Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Complementary Lag-Aligned Gauge Information and Pairwise Atmospheric Forcing for Coastal Water-Level Forecasting

Jia Rong

rong.j@kean.edu

Wenzhou-Kean University https://orcid.org/0009-0007-9772-7917

Guanchao Tong

tguancha@kean.edu

Wenzhou-Kean University https://orcid.org/0000-0002-3503-2745

Research Article

Keywords: coastal water-level forecasting, storm surge, multi-gauge learning, recurrent graph model, atmospheric forcing, extreme events

Posted Date: August 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10634756/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Complementary Lag-Aligned Gauge Information and Pairwise Atmospheric Forcing for Coastal Water-Level Forecasting

Jia Rong[1,2] Guanchao Tong[1,2]

[1]College of Science, Mathematics and Technology, Wenzhou–Kean University, 88 Daxue Road, Ouhai, Wenzhou, Zhejiang Province 325060, China
[2]College of Science, Mathematics and Technology, Kean University, 1000 Morris Avenue, Union, New Jersey 07083, USA
Jia Rong and Guanchao Tong contributed equally and are corresponding authors. Correspondence: rongj@kean.edu; tguancha@kean.edu.

Abstract

Extreme coastal water levels develop across several gauges during the same event, but the predictive value of a network depends on whether stations carry a shared atmospheric response or time-ordered information. We introduce CoastRelay, a recurrent graph framework that centres station states and represents lag-aligned gauge information and pairwise atmospheric forcing as separate inputs to a shared gated recurrent unit (GRU) forecast. Across the Thames–Southern North Sea, Charleston, and Gulf/Galveston networks, these inputs are combined in four otherwise identical graph configurations and evaluated for total water level and 95th-percentile (q95) mean absolute error over three random initialisations. The analysis reveals a regional complementarity regime: in the Thames–Southern North Sea, the joint representation adds upper-tail information beyond either single component, with an event-weighted q95 interaction of -0.0036 m (95% confidence interval, $[-0.0069\,\text{m}, -0.0003\,\text{m}]$). This finding shows that cross-gauge coupling is most useful when lag-aligned water-level states and spatial atmospheric contrasts describe distinct parts of the same high-water event. CoastRelay provides a basis for choosing network information pathways when extending coastal water-level forecasting to new gauge systems.

Keywords: coastal water-level forecasting; storm surge; multi-gauge learning; recurrent graph model; atmospheric forcing; extreme events

1 Introduction

Coastal water-level forecasts support flood preparedness, port operations, and navigation. A gauge record combines astronomical tide, local persistence, atmospheric pressure and wind response, and information shared with other gauges through the geometry and timing of the coastal system. During a high-water event, several gauges may respond to the same atmospheric disturbance, while a change at one station may precede a related change at another. These two forms of cross-gauge dependence carry different predictive content, and a useful network model should be able to examine them separately.

Multi-station neural models have improved water-level and storm-surge prediction in estuarine, coastal, and typhoon-affected waters [1–5]. Graph recurrent and attention models represent station relations through fixed, adaptive, or state-dependent connections [6–9], while atmospheric predictors supply information about the evolving disturbance [10, 11]. The usual multi-station comparison, however, changes several sources of information at once. It therefore leaves open whether an apparent graph benefit comes from synchronous station states, lag-aligned gauge states, pairwise atmospheric contrasts, or their combination.

This study asks whether lag-aligned gauge information and pairwise atmospheric forcing provide non-redundant predictive information after local station dynamics have been represented. Tide–surge

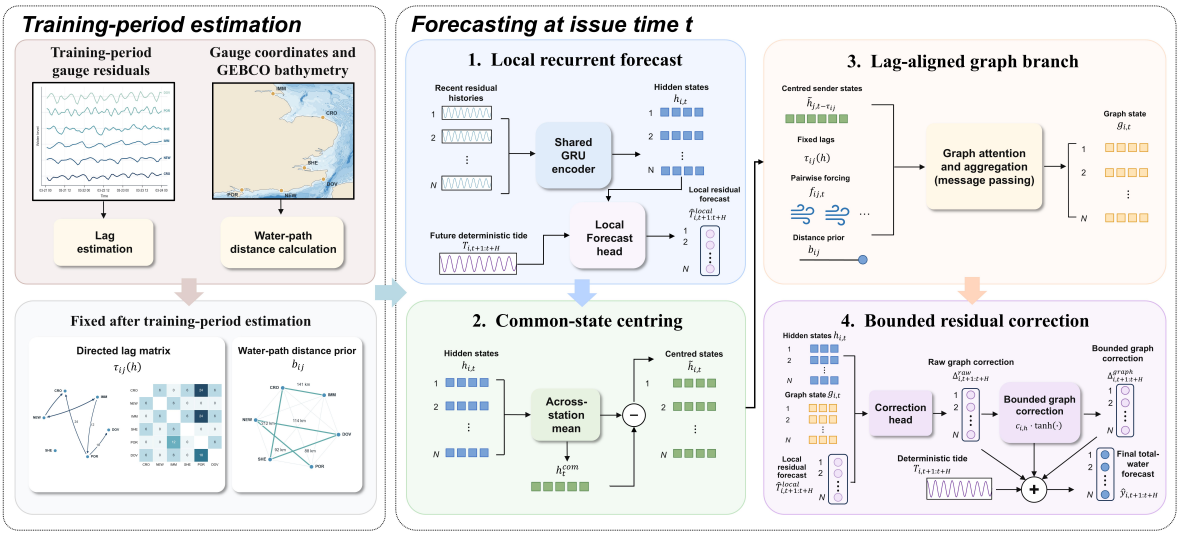


Figure 1: Method overview. Gauge residuals and bathymetry-derived water-path distances define the directed lag matrix and distance prior. At each issue time, the shared recurrent encoder produces local residual forecasts and station hidden states. Across-station centring subtracts the mean latent state before lag-aligned sender states, pairwise atmospheric forcing, and the distance prior enter graph attention. The graph contribution is bounded by the local error scale and combined with the local residual forecast and deterministic tide to obtain total water level.

interaction alters the timing and magnitude of North Sea extremes [12], while pressure, wind, tide, and coastal setting shape extreme water levels across the North Atlantic and Gulf of Mexico [10, 13, 14]. We expect the joint representation to be most informative in a network that contains both persistent nonzero-lag relations and spatially varying atmospheric conditions during high-water events.

We introduce CoastRelay to examine this question within one recurrent graph framework. A shared gated recurrent unit (GRU) encodes the local station history; centred station states provide the graph state, selected sender states enter at empirical lags, and ordered receiver–sender atmospheric features adjust the graph exchange. Four graph configurations combine zero-lag or lag-aligned exchange with the presence or absence of pairwise forcing, while the local forecast, graph structure, correction rule, and training objective remain unchanged across the three networks. Persistence and the station-wise recurrent forecast provide additional references. Figure 1 summarizes the forecasting workflow.

The central result is a regional complementarity pattern. In the Thames–Southern North Sea, lag-aligned gauge information and pairwise atmospheric forcing contribute distinct upper-tail information when used together. The three networks then provide contrasting settings in which to assess how this information relationship changes with the observed lag structure and atmospheric contrast. The primary upper-tail endpoint is the 95th-percentile (q95) event error, supported by total-water accuracy, detection, and lead-time profiles.

2 Study regions and data

2.1 Regional gauge networks

The three regional datasets cover 2015–2024 and use a common six-hour issue interval. The Thames–Southern North Sea network contains six gauges from the southern North Sea and adjacent English coast: Cromer (CRO), Newhaven (NEW), Immingham (IMM), Sheerness (SHE), Portsmouth (POR), and Dover (DOV). Charleston contains seven National Oceanic and Atmospheric Administration Center for Operational Oceanographic Products and Services (NOAA CO-OPS) gauges extending from Wilmington through Charleston and the southern Atlantic coast. Gulf/Galveston contains eight NOAA CO-OPS gauges

