Predictive Ability of Delft3D-Based Storm Surge Forecasts Using Historical and Forecasted Typhoon Tracks

Mata C.B. [1*], Duquez C.W.[1], Mata T.J.A.[1], Cabrera O.C.[2], and Alibuyog N.R.[3]

[1]Coastal Engineering and Management Research and Development Center, Mariano Marcos State University, Batac City, Ilocos Norte, Philippines

[2] Institute of Environmental Science and Meteorology, University of the Philippines, Diliman, Quezon City, Philippines

[3] College of Engineering, Mariano Marcos State University, Batac City, Ilocos Norte, Philippines

*cbmata@mmsu.edu.ph

Abstract: *Philippines, marked as the 5th longest coastline globally, is one of the storm surge-exposed countries caused by recurring typhoons. Rapid urbanization along this extensive coastline exposes population centers and economic zones. Due to variability in coastal configurations, surge behavior differs across the country and therefore requires localization. The Delft3D Flexible Mesh (D3D FM) model with an unstructured grid extending up to 10 m elevation above mean sea level inland was developed to simulate storm surge in coastal areas of Northwestern Luzon (NWL), Philippines. The objective of this research is to examine the forecasting ability of the model at different lead time during extreme events by using historical versus forecasted data. Three events were selected: Super Typhoon Mangkhut (2018), Tropical Storm Pakhar (2017), and Typhoon Hato (2017). Two sea level stations were used to validate the model: Currimao (CM) and San Fernando (SF) stations. Model's performance revealed an average NSE = 0.834, RMSE = 7.8 cm, and MAE = 6.2 cm indicating very good agreement between observed and simulated water levels at both CM and SF stations. Using forecasted tracks of the same events by Japan Meteorological Agency (JMA), the average RMSE at 72-, 48-, and 24-hr lead times are 5.6 cm, 7.5 cm, and 10.6 cm, respectively while average MAE are 4.3 cm, 6.9 cm and 8.9 cm, respectively. Forecasting ability of the model performs very high and indicates suitable for operational surge forecasting although becomes less reliable closer to landfall, especially for intense typhoons with rapid structural changes. Model performance is generally better for moderate typhoons with simpler wind and pressure fields. With this, the model reiterates the successful used of remote sensing derived data in forecasting storm surge. Overall, the model can be readily used to support disaster preparedness and early warning systems in coastal communities.*

Keywords: *Storm Surge, Typhoon Track, Lead Time, Delft3D, Forecast*

Introduction

Extreme sea levels, caused by storm surge and high tides remains one of the most devastating coastal hazards in the Philippines, frequently accompanying tropical cyclones that make landfall in the country. In topographically complex coastal regions like Northwestern Luzon (NWL), the need for high-resolution, locally calibrated modeling becomes critical. The Delft3D-FLOW model, developed by Deltares, is a flexible, hydrodynamic model capable of simulating water levels, tides, wind-driven circulation, and storm surges with high spatial resolution (Veeramony et al., 2017). It allows for more detailed representations of local coastal morphology, complex shoreline structures, and compound flooding scenarios (Laino & Iglesias, 2024) for any event-based forecast on the order of minutes (de Boer et al., 2012).

Localizing Delft3D for NWL, as a storm-prone region (Kim et al., 2019), can serve as an auxiliary system to the nationally implemented model, allowing for a one-stop command center that strengthens forecast reliability (Veeramony et al., 2017). The localized model can help address gaps in current surge warning products by offering site-specific storm surge mapping and fine-scale surge timing with varying forecast lead time. Studies have emphasized that such localization supports the government's disaster risk reduction agenda under Philippine Republic Act 10121 and aligns with global best practices in developing context-specific, impact-based forecasting systems

In summary, the proposed localization and evaluation of Delft3D for storm surge forecasting is not intended to replace national models, but to enhance the granularity, flexibility, and applicability of coastal flood forecasts at the local scale, particularly for use in high-risk, data-sparse environments like NWL.

This study thus fills a critical research gap by assessing Delft3D's predictive ability using both historical best typhoon track and forecasted typhoon track. The study also aimed to assess the predictive ability of the model during extreme events at different lead times. The outcomes are envisioned to enhance storm surge modelling practices and contribute to the ongoing modernization of climate and hazard resilience in the Philippines.

Literature Review

a. Global and regional storm surge models

Traditional storm surge models based on hydrodynamic modeling face challenges with real-time predictions, require high computational time and enormous computational resources, and perform poorly in regions with limited data. A plethora of studies shows that researchers encounter several challenges in forecasting storm surge during events. For example, although ocean numerical models such SCHISM coupled with WWM-V (Mentaschi et al., 2023), ADCIRDC (Blain et al., 1994; Probst & Franchello, 2012), SLOSH (Zhang et al., 2022)(Zhang et al., 2022) and GeoClaw (Mandli & Dawson, 2014; Vogt et al., 2024) have proven to be effective, their application is mostly limited to larger areas due to detailed data requirements. The literature shows that many researchers have delved in the use of reanalyzed inputs to assess and forecast storm surge in the absence of observed data.

In a global scope, a lot of researchers already explored the use of Global Tide and Surge Analysis (GTSR) or Global Tide and Surge Model (GTSM) as a hydrodynamic global reanalysis (Muis et al., 2016; Özkan et al., 2025; Wang et al., 2022). Although GTSR is in good agreement to many regional hydrodynamic models, low-resolution forcings lead to underestimation of extremes. Interestingly, Fan et al. (2025) found out that ERA20C neural networks performs better than GSTM despite ERA5 (the data that powers GTSM) have a higher temporal and spatial resolution than ERA-20C (the data that generates ERA20C neural network). While ERA20C shows potential in large-scale simulation, it still shows limitations in replicating extreme events.

Deterministic forecasts have become more effective in predicting risk and extending the usefulness of prediction systems over the past decade (Bernier et al., 2024; Elsberry et al., 2013; Kareem et al., 2022; Yamaguchi et al., 2012). However, forecasting surges from tropical cyclones remains a challenge due to their small-scale features and complex ocean-atmospheric coupling effects (Irish et al., 2008; Kohno et al., 2018; Tsai et al., 2020). This research demonstrates the use of Delft3D in simulating storm surge from forcings that include wind and pressure during extreme events, tide, Coriolis effect, all in one integrated platform. Exploring Delft3D as an operational tool allows for comprehensive storm surge forecasting with high flexibility for complex coastal environments.

b. Storm surge forecasting system in the Philippines

The country's meteorological agency, the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA), has taken significant strides in storm surge forecasting through its Storm Surge Advisory (SSA) system now called STORM-WAIS, which relies primarily on the Japan Meteorological Agency Storm Surge Model (JMA-SSM) and Storm Surge Warning and Inundation Module (SWIM) that employs preprocessing, quality control, numerical modeling, and post processing of outputs. Recent improvements in JMA-SSM have enhanced regional forecasting accuracy (Hiroshi et al., 2023). From 3.7 km grid resolution, mesh increased to 1.5-50 km. While this model provides national-scale coverage (0 – 50ºN, 95 – 180ºE), they are by coarser resolution and limited localized calibration. This study uses 800 m to 1.5 km mesh as it approaches land. Morevoer, a grid extension of 10-m elevation from mean sea level inland was integrated in the model for hazard mapping.

