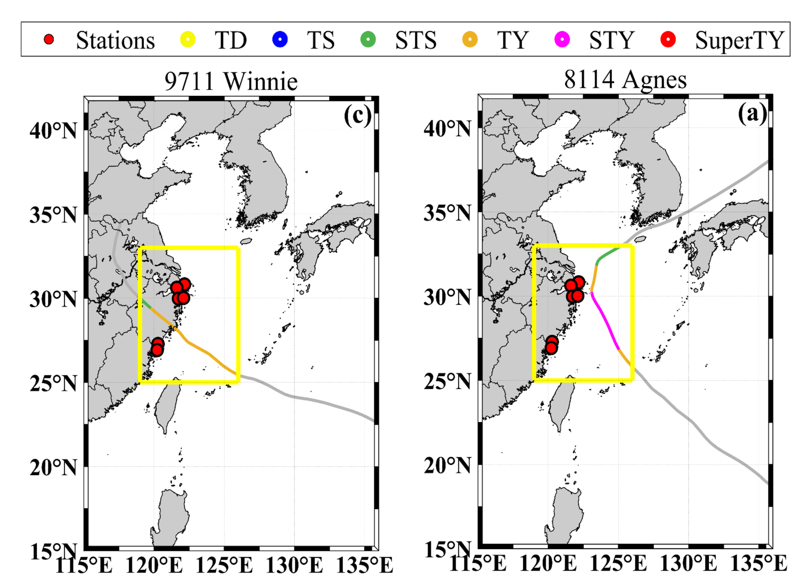
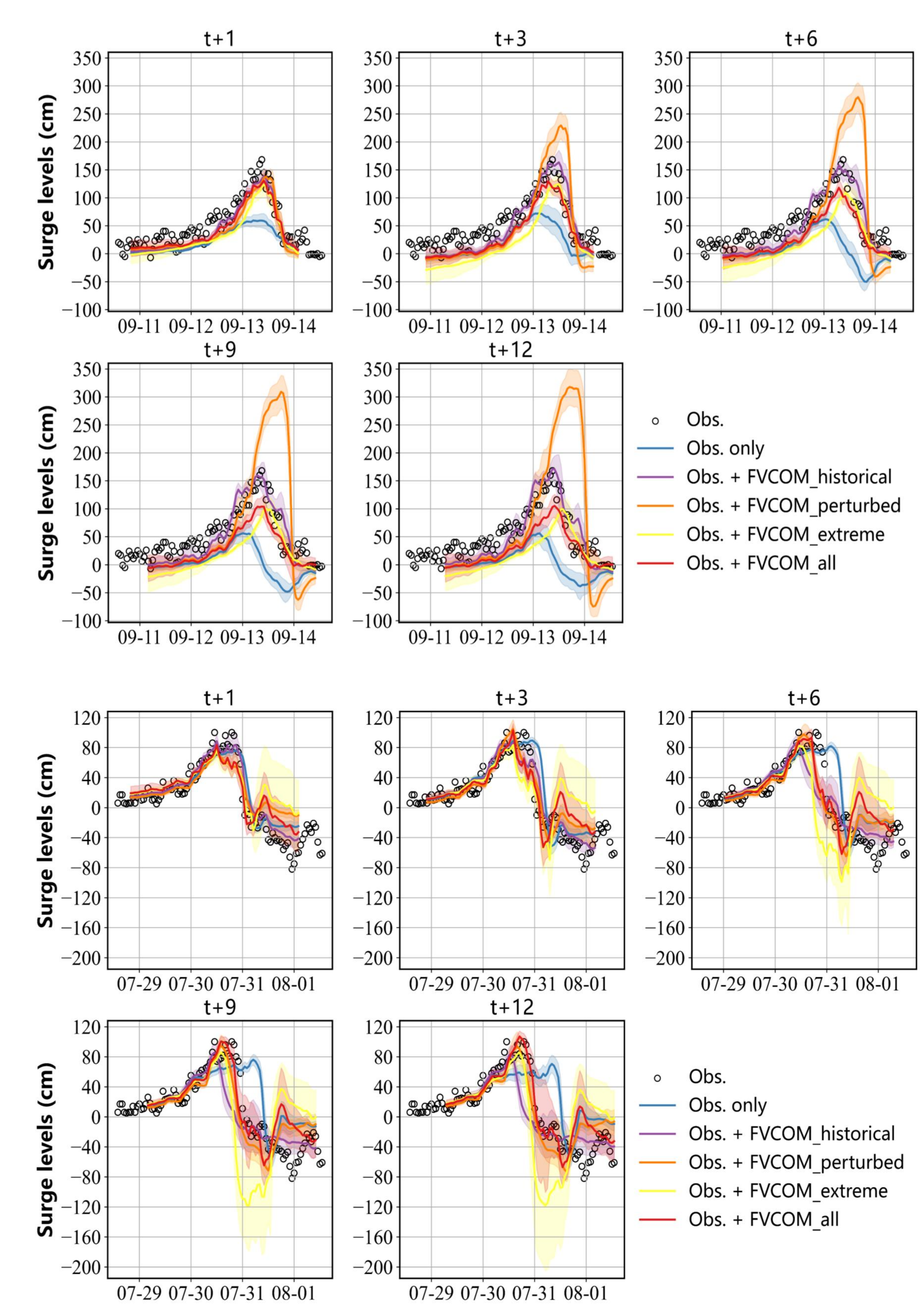


