https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Open Access Earth System Science Data Discussions

DEN-SURGE: High-frequency and multi-decadal Danish storm surge reconstruction integrating hydrodynamic modelling, observations, and machine learning

Jian Su, Jacob Woge Nielsen, Kristine Skovgaard Madsen, and Morten Andreas Dahl Larsen
Danish Meteorological Institute, Sankt Kjelds Plads 11, Copenhagen, 2100, Denmark
Correspondence: Jian Su (jis@dmi.dk)

Abstract.

Assessing coastal hazards and planning robust defence infrastructure under a changing climate requires long-term, high-frequency sea-level records. However, historical tide gauge records are often temporally fragmented, while raw hydrodynamic hindcasts tend to systematically underestimate sub-hourly storm surge peaks due to spatial model resolution limitations. This paper introduces ReSEML v1.0 (Residual Surge Ensemble Machine Learning), a machine learning framework designed to reconstruct continuous, sub-hourly extreme sea-level catalogues across complex coastlines. By training a stacked deep learning ensemble (LSTM + TCN) to exclusively predict the local physical residual between sub-hourly tide gauge observations and a hydrodynamic model baseline, we build the continuous 64-year DEN-SURGE catalogue (1961–2024) across 50 Danish stations. By eliminating sensor malfunctions, closing multi-year monitoring gaps, and correcting numerical peak underestimations, DEN-SURGE delivers an AI-ready benchmark whose quality and physical completeness surpass both raw simulations and fragmented gauge records. Due to regional characteristics, corrections are applied by partitioning the coastline into 10 clusters. Out-of-sample validation shows that ReSEML reduces nationwide mean bias during high water events to within ± 0.4 cm and lowers Root Mean Squared Errors (RMSE) for the same events below 10 cm. Cross-referencing against independent historical storm surge lists confirms the framework's reliability in capturing historical peaks and recovering missing extremes. Extending the continuous record back to 1961 stabilises Extreme Value Analysis (EVA) curves, reducing 95% confidence intervals for the 100-year return level by on average 46%. While focused here on historical catalogue homogenisation, this work also provides a methodological bridge for downscaling future climate simulations, establishing a consistent framework to quantify evolving climate risks in coastal planning.

1 Introduction

Extreme sea levels (ESLs) are a major driver of coastal inundation, one of the most dangerous and costly natural disasters across the globe (Vousdoukas et al., 2017, 2018; Ganguli and Merz, 2019; Kirezci et al., 2020). Storm surge, an abnormal increase in sea level caused by the joint action of decreased atmospheric pressure and strong onshore winds in severe storm systems (Negus, 2016; Andrée et al., 2021, 2022), is a major part of many ESL events. Storm surges are events that can inflict widespread damage to coastal infrastructure, induce severe erosion of shorelines, and present considerable hazards to human

Open Access Earth System Science Data Discussions

life and coastal ecosystems (Hallegatte et al., 2011; Clemmensen et al., 2016; Halsnæs et al., 2023). Understanding, monitoring and accurately quantifying storm surge characteristics are essential to reduce risks in coastal zones and improve the resilience of communities (Wahl et al., 2018; Hinkel et al., 2019).

High-quality, long-term, and high-frequency storm surge records are important datasets for modern coastal engineering and climate adaptation planning (Woodworth et al., 2016; Muis et al., 2020). These records underpin many applications, including coastal hazard assessments that support vulnerability mapping and the development of emergency management strategies (Jebens et al., 2016; Keskitalo et al., 2016). They also provide the statistical basis for design criteria for coastal defence structures and critical infrastructure, ensuring that these systems are robust to present and future extreme events (Arns et al., 2013; Hinkel et al., 2014). Moreover, long-term and continuous time series are essential for a rigorous Extreme Value Analysis (EVA), the statistical method used to estimate the frequency and magnitude of rare events (Hosking and Wallis, 1987; Coles, 2001; MacPherson et al., 2023). Short, fragmented or incomplete datasets can strongly reduce the reliability and accuracy of EVA results (Parent and Bernier, 2003; Calafat and Marcos, 2020).

The demand for high-quality storm surge observations is further amplified by anthropogenic climate change (IPCC, 2021; Su et al., 2021). Rising global mean sea level raises the baseline upon which storm surges are superimposed, thereby increasing the likelihood and severity of coastal flooding (Church et al., 2013; Buchanan et al., 2017; Oppenheimer et al., 2019). Moreover, there is increasing evidence that climate change is possibly intensifying storm systems and associated precipitation, which can further exacerbate surge-related impacts (Ulbrich et al., 2009; Donat et al., 2011; Zscheischler et al., 2018).

Long-term trends in storm surge behaviour can only be detected and attributed with confidence when there are consistent observational records over several decades and with sufficient temporal resolution to separate anthropogenic climate signals from natural variability (Jevrejeva et al., 2014; Marcos et al., 2015; Dangendorf et al., 2021). However, many historical records are too short, discontinuous, or incomplete to obtain a robust characterisation of these signals (Wahl et al., 2013; Woodworth, 2006). This data gap poses a significant challenge for efforts to increase global coastal resilience, and limits the ability to rigorously test and validate physics-based climate and ocean models used to project future coastal hazards (Capet et al., 2020; Hermans et al., 2020).

Storm surge information is typically obtained from two main sources, direct observations from tide gauges and simulations from physics-based hydrodynamic models (Idier et al., 2019; Muis et al., 2020). Direct in-situ measurements from tide gauges form the basis of sea level science, with global repositories such as the Global Extreme Sea Level Analysis (GESLA) dataset collating large amounts of high-frequency records (Woodworth et al., 2016). However, tide gauge networks provide sparse point measurements with large parts of the coastline unmonitored (Holgate et al., 2013). Historically, the deployment of high-frequency tide gauges across Northern Europe focused primarily on major navigation routes and vulnerable harbours (e.g., Esbjerg and Cuxhaven), resulting in records longer than 100 years. Conversely, inner straits and sheltered fjords often lack continuous digitised histories before 1990–2000, creating an observational gap that physics-guided machine learning (ML) is uniquely suited to bridge (Kierulf et al., 2014; Vestøl et al., 2019; Dangendorf et al., 2021; Jacobsen et al., 2021). To overcome these spatio-temporal limitations, the scientific community resorts to physics-based hydrodynamic models to generate continuous hindcasts of storm surge (Muis et al., 2016; Dullaart et al., 2020). However, these models are simulations

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

with significant errors for the most extreme events (Mentaschi et al., 2013; Andrée et al., 2021). Discrepancies can stem from various sources, including errors in atmospheric forcing data, poor representation of bathymetry, coarse spatial resolution, and simplified parameterisations of physical processes (She et al., 2007; Capet et al., 2020; Dullaart et al., 2020). Importantly, these model errors are often highly state-dependent, with non-linear dependence on storm severity and trajectory, rendering simple, static correction methods ineffective (Andrée et al., 2022).

The weaknesses of these data sources, sparse but locally accurate observations and spatially complete models containing systematic errors, suggest a clear need for a framework that effectively combines them (Madsen et al., 2015; Calafat and Marcos, 2020). The goal is to correct systematic errors in model hindcasts using more reliable information from tide gauges, thus producing a superior data product that is long-term, high-frequency and spatially distributed (Muis et al., 2020). Model residual correction is not a novel concept, but classical statistical methods such as the delta change method or quantile mapping often struggle to correct complex non-linear errors and perform poorly when correcting the extreme tails of the distribution (Räty et al., 2014; Laflamme et al., 2016). Also, these methods are often based on a stationarity assumption which may not hold in a changing climate, limiting their effectiveness for the events of greatest interest for hazard assessment (Kallache et al., 2011). ML has recently emerged as an innovative methodology in the Earth sciences, with a particular relevance to this residual correction problem (Dubois et al., 2023). The ability of ML algorithms to approximate highly complex, non-linear functions from data allows them to learn the complicated, state-dependent relationships among a model's inputs, its state and the error in its output (Haas et al., 2014). This skill has led to a variety of applications in oceanography and climate science, from data reconstruction to the residual correction of model outputs (Dubois et al., 2023).

This paper presents the Residual Surge Ensemble Machine Learning (ReSEML) framework, a new methodology developed for homogenising and extending historical storm surge records by predicting and correcting physics-based model residuals during high water events. The ReSEML follows two basic architectural principles: residual learning and ensemble stacking. Inspired by the success of Residual Networks (ResNets) in deep learning (He et al., 2015), ReSEML trains ML models to predict the residual error (raw model output subtracted from observations) rather than the total storm surge. Unlike end-to-end AI/ML approaches that attempt to predict water levels directly, this formulation uses the hydrodynamic model as a constrained physical baseline and redefines the ML task exclusively to learning and correcting model residuals (Chen et al., 2022; Dubois et al., 2023; Rus et al., 2025; Mik-Meyer et al., 2026). This method has already shown great promise for the correction of operational forecasts of sea levels (Irazoqui Apecechea et al., 2023). To further improve accuracy and robustness, ReSEML employs a stacking ensemble framework that combines two heterogeneous base learners and trains a meta-learner to optimise the combination of their predictions. This hierarchical approach exploits the complementary strengths of different algorithms, reducing variance and improving generalisation, a technique that has been shown to be effective in complex geoscience applications (Haas et al., 2014; Dubois et al., 2023).

The novelty of this paper lies twofold. Firstly, the development and application of the ReSEML framework is presented through a full methodological outline, including its theoretical basis and practical implementation, followed by a comprehensive case study and a performance assessment. Secondly, a resulting new sub-hourly multi-decadal (1961–present) homogenised and extended open-source storm surge dataset for the Danish coastline is produced, entitled DEN-SURGE v1.0. This dataset

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

addresses the critical limitations of the underlying observational and model data, providing a high-quality product for downstream scientific and practical applications. The ReSEML framework provides a transferable template for other coastal and geoscientific applications aiming to correct systematic numerical model errors during high water events in regions with incomplete or fragmented observational histories (Capet et al., 2020; Muis et al., 2020).

2 Data and methods

2.1 Observational sea level data and hydrodynamic hindcast simulations

The foundation for training and validating the ReSEML framework consists of sea level observations from the Danish national tide gauge network, maintained by the Danish Meteorological Institute (DMI, Danish Meteorological Institute, 2026) and the Danish Coastal Authority (KDI, Danish Coastal Authority, 2026; Poulsen et al., 2024). The observational sampling frequency varies across the multi-decadal record, transitioning from hourly measurements in earlier decades (pre-1970s) to sub-hourly (15-minute intervals) during the late 20th century and reaching modern high-frequency (10-minute) resolution in recent decades. This dataset has been extensively used for many studies on historical storm surges (Andrée et al., 2021, 2022; Andrée et al., 2023; Su et al., 2021, 2024). Initial quality control (QC), including removal of obvious spikes and errors, was performed following standard procedures and later improved by Bauer et al. (2026). We performed additional event-scale QC within the ReSEML framework; see Section 2.3. The vertical datum for all records was homogenised to the Danish Vertical Reference 1990 (DVR90) using established benchmarks and correction methods. From the national network, a subset of 50 stations was selected for this study, ensuring broad coverage of the Danish coastline, encompassing the North Sea, Skagerrak, Kattegat, and Baltic Sea regions (Figure 1). These stations together represent a wide range of tidal regimes, storm-surge patterns, and the locations of the oceanic fetch from which the storm surge originates.

A significant challenge addressed in this work is the heterogeneity in the length of these observational records. Although Denmark uniquely holds 14 long-term tide gauge records, such as Esbjerg, Hornbæk, and Gedser, possessing near-continuous records extending back beyond 1961, most pre-digital data remain in paper archives (Hansen, 2018), leaving many modern digital time series with much shorter operational windows (Figure 1). To create a consistent dataset for long-term EVA covering the period from 1961 to the present day, records shorter than approximately 65 years require extension using reliable reconstruction methods based, e.g., on adjacent stations. Prior to use in the ReSEML framework, all observational time series were corrected for Glacial Isostatic Adjustment (GIA, using the same model as in Su et al. (2021); Poulsen et al. (2024)) to ensure consistency with the geocentric sea level simulated by the hydrodynamic model.

To provide the physical basis for reconstructing historical events, we use a long-term continuous hydrodynamic simulation of sea surface elevation covering the 64-year period 1961–2024 (Andrée et al., 2021). The multi-decadal simulation was performed using the HIROMB-BOOS Model (HBM) system. HBM employs a nested grid configuration featuring a coarser parent domain covering the Northwest European Shelf and high-resolution nested grids that resolve the intricate coastlines, shallow channels, and bathymetry of the Danish inner waters, the Sound, Belts, Skagerrak, and Kattegat (DMI, 2024). For the multi-decadal baseline period spanning 1961–2018, the hydrodynamic hindcast was forced by 11 km horizontal resolution

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

atmospheric fields derived from the UERRA (Copernicus Uncertainty Estimates for Regional Reanalysis) regional reanalysis dataset for Europe. Specifically, hourly surface fields, including 10-metre horizontal wind vectors (U_{10}, V_{10}) and mean sea level pressure (MSLP), were applied to drive the coastal storm surge model. Because the UERRA reanalysis ceases in 2018, the hindcast record is seamlessly extended across the remaining six-year monitoring period (2019–2024) using continuous sea level outputs from DMI's operational ocean model system (DMI, 2024), which is driven by meteorological forcing from the high-resolution operational HARMONIE-AROME numerical weather prediction model.

High-frequency, 10-minute sea surface elevation time series were extracted at the precise geographical coordinates of the 50 coastal tide gauge stations analysed in this study (Figure 1). This composite 64-year hydrodynamic series (η_{model}) serves as the primary predictor for the ML models within the ReSEML framework, which are optimised to predict the non-linear residual correction ($\Delta\eta$) between the numerical simulation and the quality-controlled, GIA-corrected observational ground truth during high water events.

2.2 Station clustering and DEN-SURGE v1.0

The spatial distribution of the 50 coastal tide gauges and their respective historical data densities form the core structural foundation of the DEN-SURGE v1.0 dataset. Due to the complex transition of the Danish Straits, which bridge the highly dynamic, macro-tidal North Sea with the brackish, micro-tidal Baltic Sea, coastal hydrodynamics are highly localised. Rather than using a single, nationally generalised model residual correction function, we implement agglomerative hierarchical clustering of the coastal lines with Ward's minimum variance linkage on scale-normalised physical parameters (standard deviation, 99th-percentile extreme surges, and mean water levels) from the observations. Observational mean sea level referenced to DVR90 is included in the clustering feature vector to preserve spatial gradients in regional dynamic topography across the Danish transition zone. Station-specific static model-observation offsets are subsequently removed before ML training, ensuring that the neural network focuses purely on learning dynamic, time-varying storm surge residuals ($\widehat{\Delta\eta}$) rather than static datum shifts.

As shown in Figure 2, cutting the resulting dendrogram at a hierarchical distance of $d = 1.0$ partitions the 50 stations into 10 distinct coastal clusters. Sorting the stations along the vertical axis according to the dendrogram's leaf sequence groups them into distinct clusters based on local surge variability and tidal influence. The resulting clusters are indexed (Clusters 0–9) following the deterministic subtree traversal order of the agglomerative hierarchical clustering algorithm, ensuring an objective grouping purely driven by local surge variability, extreme quantiles, and observed mean sea level.

