https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

Advancing seasonal storm surge forecasting through weather pattern–informed ensemble subselection

Anna K. Miesner[1], Daniel Krieger[2], Leonard F. Borchert[1]

[1] Earth and Society Research Hub, University of Hamburg, Hamburg, 20146, Germany
[2] Max Planck Institute for Meteorology, Hamburg, 20146, Germany

Correspondence to: Anna K. Miesner (anna.miesner@uni-hamburg.de)

Abstract

This study investigates how large-scale atmospheric circulation regimes can improve seasonal storm surge predictions in the North Sea using the high-resolution Max Planck Institute Earth System Model (MPI-ESM-HR) forecast ensemble. Although the ensemble mean shows limited skill in predicting interannual variations of seasonal storm surge activity, circulation-dependent information within the ensemble can be exploited through weather-pattern-informed subselection to improve predictions beyond the full ensemble mean. The approach provides a physically interpretable framework for extending circulation-based coastal hazard forecasting towards seasonal timescales. An idealised approach using perfect knowledge of weather patterns achieves anomaly correlations of 0.78 and 0.64 for seasonal predictions of surge height and surge-event counts, respectively, substantially exceeding the forecast-based subselection (0.27 and 0.31). This gap highlights the potential to improve seasonal storm surge forecasts through better prediction of surge-relevant atmospheric regimes.

1 Introduction

Operational European storm surge forecasts currently provide valuable short-range warnings, typically two days in advance (Chatzistratis et al., 2025), while emerging decadal surge predictions provide information relevant for long-term coastal adaptation and planning (Krieger et al., 2025). However, reliable seasonal storm surge forecasts that could support intermediate planning horizons, such as early mobilisation of response measures or increased risk awareness, are currently lacking. Extending prediction capability towards seasonal timescales could therefore help bridge the gap between short-term warnings and long-term climate projections. A key challenge is to identify predictable large-scale processes that can be used to extract useful information about future storm surge activity from seasonal climate forecasts. Here, we investigate whether large-scale atmospheric circulation regimes can be used to extract additional predictive information from seasonal climate forecast ensembles and improve predictions of seasonal storm surge activity in the German Bight.

Storm surges are a prominent coastal hazard along the North Sea coast and can cause substantial damage. They are defined by the Federal Maritime and Hydrographic Agency of Germany as water levels exceeding 1.5 m above mean high water (Jensen and Müller-Navarra, 2008). The German Bight experiences some of the largest storm surge magnitudes worldwide (Apolola et al., 2025). Owing to the shallow bathymetry of the North Sea, storm surges are primarily generated through surface wind stress, whereby persistent strong north-westerly winds drive large volumes of water towards the coast. At Cuxhaven one of the longest-operating tide-gauge records globally (Dangendorf et al., 2013; Haigh et al., 2023), previous analyses have shown that meteorological forcing explains a large fraction of winter seasonal sea-level variability, with zonal wind stress accounting for much of this variability (Dangendorf et al., 2013). The resulting meteorologically induced water-level anomaly is superimposed on the astronomical tide, with the highest coastal water levels occurring when storm surges coincide with high

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

water. Under these compound conditions, total water levels can exceed 5 m (Jensen and Müller-Navarra, 2008). Because large-scale atmospheric circulation controls the wind fields that drive storm surges, it provides a physically meaningful link between seasonal atmospheric predictability and coastal storm surge activity. Hess and Brezowsky (1969) classified recurring European circulation regimes into 29 distinct weather patterns ("Großwetterlagen"), a classification framework which has recently been automated with a deep learning approach by Heinrich et al. (2025). This strong link between atmospheric circulation and seasonal storm surge variability provides a basis for using weather-pattern classifications to assess whether circulation regimes can improve seasonal storm surge prediction at Cuxhaven.

Large-scale atmospheric circulation patterns provide a physically meaningful framework for linking atmospheric variability to coastal hazards. Heinrich et al. (2023) showed that cyclonic westerly circulation patterns dominate compound flood events along the North Sea coast that are driven by the combination of extreme river discharge and storm surges, highlighting the strong control of recurring circulation regimes on coastal extremes. Beyond synoptic timescales, advances in subseasonal-to-seasonal prediction have demonstrated that slowly varying components of the climate system and the large-scale atmospheric circulation can provide sources of predictability for extreme events like heat waves and tropical cyclones (Robertson et al., 2020; Vitart and Robertson, 2018). Improved prediction of large-scale circulation features has also been shown to enhance forecasts of regional extremes, such as European windstorm activity (Degenhardt et al., 2024).