STORM-WAIS is an integrated model that includes monitoring, modeling, analysis, and forecast and warning generation which typically runs about 1.5 hrs depending on the area affected. Analysis involves forecaster's analysis but incorporates factors such as tide and local knowledge of the affected community. By exploring the potential of Delft3D to minimize processing time (~10-20 minutes) and data requirements while maintaining high forecasting accuracy, this model can complement the efforts of PAGASA.

Methodology

A storm surge event (SSE) identification process, based on Pauta Criterion and adapted from Ma & Li (2023) was coded in Python by Ilagan (2024) and was used in the study to identify abnormal increase in sea level (SL) during TC episodes. From the identified overlapping storm surge events, three typhoons of different intensity were selected and simulated to evaluate the accuracy of the model and assess its ability to forecast at different lead times.

a. Study area

NWL, bounded by the West Philippine Sea, is a hydro-meteorologically dynamic region characterized by seasonal monsoons, frequent typhoon landfalls, and storm surge events.

The region exhibits microtidal to mesotidal regimes, with diurnal tidal patterns dominating most of the year. Tidal ranges typically fall between 0.5 to 1.5 meters, but surge-tide interaction during typhoons can result in significant coastal flooding, especially in low-lying barangays along estuaries and river mouths.

The topography of the coastal shelf in NWL is generally shallow, which enhances the potential for storm surge amplification, especially when strong winds push water onshore and coincide with high tide (Lapidez et al., 2015). The unstructured grid of the study area was extended inland for terrain with 10m elevation as shown in Fig. 1.

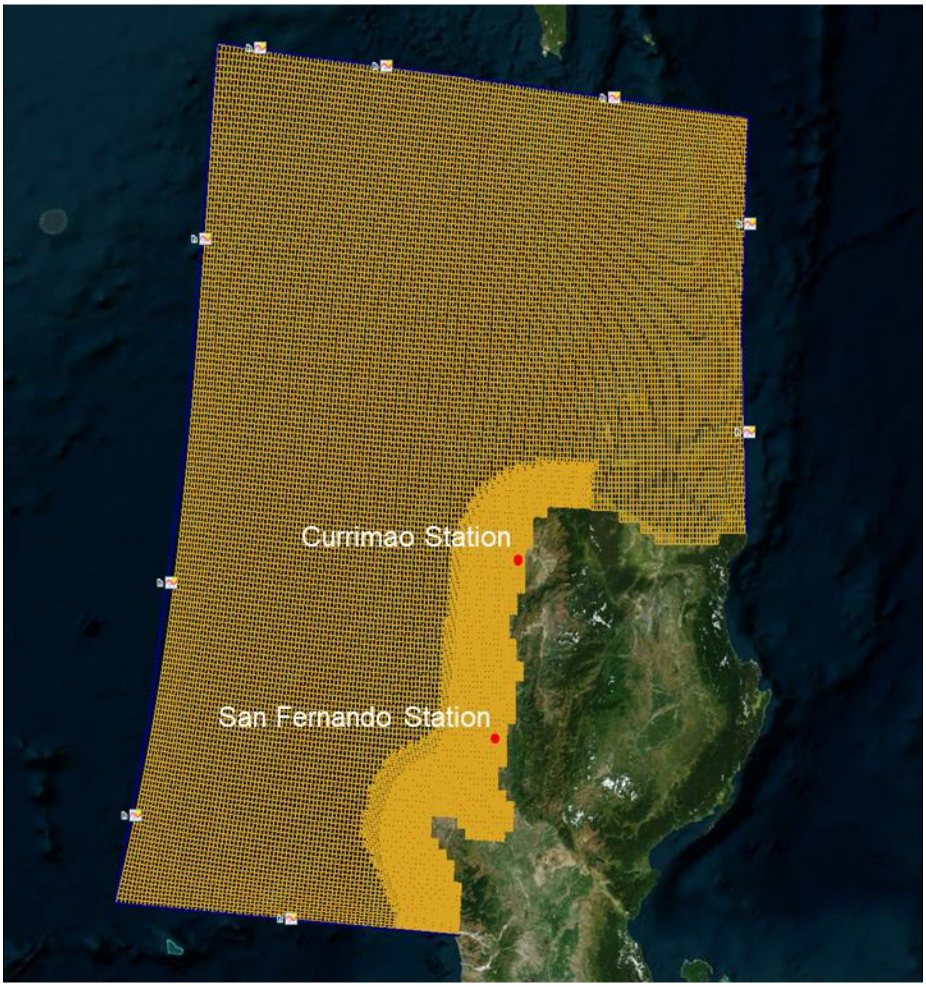


Figure 1: Map of Northwest Luzon overlain with unstructured grid of increasing spatial refinement: (1st) 5.5-6km offshore, (2nd) 1.5km, (3rd) 800m resolution. Grid is shown to extend inland to cover terrain with 10m elevation.

b. Tropical Cyclones and Storm Surge Events

The abnormal increase in sea level during tropical cyclones (TC) is known as storm surge (PAGASA, n.d.) and is measured as the increase in sea level above and beyond astronomical tides (NOAA, n.d.). The study area is located within the main typhoon pathway of the Western North Pacific, with PAGASA Area of Responsibility (PAR) records showing an average of 20 typhoons per year, about 30–40% of which either pass near or directly affect the Ilocos Region (Yumul et al., 2011). Among the 57 storm surges identified in the Philippines from 1589 to 2013, 10 affected the Ilocos Region (NOAH, 2014).

Corrected hourly observed sea level (OSL) was obtained from National Mapping and Resource Information Authority (NAMRIA), for the Currimao (CM) and San Fernando station (SF) (Figure 2). Water level at hourly resolution was subjected to tidal analysis and prediction using the UTide Python Package. The predicted levels subtracted from the observed yielded residual sea level (RSL) and subjected to an SSE detection technique that identifies outliers that exceed the mean by at least three times the standard deviation.

