Supplementary Information for

From regional to global: a typhoon-aware adaptive graph transformer for cross-basin tropical cyclone-induced storm surge forecasting

Contents

1 Training Details and Hyperparameters

All models are implemented in PyTorch with PyTorch Geometric. The full model contains 244,640 trainable parameters. Training uses the AdamW optimiser with a cosine annealing learning rate schedule. Early stopping is applied with a patience of 20 epochs based on validation RMSE. The complete hyperparameter configuration is summarised in Table 1.

Table S1: Complete hyperparameter configuration.

Component	**Setting**
Model	
Hidden dimension	64
GCN layers	2
Cross-attention layers	2
Forcing self-attention layers	2
Attention heads	4
Feed-forward dimension	256
Dropout	0.1
Residual weight α	0.1
Graph	
KNN k	3
Typhoon radius σ	500 km
Weight scale α_g	2.0
Distance decay β	0.001 km^{-1}
Loss	
Huber δ	5 cm
Extreme threshold	50 cm
λ_{NLL}	0.25
Training	
Optimiser	AdamW
Learning rate	10^{-3}
Weight decay	10^{-5}
Gradient clipping	1.0
Batch size	4
Max epochs	100
Early stopping patience	20

2 Dynamic Graph Formulation

Given a typhoon event with centre coordinates (ϕ_t, λ_t) (latitude and longitude) and normalised intensity $I_t \in [0, 1]$, the great-circle distance d_i from each station i to the typhoon centre is computed via the haversine formula:

$$d_i = 2R \cdot \arcsin \sqrt{\sin^2\left(\frac{\phi_t - \phi_i}{2}\right) + \cos\phi_t \cos\phi_i \sin^2\left(\frac{\lambda_t - \lambda_i}{2}\right)} \tag{1}$$

where $R = 6371$ km is Earth's radius and (ϕ_i, λ_i) are station coordinates. The typhoon proximity weight follows a Gaussian decay:

$$w_i = 1 + (\alpha_g - 1)\, I_t \exp\left(-\frac{d_i^2}{2\sigma^2}\right) \tag{2}$$

and the dynamic edge weight is:

$$e_{ij} = \exp(-\beta\, d_{ij})\sqrt{w_i\, w_j}. \tag{3}$$

The typhoon centre is obtained from IBTrACS best-track records by matching typhoon ID and timestamp; intensity is normalised from maximum sustained wind speed.

3 SWE-Inspired Residual Constraints

The depth-averaged shallow water equations (SWE) provide the first-order dynamical description of storm surge:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -g\frac{\partial \eta}{\partial x} + \frac{\tau_{wind,x}}{\rho H} - \frac{\tau_{bottom,x}}{\rho H} + fv, \tag{4}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -g\frac{\partial \eta}{\partial y} + \frac{\tau_{wind,y}}{\rho H} - \frac{\tau_{bottom,y}}{\rho H} - fu, \tag{5}$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial (Hu)}{\partial x} + \frac{\partial (Hv)}{\partial y} = 0, \tag{6}$$

where u, v are depth-averaged velocities, η is surge elevation, H is water depth, and f is the Coriolis parameter. A direct PINN-style residual is infeasible because tide gauges provide only η and surface forcing, without velocities, bathymetry, or bottom drag. We therefore distil two temporal correlation constraints.

Wind–surge response (continuity-inspired). Let $|U_{10}(t)| = \sqrt{u_{10}^2 + v_{10}^2}$. Normalised temporal derivatives:

$$\dot{\eta}_t = \frac{\partial_t \eta}{\sigma(\partial_t \eta) + \epsilon}, \qquad \dot{U}_t = \frac{\partial_t |U_{10}|}{\sigma(\partial_t |U_{10}|) + \epsilon}, \tag{7}$$

and the residual penalises decorrelation:

$$\mathcal{L}_{wind} = 1 - \frac{1}{T}\sum_t \dot{\eta}_t\, \dot{U}_t. \tag{8}$$

Inverse-barometer response (momentum-inspired). The IB effect gives $\eta_{IB} \approx -(p_{atm} - p_0)/(\rho g)$, approximately 1 cm per 1 hPa. We enforce anti-correlation:

$$\mathcal{L}_{IB} = 1 + \frac{1}{T}\sum_t \dot{\eta}_t\, \dot{p}_t, \tag{9}$$

where $\dot{p}_t = \partial_t\, msl/[\sigma(\partial_t\, msl) + \epsilon]$. The combined term is

$$\mathcal{L}_{swe} = \lambda_{swe}(\mathcal{L}_{wind} + 0.5\, \mathcal{L}_{IB}), \quad \lambda_{swe} = 0.01. \tag{10}$$

These residuals are disabled in the default configuration ($\lambda_{swe} = 0$) as the ablation study shows no significant improvement; they remain available as an extensible regularisation option.

