# Storm Surge Forecasting with LSTM Networks

**Villy Mik-Meyer, Francisco C. Pereira** and **Martin Drews**
*Tecnical Univercity of Denmark* vlymi@dtu.dk
**Morten Andreas Dahl Larsen**
*Danish Meteorological Institute*

## Abstract

Accurate storm surge prediction is essential for coastal risk management and climate adaptation, yet conventional hydrodynamic models are too computationally demanding for large ensemble analyses. We develop and validate a machine learning framework for the North Sea and Baltic Sea, trained on 58 years of wind and sea level data using a Long Short-Term Memory (LSTM) architecture. The approach requires only a fraction of the resources of physical models, enabling rapid forecasts across large domains and time horizons. This efficiency makes ensemble-based climate impact assessments feasible, offering a scalable alternative for projecting extreme water levels and their statistical distributions under future climate scenarios.

**Keywords:** storm surge forecasting, long short-term memory (LSTM), climate impact assessment

## 1 Introduction

Storm surges are among the most hazardous climate events (Merz et al. [2021]), posing high risks to infrastructure, ecosystems, and population safety in the North and Baltic Sea regions (Andrée et al. [2021]; Leszczyńska et al. [2025]). Accurate forecasting is essential both for short-term warning systems and long-term adaptation. However, the Baltic Sea and adjacent Danish waters present unique challenges: their shallow, semi-enclosed basins exhibit long-lasting filling and flushing effects, making surge dynamics difficult and computationally expensive to capture with conventional hydrodynamic models.

Traditional approaches divide into two categories: operational forecasts based on hydrodynamic models and long-term projections using Extreme Value Analysis (EVA). While hydrodynamic models are physically detailed, they are computationally intensive and difficult to scale. EVA, by contrast, depends heavily on observational records and often assumes stationarity, which is problematic under changing climate conditions (Slater et al. [2021]; Amante [2019]). As a result, large uncertainties remain about how surge intensity, duration, and extent will evolve in a changing climate.

To address this, we develop a data-driven model based on Long Short-Term

Memory (LSTM) networks, trained on 58 years of hindcast wind and water level data, produced by a physically based 3D hydrodynamic model called HBM. The model predicts surge levels directly from zonal and meridional wind components, maintaining accuracy across a seamless range of lead times from hours to decades. Once trained, it can generate forecasts within seconds, making ensemble-based climate impact assessments computationally feasible.

## 2 Study Area and Data

The modeling domain covers the Danish West Coast, the Inner Danish Waters, and selected locations in the Baltic Sea. This region forms a hydrodynamically complex transition zone between the North Sea and the semi-enclosed Baltic Sea, connected through the narrow Danish straits (Skagerrak, Kattegat, Øresund, and the Belts). Along the Danish West Coast, storm surges are primarily generated by extratropical cyclones, where strong westerly winds acting over long fetches produce large and relatively fast water level variations. In contrast, the water levels in the Inner Danish Waters and coastlines along the Baltic Sea are influenced by both inflows from the North Sea and local wind forcing, resulting in highly complex surge responses on longer time scales.

The study domain was divided into 15 regions (see Figure 1), chosen to reflect areas with coherent wind forcing relevant for surge generation. For each region, wind data was averaged across all grid cells, resulting in spatially aggregated time series of zonal (u) and meridional (v) wind components. This approach captures the wind density over larger areas rather than isolated grid-point fluctuations, thereby emphasizing the large-scale atmospheric patterns most relevant for storm surge dynamics.

## 3 Methods

We employed a recurrent neural network based on Long Short-Term Memory (LSTM) units to predict storm surge levels. LSTMs are well suited for this type of prediction, because they are designed to capture temporal dependencies in sequential data and can represent both short- and long-term dynamics. The model was trained using a custom loss function designed to emphasize peaks and extremes. A dynamic weight function $w_i$ was introduced to increase the importance of peak and extreme events:

$$w_i = \begin{cases} 1, & y_i \le T_{\text{peak}}, \\ \max\{1, s_{\text{peak}}(y_i - T_{\text{peak}})\}, & T_{\text{peak}} < y_i \le T_{\text{ext}}, \\ \max\{1, s_{\text{peak}}(y_i - T_{\text{peak}}) s_{\text{ext}}\}, & y_i > T_{\text{ext}}, \end{cases} \tag{1}$$

