https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

A hybrid machine-learning framework combining hydrodynamic simulation and tide-gauge observations for storm-surge nowcasting in the southern North Sea

Qiang Wang [1,2], Ludovic Lepers[1,2], Alexis Culot[2], Emmanuel Hanert[2,3], and Sebastien Legrand[1]

[1]Operational Directorate Natural Environment (OD Nature), Royal Belgian Institute of Natural Sciences (RBINS), rue Vautier 29, Brussels, 1000, Belgium
[2]Earth and Life Institute (ELI), UCLouvain, Croix du S 2, Louvain-la-Neuve, 1348, Belgium
[3]Institute of Mechanics, Materials and Civil Engineering (IMMC), UCLouvain, Place du Levant 2, Louvain-la-Neuve, 1348, Belgium

Correspondence: Qiang Wang (qwang@naturalsciences.be)

Abstract. Storm surges are a major coastal hazard that can develop within a few hours, so accurate short-term predictions are essential for early warning and emergency response. Operational hydrodynamic models remain too computationally demanding to deliver on-demand, high-frequency forecasts at the required resolution, while purely data-driven models suffer from the limited duration of tide-gauge records. We propose a hybrid machine-learning framework that combines tide-gauge observations from 22 stations along the southern North Sea coast with surge simulations from the COupled Hydrodynamical–Ecological model for REgioNal and Shelf seas (COHERENS) and ERA5 atmospheric forcing. Two architectures — a Transformer and a parallel LSTM+Transformer — are trained on this combined input and compared against (i) the same architectures trained on gauge observations alone and (ii) the COHERENS operational baseline. For 2-hour-ahead nowcasting at 10-minute resolution, the hybrid LSTM+Transformer reaches an average root-mean-square error (RMSE) of 0.055 m (0.033–0.100 m across stations), substantially better than COHERENS (0.164 m; 0.137–0.185 m) and modestly better than the pure data-driven baseline (0.061 m; 0.043–0.100 m); the hybrid Transformer reaches 0.096 m (0.067–0.121 m). The hybrid models also reproduce the magnitude and timing of extreme events well. Applied recursively, the hybrid LSTM+Transformer degrades to an average RMSE of 0.146 m at the 12-hour horizon, remaining well below the COHERENS baseline. Inference takes less than one minute on a single GPU, making the framework directly suitable for operational nowcasting. The approach should generalize to other coastal regions with comparable observational coverage.

1 Introduction

Storm surge, defined as the abnormal rise in sea level above the predicted astronomical tide, is primarily generated by strong winds and atmospheric pressure variations during severe weather events. It is widely recognized as the most severe and damaging driver of coastal inundation (Muis et al., 2016), often responsible for substantial economic losses and, in extreme cases, loss of life.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

The North Sea region is particularly vulnerable to storm surges (Gaslikova et al., 2011). Its exposure to intense extratropical storms, combined with its complex coastline and shallow bathymetry, enhances surge generation and amplification. This inherent physical susceptibility is further exacerbated by the high population density and the concentration of critical infrastructure—such as ports, industrial facilities, and urban centers—along the coastline. As a result, countries bordering the North Sea have made substantial investments in the construction and maintenance of robust coastal defense systems, including dikes, storm surge barriers, and reinforced dune systems, in order to reduce flood risk and safeguard socio-economic assets.

Despite these extensive and well-maintained sea defenses, the risk posed by extreme storm surge events cannot be completely eliminated. Under severe meteorological conditions, water levels can approach or even exceed design thresholds, leading to residual risks such as overtopping or, in rare cases, structural failure. Consequently, continuous improvements in forecasting, monitoring, and emergency preparedness remain essential to minimize the impacts of extreme storm surges in this highly exposed and densely populated region.

In Belgium, operational storm-surge forecasting systems based on hydrodynamic models have been in place since the mid-1970s (Adam, 1979). These early warning systems allow for the activation of the Belgian national contingency plan for the North Sea (Anon, 2026) up to 48 hours prior to a potentially dangerous storm surge event. This lead time enables the deployment of dynamic coastal defenses, such as the Nieuwpoort storm surge barrier, as well as the installation of temporary measures, including overtopping walls along coastal dike promenades. However, critical decisions—such as the evacuation of the most exposed areas—are typically taken closer to the event peak, often within a 6 to 12-hour window, based on human expert assessment. This article investigates how data-driven models and artificial intelligence can improve this final, yet crucial, stage of the decision-making process.

Data-driven (machine learning, ML / deep learning, DL) methods provide an alternative that alleviates the computational cost and the complexity of model setup. ML methods identify the inherent non-linear relationships between predictors and target variables directly from data, and have been applied to forecasting, classification, anomaly detection, semantic segmentation, and many other tasks. For storm-surge prediction specifically, a number of studies have used machine learning (Ayyad et al., 2022; Tadesse and Wahl, 2021; Tiggeloven et al., 2021). Ayyad et al. (2022) tested seven linear and non-linear ML models and found that the predicted return-period curves matched those generated from hydrodynamic simulations closely. Tadesse and Wahl (2021) developed a data-driven storm-surge reconstruction (DSSRT) using random forests and linear regression that performed well in extra-tropical regions but poorly in the tropics (average correlation 0.45), with no reconstructions available for the Indian coastline. Tiggeloven et al. (2021) implemented several neural-network architectures to predict storm surge globally (without astronomical tide), and found that long short-term memory (LSTM) networks generally outperform other architectures in a global setting.

Neural networks typically require large training datasets to learn complex feature relationships, which is difficult when observations are scarce or only short historical records are available — as is often the case for storm surge. Their generalization is also challenging when the training data are biased or strongly localized. Incorporating a physical model into the data-driven pipeline can help overcome these difficulties. Fu et al. (2024) estimated storm surge at a tide station using a Physics-Informed Neural Network (PINN) and (Karniadakis et al., 2021) obtained results nearly identical to those of the numerical model, with

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

an RMSE below 0.12 m. However, PINNs require a regular grid, and the computational cost grows with the complexity of the underlying physical system and the size of the observation dataset. In practice, the available observations are typically sparse compared with the simulation grids over a large domain, which limits the applicability of PINNs at the regional scale. Moreover, observation locations are usually irregular, which is awkward both for regular-grid PINNs and for operational deployment. In this paper, rather than replacing the physics with a machine-learning surrogate, we exploit the physics at the data level: we use hydrodynamic simulations as part of the DL input to compensate for the scarcity of gauge observations. Specifically, (i) we propose a hybrid input scheme that augments tide-gauge observations with COHERENS surge simulations and ERA5 atmospheric forcing; (ii) we train a parallel LSTM+Transformer model that predicts surge simultaneously at all 22 stations; and (iii) we carry out a systematic comparison against both the operational COHERENS hydrodynamic baseline and a pure data-driven LSTM+Transformer baseline.

The remainder of the paper is organized as follows. Section 2 describes the study area, the datasets, and the methodology. Section 3 presents the results in detail. Section 4 discusses the implications and limitations, and Section 5 concludes.

2 Material & Method

2.1 Study area

The study area focuses on the southern bight of the North Sea (Fig. 1). The domain spans 1°–5° E and 51°–53° N, covering the coastal zones of the Netherlands, Belgium, France and the UK. The region is dominated by tidal currents and is particularly vulnerable to flooding, as most of the coastline lies less than 5 m above mean sea level.

2.2 Datasets

Tide gauge in-situ observations, ERA5 meteorological re-analysis and storm-surge simulations produced by the COHERENS model have been used in this study.

2.2.1 Tide gauge observations

Tide gauge observations were obtained from Meetnet Vlaamse Banken (Belgium), Rijkswaterstaat (Netherlands), the French Hydrographic and Oceanographic Service (SHOM), the British Oceanographic Data Centre (BODC), and the Copernicus Marine in-situ data store (Copernicus Marine Service, 2025). Station locations are shown as green dots in Fig. 1. The records are originally provided at 5-minute, 10-minute or 1-hour intervals depending on the provider. We re-sampled the 5-minute records to 10 minutes, yielding a harmonised 10-minute dataset. Stations with less than 5 years of records were excluded to ensure that the training and validation sets cover the full range of seasonal variability. This left 22 stations in the southern North Sea, most of them located along the coastline but not evenly distributed. Each record was quality-controlled using the providers' flags; flagged values were discarded and replaced by linear interpolation when the gap was shorter than 2 hours (12 observations).

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

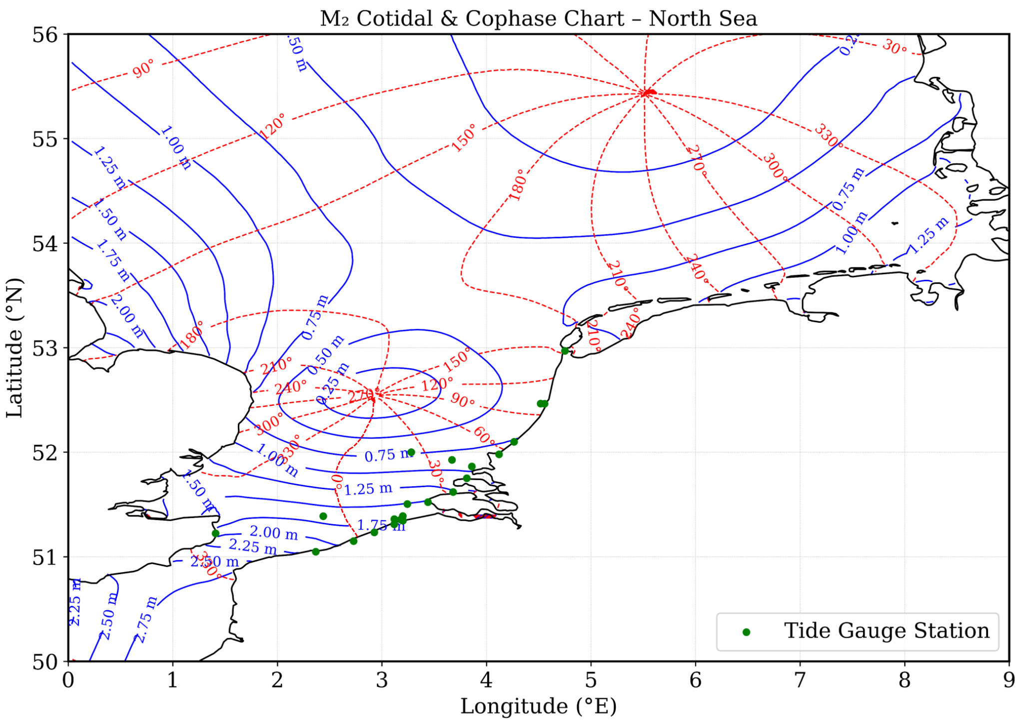

Figure 1. Distribution of the tide gauge stations (green dots) over the southern bight of the North Sea (1°–5° E and 51°–53° N) , with the M_2 cotidal (solid blue) and cophase (dashed red) chart.