Neal et al. (2018) demonstrated that probabilistic forecasts of weather patterns can identify periods with an increased likelihood of coastal flooding around the UK, and Costa et al. (2020) showed that weather-type information can improve statistical predictions of extreme storm surges by linking atmospheric circulation states to local surge responses. Together, these studies highlight the value of circulation regimes for converting atmospheric predictability into coastal risk information. Furthermore, Krieger et al. (2024) highlighted how physically informed ensemble subselection can improve seasonal predictions of German Bight storm activity by extracting additional predictive information from seasonal forecast ensembles beyond the ensemble mean. These advances suggest that combining predictable circulation regimes with ensemble-based forecast information provides a promising pathway for extending storm surge prediction beyond traditional short-range forecasts towards seasonal timescales.

A complementary approach is provided by statistical downscaling using neural networks (NNs), which can translate large-scale atmospheric information from climate models into local storm surge predictions. Krieger et al. (2025) demonstrated this approach for decadal surge predictions using an NN trained on observed surge heights and reanalysed mean sea-level pressure (MSLP), based on the framework of Harter et al. (2024) and Tiggeloven et al. (2021). Building on this approach, we apply the same NN framework to seasonal MSLP forecasts to generate local storm surge predictions at Cuxhaven. We then test whether weather-pattern-informed ensemble subselection can extract additional predictive information from these seasonal forecasts and improve the resulting local storm surge predictions. This combination links large-scale circulation predictability with a statistical translation of atmospheric forecasts into local coastal surge information.

Here, we assess the seasonal predictability of storm surge activity at Cuxhaven using retrospective forecasts from the Max Planck Institute Earth System Model in high resolution (MPI-ESM-HR) for the period 1961–2023. Storm surge activity is quantified using seasonal metrics describing the frequency and magnitude of extreme surge conditions during the main storm surge season (November–March). We first evaluate whether MPI-ESM-HR reproduces observed seasonal storm surge variability and associated large-scale circulation patterns. We then identify surge-promoting and surge-suppressing weather patterns and assess whether weather-pattern-informed ensemble subselection can extract additional predictive information beyond the ensemble mean. Finally, we evaluate whether this circulation-based subselection improves the resulting seasonal

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

storm surge predictions at Cuxhaven. By comparing the ensemble mean of the full forecast ensemble with weather-pattern-informed subselection and an idealised subselection based on observed circulation states, we further assess the potential and limitations of circulation-based approaches for seasonal storm surge prediction

2 Methods

2.1 Reanalysis data

We use hourly and daily mean sea-level pressure (MSLP), as well as daily 500 hPa geopotential height (z500) data from the ERA5 reanalysis (Hersbach et al., 2020) covering the period from 1960 - 2024 as input features to train our Neural Network**s** (NNs) and for assessing the skill of the seasonal prediction system. Before training, the data are converted into standardized anomalies, based on the available common period from 1 November 1960 to 31 May 2024. Standardization is achieved by subtracting the 1960 - 2024 mean and dividing by the standard deviation, for each grid point separately. This Z-score standardization step allows the trained NN to be applied to outputs from the seasonal prediction system, which are processed similarly to remove systematic biases with respect to ERA5.

2.2 Observations

Hourly water level observations at Cuxhaven-Steubenhöft (Fig. 1), Germany (1960 - 2024), are obtained from the Global Extreme Sea Level Analysis (GESLA) Version 4 (Caldwell et al., 2015; Haigh et al., 2023; Woodworth et al., 2016). Since these water levels include both astronomical tides and sea-level rise due to climate change, we first use *UTide* (Codiga, 2011) to estimate and subtract the tidal signal, in line with Krieger et al. (2025). We then remove the effect of sea-level rise by subtracting the annual mean water level for each year. The remaining water level time series can be considered purely atmospheric-driven and will be referred to as surge heights. The residual time series, representing atmospheric-driven variation in water level, is referred to as *surge heights*. The full observational record, including all months of the year, was used to train the neural networks and to assess the skill of the seasonal prediction system during the November–March (NDJFM) storm surge season.

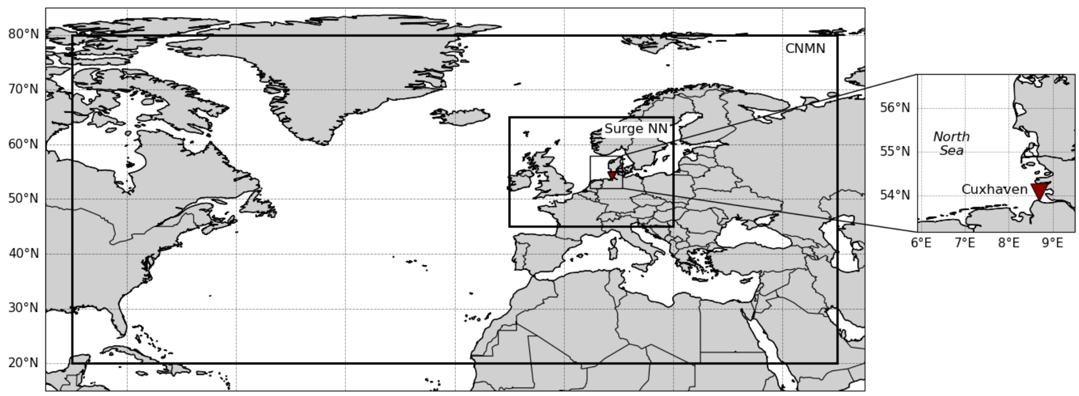

Figure 1. Study region and model domains. Location of Cuxhaven (triangle) and input domains of the neural networks (boxes): the neural network for translating MSLP into local storm surge predictions (Surge NN) and the neural network used for defining weather patterns (Convolutional Neural Network for Meteorological Pattern Recognition; CNMN). The right-hand panel provides a magnified view the German Bight where the lower point of the triangle indicates Cuxhaven.