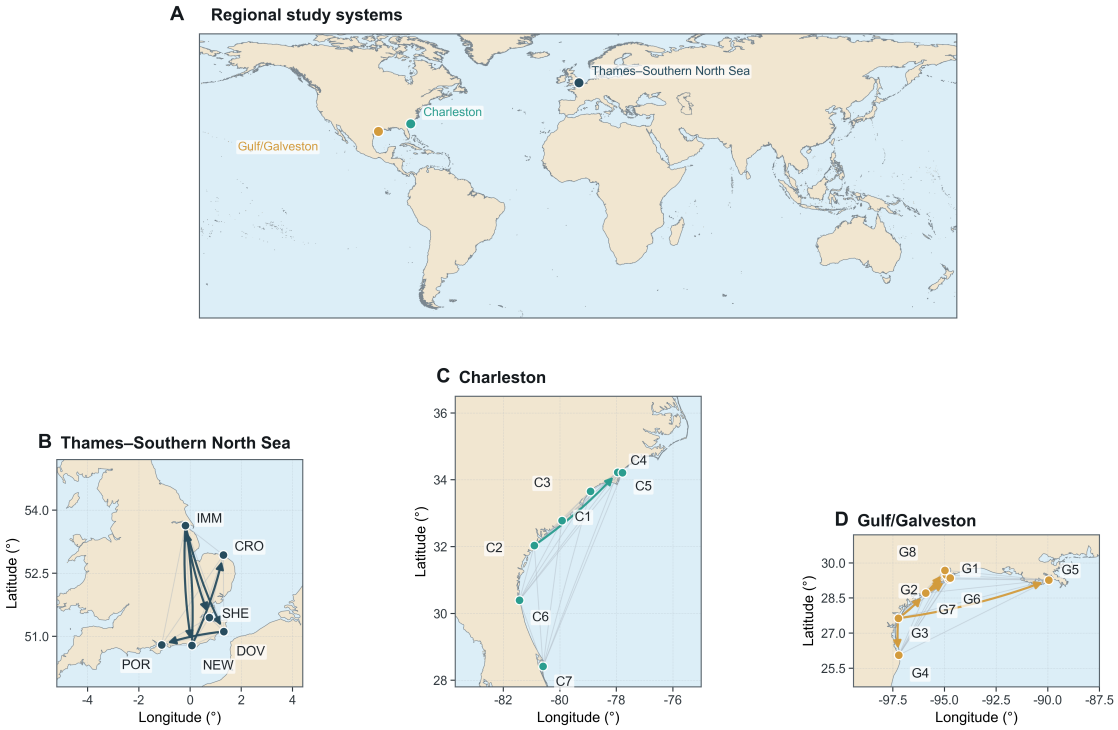


Figure 2: Geographic overview and regional network context. Panel A locates the three study systems. Panels B–D show station coordinates, Natural Earth 1:10m coastline and land polygons, and the complete distance-weighted directed graph. Pale lines show all directed non-self pairs. For each receiver with at least one positive-delay incoming pair, the coloured arrow marks the pair with the largest absolute Pearson association estimated from the training residuals. Thames labels are CRO (Cromer), NEW (Newhaven), IMM (Immingham), SHE (Sheerness), POR (Portsmouth), and DOV (Dover); the C1–C7 and G1–G8 labels are defined in Appendix Table A1.

spanning the Galveston area and the northern Gulf. The compact C1–C7 and G1–G8 labels used for the NOAA stations are defined with their station names and identifiers in Appendix Table A1. Table 1 gives the date ranges and forecast design; Figure 2 and Appendix Table A1 use the station ordering applied throughout the analysis.

Table 1: Regional data and forecast design. GESLA denotes Global Extreme Sea Level Analysis, and NOAA CO-OPS denotes the National Oceanic and Atmospheric Administration Center for Operational Oceanographic Products and Services. All dates are Coordinated Universal Time (UTC). An issue cycle is a complete 96-h history paired with its 72-h forecast target at one six-hour forecast origin. The edge count is the number of directed non-self pairs in the graph.

	Thames–Southern North Sea	Charleston	Gulf/Galveston
Water-level source	GESLA Version 4.0	NOAA CO-OPS	NOAA CO-OPS
Gauges	6	7	8
Period	2015–2024	2015–2024	2015–2024
Complete issue cycles before boundary separation	6999	13861	12704
Issue cycles excluded at split boundaries	48	24	24
Train / validation / test issue cycles	5587 / 676 / 688	11069 / 1383 / 1385	10144 / 1268 / 1268
History / horizon (h)	96 / 72	96 / 72	96 / 72
Issue stride (h)	6	6	6
Chronological split	8:1:1	8:1:1	8:1:1
Training dates (UTC)	2015-01-05 to 2021-11-22	2015-01-14 to 2023-01-30	2015-12-25 to 2023-03-29
Validation dates (UTC)	2021-11-28 to 2022-06-09	2023-02-03 to 2024-01-14	2023-04-02 to 2024-02-12
Test dates (UTC)	2022-06-15 to 2024-09-09	2024-01-17 to 2024-12-28	2024-02-16 to 2024-12-28
Boundary embargo (h)	72	72	72
95th-percentile (q95) event clusters in paired analysis	13	11	15
99th-percentile (q99) event clusters in paired analysis	6	7	6
Missing values in included windows	none	none	none
Gauge quality control	GESLA flags	NOAA CO-OPS flags	NOAA CO-OPS flags
Directed non-self graph pairs	30	42	56

2.2 Water-level observations, tide, and atmospheric fields

The six Thames observations are drawn from the Global Extreme Sea Level Analysis (GESLA) Version 4.0 release, accessed on 2 August 2026 [15, 16]. Station-specific gauge datum metadata and the provider quality-control fields are preserved before the regional timeline is formed. Charleston and Gulf/Galveston use station observations and metadata from NOAA CO-OPS [17, 18]. Source timestamps remain in Coordinated Universal Time, and the experiment uses complete windows containing native gauge histories, targets, deterministic tide, and issue-time forcing at every included station.

The target used for model fitting is the non-tidal residual

$$r_{i,t} = y_{i,t} - T_{i,t}, \tag{1}$$

where $y_{i,t}$ is the observed water level and $T_{i,t}$ is the deterministic tide at station i. For the NOAA regions, $T_{i,t}$ is the official station tide prediction supplied by NOAA CO-OPS. For Thames, the deterministic tide is obtained from a harmonic fit to the chronological training period with an intercept, linear trend, and the constituents M2, S2, N2, K1, O1, P1, Q1, and M4:

$$T_{i,t} = \beta_{i,0} + \beta_{i,1} t + \sum_{k \in \mathcal{K}} \left[a_{i,k} \cos(\omega_k t) + b_{i,k} \sin(\omega_k t) \right]. \tag{2}$$

The tide is used as an auxiliary future feature and is added back to the residual prediction for total-water metrics. Source datum information is preserved at station level: the Thames residual follows each gauge's official source datum, while the NOAA regional panels use mean sea level metadata.

Atmospheric information comes from the fifth-generation European Centre for Medium-Range Weather Forecasts reanalysis (ERA5) [19, 20]. The four forcing channels are the 10-m eastward and northward

winds u_{10} and v_{10} in $\mathrm{m\,s^{-1}}$, mean-sea-level-pressure anomaly in Pa, and wind-stress magnitude in Pa ($\mathrm{N\,m^{-2}}$). The hourly ERA5 fields remain on their native 0.25° grid. Each station is assigned its nearest grid cell, and the values valid at the corresponding issue time enter the model. The pressure anomaly is measured relative to the training-period mean at that grid cell. Wind-stress magnitude is calculated as

$$\tau_w = \rho_a C_D \left(u_{10}^2 + v_{10}^2\right), \qquad \rho_a = 1.225\ \mathrm{kg\,m^{-3}}, \quad C_D = 1.3 \times 10^{-3}. \tag{3}$$

This study is a historical hindcast using ERA5 fields valid at each forecast issue time; no atmospheric fields after that time enter the model. The processed atmospheric fields preserve their physical units. Before entering the pair-forcing encoder, each forcing channel is standardized within region using the training-period mean and standard deviation across all stations and issue cycles. The transformation is

$$\widetilde{f}_{r,t,c} = \frac{f_{r,t,c} - \mu_{r,c}^{\mathrm{train}}}{\sigma_{r,c}^{\mathrm{train}}}, \tag{4}$$

and the pair feature is $[\widetilde{\boldsymbol{f}}_{i,t}, \widetilde{\boldsymbol{f}}_{j,t}, \widetilde{\boldsymbol{f}}_{i,t} - \widetilde{\boldsymbol{f}}_{j,t}]$. The same training statistics are used for validation and test issue cycles. This region-level transformation preserves the relative atmospheric contrast between stations while placing the four channels on a common scale. A constant positive rescaling of the wind-stress channel, such as a change in the drag coefficient, cancels analytically in Eq. 4 and leaves the standardized pair feature unchanged.

The geographic background uses Natural Earth 1:10m coastline and land polygons [21]. Water-path distances used by the graph prior are derived from a General Bathymetric Chart of the Oceans (GEBCO) 2026 elevation grid [22]. The bathymetry is converted to an ocean mask, aggregated to the regional construction grid, and used in an eight-neighbour water-path calculation. All directed non-self station pairs remain in the mask, and bathymetry-derived distance sets the prior weight.

2.3 Chronological sampling and event definition

Each issue cycle contains a 96-h history and a 72-h target horizon. Issue cycles are spaced by six hours. The records are split chronologically into training, validation, and test periods in an 8:1:1 ratio, with a 72-h boundary separation. This separation excludes 48 issue cycles in Thames and 24 in each United States region; the resulting dates and partition counts are listed in Table 1. Tide parameters, pressure reference values, lag estimates, graph priors, thresholds, and correction scales are estimated from the training period.

The shorter analysed period in the Thames reflects gaps in which at least one of the six gauge records does not span a complete 96-h history and 72-h target window. Table 1 reports the available issue cycles before boundary separation and the separate number removed at split boundaries.

Station-specific 95th- and 99th-percentile thresholds (q95 and q99) are calculated from total water level in the training period, following established treatment of extreme still-water levels in coastal studies [23]. Unique hourly exceedance times are grouped into event clusters when successive exceedance times are separated by no more than 48 h; the cluster peak defines the event time. Forecast windows sharing an event identifier are aggregated within one event-level comparison, including overlapping six-hour issue cycles. The paired estimator retains only clusters with complete predictions for all four graph configurations. The full Thames test catalogue contains 13 q95 and 7 q99 clusters; the paired interaction table contains 13 q95 and 6 q99 clusters because one q99 cluster lacks complete route coverage. q99 is reported as a complementary upper-tail metric because its effective event sample is smaller in the resampling analysis.