By applying the regional, cluster-specific ReSEML stacking pipeline, comprising a dual Long Short-Term Memory (LSTM) and Temporal Convolutional Network (TCN) base estimator layer coupled with a Random Forest meta-learner across these distinct zones, the DEN-SURGE v1.0 framework is used to reconstruct high-frequency water levels across all gaps and unrecorded historical epochs. For the vast majority of the 50 stations, this hybrid physical-statistical approach extends the usable high-resolution timeline by 30 to 50 years, generating a continuous, homogenised, 64-year "virtual tide gauge" network spanning from 1961 to 2024.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

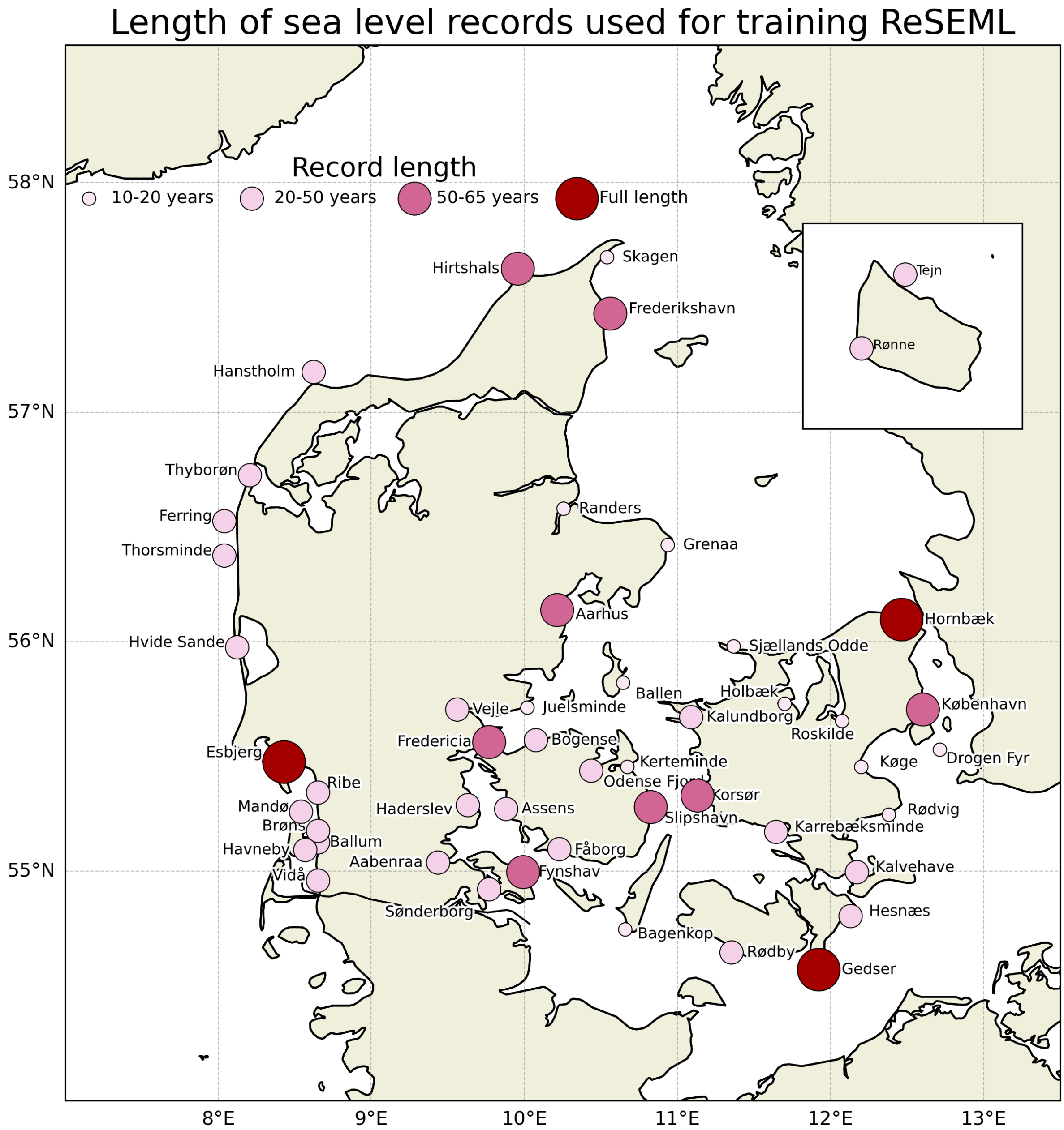

Figure 1. Spatial distribution of the 50 Danish tide gauge stations used for training ReSEML in this study. The colour and size of each circle indicate the duration (in years) of the available observational record for that station. The target period (full period) for historical storm surge reconstruction using the ReSEML framework is 1961–2024. Stations with record lengths shorter than the full period require extension to cover this full period. The three stations with the longest continuous records (Esbjerg, Hornbæk and Gedser, marked in red), together with the extensive historical peak lists from KDI (Figure 2), serve as primary benchmarks for validating the machine learning framework. The inset panel on the right highlights the Baltic island of Bornholm, which hosts two coastal tide gauge stations (Rønne and Tejn).

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Open Access Earth System Science Data Discussions

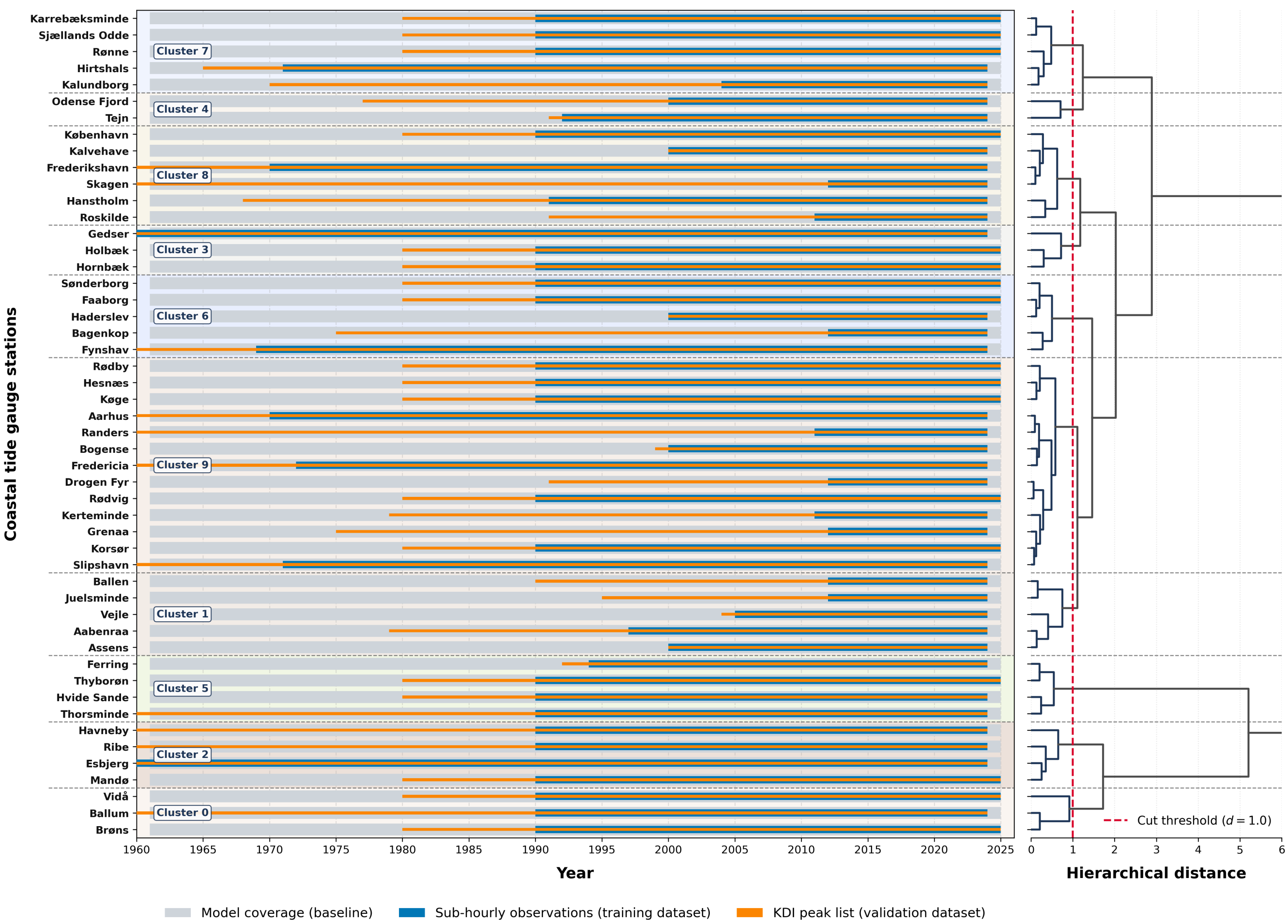

Figure 2. Unified spatial-temporal data coverage and hierarchical clustering regimes of the DEN-SURGE v1.0 network (1961–2024). Left panel: Gantt-style timeline tracking available data streams across the 50 coastal tide gauge stations. The gray bars indicate the continuous 64-year hydrodynamic model hindcast baseline (1961–2024). Dark-blue bars denote digitised, high-frequency tide gauge observations, and orange bars mark the temporal extent of manual storm surge events documented in the historical KDI peak lists. Right panel: Corresponding agglomerative hierarchical clustering dendrogram constructed from station surge variability (σ), 99th-percentile extremes, and mean sea level. The vertical red dashed line marks the cutting threshold ($d = 1.0$) defining the 10 coastal clusters. Alternating shaded background bands, horizontal dashed divider lines, and labelled cluster badges delineate the 10 clusters, ordered along the vertical axis following the dendrogram leaves.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

2.3 ReSEML machine learning framework

The ReSEML framework is structured as a pipeline architecture designed to maximise the reliability of historical storm surge reconstructions by separating the preparation of high-quality target data from the enhancement of broad-scale hydrodynamic simulations (Figure 3). The first step of the framework focuses on observation harmonisation, which aims to create a consistent and physically sound ground truth from disparate tide gauge records. Raw digital records from the national tide gauge network exhibit structural non-stationarities, including historic datum adjustments and local harbour reference shifts. To ensure multi-decadal comparability across the 64-year analysis period, all station records are aligned to DVR90 and corrected for linear trends during the study period, e.g., due to regional GIA. This procedure eliminates spurious multi-decadal trends and standardises the baseline sea level before hydrodynamic residual computation.

Beyond long-term baseline shifts, digitised records frequently contain transient instrument corruptions, such as recording flatlines, power-cut dropouts, and artificial negative spikes. Training deep recurrent and convolutional networks on unfiltered series risks memorising these unphysical artefacts. To preserve physical realism while retaining authentic, high-amplitude storm surges, synchronous observation-model pairs pass through a two-tiered QC filter prior to sequence tensorisation:

- Physical plausibility threshold: An absolute tolerance ceiling immediately rejects any sea level observation, $\eta_{\mathrm{obs}}(t)$, whose instantaneous deviation from the hydrodynamic hindcast for the same location, $\eta_{\mathrm{model}}(t)$, exceeds physical limits:

$$|\Delta\eta| = |\eta_{\mathrm{obs}}(t) - \eta_{\mathrm{model}}(t)| > \Delta\eta_{\max} \tag{1}$$

  where $\Delta\eta_{\max} = 150\,\mathrm{cm}$ based on past validation exercises.

- Local rolling IQR envelope: For transient anomalies that remain within the hard ceiling, an interquartile range (IQR) envelope is computed over a rolling 48-hour event window centred at time t:

$$\mathrm{IQR}_{48\mathrm{h}}(t) = Q_3(\Delta\eta) - Q_1(\Delta\eta) \tag{2}$$

  Measurements that diverge from the 48-hour rolling median by more than $3.5 \times \mathrm{IQR}_{48\mathrm{h}}(t)$, as well as continuous sensor flatlines spanning more than 3 hours, are flagged as instrumental artefacts and excluded from the supervised training set $\mathcal{D}_{\mathrm{train}}$.

This two-stage harmonisation ensures that the training residual target $\Delta\eta(t)$ reflects genuine meteorological and hydrodynamic discrepancies rather than corrupted sensor records.

Parallel to the observational track, the second step involves the processing and reconstruction of hydrodynamic model data (Figure 3). To reconstruct high-frequency sea levels across missing observation periods, we formulate a sub-hourly statistical residual correction model. Rather than predicting absolute sea level directly, the ML models are trained to predict the local hydrodynamic residual (Δ_t), defined as:

$$\Delta_t = \mathrm{obs\_wl\_corrected}_t - \mathrm{model\_wl}_t \tag{3}$$

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

This residual formulation ensures that the ML models focus exclusively on resolving sub-grid physical processes, local tidal phases, and peak surges that the hydrodynamic model misses, while preserving the mass-balance constraints of the underlying numerical simulation.

To focus on the dynamics under storm surge conditions rather than normal periods, training sequences are extracted from independent storm events across the synchronous observation-hindcast timeline. Candidate events for each station are identified via peak detection using a local surge threshold derived from the threshold in Poulsen et al. (2024) (or a baseline threshold of 100 cm). A temporal separation constraint of 72 hours (3 days) is enforced between consecutive peaks to guarantee physical independence between distinct weather systems. For each detected storm surge peak t_{peak}, an event window spanning 24 hours before and after the peak (a continuous 48-hour window, $[t_{\text{peak}} - 24\,\text{h},\, t_{\text{peak}} + 24\,\text{h}]$) of synchronous high-frequency observations (η_{obs}) and 10-minute numerical hindcasts (η_{model}) is extracted, forming the consolidated multi-station training datasets $\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{val}}$, see section 2.5.

Prior to model training, a data time alignment pipeline pairs the hydrodynamic baseline simulations with the corresponding sub-hourly tide gauge observations via an exact timestamp inner join. For historical records recorded at 15-minute intervals, this preserves exact synchronous points at common 30-minute steps without introducing synthetic interpolation artefacts, while modern epochs utilise the native 10-minute resolution. Although both datasets are aligned, the numerical hydrodynamic simulations are driven by hourly meteorological forcing. Consequently, the baseline model water levels exhibit a characteristically smooth trajectory that is physically unable to resolve rapid, sub-hourly storm surge fluctuations and local tide phase offsets. The overlapping temporal segments are aligned directly to construct the target residual (Δ_t), which isolates these high-frequency non-linear discrepancies. This alignment is strictly a training-phase requirement; during multi-decadal reconstruction, the trained models operate independently of active observational data, relying solely on the simulated hydrodynamic states and their derived trends to predict the sub-hourly residual correction.

2.4 Cluster-specialist stacking ensemble architecture

To prevent a single global model from smoothing out distinct regional surge peaks, ReSEML deploys an independent cluster-specialist stacking ensemble for each of the 10 identified coastal clusters (Figure 3). Under this modular framework, both the base-learner layer and the meta-learner are trained, cross-validated, and deployed for inference exclusively within their designated cluster. This ensures that model parameters remain regularised against local surge dynamics without interference from unrelated coastal regimes.

For a given cluster, a sliding temporal sequence of model water levels and engineered temporal trends ($\mathbf{X} \in \mathbb{R}^{T \times C}$, where $T = 48$ hours) is passed to a bank of parallel, heterogeneous base estimators. In this implementation, the base layer combines two complementary deep learning architectures:

- **Long Short-Term Memory (LSTM) network:** Utilises recurrent memory cells to track multi-day sequential dependencies, tidal phase memory, and low-frequency basin oscillations (e.g., Baltic seiches and slow volume exchanges through the Danish straits).