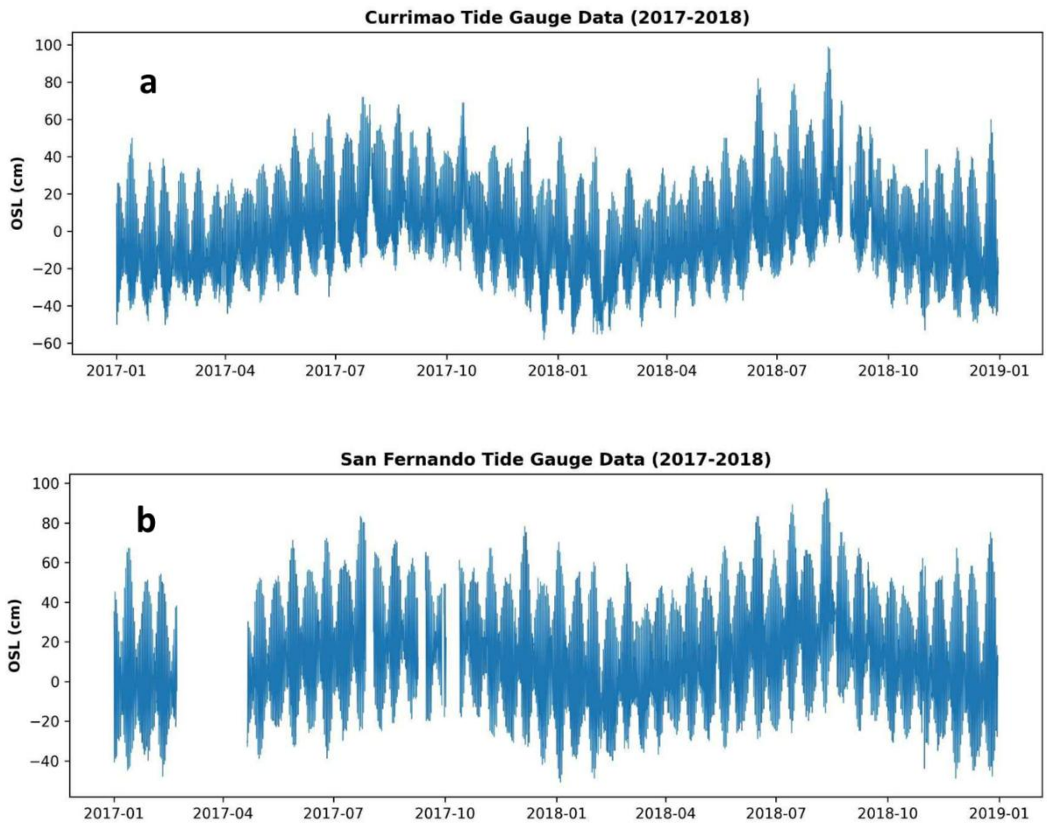


Figure 2: Observed sea level (cm) data at a) Currimao or CM and b) San Fernando or SF station.

Specifically, the RSL was calculated by calculating the difference between the OSL and the predicted tides [Eq. 1]] following the methodology of Morin et al. (2016).

$$RSL = OSL - Tide \qquad \text{[Eq. 1]}$$

Each station's per year threshold (thresh) was calculated using Ma & Li's (2023) Pauta Criterion, which assumes that random probability occurs in the range of 3σ, and values above 3σ are outliers (Ma & Li, 2023). The thresh computation is shown in Eq. 2 and Table 1 summarizes the computed threshold for 2017 and 2018.

$$thresh = Mean\ RSL + 3 * (std) \qquad \text{[Eq. 2]}$$

Table 1: Calculated annual thresholds using Pauta Criterion. Mean RSL and standard deviation for each year is also shown.

Year	CM			SF		
	Mean RSL	std	thresh	Mean RSL	std	thresh
2017	-2.69	6.74	17.52	7.97	6.14	26.4
2018	-6.25	7.83	17.24	9.03	7.11	30.38

In summary, for an RSL to be classified as a storm surge event (SSE), three requirements must be met. The RSL must, first and foremost, be higher than the yearly sea level threshold determined using the Pauta Criterion method. Second, a TC with an intensity of at least 34 knots must be present within 500 km of any of the two tidal stations during this exceedance. Third, the corresponding 6-hour Gaussian low pass filtered value must also exceed the threshold to guarantee that the exceedance is not a normal fluctuation in the RSL.

The 500 km search radius was selected because, according to historical data, even when the TC is about 400 km away, at least 0.41 m absolute storm surges have been recorded close to Metro Manila's coastlines (Morin et al., 2016). Towey et al. (2022) have pointed out that TCs more than 500 kilometers away from a site will have minimal impacts.

c. The storm surge model

Using the Delft3D FLOW software, a model has been produced to simulate the water level in NWL domain. The model was created using the Delft Dashboard software while the simulation was processed on the Delft3D FLOW software. The Delft Dashboard software was used to create the model by setting up the boundary conditions, bathymetry, inclusive dates of the occurrence of the typhoons, the tracks of the typhoon, and the areas of observation of the water level. Table 2 lists the sources of model input parameters. The inputs come from remote sensing (satellite derived bathymetry, topography), reanalysis products (wind and atmospheric pressure fields of typhoons, typhoon track, tide), and actual observations (sea level).

Table 2: Input parameters in developing the Delft3D-based storm surge.

Parameter	Value & Source	Estimated Accuracy
Bathymetry	[a]GEBCO 2019 & [b]SRTM 4.0	≈ ±2.5% error for geomorphometric metrics; vertical errors often ±150–180 m in shallow areas
Boundary Conditions	[c]TPXO9 global tide inverse model	RMSE in open/ocean ~1–2 cm; coastal/complex shelf up to ~5–10 cm
Winds	[d]SPW (spider-web) node network (user input)	Depends on resolution; wind speed accuracy ±5–10% typical
Typhoon Track	[e]JTWC Best Track	Position error often ~10–30 km; intensity ±5–10 hPa

[a](GEBCO Compilation Group, 2023), [b](NASA JPL, 2013), [c](Egbert & Erofeeva, 2002), [d](Mattocks & Forbes, 2008), [e](Knapp et al., 2010)

After the model was created, the simulations were done using the Delft3D FLOW software and results were analyzed. The model output, primarily sea level time series, was compared against observed tide data at CM and SF stations. Commonly used error metrics to quantify model accuracy applied in the study included Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Nash-Sutcliffe Efficiency (NSE). Unlike any other measures, RMSE and MAE are easy to interpret and do not indicate bias (Jackson et al., 2019; Willmott & Matsuura, 2005). We also selected NSE because it is simple in concept. It compares prediction error against the observed errors (Jackson et al., 2019). It is to note that while NSE is useful in giving information error of predicted versus observed, it is not useful in comparing forecasted versus observed (Onyutha, 2024; Teegavarapu et al., 2022). These metrics provided a comprehensive assessment of model skill in replicating both the magnitude and timing of historical storm surge events.

Acceptable values were based on storm surge modeling for most coastal sites. The accuracy of storm surge model developed were evaluated and must satisfy the following measures:

Accuracy Measure	Value	Model Performance	Interpretation	Acceptable Value
Nash-Sutcliffe Efficiency (NSE)	≥ 0.75	Excellent	Ready for operational forecasting and research	≥ 0.5[a]
	0.50–0.75:	Good	Useable, but might need refinement	
	< 0.50:	Unsatisfactory	Recalibration likely needed	
Root Mean Square Error (RMSE)	< 20 cm	Very good	High-resolution model, well-calibrated, reliable for warnings.	≤ 30 cm [b, c]
	20–40 cm:	Good	acceptable in many storm surge applications.	
	> 50 cm:	Satisfactory	Needs review	
Mean Absolute Error (MAE)	< 15 cm	Very High	Suitable for operational surge forecasting or early warning systems	< 30 cm [d]
	15–30 cm:	Good	Acceptable for research, design-level modeling, and flood hazard mapping	
	30–50 cm:	Moderate	May be acceptable depending on event magnitude and data resolution	
	> 50 cm	Low	Model needs improvement or input/boundary condition refinement	

[a] (Moriasi et al., 2007), [b] (Muis et al., 2016), [c] (Teng et al., 2017), [d] (Heemink et al., 2001)

To assess the sensitivity and forecasting performance of the storm surge model, forecasted tracks of JMA of the same typhoons were downloaded. Each forecast track included storm center coordinates, wind speed, radius of maximum wind, and central pressure. Forecast advisories were selected at different lead times including 72, 48, and 24 hours. RMSE and MAE were used to assess model performance. NSE is sensitive to peak timing mismatches which are common to storm surge forecasts that even a slight shift in peak surge timing can result in a very low value despite the peak magnitude being accurate (Jackson et al., 2019; Onyutha, 2024).