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

2.3 Seasonal prediction system

We obtain hourly and daily MSLP, as well as daily z500 predictions from seasonal retrospective forecasts (also known as re-forecasts or hindcasts) generated with the Max Planck Institute Earth System Model in -high resolution (MPI-ESM-HR). MPI-ESM-HR consists of the atmospheric component ECHAM6 (Stevens et al., 2013), configured with a horizontal resolution of ~1° on a T127 spectral grid and 95 vertical levels extending to 0.1 hPa (~80 km), and the oceanic component MPIOM (Jungclaus et al., 2013), configured with a horizontal resolution of ~0.4° on a TP04 curvilinear grid and 40 vertical levels. The model is also employed in the German climate forecasting system (Fröhlich et al., 2021). As in Sylla et al., (2026), we use an updated version of the model with 30 ensemble members, initialized every 1 November each year from 1960 to 2023 and integrated for two years. For seasonal hindcast initialization, a corresponding 30-member data assimilation experiment based on MPI-ESM-HR was created. In this experiment, the model is nudged toward monthly mean atmospheric temperature, MSLP, vorticity, and divergence fields derived from the ERA5 reanalysis (Hersbach et al., 2020) as described in Fröhlich et al. (2021). Monthly mean temperature and salinity were assimilated from EN4 profiles (Good et al., 2013), together with monthly sea-ice concentration from OSI SAF (Lavergne et al., 2019), using an ensemble Kalman filter (Brune et al., 2015; Brune and Baehr, 2020) implemented with the Parallel Data Assimilation Framework (Nerger and Hiller, 2013). All simulations were run under externally prescribed forcing consistent with the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016), with historical forcing applied through 2014 and SSP2-4.5 forcing thereafter. We standardize the output for each MPI-ESM-HR ensemble member using the gridpoint-wise mean and standard deviation calculated across all simulations of that specific member. This approach removes biases between members while preserving their individual MSLP/z500 variability.

2.4 Neural Network (NN) for translating MSLP into local storm surge predictions

In order to translate MSLP anomalies into surge heights, we apply the same NN setup as Krieger et al. (2025) applied for decadal surge predictions, based on Harter et al. (2024) and Tiggeloven et al. (2021). Accordingly, each NN has an input layer, three hidden layers with 64, 256, and 1024 neurons using Rectified Linear Unit activation (Wani et al., 2020), and a linear output layer for continuous surge height predictions. The model is trained with mean squared error loss and the Adam optimizer (Kingma and Ba, 2017), using a batch size of 64 for 80 epochs. Each NN is trained using hourly water level observations at Cuxhaven and hourly standardized MSLP anomalies from ERA5. The anomalies are interpolated onto the regular Gaussian F96 grid (horizonal resolution of ~1°) of MPI-ESM-HR and within the region bounded by 45°N–65°N and 10°W–20°E (Surge NN, Fig. 1), following the spatial extent used Krieger et al. (2025). The trained NN is then applied to standardized MSLP predictions from MPI-ESM-HR within the same spatial extent. The target dataset for the NN is hourly surge heights at Cuxhaven, which is smoothed with a centered 12-hour running mean filter to eliminate residual tidal signals from non-linear interactions not captured by harmonic analysis. In line with Krieger et al. (2025) we use a 6-hour time lag, so that each time step of MSLP anomaly matches the start of the following 12-hour average of surge heights.

2.5 Storm surge analysis and metrics

Storm surge variability is quantified using two complementary metrics describing the magnitude and frequency of extreme surge conditions. Storm-surge conditions are identified using a threshold corresponding to the 98th percentile (q98) of the hourly surge-height distribution over the common period (1 November 1960 to 31 May 2024), computed separately for observations and predictions (Fig. 3a). For ensemble predictions, each ensemble member uses its own threshold, defined as the 98th percentile of the member-specific historical hourly surge-height distribution over 1960–2023. This accounts for systematic differences in surge variability among ensemble members while maintaining a consistent exceedance probability within each member.