Since this study focuses on surge prediction, we obtained the surge residual by subtracting the astronomical tide from the observed sea level using the Unified Tidal Analysis and Prediction Functions (UTide) (Codiga, 2011). Fig. 2 shows the observed sea level, the UTide astronomical tide, and the surge residual at Oostende harbour as an example; a clear storm-surge event is visible on 8-9 December 2019.

2.2.2 Meteorological forcing

Storm surges are driven primarily by wind stress and surface-pressure anomalies during storms (Pugh and Woodworth, 2014). We therefore use meteorological fields as predictors for the machine-learning models. Specifically, we obtained the 10 m zonal and meridional wind components and the mean sea-level pressure (MSLP) from the ERA5 reanalysis (Hersbach et al., 2020, 2023), which provides global atmospheric data at 0.25° spatial resolution and hourly temporal resolution. The hourly fields were linearly interpolated to a 10-minute interval to match the gauge resolution.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

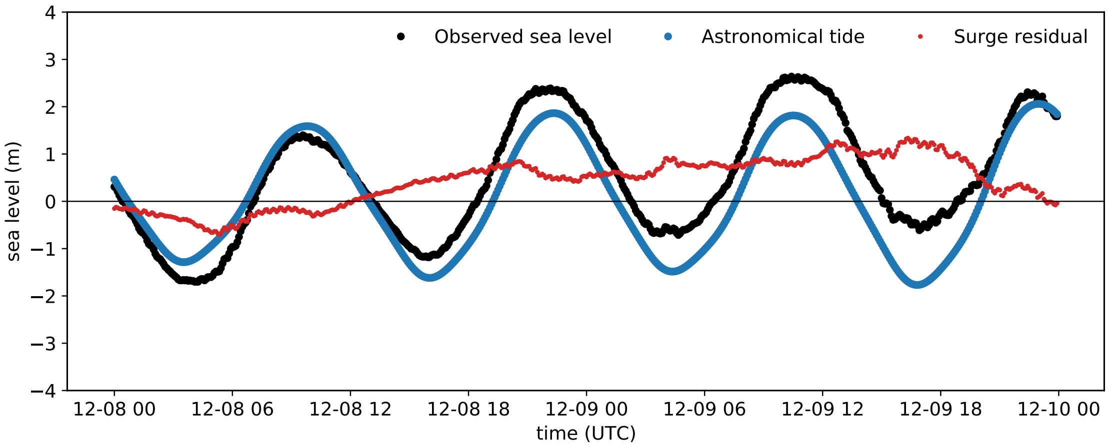

Figure 2. Observed sea level, UTide-simulated astronomical tide and surge residual at Oostende station over a 2-day interval in December 2019.

2.2.3 COupled Hydrodynamical–Ecological model for REgioNal and Shelf seas (COHERENS)

Since 2004, the RBINS marine forecasting centre has developed, maintained, and operated a hydrodynamic forecasting system based on three nested configurations of the COHERENS model (Luyten et al., 1999). The outermost domain is a 2D storm surge model covering the north-west European continental shelf, with a horizontal resolution of 5 km and a barotropic time step of 25 s. The intermediate domain consists of a 3D model of the southern North Sea and the English Channel, with a 5 km resolution, 20 sigma levels, and a baroclinic time step of 10 minutes. The innermost domain is a high-resolution 3D model of the Belgian coastal zone, with a 750 m resolution, 10 sigma levels, and a baroclinic time step of 30 s.

The forecasting system is run twice daily and is forced by UK Met Office atmospheric forecasts at 6-hour intervals. Model outputs include 10-minute time series at selected stations and hourly gridded fields, which are publicly available under a CC BY 4.0 license via the RBINS ERDDAP server (https://erddap.naturalsciences.be/erddap/index.html).

In this study, only the gridded output from the 2D storm surge configuration is used (Centre, 2023b), restricted to the southern North Sea domain. Surge residuals are computed at each grid point by subtracting the modelled harmonic astronomical tide (Centre, 2023a) from the simulated total water level. This domain corresponds to a 17×9 (longitude $\times$ latitude) grid points. After excluding land points, 102 wet grid cells are retained for the subsequent analysis.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

2.3 Methodology

2.3.1 Hybrid machine learning

A hybrid algorithm combines multiple methods, often drawn from different domains, to solve a problem more effectively than any single method alone (Azevedo et al., 2024). By integrating complementary approaches, hybrid methods can exploit the strengths of each while compensating for their weaknesses, and so improve overall performance. For example, Naeini and Snaiki (Naeini and Snaiki, 2024) obtained high accuracy and computational efficiency by combining dimensional-reduction and data-driven techniques into a hybrid model for rapid assessment of waves and storm-surge responses. In contrast to their approach, our hybridisation operates in the input-feature domain: we augment the gauge observations with COHERENS surge simulations so that the deep-learning model receives both data sources at once. The expectation is that enriching the gauge observations with high-quality hydrodynamic simulations will yield better performance than relying on gauge data alone.

2.3.2 Feature selection

Feature selection is critical for deep-learning methods, as a well-chosen feature set is a prerequisite for accurate training. A rigorous selection procedure would examine the sensitivity of the target variable to each candidate feature, but this is time-consuming, and a reasonable trade-off can be made by relying on previous studies and on domain expertise. Following the existing storm-surge literature (Ayyad et al., 2022; Hermans et al., 2025; Ebel et al., 2024; Tian et al., 2024), we use the 10 m zonal and meridional wind components and the mean sea level pressure as predictors. Other variables — latitude, longitude, central atmospheric pressure, radius of exponential pressure scale, storm forward speed and heading, and 2 m air temperature — have been considered in some prior work but are not included here; their value is left for future investigation.

2.3.3 Transformer

The Transformer is a deep-learning architecture introduced by Vaswani et al. (2017) for natural-language processing tasks. It follows an encoder–decoder design and is characterised by a self-attention mechanism that assigns attention weights across time steps. This enables the Transformer to handle long time series more naturally than long short-term memory (LSTM) networks, which must process sequences in order, and allows for efficient parallelisation during training. Both the encoder and decoder consist of multi-head self-attention layers, feed-forward layers and positional encodings. Multi-head attention captures different aspects of the time series by computing several sets of attention weights in parallel, while positional encoding preserves the order of the input sequence.

In this paper, we use a Transformer-based model for storm-surge nowcasting (Fig. 3 (a)). Here, $T = 72$ time steps (12 hours at 10-minute intervals), and $F = 583$ input features: 459 atmospheric-forcing values, 102 COHERENS surge values, and 22 gauge observations. The input first passes through a 1D convolution layer with 32 output channels, which captures local temporal patterns. Six Transformer encoder layers with 8 attention heads each follow, after which an average-pooling layer

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

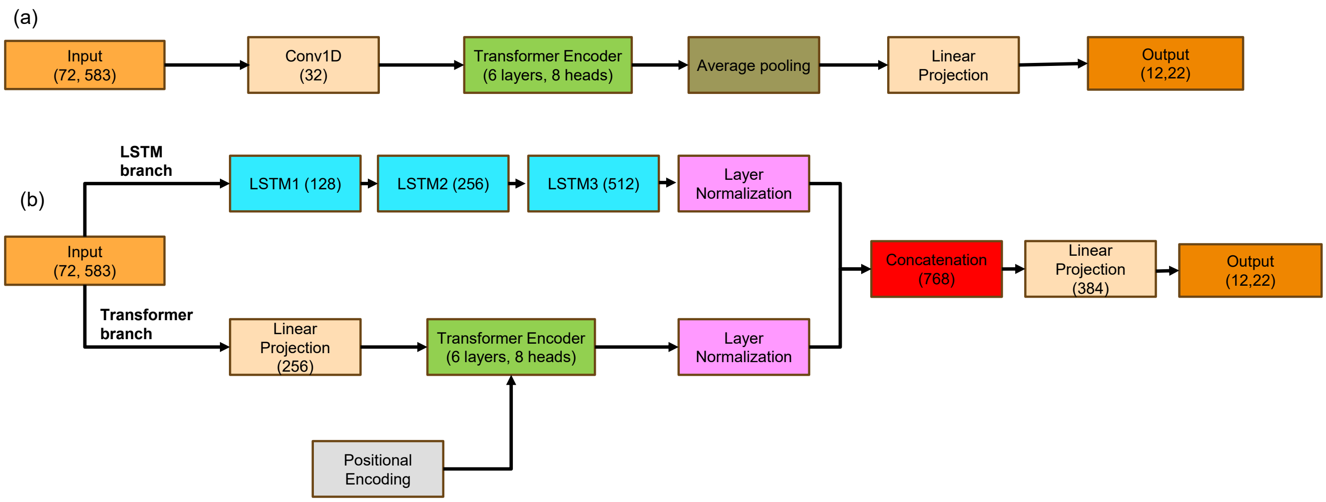

Figure 3. Architecture of (a) the hybrid Transformer and (b) the hybrid LSTM+Transformer model.

smooths residual noise and a fully connected layer projects the result onto the 22-station output. A dropout rate of 0.1 is applied throughout to improve generalisation.