Fig. 3 shows the detailed workflow of storm surge model including data sources.

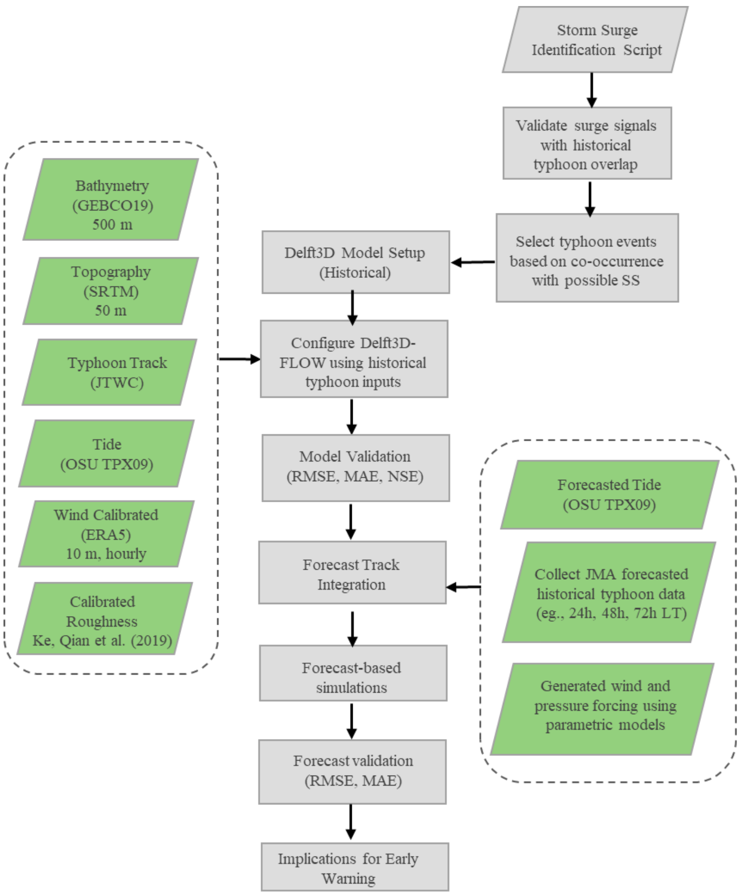


Figure 3: Flow diagram of Delft3D storm surge model

Results and Discussion

Selection of typhoon events based on co-occurrence with observed storm surge

Out of seven (7) identified possible storm surge in CM station from 2017 to 2018, only five (5) typhoons coincide (Fig. 4). Two identified possible storm surges in 2018 might be due to other anomalies in sea level due such as LPA and Southwest Monsoon. The maximum storm surge observed at CM and SF stations from 2017 to 2018 are 55 cm (7/29/2017) and

43 cm (9/15/2018), respectively. It is to note that even though a 55 cm surge might be considered low in comparison to storm surges in other places, like 78 cm in Manila Bay during TY Nesat in 2011 (Morin et al., 2016) and over 500 cm in Leyte during STY Haiyan (Soria et al., 2016), Ilocos Region is a microtidal coast. A 55 cm rise is therefore comparatively significant in this case.

Only the typhoons that co-occurred with the identified possible storm surge events from 2017-2018 were automatically identified by the script. Selection of typhoons modeled were based on the completeness of validation data in two observation stations. With this, only Super Typhoon (ST) Ompong (Mangkhut), Tropical Strom (TS) Jolina (Pakhar), and Typhoon (T) Isang (Hato).

Upon further inspection of the observed sea level, Tropical Storm Pakhar was also included because its track passes through the Northwest Luzon domain. Although it did not appear to cause a storm surge at the CM Station, it might have caused a possible storm surge in SF. However, due to the significant lack of data at that station, determining this using the Storm Surge Identification method (SSI) is not possible.

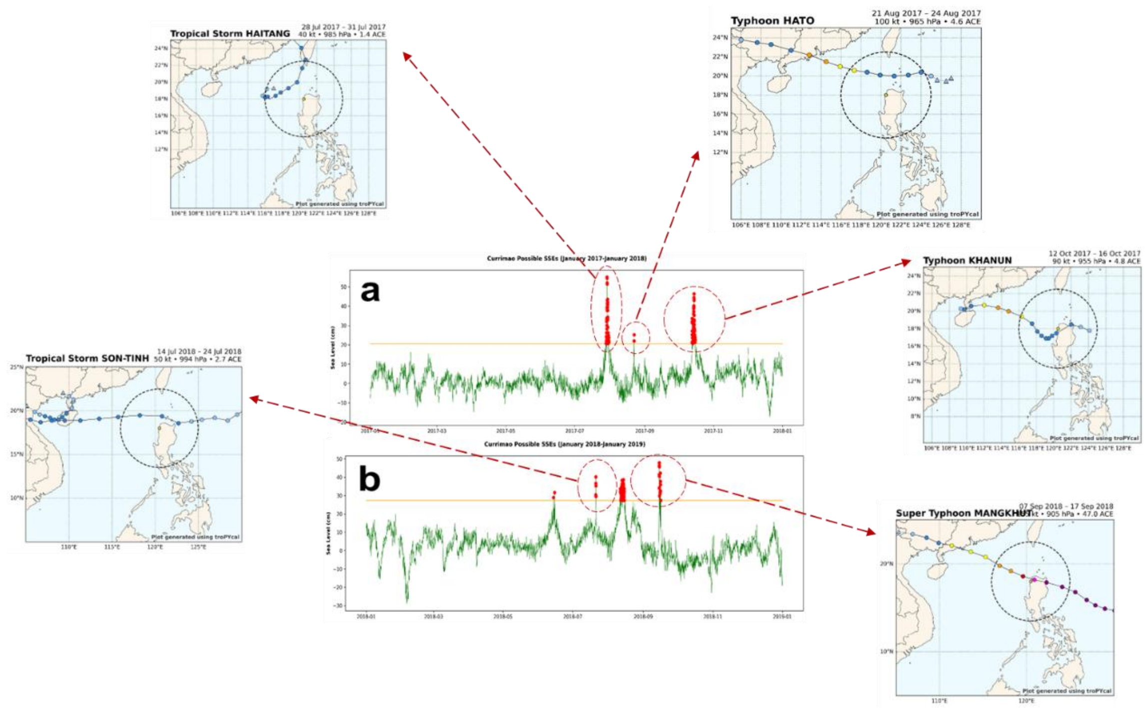


Figure 4: Identified possible storm surge events (red) at CM station from a) 2017 to b) 2018

b. Develop and evaluate the accuracy of the storm surge model under different Typhoon events

Model Setup

Delft3D Dashboard has its built-in archives of historical typhoon tracks by JTWC which makes it advantageous and useful over other software that requires manual input. Input parameters constantly used for the three typhoons are timestep (30s), bathymetry (GEBCO 19), boundary conditions (TPX09 global tide inversed model), and wind (spider web). Bottom drag coefficients/ roughness values in ocean, rivers/estuaries, and land are 0.0025, 0.0050, and 0.0075, respectively as suggested by Ke et al., (2019). In addition to his recommendation, water density used is 1,025 cm^3/s while wind drag coefficients at 5 m/s, 31.5 m/s and 60 m/s are 0.0009839, 0.002, and 0.0008013, respectively. These values were used in the calibration where the model performed best. Time frames are summarized in Table 3.