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

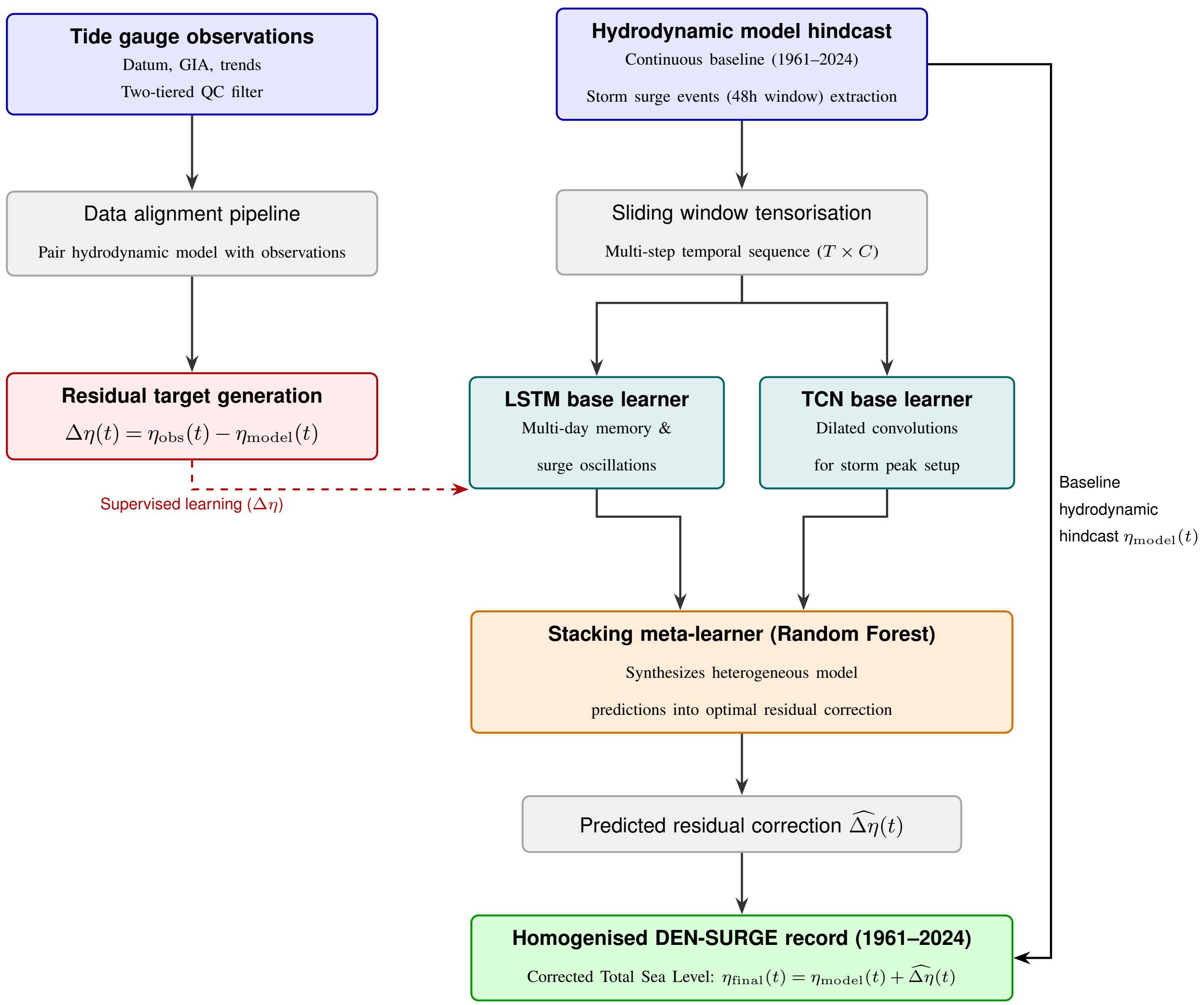

Figure 3. Schematic of the ReSEML two-fold framework for coastal storm surge homogenisation and reconstruction (Su, 2026). Raw tide gauge observations and hydrodynamic model hindcasts are harmonized and processed in parallel. High-frequency sequence tensors are fed into parallel LSTM and TCN base learners trained on the residual target $\Delta\eta$. A Random Forest meta-learner blends base model predictions to generate the continuous residual correction $\widehat{\Delta\eta}(t)$, which is superimposed onto the baseline hindcast to produce the DEN-SURGE v1.0 catalog (1961–2024).

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

– **Temporal Convolutional Network (TCN):** Leverages dilated 1D causal convolutions and residual blocks. The exponentially growing receptive field enables the network to capture high-frequency meteorological shifts and steep storm peak setup dynamics without suffering from vanishing gradients.

The architecture is modular; additional sequence estimators (e.g., Transformer or GRU layers) can be integrated into the base bank without altering downstream stacking logic.

The intermediate predictions generated by the base learners ($\hat{y}_{\mathrm{LSTM}}$ and $\hat{y}_{\mathrm{TCN}}$) form a secondary feature vector that serves as input to a cluster-level meta-learner:

$$\widehat{\Delta\eta}(t) = f_{\mathrm{meta}}\left(\hat{y}\mathrm{LSTM}(t),,\hat{y}\mathrm{TCN}(t)\right) \tag{4}$$

A Random Forest regressor is employed as the meta-learner, optimising an adaptive, non-linear blending strategy that weights each base model according to the instantaneous storm phase and peak intensity. During the reconstruction phase, inference is strictly partitioned: unobserved or historical periods for any tide gauge are generated solely by the specialist model trained on that station's parent cluster. The predicted continuous residual correction $\widehat{\Delta\eta}(t)$ is subsequently superimposed onto the baseline hydrodynamic hindcast (η_{model}) to yield the final reconstructed water level:

$$\eta_{\text{DEN-SURGE}}(t) = \eta_{\mathrm{model}}(t) + \widehat{\Delta\eta}(t) \tag{5}$$

2.5 Evaluation protocol and performance metrics

To quantitatively evaluate the reconstruction capability and generalisation skill of the ReSEML framework, we establish a standardised validation scheme and statistical benchmark protocol. For each coastal station, the temporal records containing synchronous digitised tide gauge observations (η_{obs}) and numerical hydrodynamic model hindcasts (η_{model}) during high water events are partitioned into two disjoint subsets. An in-sample training partition, comprising 80% of the observation history ($\mathcal{D}_{\mathrm{train}}$), is used to optimise the weights of the parallel base learners (LSTM and TCN) and train the cluster-specialist Random Forest meta-learner. The remaining 20% of the data ($\mathcal{D}_{\mathrm{val}}$) is strictly withheld from model optimisation to serve as an independent, out-of-sample validation partition. Model inference refers to the subsequent unconstrained forward evaluation where the trained ensemble generates continuous residual corrections ($\widehat{\Delta\eta}$) across the complete 64-year multi-decadal timeline (1961–2024), irrespective of local tide gauge record availability.

Model reliability and error reduction are quantified using three standard statistical metrics evaluated against the harmonised observation target. The instantaneous residual $\epsilon(t)$ is defined as the deviation between modelled or reconstructed water levels and the observed datum:

$$\epsilon(t) = \eta_{\mathrm{pred}}(t) - \eta_{\mathrm{obs}}(t) \tag{6}$$

where $\eta_{\mathrm{pred}}(t)$ represents either the uncorrected baseline hydrodynamic model $\eta_{\mathrm{model}}(t)$ or the final reconstructed sea level $\eta_{\text{DEN-SURGE}}(t) = \eta_{\mathrm{model}}(t) + \widehat{\Delta\eta}(t)$. Systematic mean deviations across a sample size of N observations are captured by the

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Mean Bias Error (MBE) during high water events:

$$\mathrm{MBE} = \frac{1}{N}\sum_{i=1}^{N}\left(\eta_{\mathrm{pred}}(t_i) - \eta_{\mathrm{obs}}(t_i)\right) \tag{7}$$

where a positive (negative) MBE indicates a systematic overestimation (underestimation) of the observed water level by the model.

The overall variance and magnitude of the prediction error during high water events are assessed via the Root Mean Square Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\eta_{\mathrm{pred}}(t_i) - \eta_{\mathrm{obs}}(t_i)\right)^2} \tag{8}$$

To benchmark the localised performance gain achieved by the ML framework relative to the raw numerical simulation on identical out-of-sample data points, we define the spatial error reduction metric $\Delta\mathrm{RMSE}$:

$$\Delta\mathrm{RMSE} = \mathrm{RMSE}_{\mathrm{model}}\left(\mathcal{D}_{\mathrm{val}}\right) - \mathrm{RMSE}_{\mathrm{ReSEML}}\left(\mathcal{D}_{\mathrm{val}}\right) \tag{9}$$

A positive $\Delta\mathrm{RMSE}$ signifies a net reduction in error variance and direct accuracy improvement delivered by the ReSEML ensemble over the baseline numerical hindcast.

3 Results

The evaluation and historical reconstruction of the DEN-SURGE v1.0 dataset are structured into three progressive stages. First, we examine the multi-decadal climatological performance of the uncorrected baseline hydrodynamic model across all 50 tide gauge stations (1961–2024) to diagnose persistent physical patterns in mean bias (MBE) and variance during high water events. Second, we assess the out-of-sample predictive skill and error reduction achieved by the cluster-specialist ReSEML framework on the held-out validation partition ($\mathcal{D}_{\mathrm{val}}$). Finally, to verify the physical consistency of reconstructed surge extremes during historical epochs lacking continuous digital observations, we compare the homogenised time series against the independent KDI manual storm surge peak lists and perform EVA across the network.

3.1 Baseline hydrodynamic hindcast performance across multi-decadal records

To diagnose the spatial structure of systematic errors inherent in the numerical simulations before ML reconstruction, we evaluate the baseline hydrodynamic model against all available digitised tide gauge records across its full 64-year simulation span (1961–2024). Figure 4 summarises the climatological mean bias (MBE) and RMSE across the 50 stations.

The spatial distribution of model mean bias (MBE) exhibits pronounced regional clustering driven by local bathymetry and coastline geometry (Figure 4a). Across the open waters of the Kattegat and North Sea, mean bias values remain modest (within ± 2 cm). However, severe negative mean bias exceeding -8 to -14 cm emerges systematically within semi-enclosed embayments and narrow fjord systems. This systematic underestimation stems from the finite spatial resolution of the numerical

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Open Access Earth System Science Data Discussions

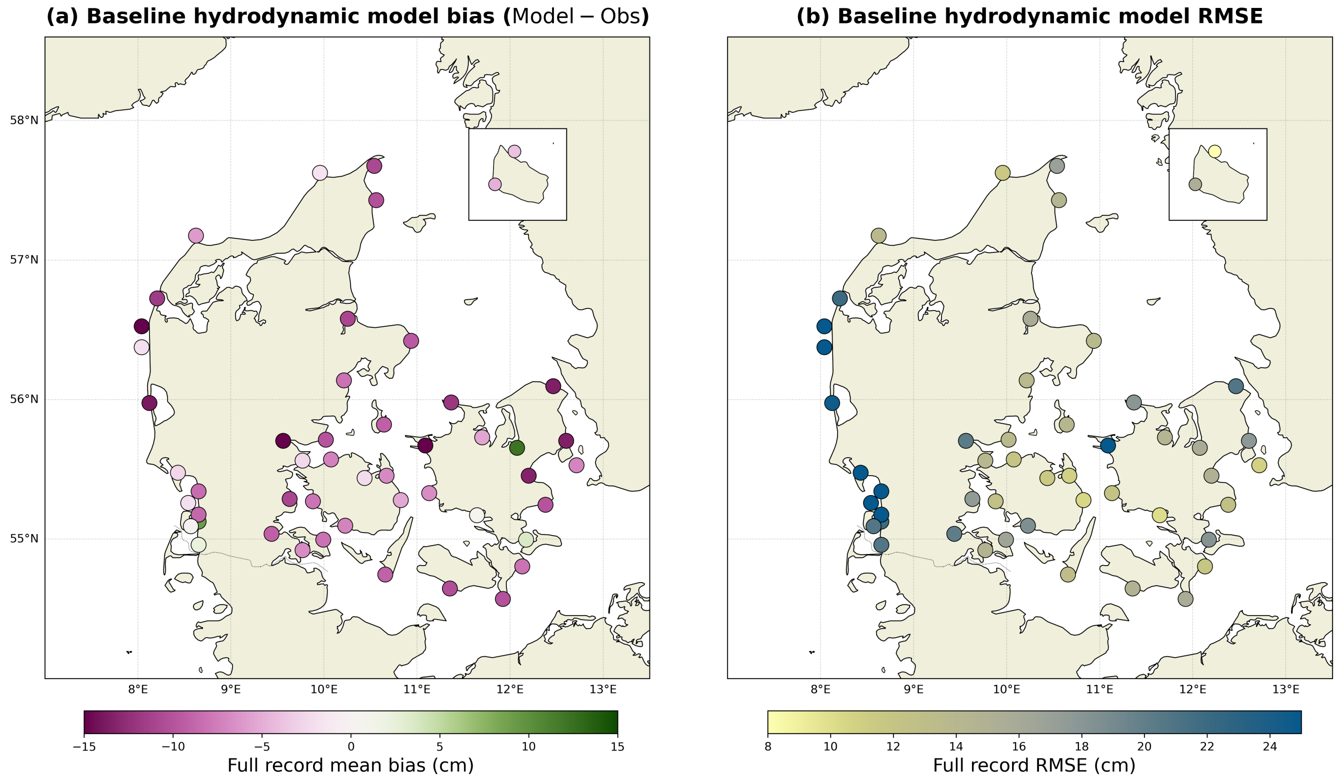

Figure 4. Climatological baseline performance of the raw hydrodynamic model across all available multi-decadal tide gauge observations (1961–2024). (a) Mean bias ($\mathrm{MBE} = \mathrm{Model} - \mathrm{Obs}$, in cm), illustrating regional underestimation in restricted inner fjords and straits. (b) Root Mean Square Error (RMSE, in cm), highlighting elevated error variance in the macro-tidal Danish Wadden Sea and complex navigational channels. Insets display the island of Bornholm.

simulation, which under-resolves narrow entrance channels, local friction, and shallow-water wind setup amplification in restricted coastal geometries.

Similarly, baseline RMSE values display strong physical regime dependence (Figure 4b). While micro-tidal Baltic stations (such as Gedser, Rødby, and Bornholm) maintain relatively low baseline errors (6–9 cm), error magnitudes escalate sharply along the Danish Wadden Sea coast (e.g., Esbjerg, Vidå, and Mandø), where RMSE values exceed $18–24$ cm. In these shallow, macro-tidal zones, non-linear shallow-water interactions, phase shifts in the astronomical tide, and boundary-layer friction introduce substantial variance that standard hydrodynamic configurations cannot fully capture. These pronounced regional patterns in mean bias and variance underscore the necessity of the cluster-specialist ReSEML framework to establish a homogenised multi-decadal baseline.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

3.2 Out-of-sample validation and error reduction across the training baseline

To rigorously assess whether the cluster-specialist ReSEML ensemble neutralises systematic local errors without overfitting, we evaluate the mean bias (MBE) on the strictly withheld 20% out-of-sample validation partition ($\mathcal{D}_{\mathrm{val}}$). Figure 5 compares the mean bias during high water events for the raw hydrodynamic model against the ReSEML ensemble on these identical, unseen records.