2.3.4 LSTM+Transformer

The LSTM (Hochreiter and Schmidhuber, 1997) is a variant of the recurrent neural network (RNN) that captures sequence-to-sequence patterns in an internal memory state, with the advantage over conventional RNNs that it can selectively retain long-term information (Shen, 2018). It addresses the vanishing-gradient problem of RNNs through a gating mechanism with input, forget and output gates, which selectively store or discard information from previous time steps. The memory cell state is updated continuously as the sequence is processed, making LSTMs well suited to time-series prediction; several recent studies have applied LSTMs to storm-surge prediction (Tiggeloven et al., 2021; Hermans et al., 2025; Ishida et al., 2020).

LSTMs and Transformers are in some sense complementary. LSTMs excel at local temporal modelling — short-term patterns and smooth long-term trends — as they update their hidden states gradually. Transformers, by contrast, capture global, long-range dependencies through attention, but can overfit noisy data. Combining the two should therefore allow both local and global structure to be captured while reducing the overfitting risk of a Transformer alone. We adopt a parallel design (Fig. 3 (b)): the input feeds into both a LSTM branch and a Transformer branch in parallel, whose outputs are then concatenated and projected onto the 22-station output. The LSTM branch consists of three stacked LSTM layers with hidden sizes 128, 256 and 512, followed by layer normalisation; the hidden-size sequence was selected by the validation performance. The Transformer branch follows the architecture described in Section 2.3.3 and a dropout rate of 0.1 is used.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

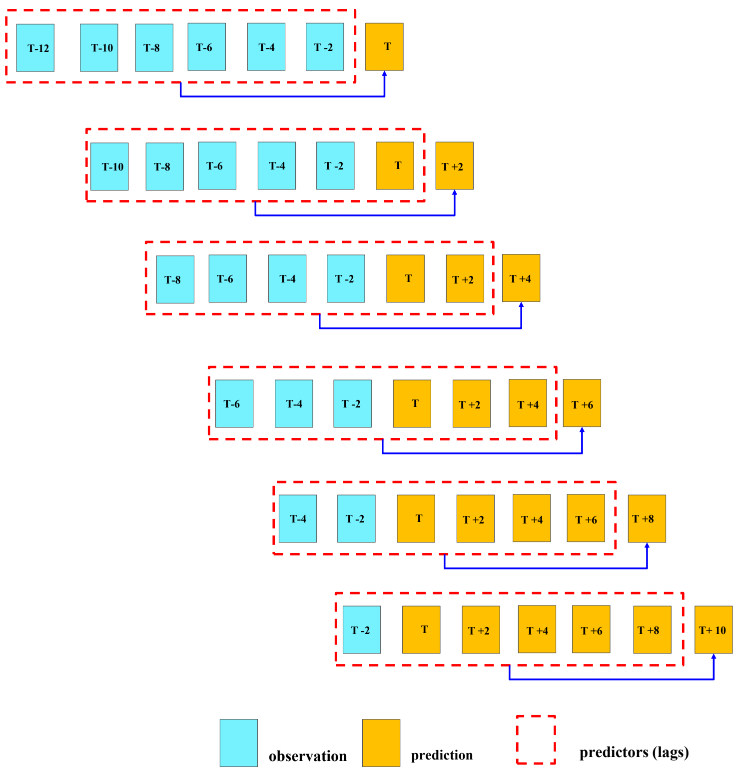

Figure 4. Illustration of recursive multi-step prediction: a model trained to predict 2 hours ahead is applied repeatedly, with each output block fed back as input, to produce forecasts up to 12 hours.

2.3.5 Multi-step forecasting

We target high-temporal-resolution nowcasting at 10-minute intervals. Predicting all 72 steps of a 12-hour window directly would substantially increase model complexity and storage requirements; furthermore, in operational practice, the input streams are typically refreshed only every 2–3 hours rather than continuously. We therefore train the model to predict the next 12 time steps (2 hours ahead) directly, and obtain longer forecast horizons by applying this 2-hour predictor recursively up to six times, feeding each output block back as input to the next prediction step. The resulting forecast covers up to 12 hours and is illustrated in Fig. 4.

2.4 Model training and hyper-parameter tuning

To assess the value of the hybrid input scheme, we trained four model configurations:

(1) Transformer with gauge observations and meteorological forcing as inputs.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

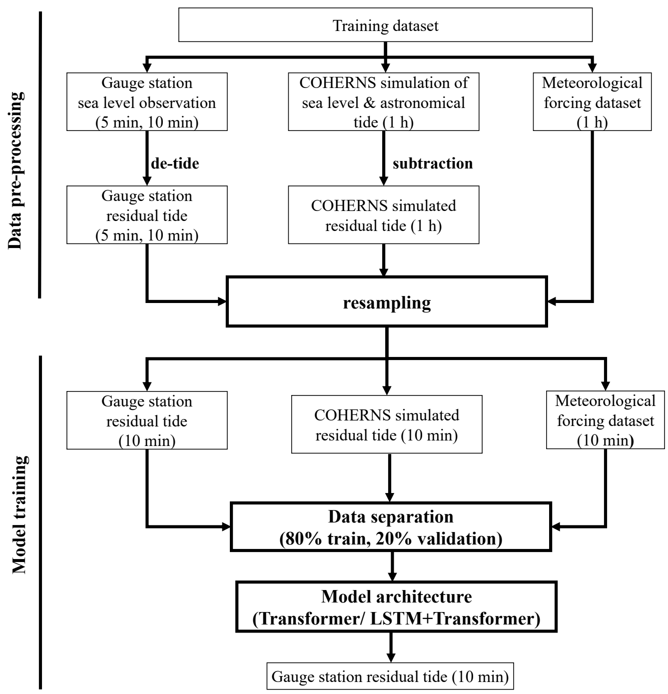

Figure 5. Overview of the training routine.

(2) hybrid Transformer with gauge observations, COHERENS surge simulations on the 17×9 grid, and meteorological forcing.

(3) LSTM+Transformer with the same inputs as (1).

(4) hybrid LSTM+Transformer with the same inputs as (2).

In all four cases, the model predicts the 2-hour-ahead surge residual simultaneously at the 22 stations.

The training procedure is summarised in Fig. 5. The gauge observations, COHERENS simulations and ERA5 forcing fields, spanning 1 March 2020 to 31 December 2024, were first harmonised to a common 10-minute interval (Section 2.2). This dataset was then split 80/20 into training and validation sets. The full year 2019, which is independent of the training period, was held out as the test set to assess the model's robustness on unseen data. To capture the causal relationship between past atmospheric and tidal forcing and the future surge response, each prediction uses the past 12 hours of forcing and gauge observations to predict the next 2 hours. Each input feature was independently z-score (zero mean, unit variance) scaled to the range $[-1, 1]$ using statistics computed on the training set; the same scaling was then applied at validation and test time.

We trained the models in PyTorch using the Adam optimiser (Kingma and Ba, 2014) with a learning rate of 10^{-4} for up to 200 epochs, with early stopping if the validation loss did not decrease for 11 consecutive epochs. The batch size was 32 and the loss function was the mean squared error. Training was carried out on a single AMD MI250x GPU on the LUMI cluster.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

2.5 Influence of the historic horizon

Our default configuration uses a 12-hour look-back window to predict 12 hours ahead by using 2-hours prediction recursively. Whether 12 hours is the optimal length is not obvious a priori: too short a window may not give the model enough temporal context, while too long a window can introduce noise and increase computational cost. To probe this trade-off, we trained an additional set of models with a 6-hour look-back window and compared their performance to the 12-hour baseline. Results are presented in Section 3.

2.6 Error metrics

We evaluate model performance using three standard metrics: the root-mean-square error (RMSE), which measures the typical magnitude of the prediction error; the coefficient of determination (R^2), which measures how much of the variance in the observations is captured by the model; and the bias, which measures the systematic offset of the model relative to the observations.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{1}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{2}$$

$$\mathrm{bias} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \tag{3}$$

where N is the number of observations, y_i the observed values, $\hat{y}_i$ the model predictions, and $\overline{y}$ the mean of the observations.

In addition to these aggregate metrics, we pay particular attention to extreme events, where both the timing and the magnitude of the surge peaks and troughs matter operationally: maxima determine flooding risk and damage potential, while minima matter for navigation, as low water reduces the draft available in shipping channels. For these events we report both the difference between predicted and observed peak (or trough) sea level, and the time difference between the predicted and observed peak (or trough).

3 Results

3.1 General performance

Table 1 summarises the station-averaged performance of each model on the 2-hour-ahead nowcast; per-station error metrics are reported in A (Table A1). All metrics are computed on the absolute sea level (surge residual plus astronomical tide).

The hybrid Transformer achieves an average RMSE of 0.096 m, compared with 0.123 m for the pure Transformer and 0.164 m for the COHERENS baseline. The R^2 values follow the same ordering: 0.988 for the hybrid Transformer, 0.978 for the pure Transformer and 0.966 for COHERENS. Across the 22 individual stations (Table A1), Europlatform is the easiest

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

Table 1. Station-averaged error metrics of the four model variants on the 2-hour-ahead nowcast, with the COHERENS operational baseline for reference. Best values per metric are in bold for RMSE and R^2.

model	RMSE (m)	R^2 (−)	bias (m)
COHERENS	0.164	0.966	0.021
Transformer	0.123	0.978	−0.051
hybrid Transformer	0.096	0.988	−0.027
LSTM+Transformer	0.061	0.995	−0.004
hybrid LSTM+Transformer	**0.055**	**0.996**	0.019

station for all three models, while the hardest is Blankenberge for COHERENS and the hybrid Transformer, and Den Helder for the pure Transformer.