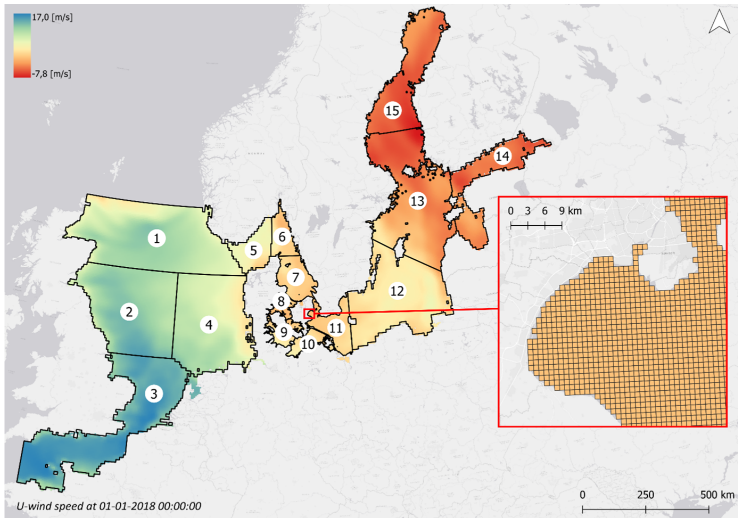


Figure 1: HBM model domain and grid (see zoomed in section) and areas (1 to 15) of regional averages.

where $T_{\text{peak}}$ and $T_{\text{ext}}$ are thresholds for peaks and extremes, and $s_{\text{peak}}$ and $s_{\text{ext}}$ are scaling factors. The total loss combined a weighted mean squared error at the median quantile with a pinball (quantile) loss for predicting the distribution around the central estimate:

$$L = \frac{1}{N}\sum_{i=1}^{N} w_i \left(y_i - \hat{y}_{i,q=0.5}\right)^2 + \sum_{\substack{q\in Q\\ q\neq 0.5}} \alpha_q \frac{1}{N}\sum_{i=1}^{N} \rho_q\left(y_i - \hat{y}_{i,q}\right), \tag{2}$$

where $Q = \{q_1, q_2, \ldots, q_K\}$ is the set of quantile levels, $\hat{y}_{i,q}$ is the model prediction at quantile $q$, $\alpha_q$ is a scale factor for quantile $q$, and $\rho_q$ is the pinball loss at quantile level $q$. To evaluate performance during Bayesian hyperparameter optimization, we applied the following Kling–Gupta efficiency (KGE) metric, slightly modified to focus on peaks:

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\beta_{\text{peak}} - 1)^2 + (\gamma - 1)^2}$$

Here, $r$ is the Pearson correlation between predictions and the target,$\gamma = \sigma(\hat{y})/\sigma(y)$ is the variability ratio, and $\beta_{\text{peak}}$ is the bias of the peaks, computed over the top 10% of observed values. This framework ensure that model training is prioritizes accurate peak representation while maintaining skill across the full distribution of storm surge levels. The modeling framework is illustrated in Figure 2.

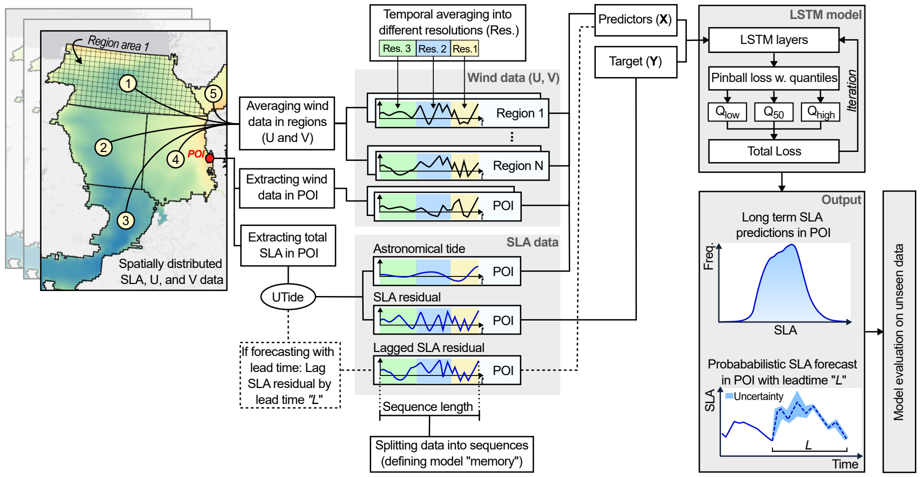


Figure 2: Conceptual diagram of LSTM model framework.

Table 1 presents a task-specific comparison: producing a 30-year, hourly water-level time series at a single location. The physically based HBM computes hydrodynamics on a spatial grid (multiple variables over many cells) on an HPC system, while the LSTM is trained once and then predicts a single-point time series on a workstation. The comparison isolates the workload for the stated task. Framed this way, the LSTM attains comparable single-point predictions with orders-of-magnitude lower computational demand, making it well suited for long-term or ensemble applications focused on point estimates.