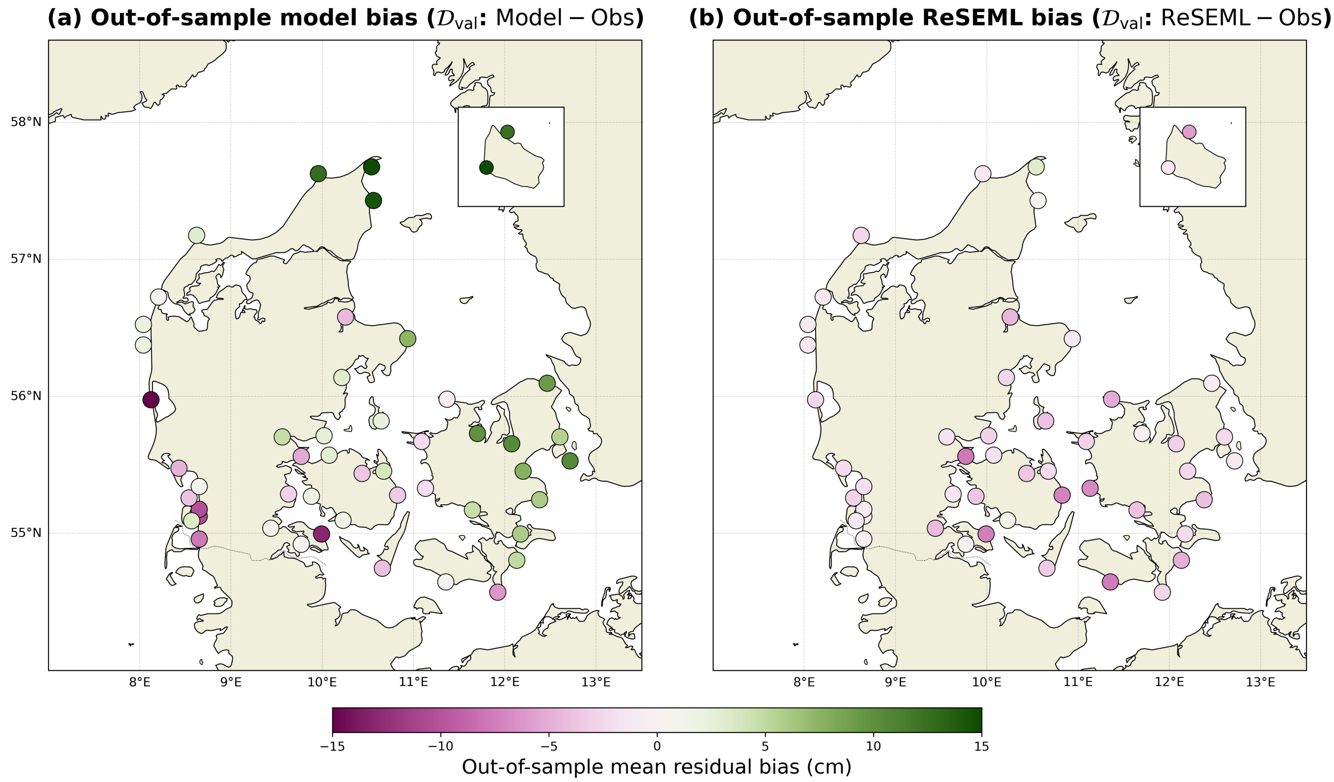

Figure 5. Spatial comparison of out-of-sample mean bias (MBE) during high water events evaluated on the independent 20% validation partition ($\mathcal{D}_{\mathrm{val}}$). (a) Baseline hydrodynamic model mean bias (Model − Obs, in cm). (b) ReSEML ensemble mean bias (ReSEML − Obs, in cm), demonstrating near-zero mean bias across all 10 coastal clusters. Insets depict the island of Bornholm.

On the validation set, the raw numerical model exhibits persistent local offsets, with severe underestimations ranging from -8 to -15 cm across restricted fjords (e.g., Vejle, Odense Fjord, and Roskilde) and southern Baltic passages. Following ReSEML reconstruction, these systematic offsets are virtually eliminated across the entire network (Figure 5b). The ensemble residual mean centres tightly around zero (± 1.0 cm at over 90% of stations), confirming that the stacking meta-learner successfully resolves unresolved shallow-water setup and friction losses without introducing geographic artifacts or drift.

The reduction in total prediction error variance is benchmarked across the network using the out-of-sample RMSE evaluated on $\mathcal{D}_{\mathrm{val}}$ (Figure 6). The baseline numerical hindcast yields an average out-of-sample RMSE exceeding 12–16 cm across the

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Kattegat and central straits, reaching peak errors of 20–24 cm in the macro-tidal Wadden Sea (Figure 6a). The ReSEML stacking architecture substantially attenuates this error variance across all stations (Figure 6b), depressing nationwide validation RMSE values down to 4–8 cm in the Baltic and inner straits and under 12 cm in the Wadden Sea.

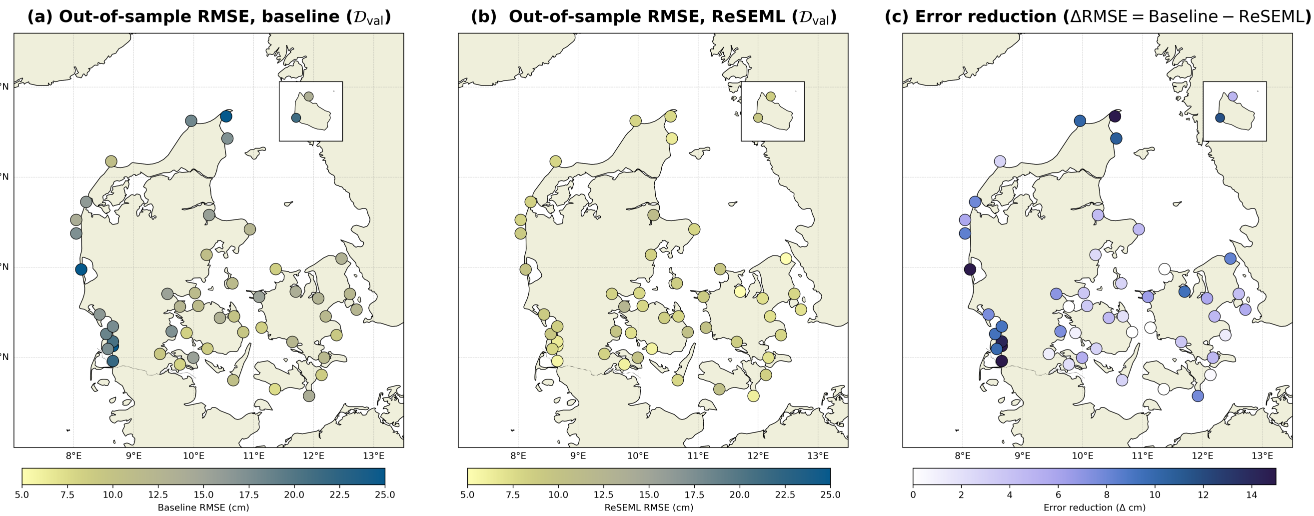

Figure 6. Out-of-sample error variance and spatial performance gain during high water events evaluated across the 20% validation partition ($\mathcal{D}_{\mathrm{val}}$). (a) Baseline hydrodynamic model RMSE (cm). (b) ReSEML ensemble RMSE (cm). (c) Spatial error reduction ($\Delta\mathrm{RMSE} = \mathrm{RMSE}_{\mathrm{model}} - \mathrm{RMSE}_{\mathrm{ReSEML}}$, in cm), highlighting nationwide accuracy gains peaking in energetic macro-tidal and transition regimes.

As quantified by the spatial error reduction metric ($\Delta\mathrm{RMSE}$, Figure 6c), every station in the 50-gauge network experiences a positive error reduction. Performance gains are especially pronounced along the west coast of Jutland (e.g., Ferring, Thorsminde, Hvide Sande) and western Wadden Sea entrances, where $\Delta\mathrm{RMSE}$ improvements reach 8–14 cm. This confirms that the parallel TCN and LSTM pathways effectively reconstruct both high-frequency storm-peak amplitudes and multi-day phase oscillations across diverse coastal regimes.

3.3 Regional error characteristics and skill metrics by coastal cluster

To provide a comprehensive overview of the reconstruction framework across the entire domain, Table 1 summarises the out-of-sample RMSE and mean bias (MBE) performance aggregated across the 10 coastal regions.

The cluster-aggregated metrics in Table 1 confirm the nationwide performance patterns illustrated in Figures 5 and 6. The uncorrected baseline hindcast exhibits structural divergences tied to local hydrodynamic behaviour rather than strict contiguous geography. In high-energy, exposed or macro-tidal regimes such as C0 (Wadden Sea) and C5 (central North Sea coast), the numerical model displays substantial baseline errors ($\mathrm{RMSE} = 22.9$ cm and 23.3 cm, respectively) alongside persistent negative mean biases during high water events (-9.5 cm in C0 and -5.6 cm in C5). Conversely, across clusters influenced by open shelf boundaries or straits such as C7 and C8, the baseline hindcast exhibits moderate variance ($\mathrm{RMSE} \approx$

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Table 1. Out-of-sample performance metrics across the 10 hydrodynamic clusters defined via hierarchical clustering ($d = 1.0$). Metrics benchmark the baseline numerical hindcast against the ReSEML stacking ensemble on the withheld 20% validation partition ($\mathcal{D}_{\text{val}}$), reporting the Root Mean Square Error (RMSE) and Mean Bias Error (MBE = Model − Obs) during high water events in cm. Absolute surge magnitudes vary substantially across regimes, such as Esbjerg compared to Hornbæk (398 cm vs. 177 cm at the 100-year return level in the DVR90 datum (Poulsen et al., 2024)).

Cluster	Stations in cluster	Baseline hindcast RMSE (cm)	Baseline hindcast MBE (cm)	ReSEML ensemble RMSE (cm)	ReSEML ensemble MBE (cm)	Net gain ΔRMSE (cm)
C0	Ballum, Brøns, Vidå	22.9	−9.5	6.0	−0.6	16.8
C1	Aabenraa, Assens, Ballen, Juelsminde, Vejle	11.1	+2.3	8.1	−3.4	3.0
C2	Esbjerg, Havneby, Mandø, Ribe	17.5	−1.0	8.7	−2.0	8.9
C3	Gedser, Holbæk, Hornbæk	13.6	+4.3	5.3	−1.4	8.4
C4	Odense Fjord, Tejn	13.4	+4.8	8.7	−4.9	4.7
C5	Ferring, Hvide Sande, Thorsminde, Thyborøn	23.3	−5.6	9.0	−1.5	14.3
C6	Bagenkop, Faaborg, Fynshav, Haderslev, Sønderborg	13.2	−4.8	8.5	−2.8	4.7
C7	Hirtshals, Kalundborg, Karrebæksminde, Sjællands Odde, Rønne	15.5	+5.2	8.9	−2.9	6.6
C8	Frederikshavn, Hanstholm, Kalvehave, København, Roskilde, Skagen	15.6	+9.2	7.5	−1.5	8.1
C9	Aarhus, Bogense, Drogden Fyr, Fredericia, Grenaa, Hesnæs, Kerteminde, Køge, Korsør, Randers, Rødby, Rødvig, Slipshavn	10.5	+1.6	9.5	−4.8	1.0

15.5 cm to 15.6 cm) coupled with positive mean biases reaching up to +9.2 cm in C8. Across sheltered and complex inner passages (e.g., C2 and C6), baseline underestimations remain notable (−1.0 cm and −4.8 cm, respectively). The ReSEML framework adapts effectively to these distinct error regimes. By deploying cluster-specialist stacking ensembles, the net error reduction (ΔRMSE) peaks in the most energetic coastal environments (16.8 cm improvement in C0 and 14.3 cm in C5), while consistently suppressing root mean square errors across all regions. Across all 10 clusters, the ReSEML ensemble confines the out-of-sample validation RMSE to below 9.5 cm everywhere (reaching 5.3 cm in C3 and 6.0 cm in C0) and significantly dampens residual mean biases, demonstrating robust generalisation across diverse hydrodynamic settings nationwide.

To further evaluate the performance and operational fidelity of the ReSEML framework, four representative storm surge events were examined across contrasting coastal regimes (Figure 7). These event-scale evaluations demonstrate how regional cluster-specialist stacking ensembles dynamically adapt to diverse physical hydrodynamic processes nationwide. In the macro-tidal Danish Wadden Sea regime during December 1999 at Ribe (Figure 7a), water level dynamics are governed by large semi-diurnal tidal ranges interacting with wind setup. Following QC filtering, the observed record provides a clean continuous profile tracking successive high-water peaks. ReSEML reproduces the tidal phase and envelope throughout the multi-tidal storm

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Open Access Earth System Science Data Discussions

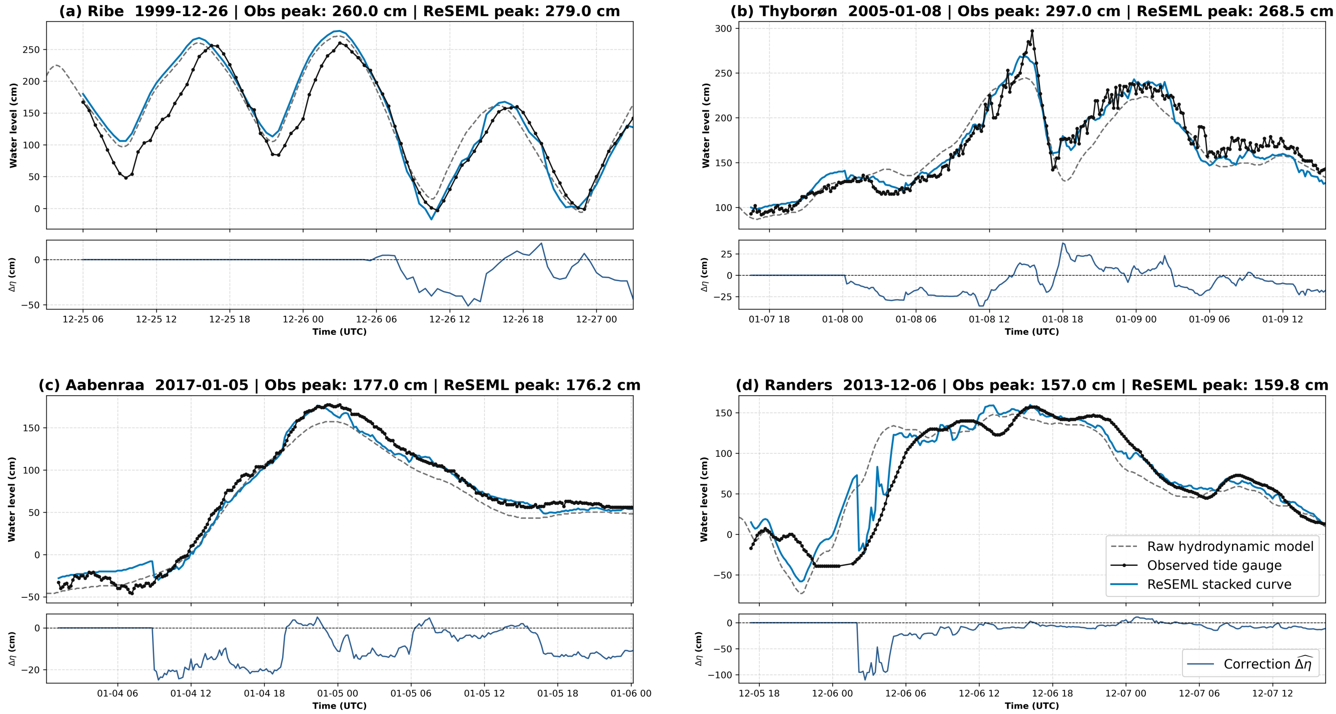

Figure 7. Representative event evaluations across distinct regional coastal regimes during observed storm periods. Comparison between quality-controlled tide gauge observations (solid black with markers), raw hydrodynamic model baseline simulations (grey dashed), and ReSEML cluster-specialist stacked reconstructions (blue with markers), alongside applied residual corrections ($\widehat{\Delta\eta}$). (a) Macro-tidal Danish Wadden Sea regime during December 1999 at Ribe (Cluster 2; Obs peak: 260 cm, ReSEML peak: 279 cm). (b) Exposed North Sea shelf surge during Storm *Erwin/Gudrun* at Thyborøn (Cluster 5; Obs peak: 297 cm, ReSEML peak: 269 cm). (c) Western Baltic/Belts wind setup event during January 2017 at Aabenraa (Cluster 1; Obs peak: 177 cm, ReSEML peak: 176 cm). (d) Kattegat fjord setup during Storm *Bodil/Xaver* at Randers (Cluster 9; Obs peak: 157 cm, ReSEML peak: 160 cm). Individual storm event time series and diagnostic plots for all 50 tide gauge stations are openly accessible in the Zenodo archive.