Adding an LSTM branch in parallel improves performance further. The pure LSTM+Transformer reaches an average RMSE of 0.061 m and the hybrid LSTM+Transformer reaches 0.055 m, with R^2 values of 0.995 and 0.996 respectively. Both are noticeably better than the corresponding Transformer-only variants. Importantly, the margin between the hybrid and the pure LSTM+Transformer (0.055 vs 0.061 m) is much smaller than between the hybrid and the pure Transformer (0.096 vs 0.123 m). This suggests that the recurrent LSTM branch already captures most of the temporal information available in the local gauge history, so the marginal value of the COHERENS input is smaller here than in the Transformer-only case — where there is no recurrent component to compress the past.

3.2 Individual extreme storm surge events

Beyond the aggregate performance, it is important to assess how the models behave during individual extreme events, since these drive the operational value of any nowcasting system. We define an extreme event as a period during which the surge residual exceeds $+1.0$ m or falls below -1.0 m. Across the 22 stations, this criterion identifies 185 extreme events in the test year; at Oostende alone there are eight positive and one negative extreme events.

We illustrate model behaviour during two events at Oostende: a 1.3 m positive surge on 9 December 2019 (Fig. 6) and a 1.7 m positive surge in early January 2019 (Fig. B1). Because the standalone and hybrid Transformer-only variants behave very similarly to their LSTM+Transformer counterparts during these events, the time series figures show only the LSTM+Transformer family. For the December event (Fig. 6), all three models capture the overall sea-level dynamics, including the surge maxima. The hybrid LSTM+Transformer follows the gauge observation most closely, the pure LSTM+Transformer second, and COHERENS the most loosely. COHERENS tends to overestimate the peak, predicting a sea surface height of 2.991 m above geoid at 10:40 on 9 December, compared with an observed value of 2.608 m.The LSTM+Transformer and hybrid LSTM+Transformer predict 2.670 m and 2.610 m, respectively, both closer to the observation. The same pattern is visible in the surge residual (Fig. 6(b)): during the period 12:00 on 9 December to 00:00 on 10 December, the mean bias is $+0.459$ m for COHERENS, $+0.044$ m for the LSTM+Transformer, and -0.007 m for the hybrid LSTM+Transformer.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

The January event (Fig. B1) shows a similar pattern. All three models reproduce the broad sea-level variations, but COHERENS exhibits a clear positive bias between the surge peak (12:00 on 8 January) and 00:00 on 10 January (mean bias +0.446 m, against +0.055 m for the LSTM+Transformer and −0.004 m for the hybrid LSTM+Transformer; Fig. B1(b)). The ML models also reproduce short-period fluctuations that COHERENS misses — for example, the oscillation between 00:00 on 7 January and 00:00 on 8 January, and a brief reversal of the trend around 00:00 and 12:00 on 9 January. The largest single COHERENS error in this event is 1.20 m at 23:40 on 8 January (Fig. B1(c)); the largest LSTM+Transformer and hybrid LSTM+Transformer errors fall in the 08:00–13:00 window of 8 January, at 0.142 and −0.134 m respectively, both well below the corresponding COHERENS error of 0.206 m. Overall, the LSTM+Transformer models capture the storm-surge dynamics under both event types, while COHERENS occasionally misses sub-cycle variations.

3.3 Impact on the high water and low water timing

Apart from the surge residual, it is also important to have a understating of high and low water timing. The positive surge residual can add up the high water and leads to high sea level, which eventually may cause the flooding across the coastal line. By contrast, a negative surge residual can even lower the sea level. When the negative surge residual coincides with the low water during tidal cycle, this can cause the danger for the ship navigation. Hereby, we further investigate the high and lower water timing as well as their corresponding sea levels predicted by various models during the tidal cycle. The error metrics obtained by averaging the available stations and listed in Table 2- 3. Overall, all models achieve an RMSE below 0.20 m for both high water and low water level predictions, demonstrating good predictive performance. Among them, the hybrid LSTM–Transformer model exhibits the closest agreement with observations, yielding RMSE values of 0.058 and 0.056 (m) for high water and low water levels, respectively. In contrast, COHERENS produces the largest errors, with RMSE values of 0.147 m for high water and 0.153 m for low water. Nevertheless, these errors remain relatively small compared with the absolute sea levels, which are typically greater than 3 m. For timing prediction, the hybrid LSTM–Transformer model again demonstrates the best performance, followed by the Transformer-based models. Although COHERENS shows the lowest accuracy among all models for timing prediction, its RMSE remains below 1 hour, indicating satisfactory predictive skill. The coefficient of determination (R^2) is consistent with the RMSE-based ranking of the models. The hybrid LSTM–Transformer achieves the highest accuracy, with an R^2 value of 0.991 (-) for both high water and low water levels. By comparison, the corresponding R^2 values for COHERENS are 0.948 and 0.934 (-), respectively. For timing prediction, all models yield exceptionally high R^2 values exceeding 0.999 (-). In terms of bias, most models exhibit absolute biases smaller than 0.03 m for both high water and low water levels when compared with observations. The only notable exception is the pure Transformer model, which underestimates low-water levels by 0.062 m. Overall, both the machine-learning models and COHERENS are capable of accurately predicting the timing and magnitude of high and low waters, with the hybrid LSTM–Transformer consistently providing the best overall performance.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

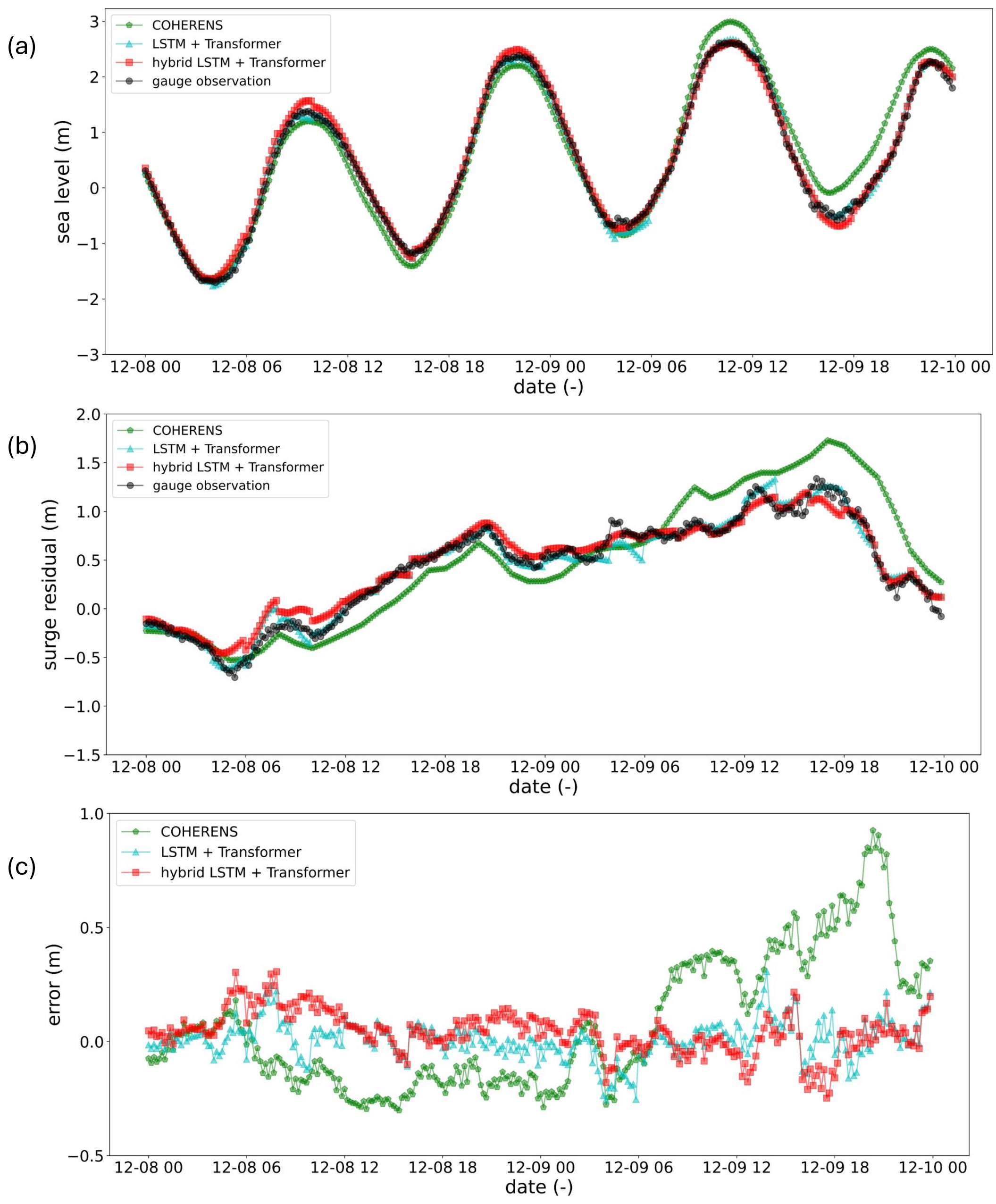

Figure 6. Sea level above geoid (a), surge residual (b), and difference between predicted and observed sea level (c) at Oostende during the storm surge of 8–9 December 2019.