3 CoastRelay methodology

3.1 Local recurrent forecast

The model receives six history channels at every station: residual water level, residual first difference, deterministic tide, tide first difference, observed water level, and observed-water-level first difference.

Water-level channels are expressed in metres, and the forcing representation is defined in Section 2.2. Let $\boldsymbol{x}_{i,t-L+1:t}$ denote the six-channel history for station i. A shared one-layer gated recurrent unit (GRU) encodes this sequence according to the standard PyTorch formulation [24]; the complete recurrence and numerical model settings are given in Appendix B.1 and Appendix Table A2. All analysed histories and targets are complete, and no target values are interpolated.

Let $\boldsymbol{h}_{i,t} \in \mathbb{R}^{d_h}$ be the final recurrent state and $\boldsymbol{T}_{i,t+1:t+H} \in \mathbb{R}^H$ the deterministic future tide. The local head concatenates these vectors and produces all residual leads in one pass:

$$\widehat{\boldsymbol{r}}^{\text{local}}_{i,t+1:t+H} = W_{\text{local}}[\boldsymbol{h}_{i,t}; \boldsymbol{T}_{i,t+1:t+H}] + \boldsymbol{b}_{\text{local}}, \quad W_{\text{local}} \in \mathbb{R}^{H\times(d_h+H)}. \tag{5}$$

Each lead has its own output weights. The local recurrent component is fitted once for each regional initialisation before the four graph configurations are optimized against the same local forecast.

3.2 Common-response centring

At each issue time, the hidden states are centred across the stations. The common latent response is

$$\boldsymbol{h}^{\text{com}}_t = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{h}_{i,t}, \qquad \widetilde{\boldsymbol{h}}_{i,t} = \boldsymbol{h}_{i,t} - \boldsymbol{h}^{\text{com}}_t. \tag{6}$$

The same centring operation is applied to the historical hidden states used for delayed lookup. The graph branch receives a station-relative state while the local forecast retains the raw recurrent state.

3.3 Graph prior and directed lag lookup

The study graph uses a receiver-by-sender hard mask M_{ij}, where $M_{ij} = 1$ denotes an admissible pair and j is the sender to receiver i. All diagonal entries are zero. The distance prior is

$$a_{ij} = M_{ij} \max\left[\exp\left(-\frac{\max(d_{ij}, 0)}{200\ \text{km}}\right), 10^{-6}\right], \tag{7}$$

where d_{ij} is the GEBCO-derived water-path distance. Row normalization and conversion to a log attention bias give

$$b_{ij} = \log\left[\max\left(\frac{a_{ij}}{\max\left(\sum_k a_{ik}, 10^{-8}\right)}, 10^{-6}\right)\right], \qquad b_{ij} = -10^9 \text{ when } M_{ij} = 0. \tag{8}$$

Directed lag estimates are calculated from training residuals. For receiver i, sender j, and candidate lag set $\mathcal{L} = \{0, 6, 12, 18, 24\}$ h, the lag score is

$$c_{ij}(\ell) = \text{corr}_{\text{train}}\left(r_i(t), r_j(t-\ell)\right), \qquad \tau_{ij} = \underset{\ell\in\mathcal{L}}{\text{argmax}}\, |c_{ij}(\ell)|. \tag{9}$$

The model receives the lag matrix in receiver-row/sender-column form; receiver i reads the centred sender state $\widetilde{\boldsymbol{h}}_{j,t-\tau_{ij}}$. Figure 4 shows all pairwise delay and lag-association values in the same orientation.

3.4 Ordered pair forcing and attention

Let $\widetilde{\boldsymbol{f}}_{i,t}$ denote the four-channel standardized forcing vector defined in Eq. 4. The ordered pair forcing is

$$\boldsymbol{f}_{ij,t} = \left[\widetilde{\boldsymbol{f}}_{i,t},\ \widetilde{\boldsymbol{f}}_{j,t},\ \widetilde{\boldsymbol{f}}_{i,t} - \widetilde{\boldsymbol{f}}_{j,t}\right]. \tag{10}$$

It has twelve channels and changes when the receiver and sender are exchanged. A multilayer perceptron maps this vector to a scalar adjustment of the attention score. The centred current state is regularised before entering the graph branch using the dropout setting in Appendix Table A2,

$$\check{\boldsymbol{h}}_{i,t} = \mathcal{D}(\widetilde{\boldsymbol{h}}_{i,t}), \tag{11}$$

where $\mathcal{D}$ denotes the dropout operator. The attention score is

$$e_{ij,t} = \frac{Q(\check{\boldsymbol{h}}_{i,t})^{\top} K(\widetilde{\boldsymbol{h}}_{j,t-\tau_{ij}})}{\sqrt{d_h}} + b_{ij} + g_\theta(\boldsymbol{f}_{ij,t}), \tag{12}$$

where d_h is the hidden-state dimension. The matched base and Delay-only configurations set the pair-forcing term to zero. Pair-forcing-only retains the same dynamic state attention and uses $\tau_{ij} = 0$ for every pair.

The distance mask removes self-connections and identifies the admissible receiver–sender pairs. The attention weights and message are

$$\alpha_{ij,t} = \mathrm{softmax}_j \left(e_{ij,t}; M_{ij}\right), \tag{13}$$

$$\boldsymbol{q}_{i,t} = \sum_j \alpha_{ij,t} V(\widetilde{\boldsymbol{h}}_{j,t-\tau_{ij}}), \tag{14}$$

$$\boldsymbol{g}_{i,t} = \mathrm{GRUCell}(\boldsymbol{q}_{i,t}, \check{\boldsymbol{h}}_{i,t}). \tag{15}$$

The query, key, and value projections operate in the attention space specified in Appendix Table A2. The dropped current state enters both the query projection and the graph-state update; the delayed history states remain centred and undropped.

3.5 Bounded residual correction and training loss

The correction head concatenates the local state, graph state, and local forecast and produces one raw correction for each forecast lead:

$$\boldsymbol{\Delta}^{\mathrm{raw}}_{i,t+1:t+H} = W_2\, \phi\Big(W_1[\boldsymbol{h}_{i,t}; \boldsymbol{g}_{i,t}; \widehat{\boldsymbol{r}}^{\mathrm{local}}_{i,t+1:t+H}] + \boldsymbol{b}_1\Big) + \boldsymbol{b}_2, \tag{16}$$

where ϕ is the activation specified in Appendix Table A2. The station-and-horizon correction scale is the 90th percentile of the absolute local forecast error in the training period:

$$c_{i,h} = Q_{0.90}\left(\left|r_{i,t+h} - \widehat{r}^{\mathrm{local}}_{i,t+h}\right|\right). \tag{17}$$

The final residual and total-water forecasts are

$$\Delta^{\mathrm{graph}}_{i,t+h} = c_{i,h} \tanh\left(\Delta^{\mathrm{raw}}_{i,t+h}\right), \tag{18}$$

$$\widehat{r}_{i,t+h} = \widehat{r}^{\mathrm{local}}_{i,t+h} + \Delta^{\mathrm{graph}}_{i,t+h}, \tag{19}$$

$$\widehat{y}_{i,t+h} = T_{i,t+h} + \widehat{r}_{i,t+h}. \tag{20}$$

The training objective is the valid-position mean squared residual error plus a tail-weighted mean squared error:

$$\mathcal{L} = \frac{1}{|\mathcal{V}|} \sum_{(i,t,h)\in\mathcal{V}} (\widehat{r}_{i,t+h} - r_{i,t+h})^2 + \lambda_{\mathrm{tail}} \frac{1}{|\mathcal{V}_{95}|} \sum_{(i,t,h)\in\mathcal{V}_{95}} (\widehat{r}_{i,t+h} - r_{i,t+h})^2, \tag{21}$$

where $\mathcal{V}$ contains all station, origin, and lead positions in the mini-batch, $\mathcal{V}_{95}$ contains positions whose total-water target exceeds the station-specific q95 threshold, and $\lambda_{\mathrm{tail}} = 0.1$. The second term is zero for a batch containing no q95 exceedance.

3.6 Four graph configurations

Table 2 summarizes the scientific role of each configuration. Persistence carries the latest residual through the horizon and restores the future tide,

$$\widehat{y}^{\mathrm{pers}}_{i,t+h} = T_{i,t+h} + r_{i,t}, \qquad h = 1, \ldots, 72. \tag{22}$$

Local-GRU is the station-wise recurrent forecast. The four graph configurations share the same recurrent encoder, dynamic attention, common-response centring, distance prior, and bounded correction. The two experimental settings are the lag lookup and the pair-forcing term. For each region and random initialisation, all four configurations use the same fitted Local-GRU checkpoint and the same initialisation of the shared graph parameters.

Table 2: Forecast configurations and the component comparison addressed by each one.