Fig. 8 The tracks of Typhoon 8114 (Agnes) and Typhoon 9711 (Winnie) in the testing data set. Only FVCOM simulation data corresponding to typhoon tracks passing through the yellow box were selected as the training/testing data.

- **Conclusions:**

(1) The incorporation of synthetic data significantly enhances predictive accuracy for extreme storm surges and their timing.

(2) For Typhoon 9711 at Zhenhai (Sea), when the **historical typhoon dataset** was added to the training set, the E-D LSTM model achieved the most significant improvement, reducing the average peak forecast error by 96 cm. For Typhoon 8114 at Tanxvdao, when **all typhoon datasets** were included, the largest error reduction was observed, with a decrease of 19 cm.



Figs. 9 (Top) & 10 (Bottom). E-D LSTM storm surge predictions for Typhoon 9711 (Winnie) at Zhenhai (Sea) and Typhoon 8114 (Agnes) at Tanxvdao across lead times. Solid line: mean of multiple runs; shaded area: 95% confidence interval.

Sensitivity Analysis of Data Uncertainty

- **WHY?** Various errors exist in real-world data, e.g., historical wind speed data provided by China Meteorological Agency are highly biased before 1970s.
- **The effect of correction of pre-1971 wind speed data**

Correcting this bias improves prediction accuracy, with ~10 cm improvement for peak prediction, indicating the significance of data quality control in building NNs-based storm surge prediction models.