Table 3: Time frame set-up for Ompong (Mangkhut) in 2018, b) Jolina (Pakhar) in 2017, and c) Isang (Hato) in 2017

Time Frame	ST Ompong (Mangkhut)	TS Jolina (Pakhar)	T Isang (Hato)
Reference Date	08/15/2018 12:00 AM	08/10/2017 12:00 AM	08/10/2017 12:00 AM
Start Date	09/07/2018 12:00 AM	08/24/2017 12:00 AM	08/20/2017 12:00 AM
End Date	09/16/2018 12:00 AM	08/27/2017 12:00 AM	08/24/2017 12:00 AM

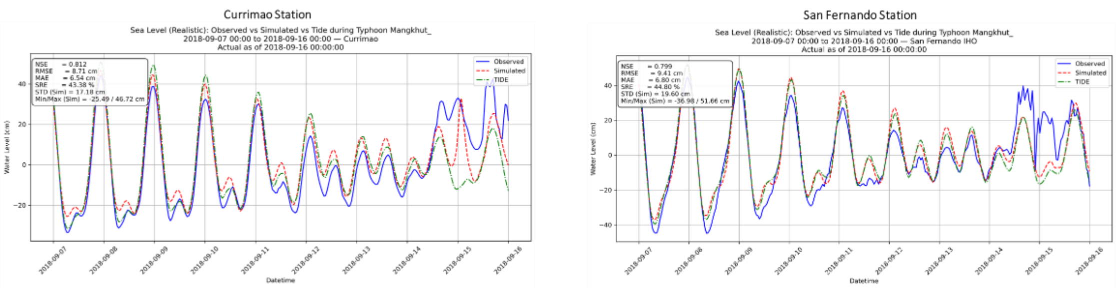
Simulation and Accuracy Results

The following sections indicate the validation results of the model. Although the model reveals high performance, the fitness between observed and simulated storm surge water levels can show overprediction or underprediction at certain points in time during typhoons (see Fig. 5) due to several physical and meteorological factors. Storm surge models rely heavily on wind speed, direction, and central pressure data to simulate water level rise. JWTC best track are being compiled after a tropical cyclone has dissipated, using all available observations (satellite imagery, aircraft reconnaissance, buoy and ship reports, etc.) and modeling tools. Detailed explanation is available at JWTC website (https://www.metoc.navy.mil/jtwc/jtwc.html). These tracks represent the most accurate estimate of the storm's position, intensity, and structure at 6-hour intervals during its lifetime. Therefore, they are not actual tracks, but a carefully constructed reanalysis. However, it is considered a high-quality reanalysis suitable for validation studies.

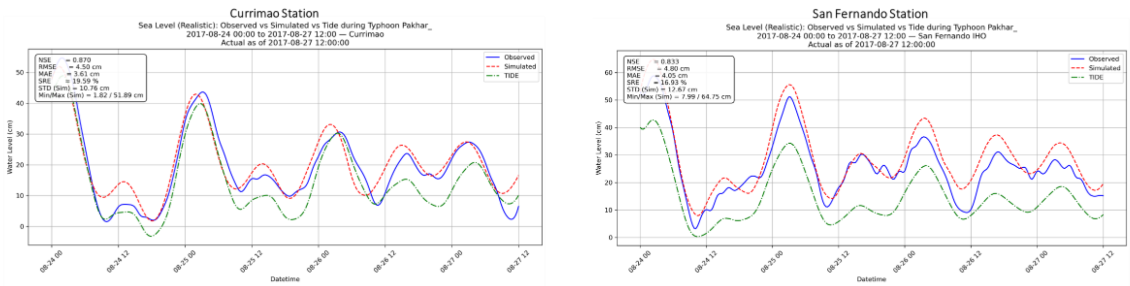
Another relevant factor is the resolution of bathymetry and topography used which can affect wave setup and water level propagation. Oversmoothed topography such as SRTM (30m resolution) may not capture swell effects especially nearshore. Moreover, fact that CM station is located in the sandy beach part of CM, south of the Pangil Coral Rock formation while SF station is located near the mouth of the semi-enclosed Lingayen Gulf where other external factors like river discharge, if not well or totally represented, may affect recorded water levels in the station leading to underprediction.

One meteorological factor is the boundary condition in which tide data from OSU TPX09 was used. The accuracy of the OSU TPXO9 global tide model (~1/30° to 1/60° grid spacing) is very high—especially for deep ocean and continental shelf regions—but it varies slightly depending on location. TPXO9 is accurate to within 1–2 cm RMSE in the open ocean, and 2–5 cm on shelves. Nesting with local or regional tide models is therefore recommended for improved accuracy. Philippines does not yet have a fully operational, high-resolution nationwide tide model that is publicly available like TPXO. Currently, PAGASA provides tide predictions for major ports based on harmonic analysis of observed data but not a hydrodynamic model.

ST Ompong (Mangkhut) in 2018



TS Jolina (Pakhar) in 2017



T Isang (Hato) in 2017

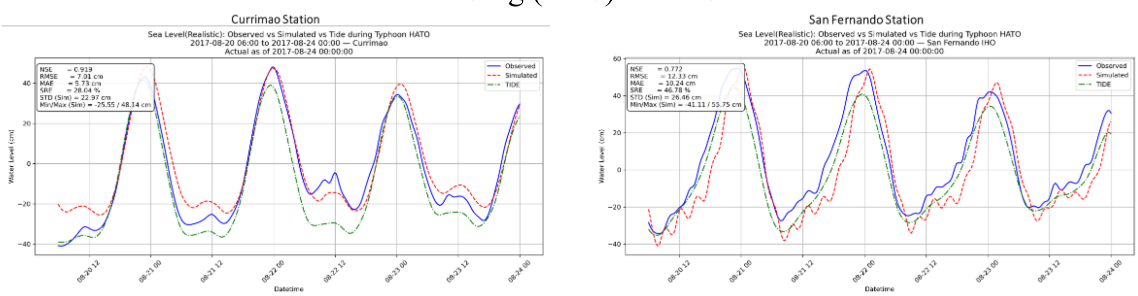


Figure 5: Sea level (cm) of observed (blue) vs simulated (red dash) vs tide (green dash) during the STY Ompong (Mangkhut) in 2018, TS Jolina (Pakhar) in 2017, and TY Isang (Hato) in 2017 in CM and SF stations

Table 4 summarizes the result of accuracy used in the study. All NSE values are ≥ 0.772, ranging from 0.772 to 0.919, with an average of 0.867 and 0.801 in SF station. Overall average is 0.834 which indicates excellent model performance (Moriasi et al., 2007) at both observation stations for all events. The highest NSE (0.919) was observed during TY Isang at CM station, suggesting that the model was particularly effective in simulating the dynamics of this event. The RMSE values ranged from 4.5 cm to 12.3 cm, with an average of 6.7 cm at CM and 8.8 cm at SF. These values are well within the acceptable range for water level simulations, especially considering that RMSE values below 30 cm are typically considered indicative of good model performance (Muis et al., 2016). Tropical Storm Pakhar yielded the lowest RMSE values at both stations (4.5 cm at CM and 4.8 cm at SF), reflecting more stable atmospheric forcing and accurate boundary conditions during this event.