Configuration	Directed delay	Pair forcing	Common mode	Dynamic attention	Correction	Role
Persistence	–	–	–	–	–	Recent local-state reference
Local-GRU	–	–	–	–	–	Station-wise recurrent reference
Matched base	0 h	–	yes	yes	bounded	Graph reference with simultaneous exchange
Delay-only	τ_{ij}	–	yes	yes	bounded	Lag-aligned inter-gauge component
Pair-forcing-only	0 h	yes	yes	yes	bounded	Ordered atmospheric component
CoastRelay	τ_{ij}	yes	yes	yes	bounded	Joint representation of the two components

4 Experimental design and evaluation

Each regional model variant is fitted from the same three random initialisations. The optimizer, learning schedule, batch size, stopping rule, and gradient handling are listed in Appendix Table A2. Optimization stops after the stated validation patience, and the checkpoint with the lowest validation loss is evaluated on the test period. The regional split, station order, history, horizon, and loss are identical across configurations.

Overall performance is measured by root mean squared error (RMSE) and mean absolute error (MAE) for total water level. Extreme-water performance is summarized with q95 and q99 MAE and the point-level F1 (F-measure) score; the precision, recall, and F-measure terminology follows standard classification definitions [25]. The threshold is station-specific and comes from the training total-water distribution. Figure 5 reports RMSE and q95 MAE by lead time at 6, 12, 24, 48, and 72 h.

For issue cycle o, station i, and lead h, let $e_{i,o,h} = \widehat{y}_{i,o,h} - y_{i,o,h}$ denote the total-water error. Over the valid positions $\mathcal{V}$,

$$\mathrm{RMSE} = \left(\frac{1}{|\mathcal{V}|} \sum_{(i,o,h)\in\mathcal{V}} e_{i,o,h}^2 \right)^{1/2}, \tag{23}$$

$$\mathrm{MAE} = \frac{1}{|\mathcal{V}|} \sum_{(i,o,h)\in\mathcal{V}} |e_{i,o,h}|. \tag{24}$$

For quantile $q \in \{0.95, 0.99\}$, let $\mathcal{V}_q = \{(i,o,h) \in \mathcal{V} : y_{i,o,h} \geq \theta_{i,q}\}$, where $\theta_{i,q}$ is the station-specific training-period total-water threshold. The pooled tail error is

$$\mathrm{MAE}_q = \frac{1}{|\mathcal{V}_q|} \sum_{(i,o,h)\in\mathcal{V}_q} |e_{i,o,h}|. \tag{25}$$

The corresponding point-level F1 score is

$$\mathrm{F1}_q = \frac{2\,\mathrm{TP}_q}{2\,\mathrm{TP}_q + \mathrm{FP}_q + \mathrm{FN}_q}, \tag{26}$$

where a positive observation satisfies $y_{i,o,h} \geq \theta_{i,q}$ and a positive forecast satisfies $\widehat{y}_{i,o,h} \geq \theta_{i,q}$. Thus q95/q99 MAE uses target-exceedance positions, whereas q95/q99 F1 classifies every valid issue-cycle,

station, and lead position. The pooled metrics retain the six-hour issue origins; event identifiers are used to aggregate the paired event effects in Eq. 27, so overlapping forecasts from the same event receive one event-level contribution.

For a candidate configuration m and a reference configuration b, the event-weighted paired effect is

$$\Delta_{95}(m, b) = \frac{1}{S} \sum_{s=1}^{S} \frac{1}{E_s} \sum_{e=1}^{E_s} \left[\mathrm{MAE}_{95,e,s}(m) - \mathrm{MAE}_{95,e,s}(b) \right], \tag{27}$$

where s indexes the three random initialisations and e indexes q95 event clusters. A negative value indicates lower event-weighted tail error for the candidate. Confidence intervals are obtained from 2,000 hierarchical resamples over random initialisation and event cluster, following standard bootstrap practice for environmental time series [26, 27]. The pooled q95 MAE first pools all valid target-exceedance positions within a random initialisation, whereas Eq. 27 first averages within each event and then gives equal weight to events and random initialisations.

The joint interaction for tail quantile q is

$$I_q = E_q(\text{CoastRelay}) - E_q(\text{Delay-only}) - E_q(\text{Pair-forcing-only}) + E_q(\text{Matched base}), \tag{28}$$

where E_q is the event-weighted tail MAE. A negative value indicates an additional error reduction when lag-aligned gauge information and pairwise forcing are combined. The interaction uses the same hierarchical resampling as the direct paired effects.

5 Results

5.1 Overall performance and tail behaviour

All four graph configurations reduce RMSE relative to Local-GRU in each network, while the clearest separation in q95 performance occurs in the Thames–Southern North Sea (Table 3). Values are the mean and standard deviation over three random initialisations; Persistence is deterministic and is reported as a single value. The comparison also shows that overall accuracy and upper-tail accuracy can favour different references, particularly in Gulf/Galveston.

In the Thames–Southern North Sea, CoastRelay has the lowest q95 MAE (0.142 m) and the highest q95 F1 among the four graph configurations. Its q95 MAE is lower than Delay-only (0.148 m) and the standardized Pair-forcing-only configuration (0.144 m). Pair-forcing-only has the lowest mean overall RMSE (0.208 m), while CoastRelay is 0.209 m and has the lowest q99 MAE (0.147 m). The joint model's clearest gain is therefore concentrated in the upper tail.

Charleston's four q95 MAE values lie between 0.080 m and 0.082 m, with the zero-lag base at the lowest mean. Gulf/Galveston also has a compact q95 range, from 0.122 m to 0.124 m; the standardized CoastRelay configuration has the lowest mean q95 MAE, while the joint interaction remains close to zero. These close rankings are resolved by the paired interaction estimates in Table 4.

5.2 Paired tail effects

The paired comparison identifies an additional q95 reduction from combining lag alignment and pair forcing in the Thames–Southern North Sea (Table 4). With standardized forcing inputs, CoastRelay lowers event-weighted q95 error by 0.0027 m relative to Pair-forcing-only. The q95 interaction is −0.0036 m, with a 95% interval of [−0.0069 m, −0.0003 m].

The corresponding effects are small in Charleston and Gulf/Galveston, and both interaction intervals include zero. Their single-component and joint configurations therefore remain closely grouped under the event-weighted q95 comparison.

At the event level, CoastRelay has a lower q95 error than the better single-component configuration in 7 of the 13 Thames events. The median effect is −0.46 mm, and the largest event-level reduction is −19.57 mm (Appendix Figure A1).

Table 3: Test-period performance averaged over three random initialisations. Pairwise atmospheric forcing uses region–channel training standardization. Graph-stage values are mean ± standard deviation; Persistence is deterministic and is reported as a single value. Lower is better for RMSE and q95/q99 MAE; higher is better for q95/q99 F1.

Region	Configuration	RMSE (m)	MAE (m)	q95 MAE (m)	q95 F1	q99 MAE (m)	q99 F1
Thames–Southern North Sea	Persistence	0.455	0.352	0.318	0.608	0.301	0.363
	Local-GRU	0.225 ± 0.002	0.170 ± 0.001	0.155 ± 0.001	0.774 ± 0.004	0.165 ± 0.005	0.603 ± 0.005
	Matched base	0.209 ± 0.002	0.156 ± 0.001	0.147 ± 0.001	0.787 ± 0.003	0.159 ± 0.003	0.622 ± 0.009
	Delay-only	0.210 ± 0.000	0.158 ± 0.000	0.148 ± 0.000	0.786 ± 0.002	0.162 ± 0.003	0.613 ± 0.003
	Pair-forcing-only	0.208 ± 0.001	0.156 ± 0.000	0.144 ± 0.000	0.792 ± 0.001	0.150 ± 0.005	0.634 ± 0.012
	CoastRelay	0.209 ± 0.001	0.156 ± 0.000	0.142 ± 0.001	0.795 ± 0.003	0.147 ± 0.002	0.629 ± 0.002
Charleston	Persistence	0.215	0.166	0.167	0.598	0.182	0.418
	Local-GRU	0.114 ± 0.000	0.084 ± 0.000	0.081 ± 0.001	0.774 ± 0.001	0.096 ± 0.002	0.682 ± 0.002
	Matched base	0.112 ± 0.000	0.082 ± 0.000	0.080 ± 0.001	0.778 ± 0.001	0.095 ± 0.002	0.691 ± 0.001
	Delay-only	0.112 ± 0.000	0.082 ± 0.000	0.080 ± 0.001	0.777 ± 0.001	0.096 ± 0.002	0.690 ± 0.002
	Pair-forcing-only	0.112 ± 0.000	0.082 ± 0.000	0.082 ± 0.000	0.778 ± 0.001	0.099 ± 0.001	0.685 ± 0.002
	CoastRelay	0.112 ± 0.000	0.082 ± 0.000	0.081 ± 0.001	0.778 ± 0.001	0.099 ± 0.002	0.683 ± 0.002
Gulf/Galveston	Persistence	0.137	0.090	0.129	0.658	0.178	0.490
	Local-GRU	0.118 ± 0.001	0.078 ± 0.001	0.131 ± 0.004	0.652 ± 0.011	0.205 ± 0.009	0.474 ± 0.024
	Matched base	0.116 ± 0.000	0.076 ± 0.000	0.124 ± 0.003	0.673 ± 0.006	0.193 ± 0.007	0.502 ± 0.016
	Delay-only	0.116 ± 0.001	0.076 ± 0.000	0.123 ± 0.002	0.673 ± 0.005	0.193 ± 0.005	0.499 ± 0.014
	Pair-forcing-only	0.115 ± 0.002	0.076 ± 0.001	0.124 ± 0.003	0.675 ± 0.008	0.194 ± 0.007	0.498 ± 0.016
	CoastRelay	0.115 ± 0.001	0.076 ± 0.001	0.122 ± 0.003	0.677 ± 0.007	0.191 ± 0.007	0.505 ± 0.016

Table 4: Paired q95 comparisons from 2,000 hierarchical random-initialisation/event-cluster resamples with standardized forcing inputs. The CoastRelay comparison with the single-component configuration having the lower pooled q95 MAE is descriptive; the joint interaction is the primary inferential contrast. The interaction is CoastRelay − Delay-only − Pair-forcing-only + Matched base. Negative values indicate lower event-weighted q95 error. Event counts are constant across the three random initialisations within each region.