Table 2. RMSE of magnitude (m) and timing (hour) for the COHERENS and four model variants with respect to the observation

parameter	COHERENS	Transformer	hybrid Transformer	LSTM+Transformer	hybrid LSTM+Transformer
magnitude of high water (m)	0.147	0.114	0.091	0.061	0.058
timing of high water (h)	0.475	0.475	0.412	0.299	0.271
magnitude of low water (m)	0.153	0.122	0.092	0.059	0.056
timing of low water (h)	0.627	0.585	0.519	0.392	0.367

Table 3. Bias of magnitude (m) and timing (hour) for the COHERENS and four model variants with respect to the observation

parameter	COHERENS	Transformer	hybrid Transformer	LSTM+Transformer	hybrid LSTM+Transformer
magnitude of high water (m)	0.021	−0.030	−0.010	0.001	0.021
timing of high water (h)	0.222	0.194	0.107	0.029	0.023
magnitude of low water (m)	0.033	−0.062	−0.033	−0.004	0.021
timing of low water (h)	0.039	0.061	0.058	0.049	0.038

3.4 Inference at extended horizons

Tables 4-–6 summarize the performances of the nowcasting models when the prediction horizon is extended to 12 hours following the recursive strategy described in section 2.3.5.

All models exhibit increasing errors with forecast horizon, but the rate of degradation differs markedly between architectures. The hybrid models show a gradual and consistent increase in RMSE. For instance, the hybrid LSTM–Transformer increases from 0.055 m at a 2 h horizon to 0.146 m at 12 h, while the hybrid Transformer rises from 0.096 to 0.152 m. Across all horizons, both hybrid configurations outperform the COHERENS baseline (RMSE of 0.164 m), while maintaining high skill with R^2 values exceeding 0.96.

The LSTM–Transformer model exhibits a similar but slightly stronger degradation, with RMSE increasing from 0.061 to 0.194 m. Although it remains competitive with COHERENS at short lead times, its performance drops below the baseline at longer horizons ($\geq$10 h), which is reflected in a decrease of R^2 to 0.937 at 12 h.

In contrast, the pure Transformer shows a substantially faster degradation. Its RMSE reaches 0.163 m at 4 h (comparable to the COHERENS baseline) and increases to 0.254 m at 12 h, clearly exceeding the baseline. This deterioration is also evident in the sharper decline of R^2, from 0.978 at 2 h to 0.878 at 12 h.

Bias values further highlight systematic differences between models. The pure Transformer exhibits a persistent negative bias that increases in magnitude with forecast horizon (from -0.051 to -0.114 m), indicating a systematic underestimation. In contrast, the hybrid Transformer maintains a small and stable negative bias (around -0.03 m), while the LSTM–Transformer

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

Table 4. RMSE (m) of the four model variants at forecast horizons from 2 to 12 hours.

horizon (h)	Transformer	hybrid Transformer	LSTM+Transformer	hybrid LSTM+Transformer
2	0.123	0.096	0.061	**0.055**
4	0.163	0.123	0.101	**0.090**
6	0.194	0.135	0.130	**0.116**
8	0.213	0.140	0.152	**0.133**
10	0.224	0.145	0.175	**0.139**
12	0.254	0.152	0.194	**0.146**

Table 5. Coefficient of determination R^2 $(-)$ of the four model variants at forecast horizons from 2 to 12 hours.

horizon (h)	Transformer	hybrid Transformer	LSTM+Transformer	hybrid LSTM+Transformer
2	0.978	0.988	0.995	**0.996**
4	0.962	0.980	0.987	**0.992**
6	0.945	0.975	0.980	**0.986**
8	0.931	0.972	0.970	**0.979**
10	0.921	0.970	0.958	**0.976**
12	0.878	0.961	0.937	**0.968**

shows near-zero bias across most horizons. The hybrid LSTM+Transformer displays a small positive bias that increases moderately with lead time (up to 0.073 m), but remains limited compared to the spread in RMSE.

The contrast between the hybrid and pure Transformer configurations suggests that the hybrid input strategy is crucial for maintaining forecast skill at longer horizons. While models relying solely on gauge observations lose predictive power as lead time increases, the inclusion of COHERENS fields provides additional physical constraints that stabilise the predictions and limit error growth.

Table 6. Bias (m) of the four model variants at forecast horizons from 2 to 12 hours.

horizon (h)	Transformer	hybrid Transformer	LSTM+Transformer	hybrid LSTM+Transformer
2	−0.051	−0.027	−0.004	0.019
4	−0.073	−0.032	0.000	0.041
6	−0.090	−0.029	0.007	0.059
8	−0.099	−0.030	0.008	0.072
10	−0.105	−0.034	0.003	0.074
12	−0.114	−0.034	−0.008	0.073

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

Table 7. Error metrics of the hybrid LSTM+Transformer trained with a 6-hour look-back window, at forecast horizons from 2 to 12 hours.

horizon (h)	RMSE (m)	R^2 (−)	bias (m)
2	0.055	0.997	−0.001
4	0.088	0.992	0.005
6	0.112	0.986	0.014
8	0.125	0.981	0.016
10	0.136	0.977	0.004
12	0.138	0.968	−0.011

The performance of the hybrid LSTM+Transformer model is illustrated in figure Fig. 7 for the storm event of 8–9 December 2019. The grey lines represent the ensemble of 12 h forecasts, each initialised every 2 h, resulting in six forecasts available at each time step. The red line corresponds to the time series obtained from successive 2 h predictions (as in Fig. 6), while the black line indicates the tide gauge observations.

The ensemble forecasts reproduce the overall sea-level dynamics and capture the main peaks, notably around 21:00 on 8 December and 18:00 on 9 December. A discrepancy between the ensemble predictions and observations is observed between 10:00 and 12:00 on 8 December, with the largest deviation occurring at 10:10. At this time, the predicted surge residual shows overestimations of 0.668, 0.684, 0.554, 0.471, 0.394, and 0.187 m for lead times of 12, 10, 8, 6, 4, and 2 h, respectively. The larger spread among ensemble members at this time indicates increased uncertainty in the model predictions.

3.5 Influence of model setting

This section investigates the sensitivity of the hybrid LSTM+Transformer agaisnt two key settings: the length of the lookback window and the number of Transformer encoder layers.

3.5.1 Look-back window

To evaluate the sensitivity to the input sequence length, we retrained the hybrid LSTM–Transformer using a 6 h look-back window instead of the 12 h baseline. The 2 h-ahead RMSE remains unchanged at 0.055 m (Table 7). At longer horizons, the 6 h configuration performs marginally better, with an RMSE of 0.138 m at 12 h compared to 0.146 m for the 12 h look-back.

A comparison of the time series (Fig. 8) shows that both configurations accurately reproduce the observed sea level (Fig. 8a) and surge (Fig. 8b). The point-wise error (Fig. 8c) is slightly larger and more variable for the 6 h model (range $(-0.18, 0.24)$ m, standard deviation 0.061 m) than for the 12 h model (range $(-0.22, 0.24)$ m, standard deviation 0.060 m), suggesting that the longer look-back window provides a modest smoothing effect on high-frequency variability in the gauge signal.

The same sensitivity experiment was conducted for the Transformer, hybrid Transformer, and LSTM–Transformer models. The corresponding performance metrics are provided in Appendix C (Tables C1–C3).

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

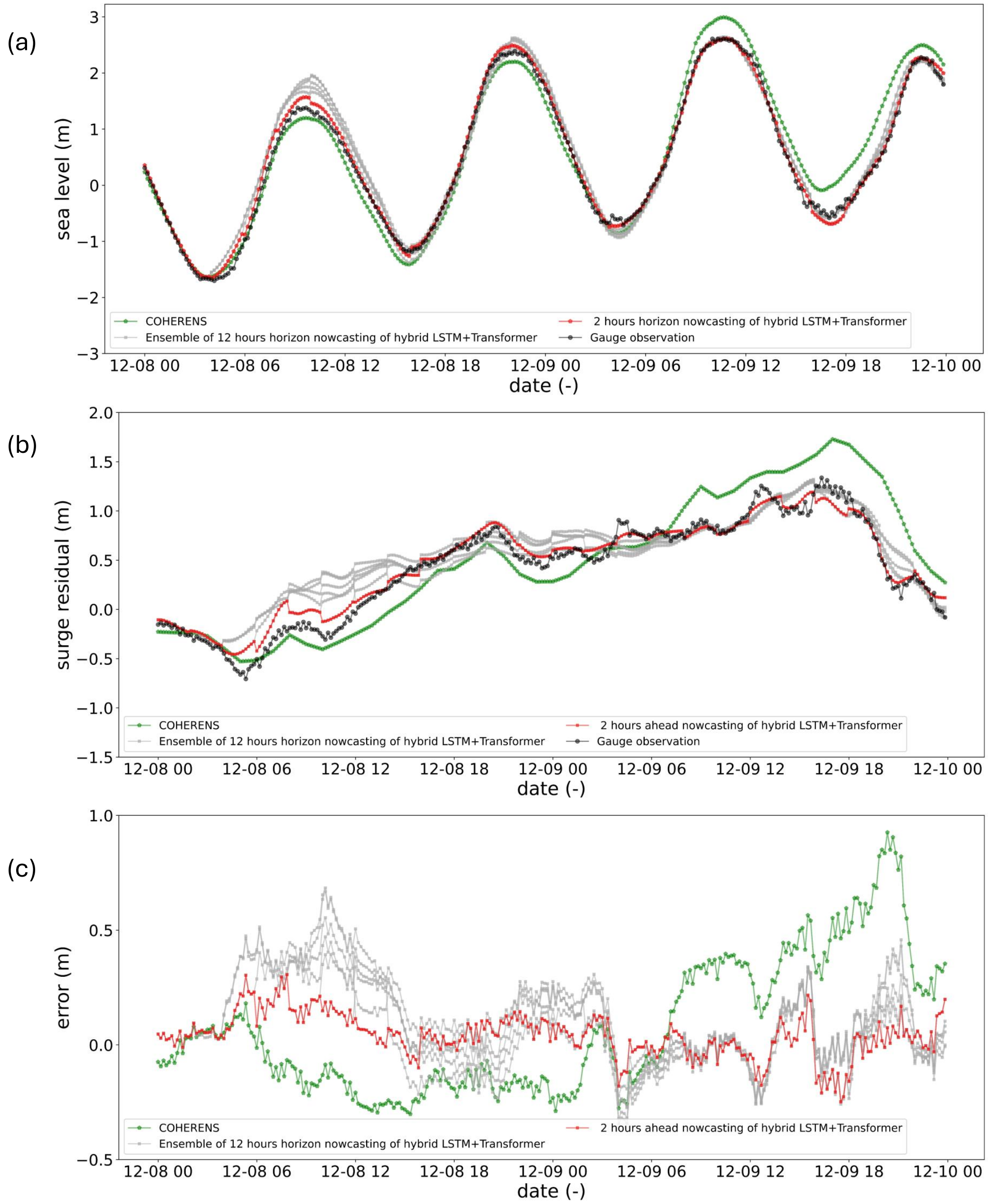

Figure 7. Sea level (a), surge residual (b), and difference between predicted and observed sea level (c) at Oostende during the storm surge of 8–9 December 2019.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