cycle, capturing the main surge peak (279 cm ReSEML vs. 260 cm observed) while applying dynamic residual corrections during low-water phases ($\widehat{\Delta\eta}$ reaching down to -40 cm) where the uncorrected hydrodynamic baseline overpredicts trough water levels. Along the exposed North Sea coast during the January 2005 Storm *Erwin/Gudrun* at Thyborøn (Figure 7b), the raw hydrodynamic baseline exhibited a pronounced peak underestimation of roughly 50 cm (245 cm baseline vs. 297 cm observed). ReSEML's dual-stream feature representation captures the rapid storm onset, elevating the reconstructed peak to 269 cm and applying dynamic positive residual adjustments ($\widehat{\Delta\eta}$ reaching $+38$ cm during the secondary surge wave) to track high-frequency observational variations closely. In the micro-tidal Danish Straits and Belts during the January 2017 storm surge event at Aabenraa (Figure 7c), the raw hydrodynamic simulation consistently underestimated water levels across the event peak (157 cm baseline vs. 177 cm observed). The ReSEML ensemble applies a steady positive residual correction across

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

the peak window, elevating the reconstructed peak to 176 cm, within 1 cm of the observed maximum, while preserving the storm's temporal duration and recession rate. Finally, in the complex Kattegat estuarine environment during the December 2013 Storm *Bodil/Xaver* at Randers (Figure 7d), the raw hydrodynamic model exhibited severe timing lag and shape distortion, overshooting during the pre-surge phase while underestimating the secondary surge peak. ReSEML dynamically suppresses the false initial surge rise ($\widehat{\Delta\eta}$ approaching -100 cm) and correctly reconstructs the true peak magnitude (160 cm ReSEML vs. 157 cm observed), producing an accurate representation of estuarine storm setup. Together, these multi-regime event reconstructions confirm that the regional cluster-specialist architecture prevents localised over-smoothing. By tailoring the stacking meta-learners to distinct physical regimes, ReSEML consistently outperforms raw hydrodynamic models across macro-tidal, exposed shelf, and fetch-restricted environments.

3.4 Historical sequence reconstruction and event consistency

Using the trained cluster-specialist networks, continuous high-frequency water level records were reconstructed back to 1961. To verify the historical reliability of the framework during decades completely lacking continuous digital observations, an event consistency analysis was conducted across six representative stations spanning the three macro-oceanographic regimes of Denmark: Ribe and Havneby (Danish Wadden Sea), Randers and Grenaa (Inner Danish Straits and Kattegat), and Køge and Kerteminde (Belt and Sound Straits). In the temporal data availability overview (Figure 2), these historical validation periods correspond to the orange bars (KDI peak lists), which precede the blue bars representing the modern sub-hourly observation records used for training (1961 to the start date of continuous digital monitoring).

Following standard KDI EVA conventions, a dynamic peak-over-threshold (POT) criterion targeting the 4-month return level ($\lambda = 3.0$ events year^{-1}) was adopted, ensuring that all moderate and severe surges were retained. Candidate storm peaks in the continuous reconstructed series were matched against the manual KDI events using a temporal co-occurrence tolerance window of ± 3 days (72 h). This multi-day window explicitly accounts for the daily timestamp resolution of older digitised archives as well as hydrodynamically driven phase lags across narrow Danish straits and fjords. Figure 8 illustrates the resulting event consistency evaluation across the six benchmark stations.

The event consistency analysis also reveals distinct types of mismatched classifications on both sides of the identity axis, which highlight the unique utility of the continuous reconstruction:

– **Missed extreme classifications ($y = 0$):** Points aligned along the horizontal axis represent historical events catalogued in the manual KDI lists where the continuous reconstructed series did not exceed the local storm surge threshold. While minor sub-grid wave setup or harbour resonance may play a role, diagnostic investigation reveals that discrepancies frequently stem from inherent uncertainties, transcription errors, or subjective local expert notations within the legacy manual catalogues themselves. For instance, at Køge, the manual KDI catalogue lists multiple distinct peak entries during January and February 2003; however, synoptic weather records and continuous regional hydrodynamics confirm that no severe storm surge meteorological conditions occurred over the western Baltic during these periods. In this context, DEN-SURGE serves not merely as a validation target but as a continuous, physically grounded reference dataset

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

to systematically audit, verify, and clean historical manual storm surge archives across the Danish coastal monitoring network.

– **Continuous catalogue discovery and archive consolidation** ($x = 0$)**:** Conversely, points aligned along the vertical axis represent reconstructed extreme surge events that are entirely absent from the historical manual KDI lists. Because older KDI peak catalogues were compiled retrospectively on an event-by-event basis, brief or regional storm events—particularly those occurring outside major municipal monitoring centres—were occasionally omitted. The continuous DEN-SURGE framework successfully recovers these uncatalogued historical extremes, bridging long-standing observation gaps. Crucially, combining the verified events from the manual archives ($y = 0$) with the newly discovered continuous extremes ($x = 0$) allows researchers to synthesise a comprehensive, fully reconciled storm surge catalogue. Complete station-by-station diagnostic matching tables and separate lists for both categories of discrepancy are archived in the Zenodo repository for open verification.

To demonstrate the framework's capability in resolving high-frequency storm hydrographs during these unobserved decades, Figure 9 highlights the highest reconstructed historical storm surge events across four representative stations spanning the 1961–1990 period. In the macro-tidal Danish Wadden Sea (Cluster 2; Figure 9a), the January 1990 storm surge at Ribe reaches a reconstructed peak of $447\,\text{cm}$, where ReSEML resolves the sharp, asymmetric tidal-surge interaction while dynamically adjusting trough levels ($\widehat{\Delta\eta}$ reaching $-52\,\text{cm}$). Along the exposed North Sea shelf coast at Thyborøn (Cluster 5; Figure 9b), the February 1990 event demonstrates the model's capacity to resolve high-frequency oscillatory dynamics, reconstructing a maximum peak of $246\,\text{cm}$ and sustaining water levels above the 4-month threshold ($153\,\text{cm}$) for over 30 hours. In the micro-tidal Western Baltic/Belts regime during the January 1976 Capella storm at Aabenraa (Cluster 1; Figure 9c), ReSEML captures the steep pre-storm drawdown below $-110\,\text{cm}$ followed by a rapid surge setup to $162\,\text{cm}$, applying dynamic phase corrections ($\widehat{\Delta\eta}$ up to $+12\,\text{cm}$) during the sharp onset. Finally, in the complex Kattegat estuarine environment (Cluster 9; Figure 9d), the February 1989 surge at Randers showcases robust setup tracking; ReSEML suppresses the numerical baseline's initial overestimation ($\widehat{\Delta\eta}$ reaching down to $-40\,\text{cm}$) while precisely capturing the primary peak at $150\,\text{cm}$. These reconstructed event profiles confirm that the cluster-specialist stacking framework robustly resolves non-linear residual dynamics and produces physically coherent surge hydrographs across distinct coastal regimes prior to the digital monitoring era.

3.5 Extreme value analysis: impact of record extension

To quantify the direct engineering impact of extending these data-scarce coastal records, comparative EVA were conducted using the POT framework fitted to the Generalised Pareto Distribution (GPD) (Figure 10). We focus specifically on three representative benchmark stations spanning contrasting hydrodynamic regimes: Brøns (Danish Wadden Sea; Cluster 0), Vejle (semi-enclosed fjord; Cluster 1), and Aabenraa (western Baltic Sea; Cluster 1). These three stations are uniquely suited for demonstrating the engineering value of the DEN-SURGE reconstruction because they possess exceptionally short modern digital observation spans and lack continuous historical digital records; evaluating their return levels reveals the severe vulnerabilities introduced by relying exclusively on truncated observational windows.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

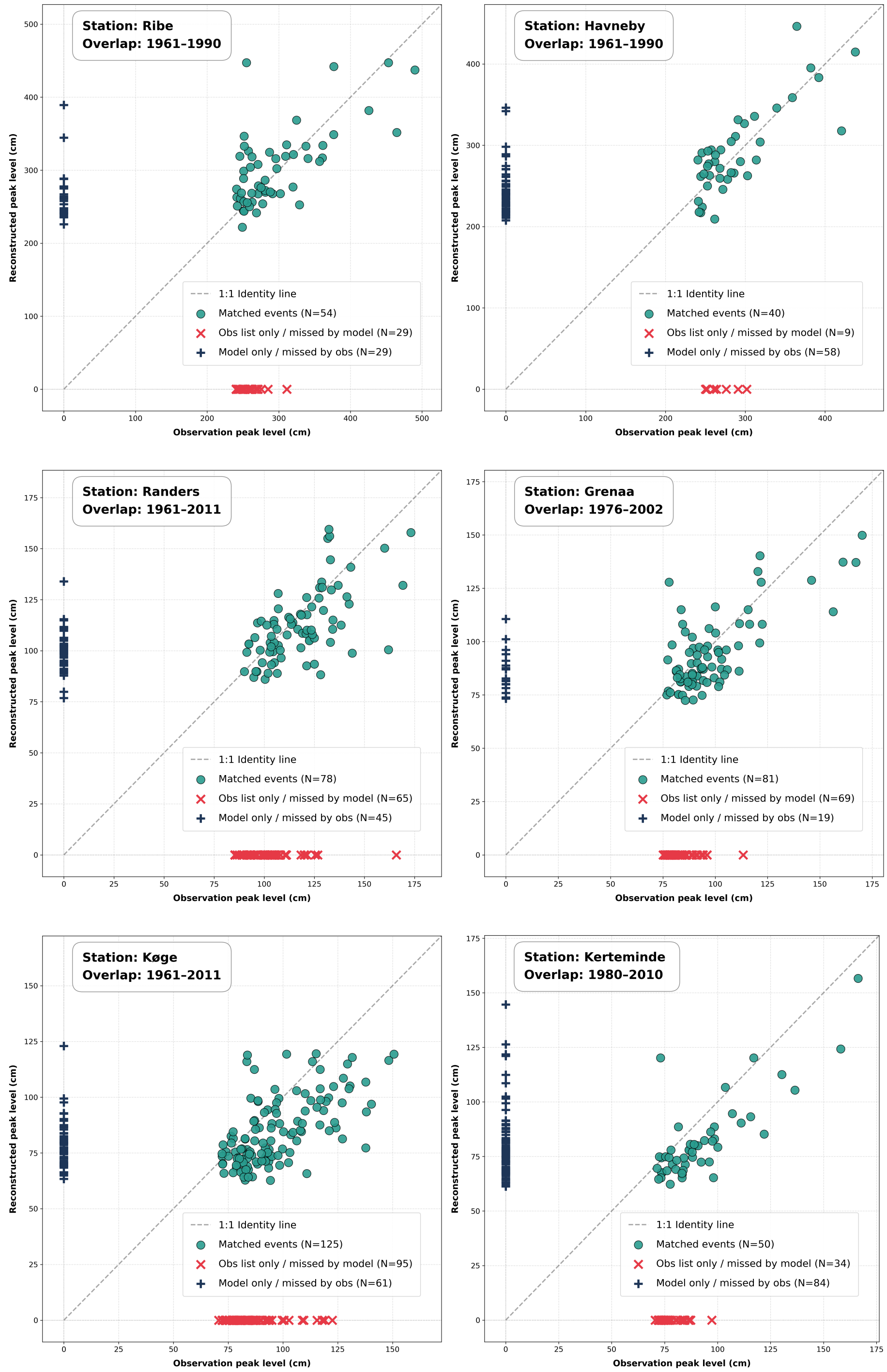

Figure 8. Historical event consistency scatter plots across six representative data-scarce stations, grouped by different oceanographic regime. Reconstructed peak surge levels (cm) are plotted against independent historical KDI observation lists. Blue-green dots represent matched extreme events, red crosses represent events unique to the historical list ($y = 0$), and dark-blue crosses show extreme events captured exclusively by the continuous model reconstruction ($x = 0$). Panels are ordered by regime: Wadden Sea (a, b), Kattegat (c, d), and Sound and Belt (e, f).