4 Evaluation Metrics and Masking Protocol

Model performance is evaluated using:

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i), \tag{11}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}, \tag{12}$$

$$R = \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum(y_i - \bar{y})^2}\sqrt{\sum(\hat{y}_i - \bar{\hat{y}})^2}}. \tag{13}$$

Event-averaged RMSE. The primary metric is the event-averaged RMSE: for each (station, typhoon) event pair, the 12-step RMSE is first computed, then averaged across all events. This avoids the pooled RMSE being dominated by quiescent stations with near-zero surge.

Typhoon-aware spatial mask. Critically, all evaluation metrics are computed exclusively on stations dynamically selected by the typhoon-aware graph mask (mask == 1). A station is masked if it satisfies both conditions: (i) the great-circle distance from the station to the typhoon centre is less than 500 km, and (ii) the observed surge amplitude exceeds 10 cm. This excludes quiescent coastal stations where the model trivially predicts near-zero surge, ensuring that the metric reflects performance on surge-relevant predictions only. Typically 5–15 stations are masked per event.

Extreme-event RMSE is computed on observations exceeding 50 cm. Peak RMSE evaluates the maximum surge within each 12-h window. Probabilistic calibration is assessed via CRPS, 90%/95% coverage, and expected calibration error (ECE).

5 Complete Ablation Study

The ablation study comprises eight configurations on the full 12,128-sample dataset (50 epochs, batch size 16, seed 42). Table 2 reports all configurations with test-set metrics. The full model adopts the optimal configuration identified through ablation: typhoon-aware dynamic graph, dual-stream architecture, and Gaussian NLL uncertainty, without static physics features or physics-informed loss terms.

Key observations. (1) The dual-stream architecture provides the largest contribution: removing the temporal branch increases RMSE by 0.91 cm (7.71 vs. 6.80), while removing the spatial branch increases it by 0.45 cm (7.25 vs. 6.80). (2) The typhoon-aware dynamic graph improves overall RMSE by 0.15 cm (6.95→6.80), enabling adaptive global forecasting without region-specific graph retraining. (3) Physics features and physics-informed loss terms do not improve performance: adding them increases RMSE from 6.80 to 6.92–6.94. (4) The Gaussian NLL head provides a consistent 0.10 cm RMSE improvement (6.90→6.80) while also yielding calibrated predictive uncertainty. The final model therefore adopts the three-term loss (data + extreme + NLL) without physics features or physics-informed constraints.

Table S2: Complete ablation results on the global test set. Part 1 compares graph structures; Part 2 evaluates stream contributions; Part 3 examines feature and loss components. Best result in **bold**.

Model	RMSE (cm)	MAE (cm)	Bias (cm)	R	Extr. RMSE (cm)	Peak Err (cm)	Params
Part 1: Graph Structure							
Static KNN Graph	6.95	3.95	−0.21	0.87	21.52	4.09	242,036
Typhoon-Aware Dynamic Graph	**6.80**	**3.91**	−0.19	**0.88**	21.80	4.10	239,560
Part 2: Stream Architecture							
Spatial Branch Only	7.71	4.28	−0.32	0.84	25.62	4.49	21,064
Temporal Branch Only	7.25	4.09	−0.15	0.86	24.08	4.30	221,704
Dual-Stream (Ours)	**6.80**	**3.91**	−0.19	**0.88**	21.80	4.10	239,560
Part 3: Feature & Loss Components (dynamic + dual-stream)							
Full Model (+ NLL, no physics)	**6.80**	**3.91**	−0.19	**0.88**	21.80	4.10	239,560
w/o NLL (Huber only)	6.90	3.90	−0.21	0.87	22.91	4.06	239,560
+ Physics Features + Physics Losses	6.92	3.93	−0.18	0.87	21.49	3.98	242,036
+ Physics Features only	6.94	3.96	−0.17	0.87	21.49	4.00	242,036
+ Physics Features + Losses, w/o NLL	6.85	3.95	−0.16	0.87	20.62	3.95	242,036