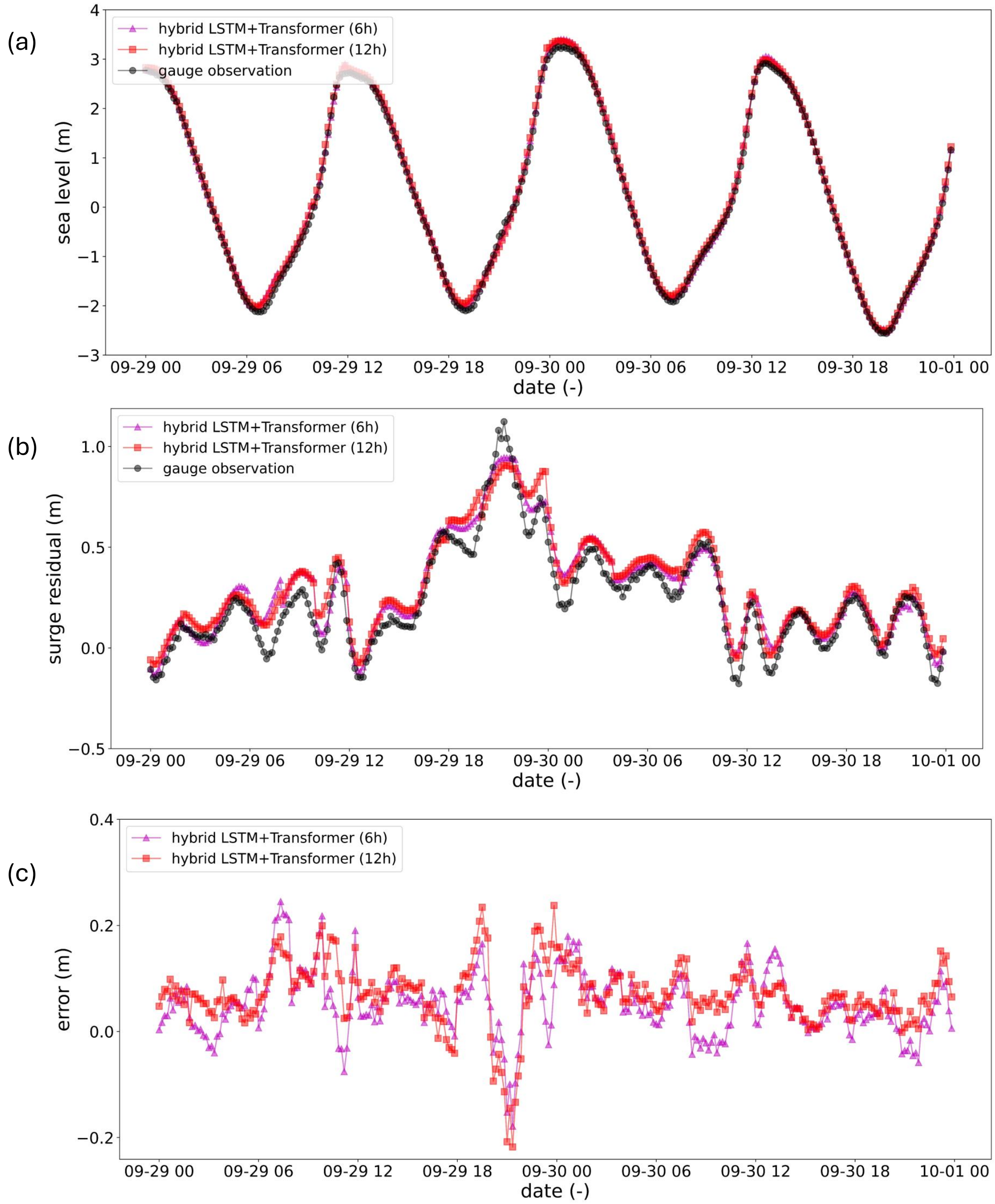

Figure 8. Sea level (a), surge residual (b), and difference between predicted and observed sea level (c) at Oostende (September 29,30 of 2019), comparing the hybrid LSTM+Transformer trained with a 6-hour and a 12-hour look-back window.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

Table 8. Station-averaged error metrics of the hybrid LSTM+Transformer with 4 and 8 Transformer encoder layers, compared with the 6-layer default. Per-station values are given in A. Best values per metric are in bold for RMSE and R^2.

layers	RMSE (m)	R^2 (−)	bias (m)
4	0.067	0.994	−0.020
6 (default)	**0.055**	**0.996**	0.019
8	0.075	0.992	−0.006

3.5.2 Number of Transformer encoder layers

The sensitivity of the hybrid LSTM–Transformer model to the Transformer encoder depth was assessed by training configurations with 4 and 8 layers in addition to the 6-layer default (Table 8). The 4-layer and 8-layer models yield average RMSE values of 0.067 m and 0.075 m, respectively, compared to 0.055 m for the 6-layer configuration. The R^2 values show a consistent ordering. While the 8-layer model exhibits a slightly smaller mean bias, the lowest RMSE and highest R^2 are obtained with 6 layers.

4 Discussion

The results of the previous section show that the hybrid LSTM+Transformer is the most accurate of the four configurations tested, reaching a 2-hour-ahead RMSE of 0.055 m on average across the 22 stations — about three times lower than the COHERENS operational baseline — and that the hybrid input scheme also brings a clear improvement at extended horizons. We discuss in turn the interpretation, limitations and operational implications of these findings.

4.1 Value of the hybrid input

The hybrid Transformer outperforms the pure Transformer configuration, with an RMSE of 0.152 m compared to 0.254 m at a 12-hour forecast horizon. The performance gap increases with lead time; the pure Transformer exceeds the COHERENS baseline beyond approximately 4 hours, whereas the hybrid model remains below it across all evaluated horizons. This suggests that COHERENS forecasts provide useful spatial and temporal information that is not fully captured by point-based tide gauge observations alone. The inclusion of model-derived sea surface elevation fields appears to improve predictive skill by providing additional hydrodynamic context.

For LSTM+Transformer architectures, the benefit of hybrid input is smaller, with an RMSE of 0.146 m compared to 0.194 m at a 12-hour forecast horizon. This may be explained because the LSTM component already captures much of the temporal variability present in the gauge data. However, the hybrid LST+Transformer still yields the best performance across the tested forecast horizons.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

Overall, combining in situ observations with physics-based model output appears beneficial, particularly for models without an explicit recurrent component. Nevertheless, the pure LSTM+Transformer demonstrates competitive performance and may remain a viable option for short-term predictions (up to 6 hours) based solely on tide gauge data.

4.2 Limitations

Several limitations should be acknowledged. First, the models are trained on a 5-year dataset, implicitly assuming that this period sufficiently captures surge regime variability for storm patterns with return periods on the order of 1–3 years. A longer training record would likely improve representation of more exceptional storm patterns. This suggests that periodic retraining—particularly following major surge events—may help maintain model robustness.

Second, the presented methodology relies on detidal signals, both in the tide gauge observation and COHERENS data set. Further methodological improvement shall have to be considered to account for waves-tide-surge interactions that sharpen the surge signal (Idier et al., 2019).

Atmospheric forcing is derived from the ERA5 reanalysis at $0.25°$ resolution. Higher-resolution products, such as the ECMWF IFS forecast ($\sim$9 km(Lang et al., 2023)), Copernicus European Regional ReAnalysis (CERRA(Ridal et al., 2024); $\sim$5.5 km) and the ALARO forecast ($\sim$2.2 km,(Termonia et al., 2018)), may better resolve coastal processes and could potentially improve model performance; this remains to be evaluated.

Finally, operational constraints, such as gauge maintenance or telemetry failures, can lead to gaps in the input data. The treatment of such missing data—whether through imputation (e.g. persistence, zero residuals, or interpolation from neighbouring stations) or alternative approaches (usage of a previous inference to fill the gap)—has not been assessed here and may influence model performance under real-world conditions.

4.3 Operational implication

The proposed framework is well suited to operational nowcasting. Inference takes less than one minute on a single GPU, which makes high-frequency cycle updates — for example every 10 minutes — entirely feasible, in contrast with the 12-hour cycle of the operational COHERENS product at RBINS. The model relies only on publicly available, quality-controlled data sources — tide-gauge observations, ERA5 forcing, and the operational COHERENS output that the centre already produces — so the additional operational footprint is minimal: the ML model can run alongside the existing physics-based system, consuming its output and refreshing the public-facing nowcast between COHERENS cycles.