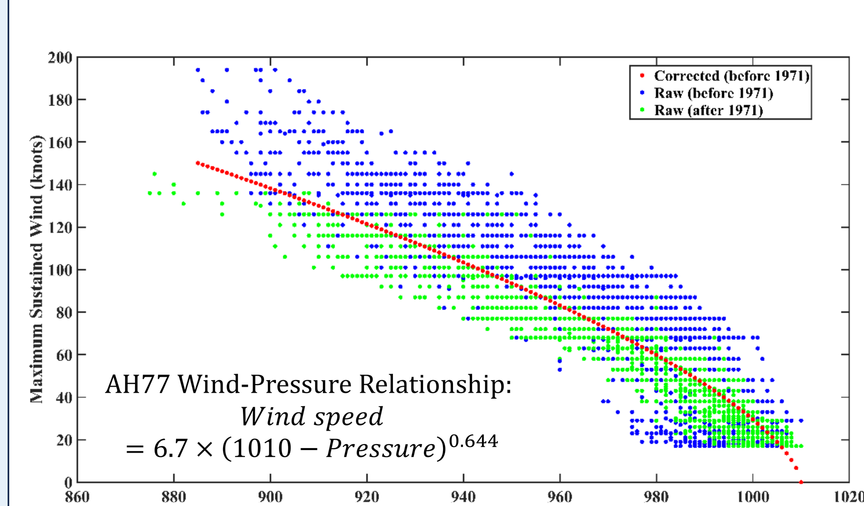


Fig. 11 Blue: original wind speed data (before 1971), green: original wind speed data (after 1971), red: corrected wind speed data (before 1971).

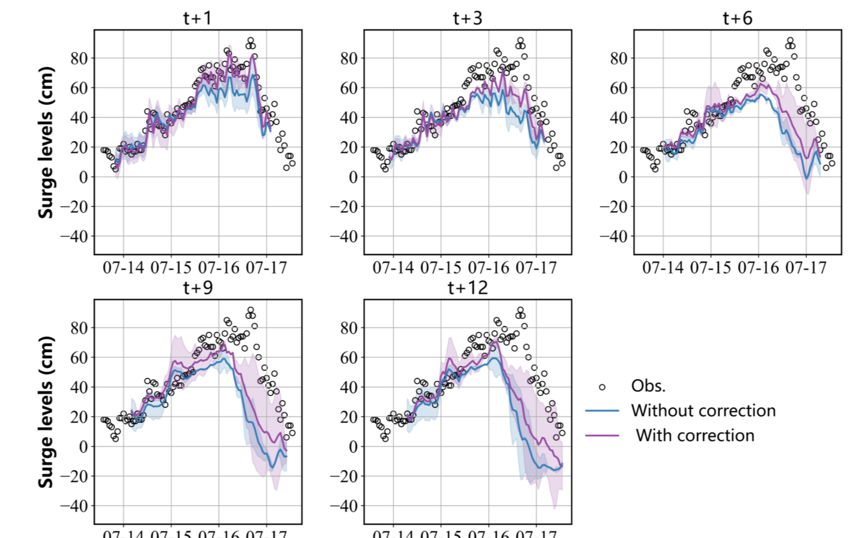


Fig. 12 Prediction results of E-D LSTM model for storm surge caused by Typhoon 0014 (Saomai) at Shacheng station using corrected/uncorrected wind speed data.

Storm Surge Database

Wind forcing (ERA5 or NCEP) + Tidal forcing

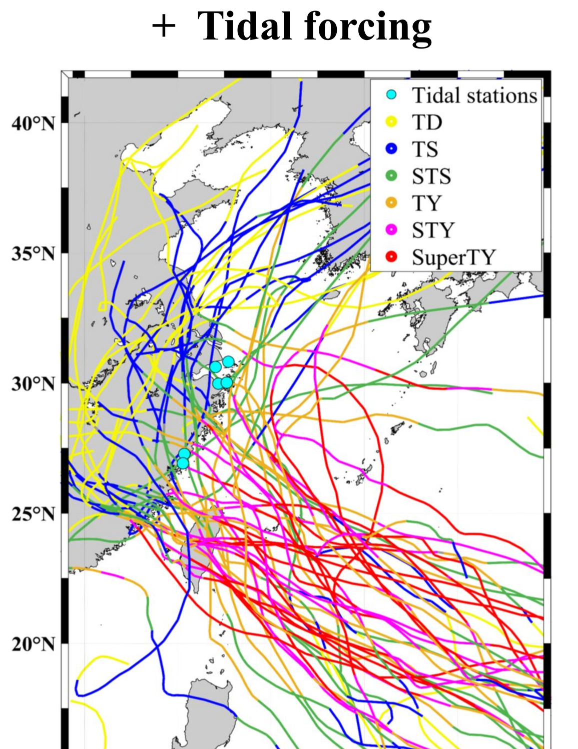


Fig. 4 Historical typhoon dataset (50)