6 Geographic Coverage and Extratropical Limitation

Figure 1 shows the global distribution of all 817 tide gauge stations overlaid with IBTrACS tropical cyclone tracks from 1993–2018. A critical observation is that the North Sea and Irish Sea (European coast) contain no tropical cyclone tracks in the IBTrACS best-track dataset. This region is dominated by extratropical cyclones (winter storms), a physically distinct mechanism characterised by larger spatial scales, longer durations, and different wind–surge coupling dynamics. Since the training dataset comprises only tropical cyclone events, the model has not learned extratropical surge dynamics. The elevated RMSE observed along the European coast in the main text (Fig. 3) is therefore attributable to the absence of tropical cyclone training data in this region, rather than a model deficiency. This limitation is explicitly acknowledged in the Discussion.

7 Additional Case Studies

7.1 Typhoon Fitow (2013)

For Fitow, the spatial-only baseline achieves RMSE 3.85 cm and $R = 0.87$; the temporal-only improves to 2.65 cm and $R = 0.91$; the full ST-GNNFormer achieves 2.43 cm and $R = 0.92$, a 36.9% improvement over spatial-only (Figure 2).

7.2 Typhoon Rananim (2004)

For Rananim, spatial-only yields RMSE 6.89 cm and $R = 0.86$; temporal-only yields 4.82 cm and $R = 0.89$; the full ST-GNNFormer achieves 4.10 cm and $R = 0.92$, a 40.5% improvement over spatial-only (Figure 3).

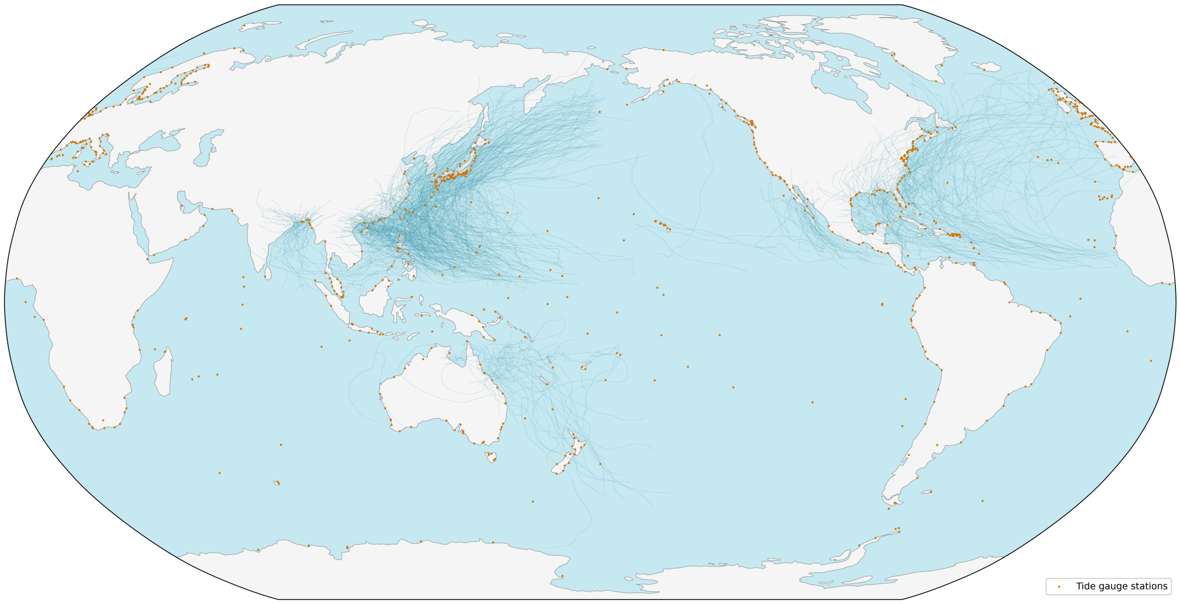

Figure S1: Global distribution of 817 tide gauge stations (blue dots) and IBTrACS tropical cyclone tracks (coloured lines) from 1993–2018. The map is centred on 150°W. Note the absence of tropical cyclone tracks in the North Sea and Irish Sea (European coast), where storm surges are dominated by extratropical cyclones.