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

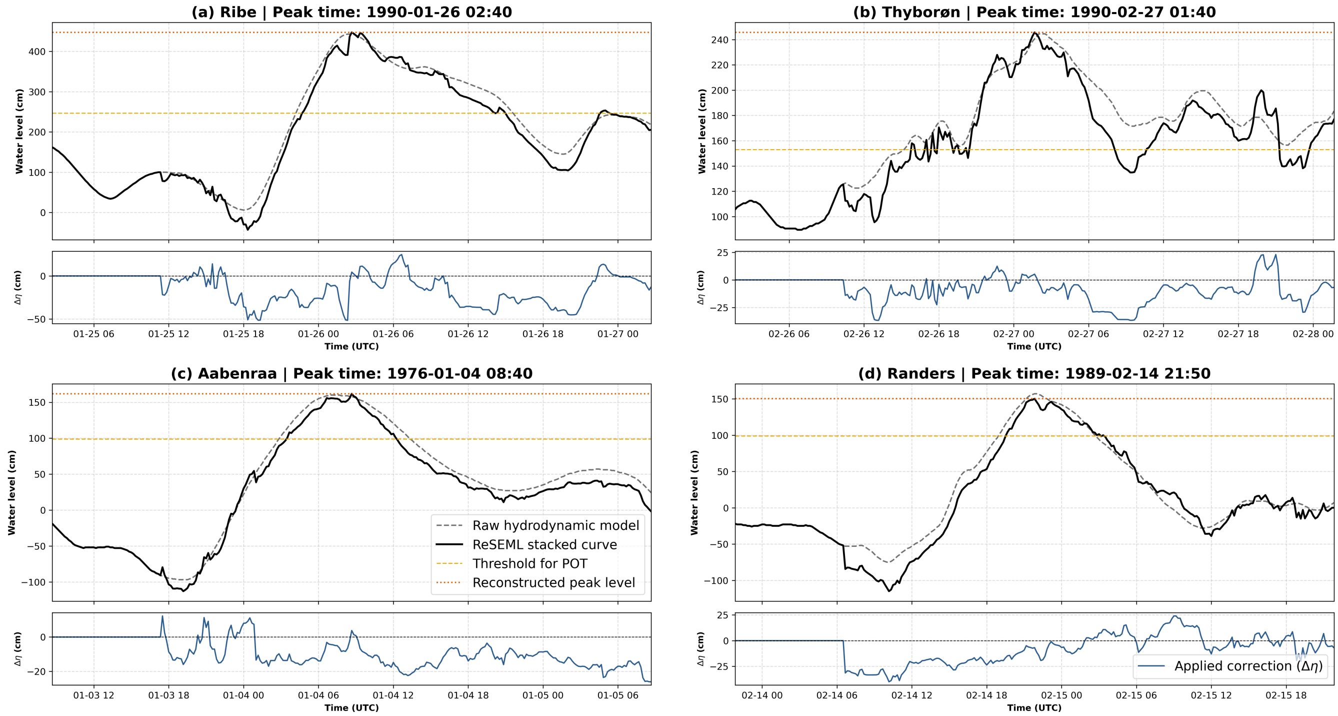

Figure 9. Unobserved storm surge events: Reconstructed continuous water level time series for the maximum historical unobserved storm surge events across four representative stations. The panels contrast the raw offset-corrected hydrodynamic model (gray dashed line) against the ReSEML stacked reconstruction (solid black line), alongside the station-specific 1-year return level threshold (yellow dashed line), the reconstructed maximum peak water level (red dotted line), and the dynamic applied residual correction ($\widehat{\Delta\eta}$, bottom panels). (a) Macro-tidal Wadden Sea surge during the January 1990 storm at Ribe (Cluster 2; Reconstructed peak: 447 cm). (b) Exposed North Sea shelf surge during the February 1990 storm at Thyborøn (Cluster 5; Reconstructed peak: 246 cm). (c) Western Baltic wind setup during the Storm Capella in January 1976 at Aabenraa (Cluster 1; Reconstructed peak: 162 cm). (d) Kattegat estuarine setup during the February 1989 storm at Randers (Cluster 9; Reconstructed peak: 150 cm). Equivalent diagnostic plots and continuous reconstruction figures for all 50 coastal stations are available in the Zenodo repository.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

The empirical EVA outcomes differ substantially across the three coastal settings. In the macro-tidal North Sea regime at Brøns, the short modern observation series (2001–2024, 24 years) yields a 100-year return level estimate of 558 cm (95% confidence interval width: 200 cm). Extending the sequence back to 1961 incorporates historical decadal storm clusters across the German Bight, adjusting the 100-year return level to 569.9 cm while compressing the 95% confidence interval width to 156 cm (a 22.2% reduction in parameter uncertainty). In the semi-enclosed fjord environment at Vejle, relying solely on the recent 19-year digital monitoring record (2006–2024) leads to an inflated 100-year return level of 275.5 cm with a wide confidence envelope (129 cm). Because the post-2006 window was dominated by recent extreme storm setups (such as the October 2023 surge), fitting a Pareto tail exclusively to this period over-projects extreme return levels. Incorporating the 43 years of reconstructed hindcast history (1961–2005) provides essential multi-decadal context, lowering the estimated 100-year design level to 197 cm and reducing parameter uncertainty by 59.3% (confidence interval width narrowed from 129 cm to 53 cm). In the western Baltic Sea regime at Aabenraa, the 28-year modern observational record (1997–2024) is similarly biased by recent storm surges, yielding an unconstrained 100-year return level of 259 cm with an uncertainty span of 102 cm. Extending the timeline to the complete 64-year baseline (1961–2024) acts as an essential statistical stabiliser, adjusting the 100-year return level to 232 cm and narrowing the 95% confidence interval width to 57 cm (a 44.7% uncertainty reduction).

Across all hydrodynamic regimes, the primary engineering benefit of the 64-year reconstructed record is a systematic, substantial reduction in extreme value parameter uncertainty. As evidenced by the shaded 95% confidence envelopes in Figure 10, extending coastal time series from truncated modern windows to the continuous 1961–2024 baseline reduces 100-year return level confidence interval widths by up to 78.6% (with an average reduction of 46.2% across all stations exhibiting constrained fits), providing an empirical foundation for coastal climate adaptation and flood defence infrastructure.

4 Discussion

The results presented in Section 3 demonstrate that the ReSEML framework is highly effective in reconstructing continuous, high-resolution coastal water levels across multi-decadal observational data gaps. However, synthesising physical hydrodynamic modelling, meteorological forcing constraints, and deep learning architectures introduces critical nuances that warrant detailed examination. Below, we discuss the methodological sensitivity and subjectivity inherent in regional hydrodynamic clustering (Section 4.1), evaluate historical observation quality limitations alongside the emerging role of ReSEML as a generative diagnostic and gap-filling engine (Section 4.2), examine the engineering implications of the extended 64-year records for extreme value statistics and coastal defence design (Section 4.3), and outline future directions for adapting this hybrid methodology to climate projection downscaling (Section 4.4).

4.1 Sensitivity and subjectivity in hydrodynamic clustering configurations

A core design principle of the ReSEML framework is the partitioning of complex regional coastlines into hydrodynamically coherent regimes, enabling cluster-specialist networks to learn localised meteorological residual dynamics rather than fitting an over-generalised global model. However, existing storm surge clustering literature remains limited and has largely been

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

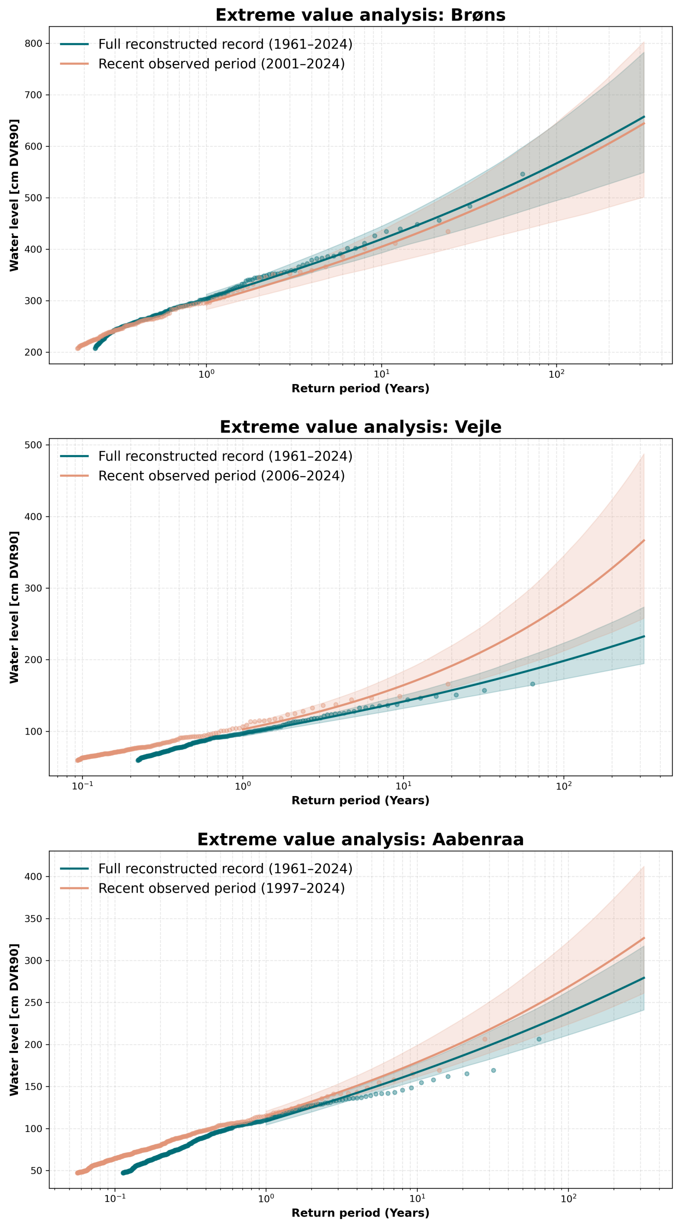

Figure 10. Extreme value analysis return level curves comparing recent observed and fully reconstructed records for the three stations. The orange curves denote the statistical fit derived exclusively from the short, modern observation windows. The blue curves show the updated fits utilising the expanded DEN-SURGE v1.0 continuous records (1961–2024). Shaded envelopes signify the corresponding 95% statistical confidence intervals. Corresponding EVA return-level plots for all remaining stations are available in the Zenodo repository (see Data Availability).

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

developed for distinct diagnostic purposes, leaving considerable flexibility in defining the regional clustering configuration (Haigh et al., 2016; Enríquez et al., 2020). While agglomerative hierarchical clustering provides an automated and statistically grounded mechanism to group tide gauge stations based on empirical surge fingerprints (such as surge variability, upper quantiles, and baseline means), the resultant spatial grouping remains inherently sensitive to several subjective configuration choices.

First, the spatial topology of the clustering dendrogram is directly conditioned by network membership and station inclusion. Because feature normalisation and pairwise Ward linkage distances are computed across the collective monitoring network, the inclusion or exclusion of individual stations located within complex hydrodynamic transition zones (e.g., the Danish Straits) can alter the relative distances between neighbouring nodes. Consequently, transitional stations may experience subtle regrouping depending on the composition of the analysed network. This sensitivity arises because coastal oceanographic processes exist along a continuous hydrodynamic spectrum rather than within discrete, isolated boundaries, making cluster assignments near regime margins inherently dependent on network geometry.

Second, the structural definition of each specialist domain is determined by methodological hyperparameters, specifically the fingerprint metrics, feature scaling, linking criteria, and the dendrogram cut threshold ($d = 1.0$). Specifically, modifying the distance threshold affects the trade-off between domain specificity and data volume. A lower distance threshold isolates unique local geometries (such as semi-enclosed shallow fjords), reducing physical variation within each group. However, it fragments the training archive, limiting the number of extreme surge sequences available for training deep LSTM and TCN layers. Conversely, a higher distance threshold maximises cross-station sequence pooling and statistical sample size during neural network optimisation, but risks blending different hydrodynamic regimes (for example, combining micro-tidal Baltic basins with tide-dominated Kattegat channels), reducing the predictive precision of specialist networks.

Ultimately, these clustering permutations propagate into the downstream reconstructions, dictating the transferability of learned residual corrections across unobserved historical periods. While regional specialist ensembles substantially outperform single un-clustered baselines, acknowledging the sensitivity to station selection and linkage distance is essential when extending the framework to other semi-enclosed seas or applying transfer learning to ungauged coastal sectors.

4.2 Observational data quality limitations and ReSEML as a generative gap-filling and diagnostic tool

A critical challenge in regional storm surge hazard analysis is the imperfection of historical tide gauge records. Although automated QC procedures, such as spike filtering, rate-of-change checks, and local residual outlier rejection, were incorporated into the training pipeline, significant quality deficiencies remain widespread across multi-decadal gauge archives (Bauer et al., 2026). Traditional QC methods in oceanography operate predominantly as binary screening mechanisms: suspect observations are identified and assigned QC flags, resulting in the exclusion or truncation of corrupted records. While flagging prevents contaminated data from distorting extreme value statistics, it cannot recover the lost physical information. Furthermore, establishing our QC thresholds (e.g., hard residual ceiling and rolling IQR multiplier) inevitably involves subjective configuration choices tailored to remove non-physical outliers from training tensors; when transferring the framework to other coastal domains, these criteria must be recalibrated to avoid being overly restrictive or overly permissive.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

The ReSEML framework developed here offers a paradigm shift: rather than serving solely as a post-processing tool for hydrodynamic models, ReSEML functions as a physics-informed, generative reconstruction engine that can both diagnose severe observational errors and seamlessly reconstruct missing extreme events (Latapy et al., 2023). This dual capability is particularly evident when diagnosing and repairing corrupted gauge sequences. Certain stations exhibit persistent observational anomalies and substantial climatological mean biases during high water events, exemplified by Kalundborg (Cluster 7, Figure 4). During the severe surge event of 2 March 2008 (Figure 11a), the tide gauge experienced a catastrophic sensor dropout, logging an unphysical, continuous flatline of -527 cm during the entire storm. While standard QC flags such an event as invalid, it leaves coastal managers without surge information during a hazardous storm. In contrast, our network leverages the 10-minute raw hydrodynamic simulation (which peaked at 145 cm) and applies the learned regional meteorological and phase corrections to reconstruct a realistic, continuous storm surge peaking at 141 cm.

Beyond repairing corrupted sensor output, the framework provides high-quality gap filling during unmonitored historical intervals. A recurrent issue in coastal monitoring networks is the presence of multi-year observational gap caused by harbour maintenance, gauge decommissioning, or sensor transitions. For instance, the Odense Fjord gauge (Cluster 4) contains a well-documented multi-year observational data gap between 2013 and 2018 (Poulsen et al., 2024). During this monitoring hiatus, Cyclone Bodil struck northern Europe on 5–6 December 2013, generating one of the most severe storm surges recorded in the inner Danish waters. As shown in Figure 11b, ReSEML synthesises the hydrodynamic hindcast with learned cluster-specific shallow-water dynamics to reconstruct the complete surge trajectory, accurately capturing the rapid set-up and resolving a peak elevation of 161 cm.

By providing physically coherent, continuous time-series reconstructions rather than simple statistical interpolation, ReSEML bridges the gap between observational quality control and hydrodynamic modelling. This capability makes the framework well-suited for digitising lost extreme records, standardising inhomogeneous multi-decadal archives, and establishing robust continuous baselines for coastal climate adaptation.

4.3 Implications for infrastructure design, extreme value statistics, and downstream AI applications

The comparative EVA highlights the significant danger of relying on short, modern digital monitoring horizons for coastal defence engineering. Short observational records are highly sensitive to multi-decadal climate variability and temporal storm clustering, which can introduce substantial structural miscalculations into long-term infrastructure planning (Zeder et al., 2023).

The regional contrasts in Section 3.5 demonstrate that truncated modern records can distort statistical hazard curves in opposite directions depending on the local hydrodynamic regime. Along the volatile North Sea shelf (e.g., Brøns), a recent decadal cluster of severe storms artificially steepens the Pareto tail when fitted exclusively to modern data, causing an overestimation of long-term return levels. Conversely, in sheltered interior fjord environments (e.g., Vejle), the modern 20-year window entirely omits the severe historical storm clusters of the late 1960s and 1980s, resulting in a dangerous underestimation of long-term flood hazards.