Table 4: Summary of accuracy for each typhoon at two stations

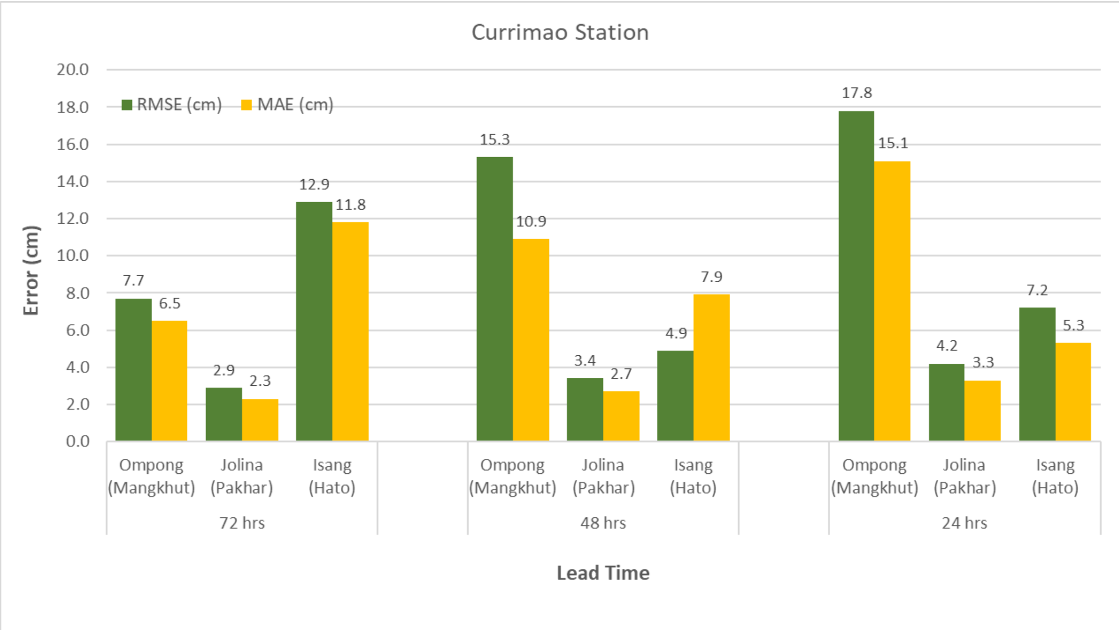
Extreme Event	**Station**	**NSE**	**RMSE (cm)**	**MAE (cm)**
ST Ompong (Mangkhut)	CM	0.812	8.7	6.5
	SF	0.799	9.4	6.8
TS Jolina (Pakhar)	CM	0.870	4.5	3.6
	SF	0.833	4.8	4.1
T Isang (Hato)	CM	0.919	7.0	5.7
	SF	0.772	12.3	10.2
Average	CM	0.867	6.7	5.3
	SF	0.801	8.8	7.0
	Average	0.834	7.8	6.2

The MAE values ranged from 3.6 cm to 10.2 cm, with CM station averaging 5.3 cm and SF staion 7.0 cm. All MAE values fall within the very high accuracy category based on benchmark studies which consider MAE values below 15–30 cm as excellent for storm surge simulations (Heemink et al., 2001). This suggests that the model's predicted water levels closely align with observations, particularly during Tropical Storm Pakhar, which recorded the lowest MAE at both stations.

For the three measures, the model performed better in CM station due to the fact that the track crossed the NWL near CM station. Same result was obtained by Rajendiran et al. (2024) wherein the discrepancy between simulated and observed difference in location between observation point and landfall location. This indicates that there is a better alignment of wind setup and tide-surge interaction at CM station. These spatial differences in accuracy are consistent with Haigh et al. (2012), who observed site-specific variability in storm surge model skill across the UK coastline due to localized coastal geometry and exposure. Overall, the discrepancy in sea level between simulated and observed are beyond acceptable range revealing that the model accurately represented sea level during typhoons at both stations and the model can be applied in further investigations related to storm surge.

c. Assess predictive capability of the storm surge model using daily typhoon forecast tracks across multiple base times (e.g., 24, 48, and 72 hours)

The performance of storm surge simulations for extreme events STY Ompong (Mangkhut), TS Jolina (Pakhar), and TY Isang (Hato) using RMSE and MAE at lead times of 72, 48, and 24 hours in CM and SF stations are shown in Fig. 6. The model revealed distinct trends in forecast accuracy depending on both typhoon characteristics and lead time.



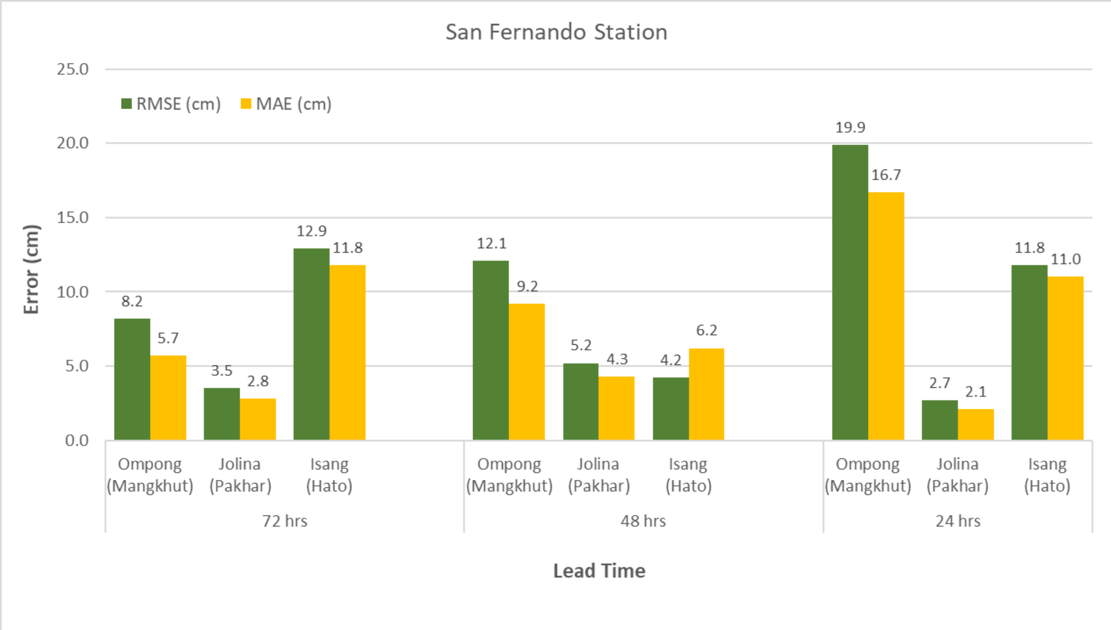


Figure 6: Model performance over different forecast lead time across events in CM and SF stations.

Generally, the more intense storms are, the higher is the error (Mandli & Dawson, 2014). For extreme events such as STY Mangkhut, the prediction grossly underpredict the storm surge size. In TS Jolina (Pakhar), there is consistently low errors across all lead times (RMSE: 2.9–4.2 cm in CM and 2.7-5.2 cm in SF), indicating better model performance for relatively weaker or less complex storms.