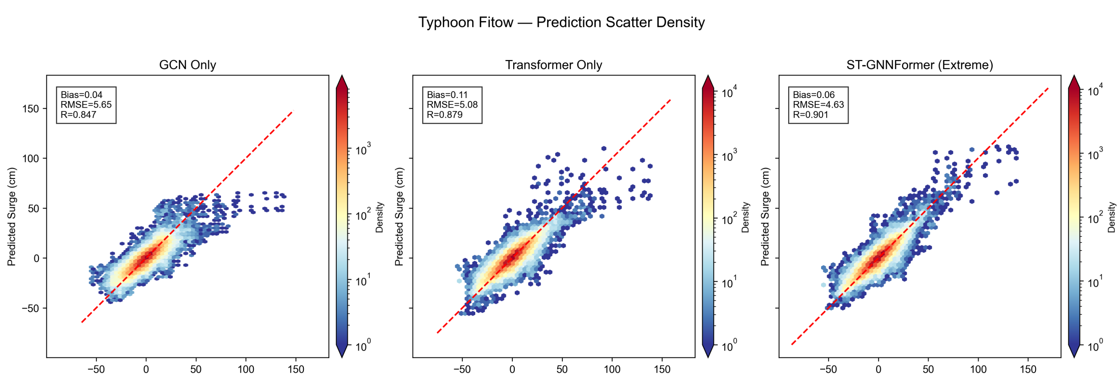

Figure S2: True-predicted scatter density for Typhoon Fitow (2013) across three model configurations.

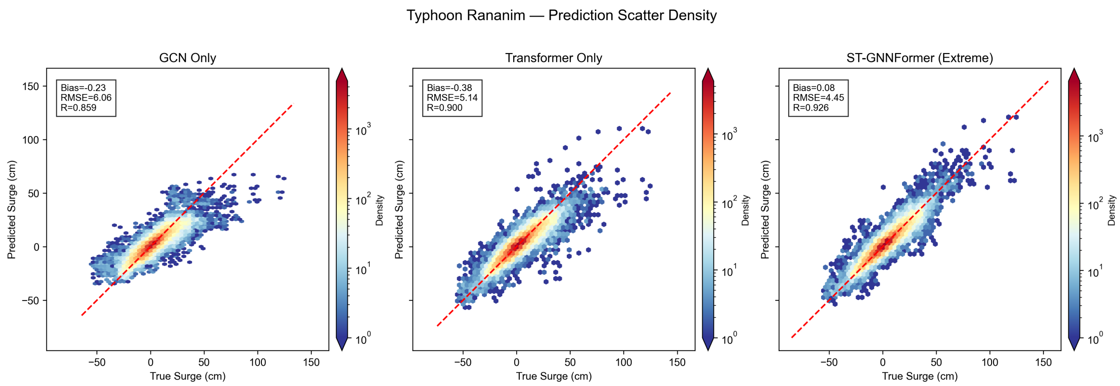

Figure S3: True-predicted scatter density for Typhoon Rananim (2004) across three model configurations.

Table S3: Statistics of selected tropical cyclones by ocean basin and maximum intensity.

Ocean Basin	Number	Prop.	Max. Intensity Category (number of storms)					
			TD	TS	Cat 1	Cat 2	Cat 3	Cat 4–5
Western North Pacific	442	56.7%	0	160	98	63	53	9
North Atlantic	184	23.6%	7	70	33	19	17	38
North Indian	65	8.3%	34	16	4	3	2	3
Eastern North Pacific	48	6.2%	2	15	5	5	4	17
South Pacific	38	4.9%	0	10	8	7	9	4
South Indian	3	0.4%	0	0	2	0	1	0
Total	780	100%	43	271	150	97	86	71

8 Dataset Statistics by Basin and Intensity

References

[1] 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. doi:10.1175/2009BAMS2755.1

[2] Muis, S., et al. (2016). A global reanalysis of storm surges and extreme sea levels. *Nature Communications*, 7, 13569. doi:10.1038/ncomms13569

[3] Hersbach, H., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. doi:10.1002/qj.3803

[4] Tadesse, T., et al. (2021). Global storm surge reconstruction (GSSR): A high-resolution dataset of historical storm surge events. *Scientific Data*, 8, 287. doi:10.1038/s41597-021-01071-5

[5] Nader, A., et al. (2026). StormNet: A graph neural network–LSTM hybrid for regional storm surge prediction. *Journal of Advances in Modeling Earth Systems*, 18, e2025MS004123. doi:10.1029/2025MS004123