Table 1: Computational performance comparison between the physically based hydrodynamic model (HBM) and the LSTM model for site-specific water level prediction. The comparison is task-oriented rather than model-to-model, as the two systems output different variables. Results are preliminary but indicate that the ML approach can achieve comparable site-specific predictions at a fraction of the computational cost.

| **Model** | **Description and performance metrics** |
|---|---|
| **Physically based model (HBM)** | ∼24 h wall-clock per simulated year on 8 nodes × 32 CPUs = 256 CPUs; equivalent to ∼6,144 core-hours per simulated year. |
| **Machine learning (ML) model** | Hardware: HP Z2 (Intel i9-14900K, 64 GB RAM). ∼2–3 minutes total to train and simulate 30 years (excluding Bayesian optimization for hyperparameter tuning); equivalent to ∼4–6 seconds per simulated year. |
| **Result (relative performance)** | ∼14,400–21,600× faster in wall-clock; ∼120,000–170,000× lower in core-hour cost (assuming full use of all 32 threads). |

# 4 Results and future work

The LSTM model reproduced storm surge dynamics across the Danish West Coast, Inner Danish Waters, and Baltic Sea with high accuracy when compared to hindcasted water levels from the HBM. Peak water levels were captured well, with KGE exceeding 0.95 at most of the 15 stations. The model effectively represented both rapid surge events in the North Sea (KGE = 0.98 in Esbjerg) and the longer filling–flushing processes characteristic of the Baltic Sea (KGE = 0.98 in Køge). Unlike the deterministic HBM model, the LSTM also estimates quantiles that showing the probability distribution around the forecast.

Importantly, predictive skill was maintained from short-term forecasts (hours to days) through long-term projections (years to decades) with a relatively small reduction in performance and without altering the model architecture. Once trained, the framework generated multi-decade time series at hourly resolution within seconds, demonstrating computational efficiency orders of magnitude higher than conventional hydrodynamic solvers. An example of a high-water level event in Køge (a coastal town around 30km south of Copenhagen) is shown in Figure 3.

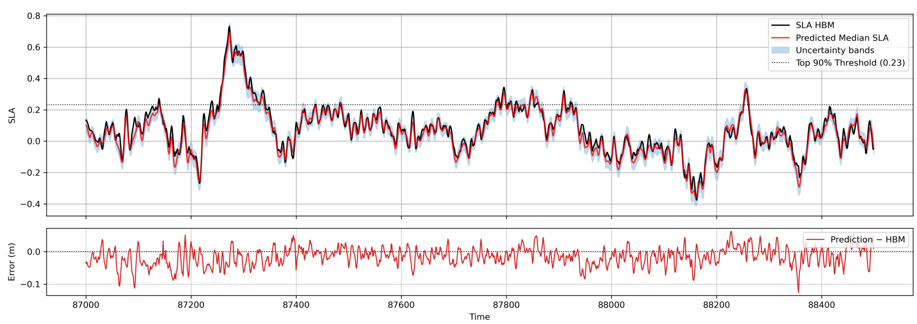


Figure 3: Predicted storm-surge time series with a 36h lead time at Køge compared with the HBM hindcast water levels. Shaded bands indicate LSTM predictive quantiles (probability distribution around the central estimate).

The framework offers several opportunities for collaboration and future development. It can be tested in operational settings to assess real-time forecasting performance. Future work includes geographically scaling the model to additional regions and conducting systematic cross-site evaluations. Extensions could couple the model with waves, atmospheric pressure, and river discharge to represent compound extremes. Finally, the computational efficiency enables climate-scenario experiments to analyse storm-surge behaviour in a warmer climate and supports adaptive planning.

# References

C. J. Amante. Uncertain seas: probabilistic modeling of future coastal flood zones. *International Journal of Geographical Information Science*, 33(11): 2188–2217, 2019. doi: 10.1080/13658816.2019.1635253.

E. Andrée, J. Su, M. A. D. Larsen, K. S. Madsen, and M. Drews. Simulating major storm surge events in a complex coastal region. *Ocean Modelling*, 162:101802, 2021. ISSN 1463-5003. doi: https://doi.org/10.1016/j.ocemod.2021.101802.

K. Leszczyńska, H. Alexanderson, L. Clemmensen, A. Giza, S. Lorenz, D. Moskalewicz, P. Oliński, D. Paprotny, A. Rosentau, A. Rutgersson, K. Stattegger, W. Szczuciński, P. Terefenko, G. Uścinowicz, S. Uścinowicz, and T. Wolski. A review of storms and marine coastal flooding in the baltic sea – insights from instrumental, historical and sedimentary record. *Earth-Science Reviews*, 266:105137, 2025. ISSN 0012-8252. doi: https://doi.org/10.1016/j.earscirev.2025.105137.

B. Merz, G. Blöschl, S. Vorogushyn, F. Dottori, J. C. J. H. Aerts, P. Bates, M. Bertola, M. Kemter, H. Kreibich, U. Lall, and E. Macdonald. Causes, impacts and patterns of disastrous river floods. *Nature Reviews Earth & Environment*, 2(9):592–609, 2021. ISSN 2662-138X. doi: 10.1038/s43017-021-00195-3.

L. J. Slater, B. Anderson, M. Buechel, S. Dadson, S. Han, S. Harrigan, T. Kelder, K. Kowal, T. Lees, T. Matthews, C. Murphy, and R. L. Wilby. Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. *Hydrology and Earth System Sciences*, 25(7): 3897–3935, 2021. doi: 10.5194/hess-25-3897-2021.