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

From the hourly surge-height time series, two daily metrics are derived. The first is the daily q98 surge height**,** defined as the 98th percentile of hourly surge heights within a given day. This metric characterises the upper tail of the daily surge-height distribution (in practice it is close to the daily maximum) and provides a measure of the magnitude of elevated surge conditions. The second metric is the daily surge-event count**,** defined as the number of independent surge peaks per day exceeding the q98 threshold. To avoid double counting of the same surge event, a minimum separation of 12 hours is imposed between consecutive peaks. Only peaks exceeding the member-specific q98 threshold are retained.

To investigate variability on seasonal timescales, daily metrics are aggregated over the storm-surge season, defined as November–March (NDJFM) when both observations and simulations show above-climatological mean storm surge activity (Fig. 3b). Seasons are assigned on a winter basis such that November and December of a given year and January–March of the following year belong to the same season (e.g. season 1960 refers to 01 November 1960 to 31 March 1961).

Daily metrics are first restricted to NDJFM months and subsequently summed to obtain seasonal values. Two seasonal metrics are considered. The seasonal cumulative q98 surge height is defined as the sum of daily q98 surge heights over all NDJFM days. This metric represents the integrated seasonal magnitude of upper-tail surge conditions. The seasonal surge-event count is defined as the total number of surge events during NDJFM, obtained by summing the daily surge-event counts over the season. Together, these metrics describe the seasonal intensity and frequency of extreme storm-surge activity.

Seasonal metrics are expressed as anomalies relative to their respective NDJFM climatological means over the common period (1960–2024). For observations, a single climatological reference is used. For ensemble predictions, anomalies are calculated separately for each ensemble member using the member-specific climatological mean over the same period. This removes systematic member biases while preserving inter-member variability. The ensemble-mean time series is treated separately and centred on its own climatology. The resulting anomaly time series are used for all subsequent analyses of seasonal storm-surge variability and forecast skill. Because MPI-ESM-HR is initialized annually in November and only the first forecast year is considered, the seasonal surge predictions represent lead month 1-5 average forecasts spanning the full NDJFM storm-surge season.

2.6 "Best Member" Selection

To evaluate the maximum potential predictive skill of MPI-ESM-HR, we identify the best ensemble member for each year in an objective, post hoc manner. For each surge metric and initialization year, the observed seasonal anomaly is compared with the corresponding predicted anomalies from all ensemble members. The member with the smallest absolute difference from the observation is selected as the best member. This identifies the ensemble member that most closely matches the observed value in a given year and provides an estimate of the maximum achievable ensemble skill. This approach is diagnostic only and does not represent real-time forecast skill, because it relies on the observed outcome, which is not available in operational forecasting.

2.7 Weather pattern-informed ensemble subselection

2.7.1 NN for defining weather patterns

To examine the dominant weather patterns affecting atmospheric variability linked to storm surges in the North Sea, we employ the Convolutional Neural Network for Meteorological Pattern Recognition (CNMN; Heinrich, 2024; Heinrich et al., 2025), which classifies weather patterns into 29 categories following Hess and Brezowsky's (1969) definition of "Großwetterlagen". Daily MSLP and z500 fields are interpolated onto a ~2.8° grid covering 20°N–80°N and 90°W–50°E (Fig. 1) and standardized using z-scores. During training, these variables are derived from ERA5 reanalysis within this domain. The trained model is then applied to MPI-ESM-HR retrospective forecasts of daily MSLP and z500 and interpolated onto the same grid, to classify

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

atmospheric circulation patterns during the forecast period. Since weather patterns are defined to last at least three days, a three-day smoothing is applied to ensure the persistence of circulation patterns.

2.7.2 Weather pattern classification and ensemble subselection

To identify weather patterns associated with storm surge events, each daily weather pattern is linked to a binary surge occurrence defined by whether at least one surge event occurs per day (daily surge-event count > 0), applied consistently to both ERA5 observations and MPI-ESM-HR predictions. The frequency of each pattern is calculated as the percentage of time it occurred during the common time period 1960–2024. Subsequently, differences in pattern frequency between surge and no-surge events are assessed, with larger positive values indicating a higher occurrence during storm surges. A threshold of ±0.5% is applied to exclude weak signals, ensuring that only substantial positive or negative anomalies are considered. Patterns exceeding the threshold in both datasets are defined as surge-promoting or surge-suppressing, while patterns below the threshold or with inconsistent signals between datasets are considered as neutral or ambiguous and not included in further analysis. This conservative approach ensures that only robust patterns that are supported by both data sets are categorized.

The resulting surge-promoting and surge-suppressing weather patterns are then used for ensemble subselection (Fig. 2): For each daily prediction, the full ensemble is analysed by counting the number of members predicting surge-promoting patterns. If more than half of the ensemble (i.e., more than 15 members) predict surge-promoting patterns, only these members are retained (Fig. 2, left pathway). Conversely, if 15 or fewer members predict surge-promoting patterns, a subset of members predicting surge-suppressing patterns is constructed (Fig. 2, right path). This subselection aims to amplify the physically meaningful signal while reducing noise from ensemble members that are physically inconsistent with the initial condition.