Region	Comparison	Effect (m)	95% interval low (m)	95% interval high (m)	Random initialisations	Event clusters
Thames–Southern North Sea	CoastRelay − Pair-forcing-only	-0.0027	-0.0052	-0.0005	3	13
	Joint interaction	-0.0036	-0.0069	-0.0003	3	13
Charleston	CoastRelay − Delay-only	-0.0001	-0.0012	0.0010	3	11
	Joint interaction	-0.0012	-0.0032	0.0008	3	11
Gulf/Galveston	CoastRelay − Delay-only	-0.0012	-0.0031	0.0006	3	15
	Joint interaction	-0.0011	-0.0031	0.0007	3	15

Table 5: Regional signal diagnostics. The nonzero-delay proportion is the fraction of directed non-self pairs with a selected delay above zero. Median nonzero delay is calculated only over pairs whose selected delay is greater than zero. Mean absolute lag correlation is averaged over all directed non-self pairs. Forcing contrast is the mean Euclidean receiver–sender contrast across forecast origins associated with q95 events after each forcing channel is scaled by its training-period standard deviation. Joint gain is the best single-component q95 MAE minus the CoastRelay q95 MAE; positive values favour the joint configuration. The primary model results use the standardized forcing representation described in Section 2.2.

Region	Nonzero-delay pairs	Median nonzero delay (h)	Mean $\lvert c_{ij} \rvert$	Forcing contrast	Issue cycles linked to q95 events	Joint gain (m)
Thames–Southern North Sea	56.7%	6	0.356	0.458	505	+0.0016
Charleston	4.8%	24	0.738	0.582	1147	-0.0015
Gulf/Galveston	35.7%	6	0.744	0.511	850	+0.0009

5.3 Regional signal diagnostics

Thames has the broadest nonzero-delay coverage and a positive pooled gain over its better single-component configuration (Table 5). Charleston combines strong lag correlations with very few nonzero selected delays, so most graph messages remain synchronous. Gulf/Galveston has broader nonzero-lag coverage and a substantial forcing contrast; its standardized joint q95 mean is close to the single-component values, and the interaction interval includes zero.

5.4 Observed event context and lag structure

The event examples show close tracking of the oscillatory Thames response and tighter grouping of the model trajectories in Charleston and Gulf/Galveston (Figure 3). At Immingham, the trajectories follow the observed peaks and the joint configuration remains near the upper-tail threshold. Wrightsville Beach has a smaller total-water range and closely grouped predictions. At Morgans Point, the model variants capture the regular component but remain more tightly grouped than the observation during the final rise.

Figure 4 displays the full receiver-by-sender lag matrices and their absolute correlation magnitudes. The selected values are discrete because the lag search evaluates 0, 6, 12, 18, and 24 h. The Thames matrix includes several nonzero delays distributed across the station order, whereas Charleston and Gulf/Galveston contain many zero-hour entries alongside a smaller number of positive delays. Figure 2 highlights one positive-delay incoming pair per receiver; Figure 4 shows every matrix entry.

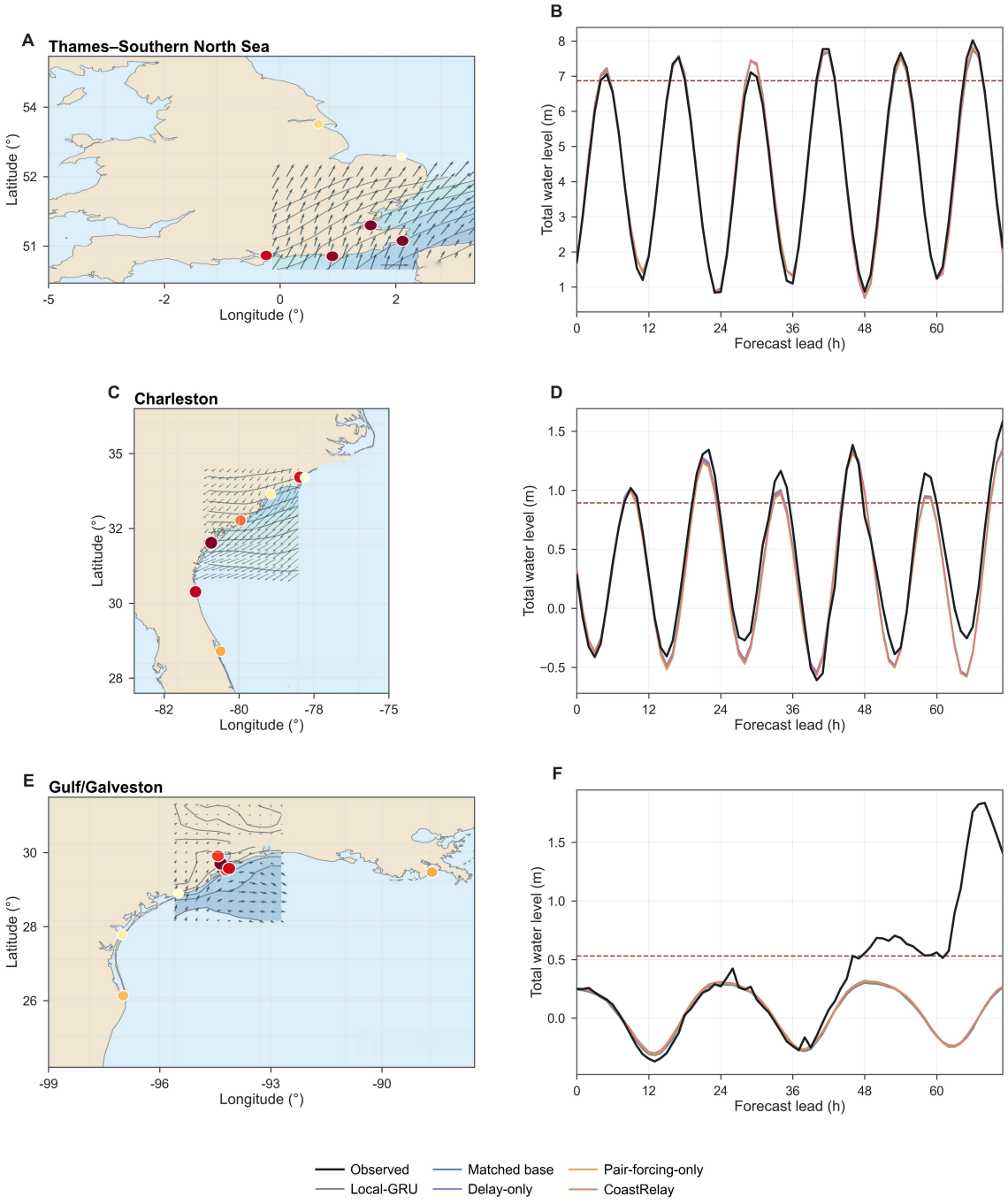


Figure 3: Representative test-period q95 event responses. The examples are 21 March 2023 00:00 UTC at Immingham, 13 November 2024 12:00 UTC at Wrightsville Beach, and 5 July 2024 18:00 UTC at Morgans Point. Panels A, C, and E show the regional coastline, gauge locations, and ERA5 10-m wind vectors at the selected issue time. Panels B, D, and F show observed total water level in metres and mean predictions across three random initialisations. Thames water levels retain the station-specific official gauge datum; NOAA records use mean sea level. Shaded bands indicate one standard deviation for the graph configurations, and dashed lines mark station-specific q95 thresholds.