The DEN-SURGE catalogue resolves these empirical distortions by embedding the 50 coastal stations within a 64-year high-resolution physical baseline (1961–2024). Beyond mitigating directional bias, the primary engineering benefit of this

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

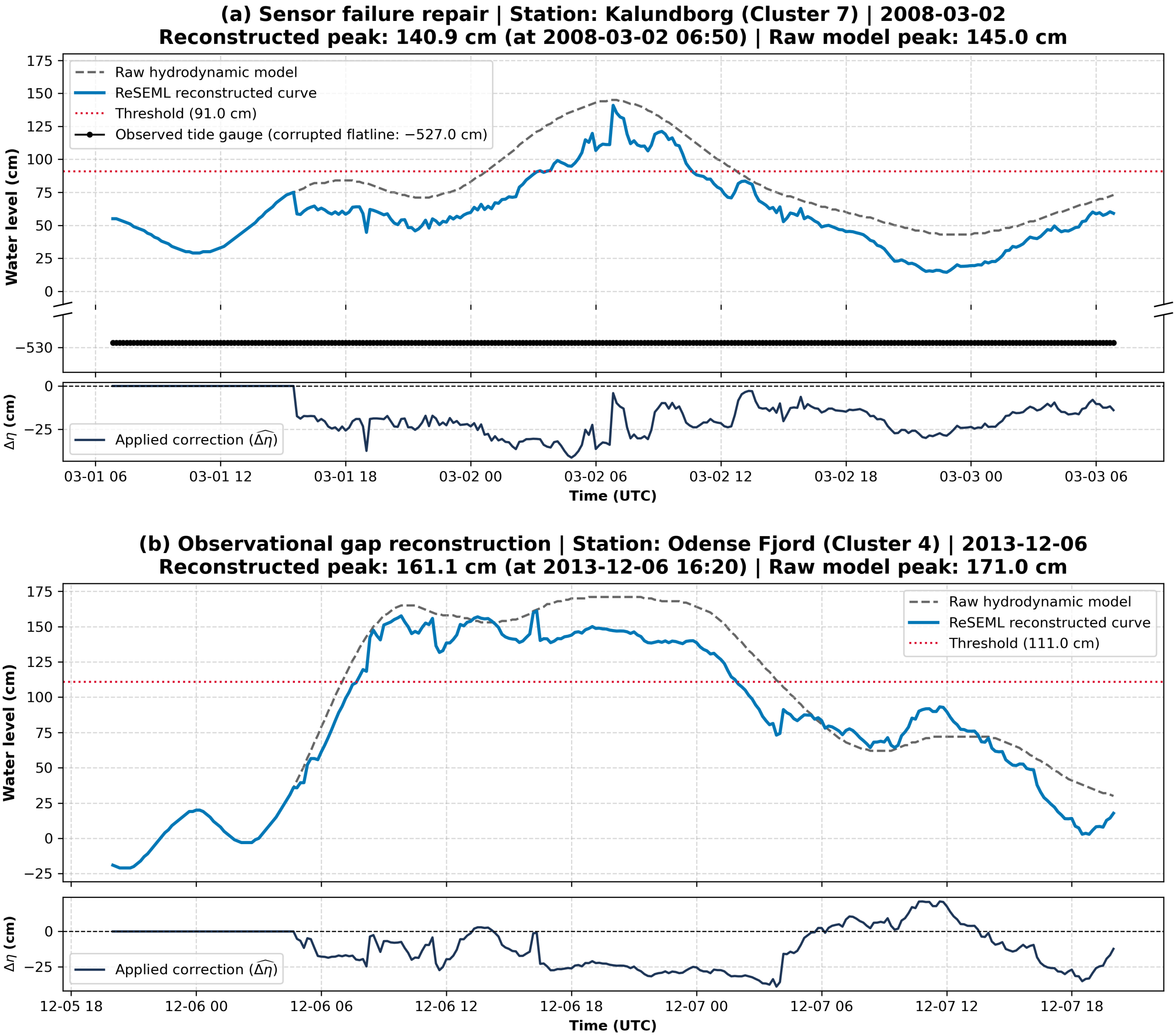

Figure 11. (a) Repair of a severe gauge sensor malfunction at Kalundborg (Cluster 7) during the storm surge of 2 March 2008, where the gauge recorded a persistent unphysical flatline at -527 cm (black points). ReSEML bypasses the corrupted sensor data by applying regional stacked residual corrections ($\widehat{\Delta\eta}$, lower panel) to the raw hydrodynamic simulation (dashed line), synthesising a physically consistent storm surge (solid blue curve) peaking at 141 cm. (b) High-frequency reconstruction of Cyclone Bodil (6 December 2013) across an unmonitored monitoring gap (2013–2018) at Odense Fjord (Cluster 4), accurately recovering the complete surge curve and resolving an extreme peak of 161 cm despite the absence of observational gauge data.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

continuous record is the substantial, systematic reduction in statistical parameter uncertainty. Narrowing 95% confidence intervals at the 100-year return horizon by up to 78.6% (and by an average of 46.2% across constrained stations) removes the necessity for excessively conservative, cost-prohibitive safety multipliers in marine civil engineering. This provides a cost-effective, physically grounded foundation for municipal flood risk management, sea-level rise adaptation schemes, and hard coastal defence design nationwide.

Beyond conventional extreme value statistics, the refined DEN-SURGE catalogue serves as an open-access, AI-ready benchmark tailored for next-generation data-driven coastal intelligence (Mik-Meyer et al., 2026). Because the dataset provides continuous, 10-minute sea-level sequences that have been cleansed of instrument dropouts, harmonised across vertical datums (DVR90), and spatially indexed into coherent hydrodynamic clusters, it removes the extensive data pre-processing bottlenecks that traditionally hinder deep learning applications in oceanography. Operationally, this 64-year archive can serve as ground truth for training real-time surge forecasting emulators, enabling millisecond-latency peak inference during severe North Sea extratropical cyclones. Furthermore, the dataset establishes a standardised foundation for deriving AI-driven early-warning indicators, compound hazard indices, and localised flood risk metrics, accelerating the deployment of ML into operational warning centres and municipal coastal protection workflows (Rabier et al., 2026).

4.4 The climate application pivot: future directions for climate model downscaling

While the current deployment of ReSEML v1.0 is dedicated to the historical reconstruction and homogenisation of the 1961–2024 surge catalogue, the underlying mathematical framework was specifically architected to facilitate a direct pivot toward future climate change impact assessments (Somot et al., 2025). Projecting extreme sea-level frequencies under transient greenhouse gas emission pathways remains a critical bottleneck in coastal climate adaptation. Standard global and regional climate models (GCMs/RCMs) and regional shelf hydrodynamics cannot resolve fine-scale coastal bathymetry, localised wind-stress setup inside restrictive fjord branches, or shallow-water shoaling along complex straits.

The residual formulation of ReSEML v1.0 provides a computationally efficient mechanism to bridge this resolution gap. Because the deep learning stacking ensemble predicts the localised hydrodynamic residual ($\Delta\eta$) conditioned purely on numerical model states, temporal rates of change ($\mathrm{trend}_{1\mathrm{h}}$, $\mathrm{trend}_{3\mathrm{h}}$, $\mathrm{trend}_{6\mathrm{h}}$), and cyclical tidal phases, it operates independently of real-time observational inputs at runtime. Consequently, pre-trained cluster-specialist networks can be deployed directly as an emulator layer over century-scale hydrodynamic climate projections forced by GCM meteorology.

A fundamental methodological consideration in this downscaling pivot is the stationarity of the learned residual function under a warmer climate state. Transferring historical residual corrections to future projections rests on the premise that numerical model discrepancies are primarily driven by unresolved, static coastal features, specifically sub-grid bathymetric friction, complex coastline geometries, and narrow channel constrictions, which remain invariant over multi-decadal planning horizons (Krinner et al., 2020). Furthermore, because the ensemble learns state-dependent, non-linear physical interactions rather than static additive offsets, the network inherently scales its adjustments when forced with higher or more rapidly shifting storm surge profiles.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Nonetheless, severe mean sea-level rise (SLR) and localised morphodynamic evolutions may induce non-stationary changes in shallow-water resonance and non-linear tide-surge interactions, particularly across macro-tidal environments such as the Danish Wadden Sea. In near-term future research, we aim to couple ReSEML with dynamic SLR scenario baselines and non-stationary EVA. By projecting deep-learning-corrected residuals onto 21st-century physical simulations, this framework will enable high-resolution projections of shifting return periods, compound flood hazards, and localised adaptation thresholds across the Danish coastal zone.

5 Conclusion

This paper has presented the development, validation, and empirical application of the ReSEML v1.0 framework, producing the unified, continuous 10-minute DEN-SURGE storm surge catalogue for 50 Danish coastal stations spanning 1961–2024. By training a stacked deep learning ensemble (LSTM + TCN) to predict local physical residuals relative to a numerical hydrodynamic baseline, the framework effectively overcomes the challenge of severe data scarcity and historical fragmentation along complex coastlines. By repairing corrupted sensor records, bridging multi-year observational hiatuses, and resolving sub-grid hydrodynamic setups, the resulting dataset provides an AI-ready, gap-free foundation of higher physical fidelity than either raw numerical simulations or noisy, fragmented gauge records.

The division of the coastline into 10 regional physical clusters successfully isolated distinct storm surge dynamics, ranging from macro-tidal wind setup along the Wadden Sea shelf to slow basin-scale pre-filling within the Baltic Sea. Nationwide out-of-sample validation confirmed that the framework minimised systematic spatial mean bias (MBE) during high water events to within ± 0.4 cm and compressed errors below 10 cm RMSE. Cross-referencing against independent historical lists spanning 1961 to the onset of continuous digital monitoring verified that ReSEML v1.0 reproduces historical storm peaks with high fidelity while successfully recovering regional extremes omitted from manual records. Furthermore, expanding the continuous record to 64 years stabilised long-term EVA return curves and halved statistical parameter uncertainty, narrowing the average 95% confidence interval width at the 100-year return horizon from 101.0 cm to 54.5 cm (by up to 224.3 cm at Roskilde), representing a mean (median) uncertainty reduction of 46.2% (50.7%) across stations with constrained fits (and an overall dataset-wide mean reduction of 35.1%, reaching up to 78.6%).

Crucially, the success of this historical reconstruction establishes a validated, mathematically bounded baseline for our next research phase on climate adaptation. Because the residual learning architecture operates independently of real-time gauge inputs during inference, it is uniquely suited for integration with climate forcing models. Our upcoming work will pivot this framework toward future climate simulations, utilising the trained cluster specialist networks to downscale regional climate projections through the end of the 21st century. DEN-SURGE and ReSEML v1.0 thus provide a critical methodological bridge, shifting coastal engineering from an era of fragmented historical observations to a continuous, high-resolution framework capable of quantifying evolving extreme sea-level hazards under a changing climate.

Code availability

The open-source ReSEML v1.0 software framework (PyTorch/Scikit-Learn training and evaluation code) is available on Zenodo at https://doi.org/10.5281/zenodo.21719179 (Su, 2026).

Data availability

The DEN-SURGE v1.0 dataset, including reconstructed high-frequency water level time series, extracted event catalogues, and supplementary evaluation plots for all 50 Danish tide gauge stations (1961–2024), is openly archived on Zenodo at https://doi.org/10.5281/zenodo.21719179 (Su, 2026).

Author contribution

JS conceptualised the ML framework, developed the methodology and software, performed the analysis, made visualizations and wrote the initial draft. JWN conceptualised the initial idea, handled data acquisition and data analysis. KSM data quality, data analysis and interpretation. MADL acquired funding and coordinated. All authors revised and edited the manuscript. All authors have read and approved the manuscript.

Competing interests

The contact author has declared that none of the authors have any competing interests.

Acknowledgements

The work is partly funded by the ECCO project, which received funding from Innovation Fund Denmark and the European Union's Horizon Europe Programme under the 2022 Joint Transnational Call of the European Partnership Water4All (Grant agreement ID: 101060874). This work is also supported by Danish Klimaatlas and the Danish National Centre for Climate Research (NCKF) projects, funded by the Danish State. The authors express their sincere gratitude to the dedicated technical staff at the Danish Meteorological Institute (DMI), the Danish Coastal Authority (Kystdirektoratet, KDI), and municipal port authorities whose sustained, multi-decadal efforts in maintaining coastal tide gauge infrastructure made this study possible. We also gratefully acknowledge the DMI Open Data initiative for providing open-access distribution of the tide-gauge observations.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

References

Andrée, E., Su, J., Dahl Larsen, M. A., Drews, M., Stendel, M., and Skovgaard Madsen, K.: The role of preconditioning for extreme storm surges in the western Baltic Sea, Natural Hazards and Earth System Sciences, 23, 1817–1834, https://doi.org/10.5194/nhess-23-1817-2023, 2023.

Andrée, E., Su, J., Larsen, M. A. D., Madsen, K. S., and Drews, M.: Simulating major storm surge events in a complex coastal region, Ocean Modelling, 162, 101 802, https://doi.org/10.1016/j.ocemod.2021.101802, 2021.

Andrée, E., Drews, M., Su, J., Larsen, M. A. D., Drønen, N., and Madsen, K. S.: Simulating wind-driven extreme sea levels: Sensitivity to wind speed and direction, Weather and Climate Extremes, 36, 100 422, https://doi.org/https://doi.org/10.1016/j.wace.2022.100422, 2022.

Arns, A., Wahl, T., Haigh, I., Jensen, J., and Pattiaratchi, C.: Estimating extreme water level probabilities: A comparison of the direct methods and recommendations for best practise, Coastal Engineering, 81, 51–66, https://doi.org/10.1016/j.coastaleng.2013.07.003, 2013.

Bauer, F., Su, J., Madsen, K. S., and Larsen, M. A. D.: SeaQC-X: Transferability of a machine learning-based sea level quality control framework, Applied Ocean Research, 170, 105 065, https://doi.org/10.1016/j.apor.2026.105065, 2026.

Buchanan, M. K., Oppenheimer, M., and Kopp, R. E.: Amplification of flood frequencies with local sea level rise and emerging flood regimes, Environmental Research Letters, 12, 064 009, https://doi.org/10.1088/1748-9326/aa6cb3, 2017.

Calafat, F. M. and Marcos, M.: Probabilistic reanalysis of storm surge extremes in Europe, Proceedings of the National Academy of Sciences, 117, 1877–1883, https://doi.org/10.1073/pnas.1913049117, 2020.

Capet, A., Fernández, V., She, J., Dabrowski, T., Umgiesser, G., Staneva, J., Mészáros, L., Campuzano, F., Ursella, L., Nolan, G., and El Serafy, G.: Operational Modeling Capacity in European Seas—An EuroGOOS Perspective and Recommendations for Improvement, Frontiers in Marine Science, 7, 129, https://doi.org/10.3389/fmars.2020.00129, 2020.

Chen, K., Kuang, C., Wang, L., Chen, K., Han, X., and Fan, J.: Storm Surge Prediction Based on Long Short-Term Memory Neural Network in the East China Sea, Applied Sciences, 12, https://doi.org/10.3390/app12010181, 2022.

Church, J., Clark, P., Cazenave, A., Gregory, J., Jevrejeva, S., Levermann, A., Merrifield, M., Milne, G., Nerem, R., Nunn, P., et al.: Sea level change. Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1137–1216, 2013.

Clemmensen, L. B., Glad, A. C., and Kroon, A.: Storm flood impacts along the shores of micro-tidal inland seas: A morphological and sedimentological study of the Vesterlyng beach, the Belt Sea, Denmark, Geomorphology, 253, 251–261, https://doi.org/https://doi.org/10.1016/j.geomorph.2015.10.020, 2016.

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer London, https://doi.org/https://doi.org/10.1007/978-1-4471-3675-0, 2001.

Dangendorf, S., Frederikse, T., Chafik, L., Klinck, J. M., Ezer, T., and Hamlington, B. D.: Data-driven reconstruction reveals large-scale ocean circulation control on coastal sea level, Nature Climate Change, 11, 514–520, https://doi.org/10.1038/s41558-021-01046-1, 2021.

Danish Coastal Authority: Frie data, https://kyst.dk/hav-og-anlaeg/maalinger-og-data/vandstandsmaalinger/arkiv, accessed: 2026-01-01, 2026.

Danish Meteorological Institute: Frie data, https://www.dmi.dk/frie-data, accessed: 2026-01-01, 2026.

DMI: Storm surge model (DKSS), Available at: https://opendatadocs.dmi.govcloud.dk/Data/Forecast_Data_Storm_Surge_Model_DKSS, [Accessed: 01-01-2025], 2024.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Donat, M. G., Leckebusch, G. C., Wild, S., and Ulbrich, U.: Future changes in European winter storm losses and extreme wind speeds inferred from GCM and RCM multi-model simulations, Natural Hazards and Earth System Sciences, 11, 1351–1370, https://doi.org/10.5194/nhess-11-1351-2011, 2011.