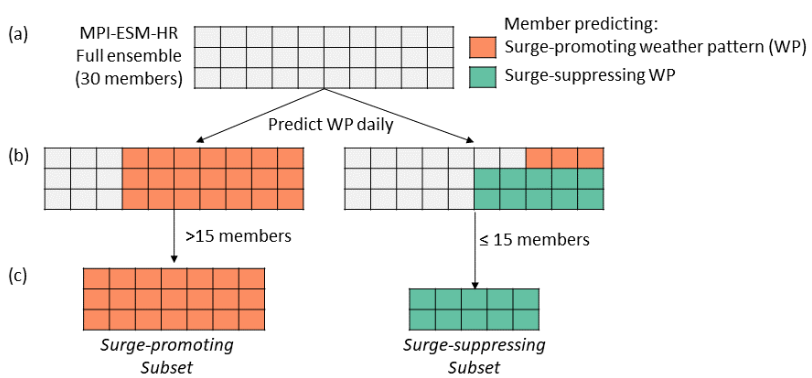

Figure 2. Illustration of the weather pattern conditioned ensemble subselection. a) For each prediction with MPI-ESM-HR, we examine the full ensemble and count how many ensemble members predict surge-promoting weather patterns (based on the WP defined in Fig. 6). b) If the majority (> 15 ensemble members) predict surge-promoting patterns, only those members are retained in the subset (c, surge-promoting subset). b) Conversely, if 15 or fewer members predict surge-promoting patterns, we construct a subset consisting only of ensemble members predicting surge-suppressing patterns (c).

2.7.3 Idealized Ensemble Subselection (ERA-Matched subset, ERA-Sub)

To assess the upper bound of predictive skill achievable through weather-pattern-based conditioning, we construct an idealised subselection. In this approach, only those MPI-ESM-HR ensemble members whose simulated weather patterns match the corresponding observed weather pattern for each day are retained. This yields an ensemble subset that is dynamically consistent with the ERA5 large-scale atmospheric state at each time step. The procedure is strictly diagnostic and not suitable for operational forecasting, as it relies on perfect knowledge of the observed circulation during storm surge season. Nevertheless,

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

it provides a benchmark for the maximum potential skill attainable if the correct weather-pattern-conditioned ensemble members could be identified a priori.

2.8 Skill Metrics and Forecast Verification

Forecast performance is evaluated within a unified verification framework based on deterministic, probabilistic, and binary (categorical) metrics applied to seasonal (NDJFM) anomaly time series of surge-related variables. For consistency across verification frameworks, all deterministic and binary metrics are computed using the ensemble mean of each model and evaluated against observations, while probabilistic metrics used the full ensemble distribution.

Deterministic skill is assessed using the Pearson correlation coefficient and the root-mean-square error (RMSE). For visualizing forecast skill compared to a climatology reference forecast (as in Fig. 7) a year-wise categorical skill indicator is defined relative to a climatological benchmark. For each year, the absolute forecast error is compared with the absolute deviation of the observation from the climatological mean (zero-anomaly baseline). A forecast is classified as skilful when its error was smaller than that of the climatological reference forecast.

Ensemble probabilistic performance is quantified using the continuous ranked probability score (CRPS) computed from the full ensemble distribution, providing a proper scoring rule for evaluating the entire predictive distribution.

Categorical forecast performance is evaluated using a contingency-table framework to assess the ability of MPI-ESM-HR to discriminate between positive and non-positive surge anomalies. Seasonal surge anomalies from both MPI-ESM-HR and ERA5 are binarized using a threshold of zero, with anomalies >0 classified as positive anomaly conditions representing above-climatological surge conditions, and anomalies ≤0 classified as non-positive anomaly conditions. The resulting binary forecasts are compared with ERA5 using a confusion matrix, which summarizes true positives (TP; correctly predicted positive anomalies), false positives (FP; predicted positive anomalies that are not observed), false negatives (FN; observed positive anomalies that are not captured by the forecast), and true negatives (TN; correctly predicted non-positive anomalies). Forecast performance is quantified using accuracy, positive predictive value (PPV), negative predictive value (NPV), and the True Skill Statistic (TSS). Accuracy measures the overall fraction of correctly classified seasons on a deterministic level:

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (1)$$

PPV represents the proportion of forecast positive anomaly conditions that correspond to observed positive anomalies on a probabilistic level:

$$PPV = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (2)$$

whereas NPV represents the proportion of forecast non-positive anomaly conditions that correspond to observed non-positive anomalies in a binary way:

$$NPV = \frac{\mathrm{TN}}{\mathrm{TN+FN}} \quad (3)$$

TSS evaluates the ability of the forecast to correctly distinguish positive anomaly conditions from non-positive anomaly conditions while accounting for random agreement. It is calculated as the sum of sensitivity and specificity minus one:

$$TSS = \frac{\mathrm{TP}}{\mathrm{TP+FN}} + \frac{\mathrm{TN}}{\mathrm{TN+FP}} - 1 \quad (4)$$

Together, these three verification levels (deterministic, probabilistic, binary) provide a hierarchical assessment of forecast quality from deterministic accuracy to probabilistic calibration and event-based discrimination (Jolliffe and Stephenson, 2012; Wilks, 2011). Uncertainty in all skill metrics is quantified using a non-parametric bootstrap with 5000 resamples, in which annual NDJFM anomalies are independently sampled with replacement. This is supported by preliminary diagnostics showing

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

negligible interannual autocorrelation and an effective sample size equal to the nominal sample size (n = 32). Ninety-five percent confidence intervals were derived from the empirical 2.5th and 97.5th percentiles of the bootstrap distributions.