In TY Isang (Hato), CM station revealed RMSE at 72 hrs is 12.9 cm, drops to 4.2 cm at 48 hrs, then increases to 11.8 cm at 24 hrs, showing a fluctuating trend like in CM station, likely due to complex regional effects. At SF station, RMSE is identical at 72-hrs in both stations (12.9 cm), but diverges at 24- hrs—CM station records 7.2 cm while SF station has 11.8 cm. This variation may stem from differences in observed surge timing or bathymetric sensitivity.

Moreover, a clear trend of decreasing accuracy with shorter lead times was observed, particularly evident in STY Ompong. For instance, Ompong's RMSE increased from 7.7 cm (72 hrs) to 17.8 cm (24 hrs), while MAE rose from 6.5 cm to 15.1 cm. This suggests an increasing deviation between simulated and observed storm surge values as the storm approached landfall, likely due to intensifying meteorological and oceanic variability not fully captured in the model input (Kohno et al., 2018; Rajendiran et al., 2024). It should be noted that the accuracy of model skill depends heavily on the realism of atmospheric settings, including typhoon tracks like those from JMA. The Digital Typhoon archive (Kitamoto et al., n.d.) provides best track and forecast error data. It shows the error statistics of historical JMA typhoon forecast illustrating uncertainties at different lead times.

Conclusion and Recommendation

The model demonstrates good predictive skill, with particularly strong performance at CM station. Overall, the strong correlation between observed and simulated storm surge indicates that Delft3D, and its Dashboard with its archived datasets including boundary conditions, can be used for local models. The use of statistical metrics supports its applicability for operational and research-based storm surge forecasting in the northwestern coast of the Philippines.

However, performance inconsistencies, especially in SF station, highlight the need for 1) improved local input data (e.g., bathymetry, tide data), 2) site-specific calibration, and 3) further model tuning based on typhoon characteristics (e.g., intensity, forward speed, track). Overall, the results confirm that the model can be reliably used for risk assessment, early warning systems, and coastal planning, with recommendations for continued validation and refinement to increase robustness across different coastal settings.

For future works, it is recommended to first, compare the forecast capabilities of the model developed to the STORM-WAIS model that is currently being used by PAGASA for storm surge warnings. Secondly, we recommend to explore coupling with Delft3D-wave to determine wave height for total impact assessment or hazard mapping. And third, explore its applicability in other regions in the Philippines and other areas globally.

References

Bernier, N. B., Hemer, M., Mori, N., Appendini, C. M., Breivik, O., de Camargo, R., Casas-Prat, M., Duong, T. M., Haigh, I. D., Howard, T., Hernaman, V., Huizy, O., Irish, J. L., Kirezci, E., Kohno, N., Lee, J. W., McInnes, K. L., Meyer, E. M. I., Marcos, M., … Zhang, Y. J. (2024). Storm surges and extreme sea levels: Review, establishment of model intercomparison and coordination of surge climate projection efforts (SurgeMIP). In *Weather and Climate Extremes* (Vol. 45, pp. 1–11). Elsevier B.V. https://doi.org/10.1016/j.wace.2024.100689

Blain, C. A., Westerink, J. J., Luettich, R. A. , Jr., & Scheffner, N. W. (1994). *ADCIRC: An advanced three-dimensional circulation Model for shelves, coasts, and estuaries. Report 4. Hurricane storm surge modeling using large domains*. U.S. Army Engineer Waterways Experiment Station. https://www.researchgate.net/publication/235039076

de Boer, G. J., Bruens, A., Damsma, T., van Geer, P., Grasmeijer, B. T., Heijer, C. (Kees), & van Koningsveld, M. (2012, February 27). OpenEarth: using Google Earth as outreach for NCK's data. *NCK-Days 2012: Crossing Borders in Coastal Research: Jubilee Conference Proceedings*. https://doi.org/10.3990/2.177

Egbert, G. D., & Erofeeva, S. Y. (2002). Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology*, *19*(2), 183–204.

Elsberry, R. L., Chen, L., Davidson, J., Rogers, R., Wang, Y., & Wu, L. (2013). Advances in understanding and forecasting rapidly changing phenomena in tropical cyclones. *Tropical Cyclone Research and Review*, *2*(1), 13–24. https://doi.org/10.6057/2013TCRR01.02

Fan, M., Dang, W., Feng, J., Li, D., & Zhao, L. (2025). Comparative analysis of storm surge reconstructions in the Western North Pacific: Implications for coastal flood risk assessment. *International Journal of Disaster Risk Science*, *16*(3), 446–463. https://doi.org/10.1007/s13753-025-00647-0

GEBCO Compilation Group. (2023). *The GEBCO 2023 Grid – a continuous terrain model of the global oceans and land*. https://www.gebco.net

Haigh, I. D., Wijeratne, E., Macpherson, L. R., & Mason, M. S. (2012). *Estimating present day extreme total water level exceedance probabilities around the coastline of Australia*. www.acecrc.org.au

Heemink, A. W., Bolding, K., Bruggeman, B., & Verlaan, M. (2001). Storm surge forecasting using Kalman filtering. In *Article in Journal of the Meteorological Society of Japan Ser II*. https://www.researchgate.net/publication/2362995

Hiroshi, H., Jumpei, S., Takashi, F., & Masakazu, H. (2023). *Upgrade of JMA's Storm Surge Prediction for the WMO Storm Surge Watch Scheme (SSWS) in 2022*. https://www.jma.go.jp/jma/index.html

Ilagan, A. A. (2024). *Meteorological drivers of storm surge events in the Ilocos Region*. MSc Thesis, University of the Philippines, Diliman, Quezon City.

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

Jackson, E. K., Roberts, W., Nelsen, B., Williams, G. P., Nelson, E. J., & Ames, D. P. (2019). Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption. In *Environmental Modelling and Software* (Vol. 119, pp. 32–48). Elsevier Ltd. https://doi.org/10.1016/j.envsoft.2019.05.001

Kareem, K. Y., Seong, Y., Kim, K., & Jung, Y. (2022). A case study of tidal analysis Using theory-based artificial intelligence techniques for disaster management in Taehwa River, South Korea. *Water (Switzerland)*, *14*(14). https://doi.org/10.3390/w14142172

Ke, Q., Bricker, J., & Ye, T. (2019). Storm surge modelling by Delft3D FM- a case study in Shanghai area. *EGU General Assembly 2019*.