5 Conclusions

We have presented a hybrid machine-learning framework for storm-surge nowcasting in the southern North Sea, in which simulations from the operational COHERENS hydrodynamic model are combined with observations from 22 tide gauges and ERA5 atmospheric forcing as inputs to a deep neural network. Two architectures were evaluated: a standalone Transformer and a parallel LSTM+Transformer, each in a pure (gauge-only) and a hybrid (gauge + COHERENS) configuration. Trained

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

on data from March 2020 to December 2024 and tested on the year 2019, the hybrid LSTM+Transformer achieves a station-averaged RMSE of 0.055 m (0.033–0.100 m across stations) for 2-hour-ahead surge predictions at 10-minute resolution, with R^2 between 0.991 and 0.999. This is substantially below the COHERENS operational baseline (average 0.164 m; 0.137–0.185 m) and modestly below the corresponding pure data-driven model (0.061 m; 0.043–0.100 m). The hybrid Transformer reaches 0.096 m (0.067–0.121 m). The hybrid models also reproduce the magnitude and timing of extreme events well, both for surge peaks and surge troughs.

Applied recursively, the hybrid models extend to longer horizons with smooth degradation. The hybrid LSTM+Transformer reaches an average RMSE of 0.146 m at 12 hours, and the hybrid Transformer 0.152 m — both well below the COHERENS 2-hour baseline. The pure LSTM+Transformer also remains below the baseline up to 8 hours, reaching 0.194 m at 12 hours. By contrast, the pure Transformer degrades to 0.254 m at 12 hours, exceeding the COHERENS baseline beyond roughly 4 hours and underscoring the value of incorporating the physical simulation as an input feature when the data-driven backbone lacks a recurrent component.

Two ablation experiments show that the model is robust to its principal design choices. Reducing the look-back window from 12 to 6 hours changes the 2-hour RMSE only at the third decimal place, and the 6-layer Transformer encoder used in the LSTM+Transformer branch outperforms both 4- and 8-layer variants.

Inference takes less than one minute on a single AMD MI250x GPU. The framework is therefore directly suitable for high-frequency operational nowcasting, and could be deployed alongside the existing physics-based system to refresh the public-facing nowcast between COHERENS cycles.

In future work, we plan to extend the framework to a larger domain, to add waves infomation as input, to explore Bayesian extensions of the Transformer for probabilistic surge forecasting, and to incorporate higher-resolution atmospheric forcing such as CERRA or ALARO — and possibly satellite altimetry — to further improve accuracy.

Code and data availability. The data used and models discussed in this paper are available to academic institutions for non-commercial research upon reasonable request by contacting the corresponding author.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

Appendix A: Per-station error metrics

Table A1. Per-station error metrics for the four ML configurations and the COHERENS baseline, for 2-hour-ahead nowcasting. Best values per station are in bold for RMSE and R^2; bias is reported without bolding since the smallest absolute bias does not necessarily correspond to the best-performing model.

stations	RMSE (m)					R^2 (−)					bias (m)				
	COHERENS	Transformer	hybrid Transformer	LSTM + Transformer	hybrid LSTM + Transformer	COHERENS	Transformer	hybrid Transformer	LSTM + Transformer	hybrid LSTM + Transformer	COHERENS	Transformer	hybrid Transformer	LSTM + Transformer	hybrid LSTM + Transformer
Zeebruggeleopoldiidam	0.166	0.115	0.088	**0.052**	0.058	0.983	0.992	0.995	**0.998**	**0.998**	0.028	−0.042	−0.019	0.015	0.033
ZeebruggeWielingendok	0.164	0.113	0.087	**0.051**	0.055	0.984	0.992	0.995	**0.998**	**0.998**	0.007	−0.041	−0.018	0.011	0.028
Nieuwpoort	0.176	0.120	0.101	**0.070**	0.078	0.986	0.994	0.995	**0.998**	0.997	0.018	−0.013	0.018	0.041	0.054
Oostende	0.170	0.119	0.093	**0.054**	0.060	0.985	0.993	0.996	**0.999**	0.998	0.012	−0.036	0.006	0.016	0.032
Blankenberge	0.185	0.134	0.121	**0.100**	**0.100**	0.983	0.991	0.993	**0.995**	**0.995**	0.029	−0.043	−0.018	0.002	0.018
Westhinder	0.162	0.097	0.084	**0.060**	0.063	0.984	0.994	0.996	**0.998**	**0.998**	−0.017	−0.025	0.006	0.011	0.022
A2	0.165	0.115	0.090	**0.048**	0.061	0.984	0.992	0.995	**0.999**	0.998	0.022	−0.043	−0.027	0.012	0.039
Bol van heist	0.163	0.120	0.091	**0.046**	0.049	0.983	0.991	0.995	**0.999**	0.998	0.020	−0.057	−0.032	−0.004	0.018
LichteilandGoeree	0.147	0.111	0.084	0.051	**0.033**	0.957	0.975	0.986	0.995	**0.997**	0.029	−0.060	−0.041	−0.015	0.004
Europlatform	0.137	0.089	0.067	0.062	**0.056**	0.949	0.979	0.988	0.990	**0.992**	0.021	−0.038	−0.017	0.019	0.033
VlakteVdRaan	0.161	0.128	0.102	0.057	**0.045**	0.981	0.988	0.992	0.998	**0.999**	−0.006	−0.085	−0.067	−0.035	−0.015
Brouwershavensegat8	0.162	0.144	0.116	0.062	**0.045**	0.966	0.974	0.983	0.995	**0.997**	0.005	−0.081	−0.065	−0.033	0.003
RoompotBuiten	0.163	0.122	0.100	0.048	**0.041**	0.976	0.986	0.991	**0.998**	**0.998**	0.028	−0.064	−0.045	−0.017	0.014
Hoek van holland	0.157	0.120	0.092	0.052	**0.049**	0.945	0.968	0.981	0.994	**0.995**	0.033	−0.050	−0.030	0.002	0.029
Scheveningen	0.161	0.130	0.097	0.058	**0.046**	0.937	0.959	0.977	0.992	**0.995**	0.042	−0.065	−0.041	−0.016	0.016
Ijmuiden	0.175	0.146	0.107	0.066	**0.049**	0.914	0.940	0.968	0.988	**0.993**	0.055	−0.072	−0.040	−0.022	0.011
Ijmondstroompaal	0.174	0.150	0.110	0.070	**0.048**	0.915	0.937	0.966	0.986	**0.994**	0.045	−0.076	−0.046	−0.031	−0.001
Den helder	0.155	0.150	0.104	0.077	**0.052**	0.924	0.929	0.966	0.981	**0.991**	0.040	−0.085	−0.050	−0.042	−0.011
Haringvliet10	0.163	0.123	0.098	0.062	**0.049**	0.956	0.975	0.984	0.994	**0.996**	0.038	−0.058	−0.041	−0.023	0.004
Dunkerque	0.175	0.113	0.099	**0.076**	0.078	0.989	0.995	0.996	**0.998**	**0.998**	−0.014	0.024	0.033	0.044	0.052
DealPier	0.154	0.107	0.093	0.069	**0.061**	0.991	0.995	0.997	0.998	**0.999**	−0.001	−0.021	−0.005	0.000	0.011
Westkapelle	0.171	0.125	0.096	0.043	**0.039**	0.980	0.989	0.993	**0.999**	**0.999**	0.021	−0.063	−0.043	−0.015	0.013
average	0.164	0.123	0.096	0.061	**0.055**	0.966	0.978	0.988	0.995	**0.996**	0.021	−0.051	−0.027	−0.004	0.019

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

Appendix B: Sea level above geoid, surge residual, and difference between predicted and observed sea level at Oostende during a storm surge event