3. Results

3.1 Defining storm surges and storm surge season

Seasonal predictions with MPI-ESM-HR, when compared to ERA5 observations, tend to slightly underestimate the overall surge height (Fig. 3a) as well as the mean monthly number of surges (Fig. 3b). Most notably, the 98th percentile of surge heights differs between the observations and predictions (Fig. 3a). In ERA5 observations, surges above 1.49 m are considered storm surges, whereas in MPI-ESM-HR, a surge is defined at a lower threshold of above 1.22 m (Fig. 3a). Storm surges in Cuxhaven exhibit a clear seasonal cycle, with the highest frequency in winter and the lowest in summer (Fig. 3b). The period from November to March (NDJFM) is identified as the storm surge season, as both observations and simulations exceed their climatological means during these months. Overall, observed and simulated surge frequencies are consistent, with the largest discrepancy occurring in October. In that month, observed surges exceed the climatological mean (2.08 surges yr^{-1}), whereas MPI-ESM-HR shows lower values (Fig. 3b). Although some ensemble members lie near or slightly above the climatological mean (1.17 surges per season), the ensemble-mean October value remains below the long-term climatology (Fig. 3b). The corresponding negative bias of 0.91 surges per season, indicates that MPI-ESM-HR underestimates surge frequency throughout the year. The climatological NDJFM mean over 1960–2024 shows 18.33 surge events per season in ERA5 and 11.86 in MPI-ESM-HR. Mean daily q98 surge heights during NDJFM are likewise lower in MPI-ESM-HR (0.24 m) than in ERA5 (0.34 m), corresponding to cumulative seasonal q98 surge heights of 36.07 m and 51.23 m, respectively, over the 151-day season. These climatological values serve as the baseline for computing seasonal anomalies. Anomalies are used to reduce systematic mean differences between ERA5 and MPI-ESM-HR and to enable a clearer assessment of interannual variability during the storm surge season.

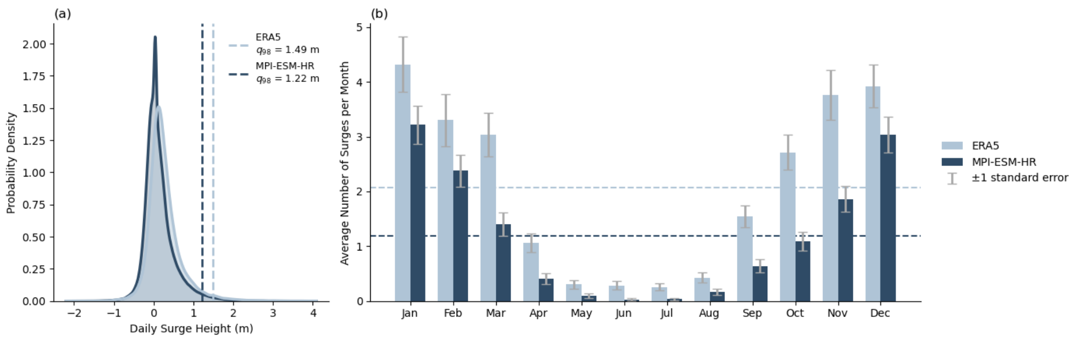

Figure 3. Probability density of daily surge heights (a) and monthly climatology of storm surge frequency (b) from ERA5 observations (light colour) and MPI-ESM-HR lead year one predictions (dark colour). In (a) MPI-ESM-HR is aggregated across all ensemble members and the dashed vertical lines indicate the 98th percentile surge height threshold (q98) for each dataset, where surge heights to the right of these lines represent the extreme tail of the distribution and are defined as storm surge events. In (b) bars show monthly mean counts of storm surge events, and error bars indicate ±1 standard error of the mean, estimated from interannual variability in ERA5 and from interannual variability across MPI-ESM-HR ensemble members. In (b), dashed horizontal lines denote the respective climatological mean number of surges for each dataset. Panel (a) uses the full daily record over the common period 1960–2024, whereas panel (b) is based on monthly aggregated values covering the period January 1961–December 2023.