Wind forcing (Parametric wind field) + No Tidal forcing

$$\vec{V_M} = c_1\vec{V_E} + c_2 V_g \begin{cases} -[(x_m-x_0)\sin\theta+(y_m-y_0)\cos\theta]\vec{i} \\ +[(x_m-x_0)\cos\theta-(y_m-y_0)\sin\theta]\vec{j} \end{cases}$$

$\vec{V_M}$: Typhoon model wind field

$\vec{V_E}$: Ueno forward wind field

V_g: Symmetric circular gradient wind field

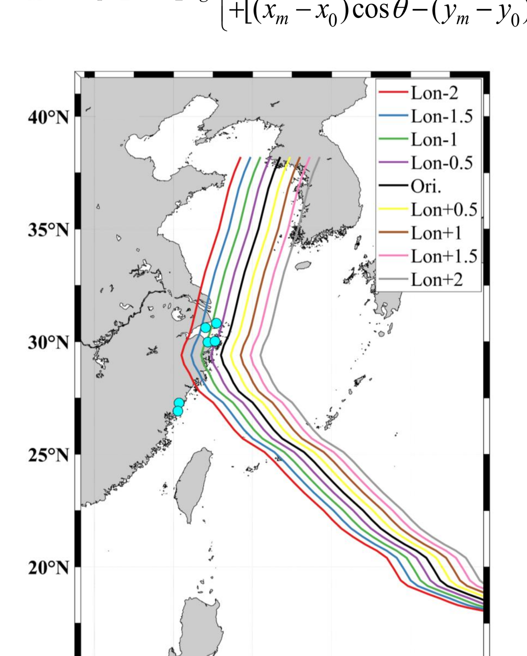


Fig. 5 Perturbed typhoon dataset (27)

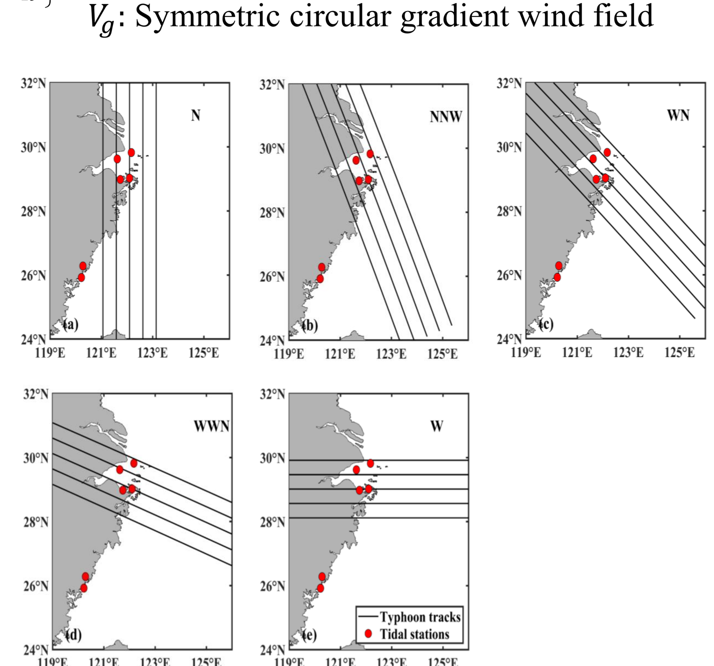


Fig. 6 Hypothetical extreme typhoon dataset (25)

Affiliations:

1 Ocean College, Zhejiang University, Zhoushan, China. (Now at Department of Earth Science and Engineering, Imperial College London, UK)

2 State Key Laboratory for Estuarine and Coastal Research, East China Normal University, Shanghai, China.

3 River Research Department of Changjiang River Scientific Research Institute, Wuhan, China.

4 Zhejiang Provincial Ocean and Fisheries Bureau, Department of Natural Resources of Zhejiang Province, Hangzhou, China.