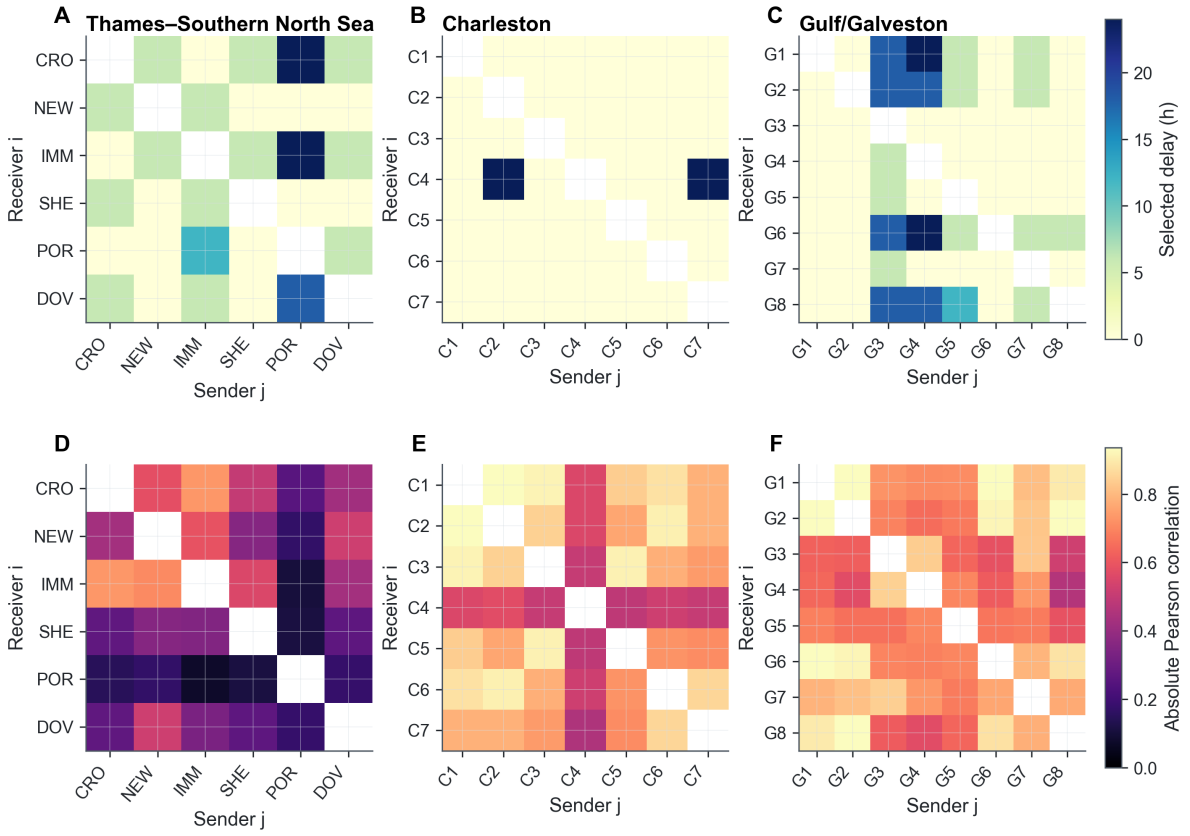


Figure 4: Directed lag structure for the three regional networks. For each receiver–sender pair, the selected delay maximizes the absolute Pearson association of training-period residuals over 0, 6, 12, 18, and 24 h. Panels A–C show the selected delays with receiver rows and sender columns; Panels D–F show the corresponding association magnitudes. Diagonal entries are omitted. In the Thames panels, CRO denotes Cromer, NEW Newhaven, IMM Immingham, SHE Sheerness, POR Portsmouth, and DOV Dover; the C1–C7 and G1–G8 labels are defined in Appendix Table A1.

5.5 Dependence on forecast lead time

Figure 5 shows that all graph configurations have their largest RMSE improvement at the shortest reported lead. In the Thames–Southern North Sea, the matched base has the largest short-lead improvement, while Delay-only and CoastRelay remain positive through 72 h. CoastRelay has the highest or near-highest q95 improvement at 24–72 h. Charleston has a smaller and steadily declining RMSE gain, with CoastRelay below the Matched base and Delay-only curves at longer leads. Gulf/Galveston retains positive RMSE improvement across the forecast horizon. Under standardized forcing, CoastRelay gives the largest or near-largest improvement at the middle and longer leads, with Pair-forcing-only remaining close. The q95 curves remain clustered after 48 h.

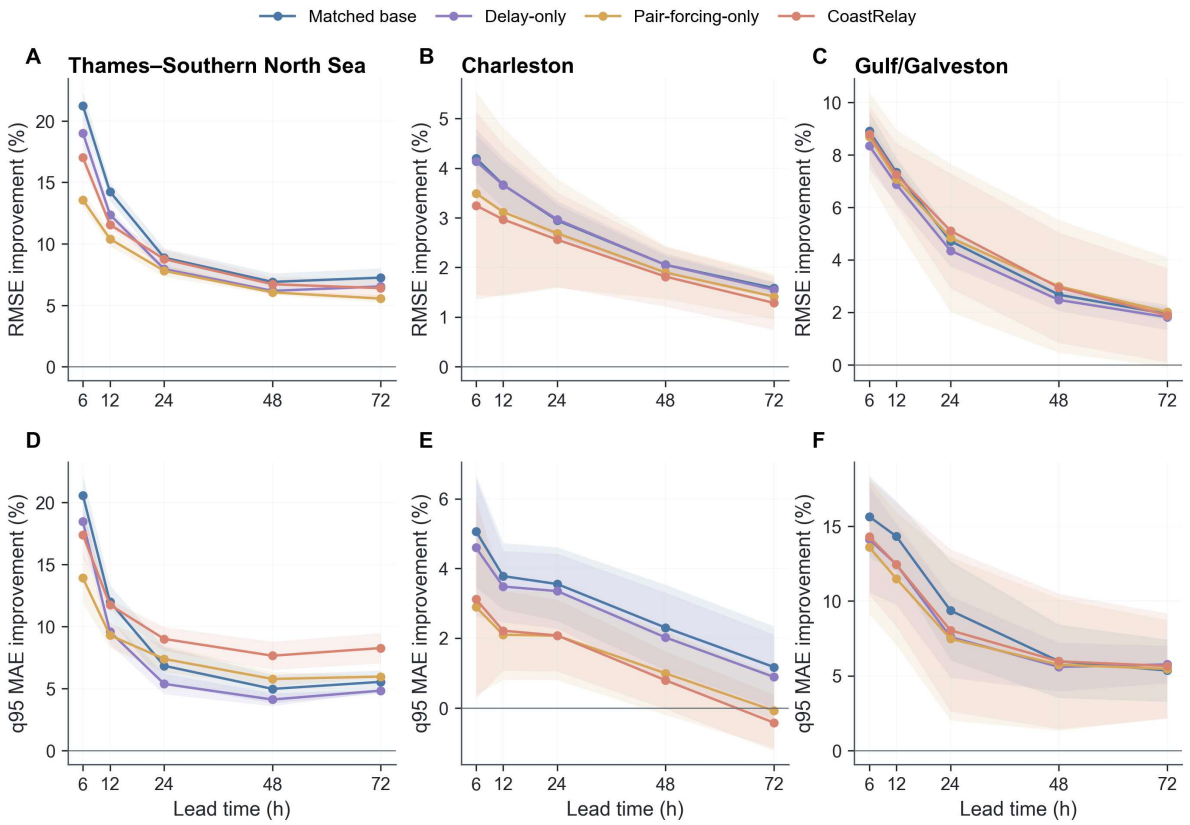


Figure 5: Test-period forecast lead-time profiles at 6, 12, 24, 48, and 72 h. Panels A–C show RMSE improvement of the four graph configurations relative to Local-GRU; Panels D–F show q95 mean absolute error improvement. For each random initialisation s, improvement is $100(E_{\text{Local-GRU},s} - E_{m,s})/E_{\text{Local-GRU},s}$, where E denotes the lead-specific error and m a graph configuration. Positive values favour the graph configuration. Lines and ribbons show the mean and sample standard deviation, respectively, of the improvements obtained from three random initialisations.

6 Discussion

6.1 What the separation reveals

The Thames results show how the two cross-gauge inputs acquire joint value. More than half of the directed station pairs select a nonzero lag, and the pair-forcing component reduces q95 error before the two components are combined. CoastRelay then has lower q95 error than either single-component configuration, giving a direct example of upper-tail complementarity. Pair-forcing-only retains the lowest overall RMSE, whereas CoastRelay has the lowest q95 MAE; the joint benefit is concentrated in the high-water tail rather than spread uniformly across the error distribution.

The 0–24 h lag grid aligns station states according to their observed lead–lag associations. Water-path distance and pairwise atmospheric contrasts then organize the contribution from neighbouring gauges after station states have been centred. The Thames pattern indicates that the gain comes from combining two distinct predictive views of the same event: the sequence carried by the gauge network and the spatial contrast in the atmospheric field.

6.2 Regional transfer of the result

Charleston and Gulf/Galveston provide contrasting network settings for this interpretation. Charleston has strong absolute lag correlations but few selected nonzero lags, so most graph messages remain synchronous.

Gulf/Galveston has broader nonzero-lag coverage and a substantial station-pair forcing contrast, while the single-component and joint configurations remain close in upper-tail error. Together, the three networks show that the value of a joint graph representation is governed by the relationship between lag structure and atmospheric contrast, rather than by the addition of cross-station inputs alone.

This gives the method a practical use beyond the three study systems. For a new gauge network, the lag matrix and event-period forcing contrasts can be estimated before model training. Their joint presence provides a principled reason to retain both information pathways; a network dominated by one pattern can be represented with the corresponding single component. The resulting diagnostic links model design to measurable network properties instead of relying on one universal graph architecture.

6.3 Forecast setting and scope

The selected lags are empirical predictive alignments within the three regional networks. The atmospheric experiment uses issue-time ERA5 reanalysis; an operational implementation can supply the same variables from numerical-weather-prediction fields. The smaller q99 event samples make q95 the primary upper-tail endpoint, with q99 characterising the most extreme observations. Applying the diagnostic to additional gauge networks and forecast atmospheric fields will test how widely the identified complementarity pattern transfers.

7 Conclusions

This study identifies two distinct sources of cross-gauge predictive dependence in coastal water-level forecasting: time-ordered gauge information and pairwise atmospheric contrast. CoastRelay makes their contributions explicit within one recurrent graph forecast by centring station states, aligning sender states with the estimated lag structure, and allowing the ordered atmospheric pair to modify graph attention.