3.2 Seasonal storm surge forecast skill

Seasonal (NDJFM) storm surge forecast skill is assessed by comparing anomalies of the seasonal cumulative q98 surge height and seasonal surge-event count derived from MPI-ESM-HR with those from ERA5. The ensemble mean derived from MPI-ESM-HR exhibits weak agreement with ERA5-derived storm surge metrics over the surge season, with low and statistically

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

non-significant correlations for both seasonal cumulative q98 surge height (r = 0.213, p = 0.090, Fig. 4a) and seasonal surge-event count (r = 0.229, p = 0.070, Fig. 4b). Consistent with these modest correlations, the ensemble mean shows relatively large prediction errors, with anomaly-based RMSE values of ~19 m for seasonal cumulative q98 surge height and ~8.5 surges for seasonal surge surge-event count (Fig. 4), corresponding to 37% and 47% of the observed seasonal mean values of ~ 51 m and 18 events, respectively. Despite the limited skill of the ensemble mean, the ensemble spread encompasses much of the observed variability, with only 6 of 63 observations falling outside the ensemble distribution for seasonal cumulative q98 surge height and 14 of 63 for surge-event counts (Fig. 4).

To estimate the upper bound of forecast skill available within the ensemble, we evaluate the correlation between observations and the ensemble member that most closely matched each observed year (‘Best Member’). This Best Member exhibits near-perfect agreement with observations, with correlations of 0.984 for seasonal cumulative q98 surge height (Fig. 4a) and 0.985 for seasonal surge-event count (Fig. 4b). Correspondingly, RMSE is reduced to ~3.6 m and ~1.9 surges, respectively (Fig. 4). These results indicate that MPI-ESM-HR contains ensemble members that closely track observed variability, indicating that the ensemble contains substantial predictive information, but that the signal is relatively low compared to the noise within the ensemble. The large disparity between the skill of the ensemble means and that of the Best Member suggests that forecast improvements may be achievable through physically informed ensemble subselection.

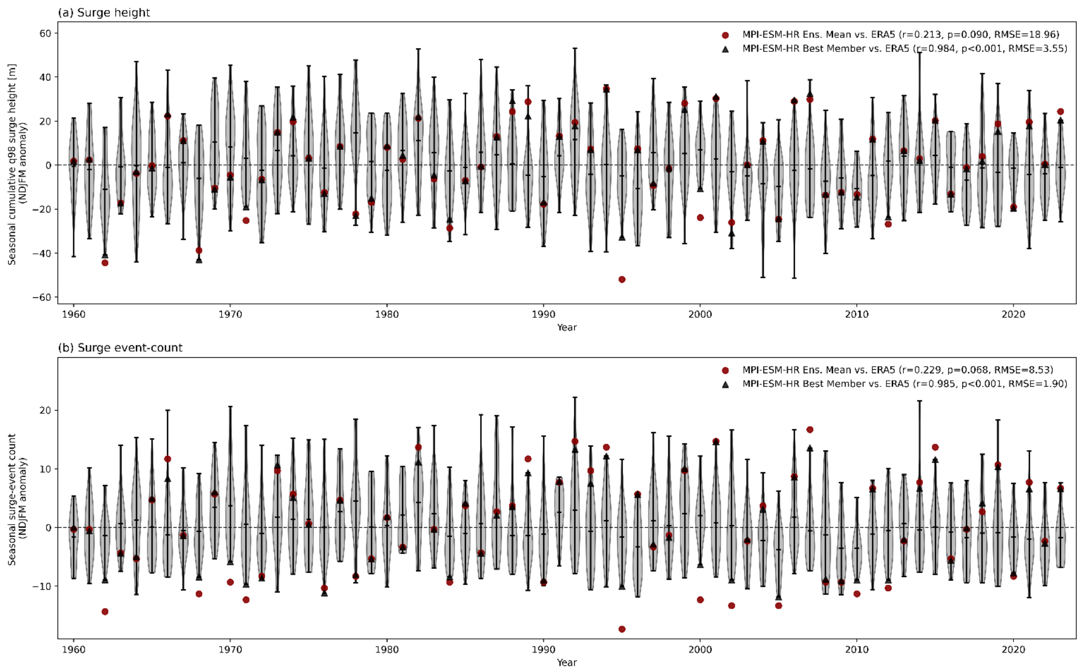

Figure 4. Annual anomalies of seasonal cumulative q98 surge height (a) and seasonal surge-event count (b) during NDJFM, comparing MPI-ESM-HR five-month mean ensemble predictions with ERA5 observations. MPI-ESM-HR predictions are shown as ensemble-member distributions using violins to illustrate data density. Observed values from ERA5 are indicated as red points, and the ‘Best Member’, the ensemble member closest to observations, is marked by a black triangle. The legend reports Pearson correlation/Anomaly correlation coefficients (r) and corresponding p-values (p), quantifying the agreement between observed and predicted anomalies. The year denotes the November to March period; e.g. 1960 refers to 01 November 1960 to 31 March 1961.