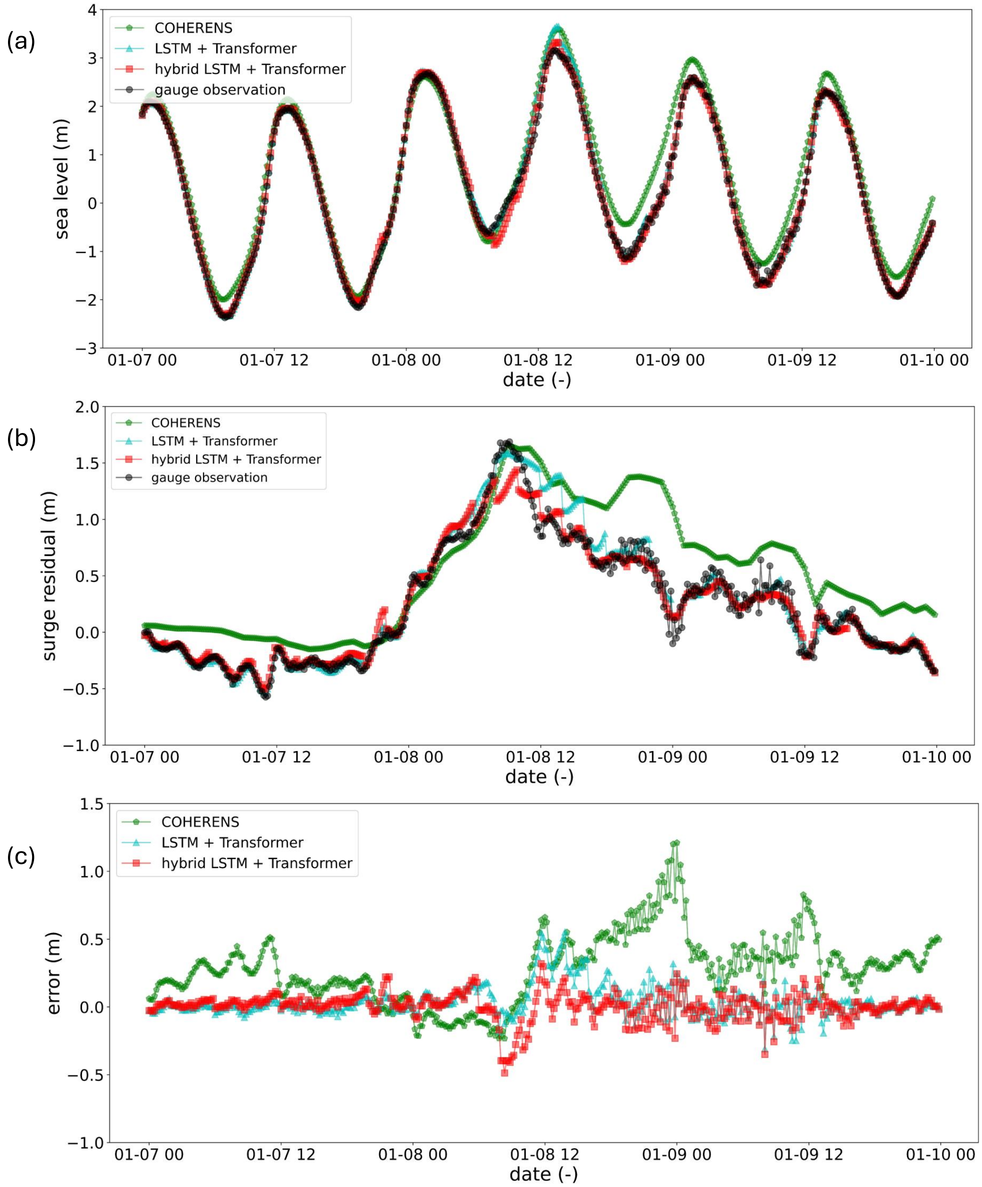

Figure B1. Sea level (a), surge residual (b), and difference between predicted and observed sea level (c) at Oostende during the storm surge of 7–9 January 2019.

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

Appendix C: Error metrics of various models trained with 6-hour look-back window, at forecast horizons from 2 to 12 hours

Table C1. Error metrics of the hybrid Transformer trained with a 6-hour look-back window, at forecast horizons from 2 to 12 hours.

horizon (h)	RMSE (m)	R^2 (−)	bias (m)
2	0.074	0.993	−0.008
4	0.108	0.986	−0.003
6	0.129	0.979	0.008
8	0.143	0.973	0.011
10	0.155	0.967	0
12	0.157	0.966	−0.010

Table C2. Error metrics of the Transformer trained with a 6-hour look-back window, at forecast horizons from 2 to 12 hours.

horizon (h)	RMSE (m)	R^2 (−)	bias (m)
2	0.071	0.994	−0.014
4	0.104	0.988	−0.022
6	0.133	0.979	−0.030
8	0.154	0.970	−0.040
10	0.172	0.962	−0.055
12	0.171	0.961	−0.067

Table C3. Error metrics of the LSTM+Transformer trained with a 6-hour look-back window, at forecast horizons from 2 to 12 hours.

horizon (h)	RMSE (m)	R^2 (−)	bias (m)
2	0.068	0.994	0.001
4	0.114	0.984	0.009
6	0.151	0.973	0.016
8	0.180	0.959	0.019
10	0.197	0.948	0.016
12	0.193	0.951	0.015

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

Author contributions. **Qiang Wang**: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft. **Ludovic Lepers**: Writing – review & editing. **Alexis Culot**: Writing – review & editing. **Emmanuel Hanert**: Writing – review & editing, Funding acquisition, Supervision. **Sebastien Legrand**: Conceptualization, Writing – review & editing, Funding acquisition, Supervision.

Competing interests. The authors declare no competing interests.

Acknowledgements. This study was supported by the FED-tWIN programme (Prf-2022-021). We acknowledge LUMI-BE for awarding this project access to the LUMI supercomputer, owned by the EuroHPC Joint Undertaking, hosted by CSC (Finland) and the LUMI consortium through a LUMI-BE Regular Access call. LUMI-BE is a joint effort from BELSPO (federal), SPW Économie, Emploi, Recherche (Wallonia), Department of Economy, Science & Innovation (Flanders) and Innoviris (Brussels).

https://doi.org/10.5194/egusphere-2026-4129
Preprint. Discussion started: 22 July 2026

References

Adam, Y.: Belgian real-time system for the forecasting of currents and elevations in the North Sea, in: Elsevier Oceanography Series, vol. 25, pp. 411–425, 1979.

Anon: Algemeen Nood- en Interventieplan Noordzee, FDOH, federale diensten Gouverneur West-Vlaanderen, FOD Binnenlandse Zaken, 2026.

Ayyad, M., Hajj, M. R., and Marsooli, R.: Machine learning-based assessment of storm surge in the New York metropolitan area, Scientific Reports, 12, 19 215, 2022.

Azevedo, B. F., Rocha, A. M. A., and Pereira, A. I.: Hybrid approaches to optimization and machine learning methods: A systematic literature review, Machine Learning, 113, 4055–4097, 2024.

Centre, M. F.: Harmonic Astronomical Tide - Continental Shelf - COHERENS, https://doi.org/10.24417/Harmonic_CSM0, 2023a.

Centre, M. F.: Tide - Continental Shelf - COHERENS UKMO, https://doi.org/10.24417/Tide_CSM6, 2023b.

Codiga, D. L.: Unified Tidal Analysis and Prediction Using the UTide Matlab Functions, Tech. Rep. 2011-01, Graduate School of Oceanography, University of Rhode Island, https://doi.org/10.13140/RG.2.1.3761.2008, 2011.

Copernicus Marine Service: Atlantic European North West Shelf Ocean In-Situ Near Real Time Observations, https://doi.org/10.48670/moi-00045, 2025.

Ebel, P., Victor, B., Naylor, P., et al.: Implicit Assimilation of Sparse In Situ Data for Dense and Global Storm Surge Forecasting, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 471–480, 2024.

Fu, C., Xiong, J., and Yu, F.: Storm surge forecasting based on physics-informed neural networks in the Bohai Sea, Journal of Physics: Conference Series, 2718, 012 057, 2024.

Gaslikova, L., Schwerzmann, A., Raible, C. C., and Stocker, T. F.: Future storm surge impacts on insurable losses for the North Sea region, Natural Hazards and Earth System Sciences, 11, 1205–1216, https://doi.org/10.5194/nhess-11-1205-2011, 2011.

Hermans, T. H., Ben Hammouda, C., Treu, S., et al.: Computing extreme storm surges in Europe using neural networks, Natural Hazards and Earth System Sciences, 25, 4593–4612, 2025.

Hersbach, H., Bell, B., Berrisford, P., et al.: The ERA5 Global Reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hersbach, H., Bell, B., Berrisford, P., , and others.: ERA5 hourly data on single levels from 1940 to present, https://doi.org/10.24381/cds.adbb2d47, 2023.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Computation, 9, 1735–1780, 1997.

Idier, D., Bertin, X., Thompson, P., and Pickering, M. D.: Interactions Between Mean Sea Level, Tide, Surge, Waves and Flooding: Mechanisms and Contributions to Sea Level Variations at the Coast, Surveys in Geophysics, 40, 1603–1630, https://doi.org/10.1007/s10712-019-09549-5, 2019.

Ishida, K., Tsujimoto, G., Ercan, A., et al.: Hourly-scale coastal sea level modeling in a changing climate using long short-term memory neural network, Science of the Total Environment, 720, 137 613, 2020.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., et al.: Physics-informed machine learning, Nature Reviews Physics, 3, 422–440, 2021.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980, 2014.

Lang, S., Schepers, D., and Rodwell, M.: IFS Upgrade Brings Many Improvements and Unifies Medium-Range Resolutions, ECMWF Newsletter, 176, 6–13, 2023.

Luyten, P. J., Jones, J. E., Proctor, R., et al.: COHERENS: A Coupled Hydrodynamical-Ecological Model for Regional and Shelf Seas: User Documentation, Tech. rep., Management Unit of the Mathematical Models of the North Sea, 1999.

Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H., and Ward, P. J.: A global reanalysis of storm surges and extreme sea levels, Nature Communications, 7, 11 969, https://doi.org/10.1038/ncomms11969, 2016.

Naeini, S. S. and Snaiki, R.: A novel hybrid machine learning model for rapid assessment of wave and storm surge responses over an extended coastal region, Coastal Engineering, 190, 104 503, 2024.

Pugh, D. and Woodworth, P.: Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes, Cambridge University Press, 2014.

Ridal, M., Bazile, E., Le Moigne, P., Randriamampianina, R., Schimanke, S., Andrae, U., Berggren, L., Brousseau, P., Dahlgren, P., Edvinsson, L., et al.: CERRA, the Copernicus European Regional Reanalysis System, Quarterly Journal of the Royal Meteorological Society, 150, 3385–3411, https://doi.org/10.1002/qj.4764, 2024.

Shen, C.: A transdisciplinary review of deep learning research and its relevance for water resources scientists, Water Resources Research, 54, 8558–8593, 2018.

Tadesse, M. G. and Wahl, T.: A database of global storm surge reconstructions, Scientific Data, 8, 125, 2021.

Termonia, P., Fischer, C., Bazile, E., Bouyssel, F., Brožková, R., Caian, M., Giot, O., et al.: The ALADIN System and Its Canonical Model Configurations AROME CY41T1 and ALARO CY40T1, Geoscientific Model Development, 11, 257–281, https://doi.org/10.5194/gmd-11-257-2018, 2018.

Tian, Q., Luo, W., Tian, Y., et al.: Prediction of storm surge in the Pearl River Estuary based on data-driven model, Frontiers in Marine Science, 11, 1390 364, 2024.

Tiggeloven, T., Couasnon, A., van Straaten, C., et al.: Exploring deep learning capabilities for surge predictions in coastal areas, Scientific Reports, 11, 17 224, 2021.

Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention Is All You Need, in: Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017.