Kitamoto, A., Hwang, J., Vuillod, B., Gautier, L., Tian, Y., & Clanuwat, T. (n.d.). *Digital Typhoon: Typhoon Images and Information*. National Institute of Informatics (NII). Retrieved June 3, 2025, from https://agora.ex.nii.ac.jp/digital-typhoon/

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). *Bulletin of the American Meteorological Society*, *91*(3), 363–376. https://doi.org/10.1175/2009BAMS2755.1

Kohno, N., Dube, S. K., Entel, M., Fakhruddin, S. H. M., Greenslade, D., Leroux, M.-D., Rhome, J., & Ba Thuy, N. (2018). Recent progress in storm surge forecasting. *Tropical Cyclone Research and Review*, *7*(2), 128–139. https://doi.org/10.6057/2018TCRR02.04

Laino, E., & Iglesias, G. (2024). Beyond coastal hazards: A comprehensive methodology for the assessment of climate-related hazards in European coastal cities. *Ocean and Coastal Management*, *257*. https://doi.org/10.1016/j.ocecoaman.2024.107343

Lapidez, J. P., Tablazon, J., Dasallas, L., Gonzalo, L. A., Cabacaba, K. M., Ramos, M. M. A., Suarez, J. K., Santiago, J., Lagmay, A. M. F., & Malano, V. (2015). Identification of storm surge vulnerable areas in the Philippines through the simulation of Typhoon Haiyan-induced storm surge levels over historical storm tracks. *Natural Hazards and Earth System Sciences*, *15*(7), 1473–1481. https://doi.org/10.5194/nhess-15-1473-2015

Ma, P., & Li, G. (2023). Comparison and analysis of detection methods for typhoon-storm surges based on tide-gauge data—Taking coasts of China as examples. *International Journal of Environmental Research and Public Health*, *20*(4). https://doi.org/10.3390/ijerph20043253

Mandli, K. T., & Dawson, C. N. (2014). Adaptive mesh refinement for storm surge. *Ocean Modelling*, *75*, 36–50. https://doi.org/10.1016/j.ocemod.2014.01.002

Mattocks, C., & Forbes, C. (2008). A real-time, event-triggered storm surge forecasting system for the state of North Carolina. *Ocean Modeling*, *25*(3–4), 95–119. https://doi.org/https://doi.org/10.1016/j.ocemod.2008.06.008

Mentaschi, L., Vousdoukas, M. I., García-Sánchez, G., Fernández-Montblanc, T., Roland, A., Voukouvalas, E., Federico, I., Abdolali, A., Zhang, Y. J., & Feyen, L. (2023). A global unstructured, coupled, high-resolution hindcast of waves and storm surge. *Frontiers in Marine Science*, *10*. https://doi.org/10.3389/fmars.2023.1233679

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153

Morin, V., Warnitchai, P., & Weesakul, S. (2016). Storm surge hazard in Manila Bay: Typhoon Nesat (Pedring) and the SW monsoon. *Natural Hazards*, *81*(3), 1569–1588.

Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H., & Ward, P. J. (2016). A global reanalysis of storm surges and extreme sea levels. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11969

NASA JPL. (2013). *NASA Shuttle Radar Topography Mission Global 1 arc second [Data set]*. https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003

National Oceanic and Atmospheric Administration (NOAA). (n.d.). *What is storm surge?* Retrieved August 26, 2025, from https://oceanservice.noaa.gov/facts/stormsurge-stormtide.html

Nationwide Operational Assessment of Hazards (NOAH). (2014). *Compilation of Storm Surge Occurrences in the Philippines. Project NOAH Open-File Reports*.

Onyutha, C. (2024). Pros and cons of various efficiency criteria for hydrological model performance evaluation. *Proceedings of the International Association of Hydrological Sciences*, *385*, 181–187. https://doi.org/10.5194/piahs-385-181-2024

Özkan, F. N., Verlaan, M., Muis, S., & Zijl, F. (2025). Sensitivity of global storm surge modelling to sea surface drag. *Ocean Dynamics*, *75*(8). https://doi.org/10.1007/s10236-025-01713-3

Philippine Atmospheric, G. and A. S. A. (PAGASA). (n.d.). *Storm Surge*. Retrieved August 26, 2025, from https://www.pagasa.dost.gov.ph/information/storm-surge

Probst, P., & Franchello, G. (2012). *Global storm surge forecast and inundation modeling*. https://doi.org/10.2788/14951

Rajendiran, S., Beulah, H. R., & Balaji, R. (2024). Comparative Analysis of Different Meteorological Parameters on Storm Surge Simulation during Tropical Cyclone (TC) Tauktae. *Journal of Coastal Research*, *113*(sp1). https://doi.org/10.2112/jcr-si113-199.1

Soria, J. L. A., Switzer, A. D., Villanoy, C. L., Fritz, H. M., Bilgera, P. H. T., Cabrera, O. C., Siringan, F. P., Maria, Y. Y. S., Ramos, R. D., & Fernandez, I. Q. (2016). Repeat storm surge disasters of Typhoon Haiyan and its 1897 predecessor in the Philippines. *Bulletin of the American Meteorological Society*, *97*(1), 31–48. https://doi.org/10.1175/BAMS-D-14-00245.1

Teegavarapu, R. S. V., Sharma, P. J., & Lal Patel, P. (2022). Frequency-based performance measure for hydrologic model evaluation. *Journal of Hydrology*, *608*. https://doi.org/10.1016/j.jhydrol.2022.127583

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. In *Environmental Modelling and Software* (Vol. 90, pp. 201–216). Elsevier Ltd. https://doi.org/10.1016/j.envsoft.2017.01.006

Towey, K. L., Booth, J. F., Rodriguez Enriquez, A., & Wahl, T. (2022). Tropical cyclone storm surge probabilities for the east coast of the United States: A cyclone-based perspective. *Natural Hazards and Earth System Sciences*, *22*(4), 1287–1300. https://doi.org/10.5194/nhess-22-1287-2022

Tsai, Y. L., Wu, T. R., Lin, C. Y., Lin, S. C., Yen, E., & Lin, C. W. (2020). Discrepancies on storm surge predictions by parametric wind model and numerical weather prediction model in a semi-enclosed bay: Case study of typhoon Haiyan. *Water (Switzerland)*, *12*(12). https://doi.org/10.3390/w12123326

Veeramony, J., Condon, A., & van Ormondt, M. (2017). Forecasting storm surge and inundation: Model validation. *Weather and Forecasting*, *32*(6), 2045–2063. https://doi.org/10.1175/WAF-D-17-0015.1

Vogt, T., Treu, S., Mengel, M., Frieler, K., & Otto, C. (2024). Modeling surge dynamics improves coastal flood estimates in a global set of tropical cyclones. *Communications Earth and Environment*, *5*(1). https://doi.org/10.1038/s43247-024-01707-x

Wang, X., Verlaan, M., Apecechea, M. I., & Lin, H. X. (2022). Parameter estimation for a global tide and surge model with a memory-efficient order reduction approach. *Ocean Modelling*, *173*. https://doi.org/10.1016/j.ocemod.2022.102011

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79–82. https://doi.org/https://doi.org/10.3354/cr030079

Yamaguchi, M., Nakazawa, T., & Aonashi, K. (2012). Tropical cyclone track forecasts using JMA model with ECMWF and JMA initial conditions. *Geophysical Research Letters*, *39*(9). https://doi.org/10.1029/2012GL051473

Yumul, G. Jr., Cruz, N., Servando, N., & Dimalanta, C. (2011). Extreme weather events and related disasters in the Philippines, 2004-08: a sign of what climate change will mean? *Disasters*, *35*(2), 362–382. https://doi.org/10.1111/j.1467-7717.2010.01216.x.

Zhang, M., Zhou, C., Zhang, J., Zhang, X., & Tang, Z. (2022). Numerical simulation and analysis of storm surges under different extreme weather event and typhoon experiments in the South Yellow Sea. *Journal of Ocean University of China*, *21*(1). https://doi.org/10.1007/s11802-022-4750-7