Dubois, K. A. D., Dahl Larsen, M. A., Drews, M., Nilsson, E., and Rutgersson, A.: Technical note: Extending sea level time series for extremes analysis with machine learning and neighbouring station data, EGUsphere, 2023, 1–14, https://doi.org/10.5194/egusphere-2023-1159, 2023.

Dullaart, J. C. M., Muis, S., Bloemendaal, N., and Aerts, J. C. J. H.: Advancing global storm surge modelling using the new ERA5 climate reanalysis, Climate Dynamics, 54, 1007–1021, https://doi.org/10.1007/s00382-019-05044-0, 2020.

Enríquez, A. R., Wahl, T., Marcos, M., and Haigh, I. D.: Spatial Footprints of Storm Surges Along the Global Coastlines, Journal of Geophysical Research: Oceans, 125, e2020JC016 367, https://doi.org/10.1029/2020JC016367, e2020JC016367 2020JC016367, 2020.

Ganguli, P. and Merz, B.: Extreme Coastal Water Levels Exacerbate Fluvial Flood Hazards in Northwestern Europe, Scientific Reports, 9, 13 165, https://doi.org/10.1038/s41598-019-49822-6, 2019.

Haas, R., Pinto, J. G., and Born, K.: Can dynamically downscaled windstorm footprints be improved by observations through a probabilistic approach?, Journal of Geophysical Research: Atmospheres, 119, 713–725, https://doi.org/10.1002/2013JD020882, 2014.

Haigh, I. D., Wadey, M. P., Wahl, T., Ozsoy, O., Nicholls, R. J., Brown, J. M., Horsburgh, K., and Gouldby, B.: Spatial and temporal analysis of extreme sea level and storm surge events around the coastline of the UK, Scientific Data, 3, 160 107, https://doi.org/10.1038/sdata.2016.107, 2016.

Hallegatte, S., Ranger, N., Mestre, O., Dumas, P., Corfee-Morlot, J., Herweijer, C., and Wood, R. M.: Assessing climate change impacts, sea level rise and storm surge risk in port cities: a case study on Copenhagen, Climatic Change, 104, 113–137, https://doi.org/10.1007/s10584-010-9978-3, 2011.

Halsnæs, K., Larsen, M. A. D., Sunding, T. P., and Dømgaard, M. L.: The value of advanced flood models, damage costs and land use data in cost-effective climate change adaptation, Climate Services, 32, 100 424, https://doi.org/10.1016/j.cliser.2023.100424, 2023.

Hansen, L.: Sea Level data 1889-2017 from 14 stations in Denmark Mean, maximum and minimum values calculated on monthly and yearly basis including plots of mean values, Tech. Rep. 18-16, Danish Meteorological Institute, ISSN 1399-1949, https://www.dmi.dk/fileadmin/Rapporter/2018/DMI_Report_18-16.pdf, 2018.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, https://arxiv.org/abs/1512.03385, 2015.

Hermans, T. H. J., Tinker, J., Palmer, M. D., Katsman, C. A., Vermeersen, B. L. A., and Slangen, A. B. A.: Improving sea-level projections on the Northwestern European shelf using dynamical downscaling, Climate Dynamics, 54, 1987–2011, https://doi.org/10.1007/s00382-019-05104-5, 2020.

Hinkel, J., Lincke, D., Vafeidis, A. T., Perrette, M., Nicholls, R. J., Tol, R. S. J., Marzeion, B., Fettweis, X., Ionescu, C., and Levermann, A.: Coastal flood damage and adaptation costs under 21st century sea-level rise, Proceedings of the National Academy of Sciences, 111, 3292–3297, https://doi.org/10.1073/pnas.1222469111, 2014.

Hinkel, J., Church, J. A., Gregory, J. M., Lambert, E., Le Cozannet, G., Lowe, J., McInnes, K. L., Nicholls, R. J., van der Pol, T. D., and van de Wal, R.: Meeting User Needs for Sea Level Rise Information: A Decision Analysis Perspective, Earth's Future, 7, 320–337, https://doi.org/10.1029/2018EF001071, 2019.

Holgate, S. J., Matthews, A., Woodworth, P. L., Rickards, L. J., Tamisiea, M. E., Bradshaw, E., Foden, P. R., Gordon, K. M., Jevrejeva, S., and Pugh, J.: New Data Systems and Products at the Permanent Service for Mean Sea Level, Journal of Coastal Research, 29, 493–504, https://doi.org/10.2112/JCOASTRES-D-12-00175.1, 2013.

Hosking, J. R. and Wallis, J. R.: Parameter and Quantile Estimation for the Generalized Pareto Distribution, Technometrics, 29, 339–349, https://doi.org/10.1080/00401706.1987.10488243, 1987.

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.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/9781009157896, 2021.

Irazoqui Apecechea, M., Melet, A., and Armaroli, C.: Towards a pan-European coastal flood awareness system: Skill of extreme sea-level forecasts from the Copernicus Marine Service, Frontiers in Marine Science, 9, https://doi.org/10.3389/fmars.2022.1091844, 2023.

Jacobsen, T., Sørensen, C., Woge Nielsen, J., and Su, J.: Historical extreme high water levels along the coastline of Denmark, DMI Report 21-28, Danish Meteorological Institute; Danish Coastal Authority, 2021.

Jebens, M., Sorensen, C., and Piontkowitz, T.: Danish risk management plans of the EU Floods Directive, E3S Web of Conferences, 7, 23 005, https://doi.org/https://doi.org/10.1051/e3sconf/20160723005, 2016.

Jevrejeva, S., Moore, J., Grinsted, A., Matthews, A., and Spada, G.: Trends and acceleration in global and regional sea levels since 1807, Global and Planetary Change, 113, 11–22, https://doi.org/https://doi.org/10.1016/j.gloplacha.2013.12.004, 2014.

Kallache, M., Vrac, M., Naveau, P., and Michelangeli, P.-A.: Nonstationary probabilistic downscaling of extreme precipitation, Journal of Geophysical Research, 116, https://doi.org/10.1029/2010jd014892, 2011.

Keskitalo, E. C. H., Juhola, S., Baron, N., Fyhn, H., and Klein, J.: Implementing Local Climate Change Adaptation and Mitigation Actions: The Role of Various Policy Instruments in a Multi-Level Governance Context, Climate, 4, https://doi.org/10.3390/cli4010007, 2016.

Kierulf, H. P., Steffen, H., Simpson, M. J. R., Lidberg, M., Wu, P., and Wang, H.: A GPS velocity field for Fennoscandia and a consistent comparison to glacial isostatic adjustment models, Journal of Geophysical Research: Solid Earth, 119, 6613–6629, https://doi.org/10.1002/2013JB010889, 2014.

Kirezci, E., Young, I. R., Ranasinghe, R., Muis, S., Nicholls, R. J., Lincke, D., and Hinkel, J.: Projections of global-scale extreme sea levels and resulting episodic coastal flooding over the 21st Century, Scientific Reports, 10, 11 629, https://doi.org/10.1038/s41598-020-67736-6, 2020.

Krinner, G., Kharin, V., Roehrig, R., Scinocca, J., and Codron, F.: Historically-based run-time bias corrections substantially improve model projections of 100 years of future climate change, Communications Earth & Environment, 1, 29, https://doi.org/10.1038/s43247-020-00035-0, 2020.

Laflamme, E. M., Linder, E., and Pan, Y.: Statistical downscaling of regional climate model output to achieve projections of precipitation extremes, Weather and Climate Extremes, 12, 15–23, https://doi.org/10.1016/j.wace.2015.12.001, 2016.

Latapy, A., Ferret, Y., Testut, L., Talke, S., Aarup, T., Pons, F., Jan, G., Bradshaw, E., and Pouvreau, N.: Data rescue process in the context of sea level reconstructions: An overview of the methodology, lessons learned, up-to-date best practices and recommendations, Geoscience Data Journal, 10, 396–425, https://doi.org/10.1002/gdj3.179, 2023.

MacPherson, L. R., Arns, A., Fischer, S., Méndez, F. J., and Jensen, J.: Bayesian extreme value analysis of extreme sea levels along the German Baltic coast using historical information, EGUsphere, 2023, 1–20, https://doi.org/10.5194/egusphere-2023-1122, 2023.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Open Access Earth System Science Data Discussions

Madsen, K. S., Høyer, J. L., Fu, W., and Donlon, C.: Blending of satellite and tide gauge sea level observations and its assimilation in a storm surge model of the North Sea and Baltic Sea, Journal of Geophysical Research: Oceans, 120, 6405–6418, https://doi.org/10.1002/2015JC011070, 2015.

Marcos, M., Calafat, F. M., Berihuete, Á., and Dangendorf, S.: Long-term variations in global sea level extremes, Journal of Geophysical Research: Oceans, 120, 8115–8134, https://doi.org/10.1002/2015JC011173, 2015.

Mentaschi, L., Besio, G., Cassola, F., and Mazzino, A.: Problems in RMSE-based wave model validations, Ocean Modelling, 72, 53–58, https://doi.org/https://doi.org/10.1016/j.ocemod.2013.08.003, 2013.

Mik-Meyer, V., Pereira, F. C., Larsen, M. A. D., Su, J., and Drews, M.: Storm surge prediction using a surrogate hydrodynamic model based on long short-term memory networks, Applied Ocean Research, 174, 105 197, https://doi.org/10.1016/j.apor.2026.105197, 2026.

Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C., 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.

Muis, S., Apecechea, M. I., Dullaart, J., de Lima Rego, J., Madsen, K. S., Su, J., Yan, K., and Verlaan, M.: A High-Resolution Global Dataset of Extreme Sea Levels, Tides, and Storm Surges, Including Future Projections, Frontiers in Marine Science, 7, 263, https://doi.org/10.3389/fmars.2020.00263, 2020.

Negus, A. R. A.: Building resilience to extremity and climatic changes investigating the phenomena of Compound Events, Master's thesis, University of Copenhagen, 2016.

Oppenheimer, M., Glavovic, B., Hinkel, J., van de Wal, R., Magnan, A. K., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., Deconto, R. M., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., and Sebesvari, Z.: Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities, The Intergovernmental Panel on Climate Change, in press, ISBN ???, www.ipcc.ch/report/srocc/, 2019.

Parent, E. and Bernier, J.: Bayesian POT modeling for historical data, Journal of Hydrology, 274, 95–108, https://doi.org/https://doi.org/10.1016/S0022-1694(02)00396-7, 2003.

Poulsen, B., Grupe, C., Frank, S. F., Piontkowitz, T., and Bonde, L.: Højvandsstatistikker 2024, Tech. rep., Kystdirektoratet, ISBN 978-87-7179-004-7, https://kyst.dk/media/ikykmzej/hoejvandsstatistikker-2024-05-11-2024.pdf, 2024.

Rabier, F., Brown, A., Chantry, M., and Pappenberger, F.: Weather forecasting in a changing climate: the rise of AI and Machine learning?, Journal of the European Meteorological Society, 4, 100 040, https://doi.org/10.1016/j.jemets.2026.100040, 2026.

Räty, O., Räisänen, J., and Ylhäisi, J. S.: Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations, Climate Dynamics, 42, 2287–2303, https://doi.org/10.1007/s00382-014-2130-8, 2014.

Rus, M., Mihanović, H., Ličer, M., and Kristan, M.: HIDRA3: a deep-learning model for multipoint ensemble sea level forecasting in the presence of tide gauge sensor failures, Geoscientific Model Development, 18, 605–620, https://doi.org/10.5194/gmd-18-605-2025, 2025.

She, J., Berg, P., and Berg, J.: Bathymetry impacts on water exchange modelling through the Danish Straits, Journal of marine systems, 65, 450–459, 2007.

Somot, S., Melet, A., Drenkard, E., Estrada-Allis, S., Herrmann, M., Holt, J., Iovino, D., Meier, H. E. M., Orr, A., Thatcher, M., Urakawa, S., and Zhang, X.: Report of the CORDEX Task Force on Regional Ocean Modelling and Climate Projections , https://doi.org/10.5281/zenodo.20290498, 2025.

Su, J.: ReSEML v1.0: Residual stacking ensemble for coastal sea level reconstruction, https://doi.org/10.5281/zenodo.21719179, 2026.

Su, J., Andrée, E., Nielsen, J. W., Olsen, S. M., and Madsen, K. S.: Sea Level Projections From IPCC Special Report on the Ocean and Cryosphere Call for a New Climate Adaptation Strategy in the Skagerrak-Kattegat Seas, Frontiers in Marine Science, 8, 471, https://doi.org/10.3389/fmars.2021.629470, 2021.

https://doi.org/10.5194/essd-2026-693
Preprint. Discussion started: 3 September 2026

Su, J., Poulsen, B., Nielsen, J. W., Sørensen, C. S., and Larsen, M. A. D.: Integrating historical storm surge events into flood risk security in the Copenhagen region, Weather and Climate Extremes, 45, 100 713, https://doi.org/https://doi.org/10.1016/j.wace.2024.100713, 2024.

Ulbrich, U., Leckebusch, G., and Pinto, J. G.: Extra-tropical cyclones in the present and future climate: a review, Theoretical and Applied Climatology, 96, 117–131, https://doi.org/10.1007/s00704-008-0083-8, 2009.

Vestøl, O., Ågren, J., Steffen, H., Kierulf, H., and Tarasov, L.: NKG2016LU: a new land uplift model for Fennoscandia and the Baltic Region, Journal of Geodesy, 93, 1759–1779, https://doi.org/10.1007/s00190-019-01280-8, 2019.

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., and Feyen, L.: Extreme sea levels on the rise along Europe's coasts, Earth's Future, 5, 304–323, https://doi.org/10.1002/2016EF000505, 2017.

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., Jevrejeva, S., Jackson, L. P., and Feyen, L.: Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard, Nature Communications, 9, 2360, https://doi.org/10.1038/s41467-018-04692-w, 2018.

Wahl, T., Haigh, I. D., Woodworth, P. L., Albrecht, F., Dillingh, D., Jensen, J., Nicholls, R. J., Weisse, R., and Wöppelmann, G.: Observed mean sea level changes around the North Sea coastline from 1800 to present, Earth-Science Reviews, 124, 51–67, 2013.

Wahl, T., Ward, P., Winsemius, H., AghaKouchak, A., Bender, J., Haigh, I., Jain, S., Leonard, M., Veldkamp, T., and Westra, S.: When Environmental Forces Collide, Eos, 99, https://doi.org/10.1029/2018EO099745, 2018.

Woodworth, P. L.: Some important issues to do with long-term sea level change, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 364, 787–803, https://doi.org/10.1098/rsta.2006.1737, 2006.

Woodworth, P. L., Hunter, J. R., Marcos, M., Caldwell, P., Menéndez, M., and Haigh, I.: Towards a global higher-frequency sea level dataset, Geoscience Data Journal, 3, 50–59, https://doi.org/10.1002/gdj3.42, 2016.

Zeder, J., Sippel, S., Pasche, O. C., Engelke, S., and Fischer, E. M.: The Effect of a Short Observational Record on the Statistics of Temperature Extremes, Geophysical Research Letters, 50, e2023GL104 090, https://doi.org/10.1029/2023GL104090, e2023GL104090 2023GL104090, 2023.

Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., and Zhang, X.: Future climate risk from compound events, Nature Climate Change, 8, 469–477, https://doi.org/10.1038/s41558-018-0156-3, 2018.