The central scientific finding is that the two sources can be complementary at the upper tail. In the Thames–Southern North Sea, the joint representation adds q95 information beyond either component alone, linking the improvement to a network in which nonzero-lag gauge relations and event-scale atmospheric contrasts coexist. The contrasting regional results show why this distinction matters: the value of a combined graph input depends on the information carried by the network, not simply on using more stations or more variables.

For future coastal forecasting systems, the study suggests a simple design sequence. Estimate the empirical lag structure and the spatial atmospheric contrasts of the target network, then retain joint graph coupling where both sources describe distinct parts of high-water events. This principle can guide extension to new gauges and to numerical-weather-prediction forcing while preserving the same local forecast and evaluation framework.

Data and code availability

The water-level observations and station metadata originate from GESLA and NOAA CO-OPS. Atmospheric fields are from ERA5, and the geographic prior uses GEBCO bathymetry with Natural Earth coastline vectors. Source data are available from their respective providers. Code and processed data used in the analysis are available from the corresponding authors upon reasonable request.

Competing interests

The authors declare no competing interests.

Acknowledgements

This work was partially supported by the Zhejiang Provincial Department of Education under grant Y202456288, the Wenzhou Science and Technology Program under grant GC20250005, and the 2023 Wenzhou–Kean University Research Start-up Project under grant ISRG2023023.

Author contributions

Jia Rong and Guanchao Tong contributed equally. Both authors contributed to study design, methodology, analysis, and manuscript preparation.

References

[1] Zhuo Zhang, Lu Zhang, Songshan Yue, Dong Zhang, Zhaoyuan Yu, Di Hu, and Peng Chen. Short-term prediction of storm surges in estuarine and coastal waters via a multipoint deep learning neural network with limited training samples. *International Journal of Digital Earth*, 18(1):2536074, 2025. doi: 10.1080/17538947.2025.2536074.

[2] Xianwu Shi, Peng Chen, Zuchao Ye, Xinlong Zhang, and Weiping Wang. Tide level prediction during typhoons based on variable topology in graph convolution recurrent neural networks. *Ocean Engineering*, 312:119228, 2024. doi: 10.1016/j.oceaneng.2024.119228.

[3] Wenjun Jiang, Jize Zhang, Yuerong Li, Dongqin Zhang, Gang Hu, Huanxiang Gao, and Zhongdong Duan. Advancing storm surge forecasting from scarce observation data: A causal-inference based spatio-temporal graph neural network approach. *Coastal Engineering*, 190:104512, 2024. doi: 10.1016/j.coastaleng.2024.104512.

[4] Zhangke Huang, Dantong Li, Ning Jia, Xianwu Shi, Wei Zhai, and Weiping Wang. Prediction of the typhoon-induced tidal level at multiple stations using dynamic graph neural network. *Ocean Engineering*, 349:124180, 2026. doi: 10.1016/j.oceaneng.2026.124180.

[5] Hai Sun, Yue Yan, Yuan Qu, Wanxiang Yao, Huiqian Wang, Yanan Chu, and Bingchen Liang. Causality-aware spatiotemporal graph attention modeling for multi-station storm surge forecasting under typhoon forcing. *Coastal Engineering*, 208:105020, 2026. doi: 10.1016/j.coastaleng.2026.105020.

[6] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. International Conference on Learning Representations, OpenReview, 2018. URL https://openreview.net/forum?id=SJiHXGWAZ.

[7] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pages 3634–3640, 2018. doi: 10.24963/ijcai.2018/505.

[8] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph WaveNet for deep spatial-temporal graph modeling. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pages 1907–1913, 2019. doi: 10.24963/ijcai.2019/264.

[9] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. International Conference on Learning Representations, OpenReview, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

[10] Sanne Muis, Martin Verlaan, Hessel C. Winsemius, Jeroen C. J. H. Aerts, and Philip J. Ward. A global reanalysis of storm surges and extreme sea levels. *Nature Communications*, 7:11969, 2016. doi: 10.1038/ncomms11969.

[11] Job C. M. Dullaart, Sanne Muis, Nadia Bloemendaal, and Jeroen C. J. H. Aerts. Advancing global storm surge modelling using the new ERA5 climate reanalysis. *Climate Dynamics*, 54(1–2): 1007–1021, 2020. doi: 10.1007/s00382-019-05044-0.

[12] Kevin Horsburgh and Chris Wilson. Tide-surge interaction and its role in the distribution of surge residuals in the North Sea. *Journal of Geophysical Research: Oceans*, 112:C08003, 2007. doi: 10.1029/2006JC004033.

[13] Marta Marcos and Philip L. Woodworth. Spatiotemporal changes in extreme sea levels along the coasts of the North Atlantic and the Gulf of Mexico. *Journal of Geophysical Research: Oceans*, 122 (9):7031–7048, 2017. doi: 10.1002/2017JC013065.

[14] Michalis I. Vousdoukas, Lorenzo Mentaschi, Evangelos Voukouvalas, Martin Verlaan, Svetlana Jevrejeva, Luke P. Jackson, and Luc Feyen. Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard. *Nature Communications*, 9:2360, 2018. doi: 10.1038/s41467-018-04692-w.

[15] Ivan D. Haigh, Marta Marcos, Stefan A. Talke, Philip L. Woodworth, John R. Hunter, Bruce S. Hague, Arne Arns, Emma Bradshaw, and Phil Thompson. GESLA version 3: a major update to the global higher-frequency sea-level dataset. *Geoscience Data Journal*, 10(3):293–314, 2023. doi: 10.1002/gdj3.174.

[16] GESLA Consortium. Global extreme sea level analysis version 4.0: data, metadata, and access services, 2026. URL `https://gesla787883612.wordpress.com/downloads/`. Official GESLA download page; accessed 2 August 2026.

[17] NOAA National Ocean Service. NOAA tides and currents: water-level station metadata and observations, 2026. URL `https://tidesandcurrents.noaa.gov/`. Accessed 24 July 2026; dataset and metadata resource.

[18] NOAA Center for Operational Oceanographic Products and Services. Station metadata and observations application programming interface, 2026. URL `https://api.tidesandcurrents.noaa.gov/mdapi/prod/`. Official station metadata service; accessed 24 July 2026.

[19] Hans Hersbach, Bill Bell, Paul Berrisford, Shinya Hirahara, András Horányi, Joaquín Muñoz Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Iryna Rozum, Dinand Schepers, Adrian Simmons, Cornel Soci, Sharif Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Giacomo Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leopold Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi: 10.1002/qj.3803.

[20] Copernicus Climate Change Service. ERA5 hourly data on single levels from 1940 to present, 2026. URL `https://doi.org/10.24381/cds.adbb2d47`. Dataset access and metadata; accessed 24 July 2026.

[21] Natural Earth. 1:10m physical vectors: coastline and land polygons, 2026. URL `https://www.naturalearthdata.com/downloads/10m-physical-vectors/`. Version 5.1.1 page; accessed 24 July 2026.

[22] GEBCO Bathymetric Compilation Group. The GEBCO_2026 grid: a continuous terrain model for oceans and land at 15 arc-second intervals, 2026. [dataset].

[23] Ivan D. Haigh, Robert J. Nicholls, and Neil Wells. A comparison of the main methods for estimating probabilities of extreme still water levels. *Coastal Engineering*, 57(9):838–849, 2010. doi: 10.1016/j.coastaleng.2010.04.002.

[24] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1724–1734, 2014. doi: 10.3115/v1/D14-1179.

[25] David M. W. Powers. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.

[26] Anthony C. Davison and David V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge, 1997. doi: 10.1017/CBO9780511802843.

[27] Daniel S. Wilks. *Statistical Methods in the Atmospheric Sciences*. Academic Press, Oxford, 3 edition, 2011. ISBN 978-0-12-385022-5.

A Station information

Table A1: Stations used in the three regional networks. Longitude and latitude are in decimal degrees. Coordinates are those used to construct the graph and extract atmospheric forcing. Map labels correspond to Figures 2 and 4; datum descriptions follow the source metadata.

Region	Map label	Station ID	Station name	Longitude	Latitude	Source	Datum
Thames–Southern North Sea	CRO	cromer-cro-gbr-cmems	Cromer	1.300000	52.933300	GESLA Version 4.0	station-specific official gauge datum
	NEW	newhaven-new-gbr-cmems	Newhaven	0.066670	50.783298	GESLA Version 4.0	station-specific official gauge datum
	IMM	imingham-imm-gbr-cmems	Immingham	-0.183330	53.633301	GESLA Version 4.0	station-specific official gauge datum
	SHE	sheerness-she-gbr-cmems	Sheerness	0.750000	51.450001	GESLA Version 4.0	station-specific official gauge datum
	POR	portsmouth-por-gbr-cmems	Portsmouth	-1.110200	50.800400	GESLA Version 4.0	station-specific official gauge datum
	DOV	dover-dov-gbr-cmems	Dover	1.316670	51.116699	GESLA Version 4.0	station-specific official gauge datum
Charleston	C1	8665530	Charleston	-79.923615	32.780834	NOAA CO-OPS	mean sea level
	C2	8670870	Fort Pulaski	-80.903030	32.034695	NOAA CO-OPS	mean sea level
	C3	8661070	Springmaid Pier	-78.918300	33.655000	NOAA CO-OPS	mean sea level
	C4	8658120	Wilmington	-77.953300	34.226700	NOAA CO-OPS	mean sea level
	C5	8658163	Wrightsville Beach	-77.786700	34.213306	NOAA CO-OPS	mean sea level
	C6	8720218	Mayport	-81.427890	30.398167	NOAA CO-OPS	mean sea level
	C7	8721604	Trident Pier	-80.593100	28.415800	NOAA CO-OPS	mean sea level
Gulf/Galveston	G1	8771450	Galveston Pier 21	-94.793300	29.310000	NOAA CO-OPS	mean sea level
	G2	8771013	Eagle Point, Galveston Bay	-94.917250	29.481306	NOAA CO-OPS	mean sea level
	G3	8775792	Packery Channel	-97.236700	27.633300	NOAA CO-OPS	mean sea level
	G4	8779770	Port Isabel	-97.215530	26.061167	NOAA CO-OPS	mean sea level
	G5	8761724	Grand Isle	-89.956700	29.263300	NOAA CO-OPS	mean sea level
	G6	8771341	Galveston Bay Entrance	-94.724724	29.357462	NOAA CO-OPS	mean sea level
	G7	8773146	Matagorda City	-95.913900	28.710000	NOAA CO-OPS	mean sea level
	G8	8770613	Morgans Point, Barbours Cut	-94.985000	29.681667	NOAA CO-OPS	mean sea level

B Model settings

Table A2 gives the model and training settings used in the comparison. The four graph configurations share the same architecture; the two factors control the lag lookup and pair-forcing term.

B.1 Gated recurrent unit equations

For input $\boldsymbol{x}_t$ and previous state $\boldsymbol{h}_{t-1}$, the shared encoder follows the PyTorch gated recurrent unit recurrence

$$\boldsymbol{z}_t = \sigma(W_z \boldsymbol{x}_t + U_z \boldsymbol{h}_{t-1} + \boldsymbol{b}_z), \tag{29}$$

$$\boldsymbol{u}_t = \sigma(W_u \boldsymbol{x}_t + U_u \boldsymbol{h}_{t-1} + \boldsymbol{b}_u), \tag{30}$$

$$\boldsymbol{n}_t = \tanh(W_n \boldsymbol{x}_t + \boldsymbol{b}_n + \boldsymbol{u}_t \odot (U_n \boldsymbol{h}_{t-1} + \boldsymbol{c}_n)), \tag{31}$$

$$\boldsymbol{h}_t = (1 - \boldsymbol{z}_t) \odot \boldsymbol{n}_t + \boldsymbol{z}_t \odot \boldsymbol{h}_{t-1}, \tag{32}$$

where $\boldsymbol{z}_t$ and $\boldsymbol{u}_t$ are the update and reset gates.

B.2 Atmospheric-input scale sensitivity

The pair-forcing configurations were also evaluated with the atmospheric channels kept in their physical units. The physical-unit sensitivity runs used the same data partitions, random initialisations, fitted Local-GRU checkpoints, and the same initialisation of shared graph parameters as the standardized analysis. The resulting interaction estimates are reported alongside the standardized primary analysis in Table A3. The regional interpretation is unchanged: Thames remains negative with an interval below zero, whereas the Charleston and Gulf/Galveston intervals include zero.

B.3 Event-level upper-tail effects

The event-level distribution in Appendix Figure A1 complements the regional average by showing how the joint representation behaves across the individual Thames q95 events. The calculation uses the existing predictions from the three random initialisations and the standardized forcing analysis.

Table A2: Model and training settings used in the three-region comparison.

Item	Setting
History / horizon	96 h / 72 h
Input channels	residual, residual difference, tide, tide difference, observed level, observed difference
Forcing channels	u_{10}, v_{10}, mean-sea-level-pressure anomaly, wind-stress magnitude
Input representation	Water-level history in m; forcing channels standardized by region and channel using training-period statistics
Scaling transfer	Training means and standard deviations fixed for validation and test partitions
Input windows	Complete, with no target interpolation
ERA5 extraction	Nearest native 0.25° grid cell; issue-time alignment
GRU encoder	One layer, hidden dimension 96, recurrent dropout 0
Local forecast head	$168 \rightarrow 72$ linear map
Attention	One head; $Q/K/V : 96 \rightarrow 96$
Pair-forcing multilayer perceptron	$12 \rightarrow 96 \rightarrow 1$, Gaussian error linear unit (GELU)
Correction head	$264 \rightarrow 96 \rightarrow 72$, Gaussian error linear unit (GELU)
Graph dropout	0.1 after common-response centring
Candidate delay set	0, 6, 12, 18, 24 h
Delay estimator	Absolute Pearson correlation over the candidate set
Delay orientation	Receiver rows, sender columns
Distance prior scale	200 km
Optimizer	Adam with decoupled weight decay (AdamW)
Learning rate	3×10^{-4}
Weight decay	1×10^{-4}
Maximum epochs / patience	300 / 30
Dropout	0.1
Gradient-norm clip	1
Tail-loss weight	0.1
Batch size	512
Random initialisations	3
Local recurrent forecast parameters	42,216
Cross-station graph and correction parameters	117,385
CoastRelay total parameters	159,601
Bootstrap resamples	2,000

Table A3: Sensitivity of the q95 joint interaction to atmospheric-input scaling. The primary analysis uses region–channel training standardization; the physical-unit values provide a scale-sensitivity comparison. Negative values indicate additional error reduction from the joint representation.

Region	Standardized interaction (m)	Standardized 95% interval (m)	Physical-unit interaction (m)	Physical-unit 95% interval (m)
Thames–Southern North Sea	-0.0036	[-0.0069,-0.0003]	-0.0070	[-0.0113,-0.0030]
Charleston	-0.0012	[-0.0032,0.0008]	0.0003	[-0.0002,0.0007]
Gulf/Galveston	-0.0011	[-0.0031,0.0007]	0.0003	[-0.0012,0.0022]

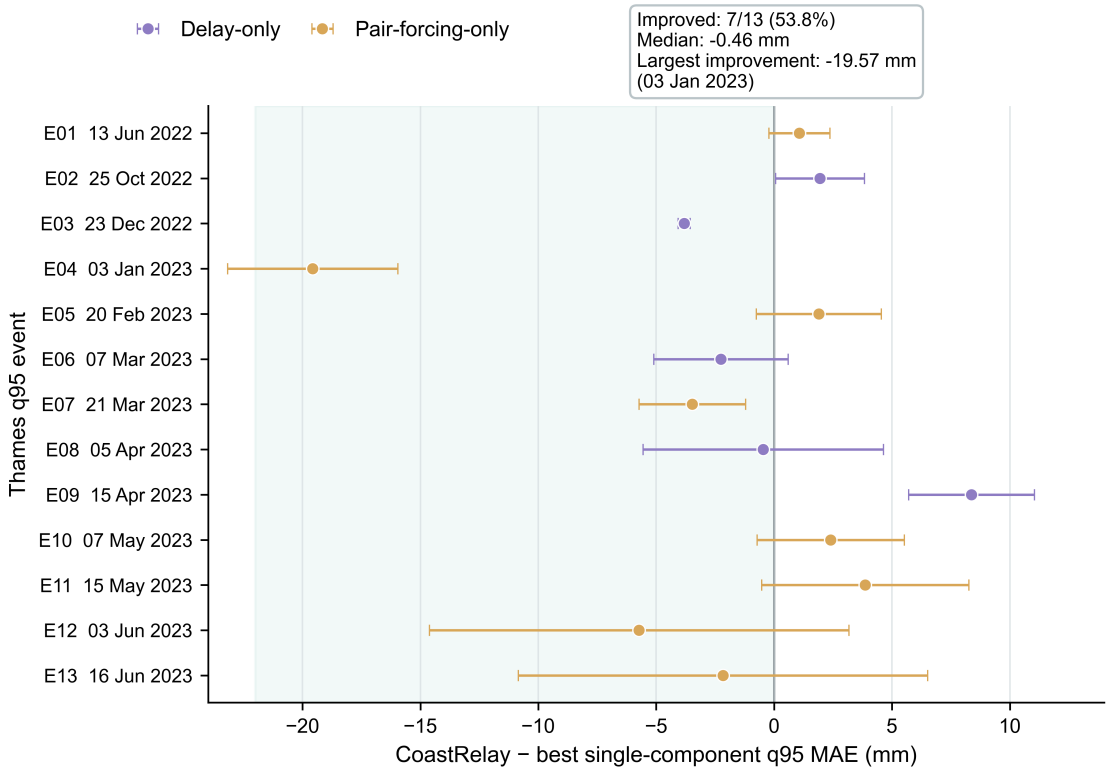


Figure A1: Event-level q95 effects in the Thames–Southern North Sea. For each q95 event cluster linked to the paired test-period analysis, the point is the mean across three random initialisations of CoastRelay q95 mean absolute error minus the lower of the Delay-only and Pair-forcing-only event q95 errors; horizontal bars show the sample standard deviation across the three random initialisations. Negative values indicate lower event error for CoastRelay. The summary reports the proportion of events with a negative effect, the median effect, and the largest event-level reduction.