3.3 Identifying and classifying weather patterns connected to storm surges

The climatological frequency distribution of large-scale weather patterns is well reproduced by MPI-ESM-HR compared with ERA5 (Fig. 5). Across the full analysis period (1961–2023), both datasets are dominated by westerly circulation types, with

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

the Cyclonic Westerly regime (WZ) being the most frequent pattern. WZ accounts for 27.5% of all events in ERA5 (Fig. 5a) and 29.3% in MPI-ESM-HR (Fig. 5b), indicating close agreement between the observed and simulated climatology. This is followed by high-pressure conditions (HM: 9.7% and 12.2%) and a zonal ridge structure across Central Europe (BM: 7.9 and 5.6). The model exhibits a slightly higher occurrence of WZ and HM, relative to ERA5, while preserving the overall ranked structure of dominant weather patterns (Fig. 5a,b).

During storm-surge events, the frequency distribution of weather patterns associated with strong westerly (onshore) wind conditions increase substantially with consistent behaviour in ERA5 and MPI-ESM-HR. WZ increases from 27.5% to 62.7% in ERA5 and from 29.3% to 65.2% in MPI-ESM-HR (Fig. 5c-d). Cyclonic North-Westerly (NWZ) becomes the second most dominant regime and increases from ~4% to 12.9 in ERA5 and 15.1% in MPI-ESM-HR (Fig. 5c-d). The anticyclonic westerly regime (WA) is present during 6% of days in ERA5 (Fig. 5c) but only during 3% of days in MPI-ESM-HR.

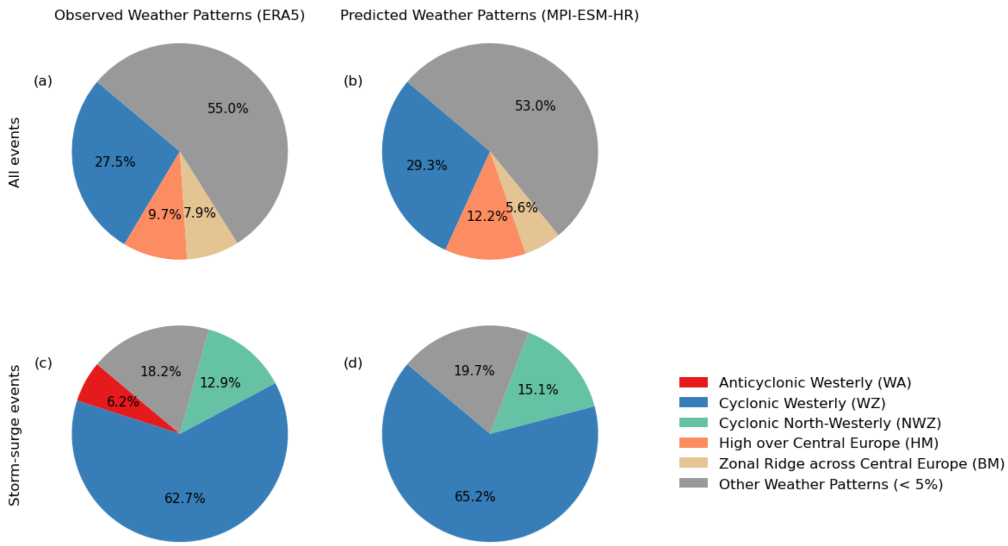

Figure 5. Climatological frequency distribution (%) of large-scale weather patterns associated with all events (a–b) and storm-surge events (c–d), derived from daily ERA5 observations (left column) and MPI-ESM-HR lead year-one predictions (right column), covering the period January 1961 to December 2023. Weather patterns occurring on at least 5% of days are shown, all other patterns are combined as 'Others'. For MPI-ESM-HR all ensemble members are pooled to construct a single probabilistic distribution of pattern occurrences, thereby representing the full ensemble climatology rather than a single deterministic realisation or ensemble mean state.

The direct comparison of weather patterns connected to surge-events and no surge-events enables a more thorough exploration of differences (Fig. 6). Consistent with previous results (Fig. 5) there is a shift towards westerly circulation types during surge-events: WZ increases strongest by +37.4 percentage points in ERA5 and 37.3 in MPI-ESM-HR, while NWZ increases by +9.4 and +11.9 percentage points, respectively, with a slightly stronger signal in the model (orange bars, Fig. 6). Since WZ and NWZ dominate during surge-events we define them as surge-promoting weather patterns in the subsequent ensemble sub-selection.

In contrast, during non-surge conditions blocked or ridge-dominated weather patterns dominate. HM shows the largest negative anomaly (−9.5% in ERA5 and −11.9% in MPI-ESM-HR), followed by BM (−7.3% and −4.9 % , respectively). These and all other weather patterns occurring more frequently under no-surge conditions (green bars, Fig. 6) form the surge-suppressing subset used for ensemble subselection. Intermediate or transitional regimes with weak or inconsistent relationships with surge occurrence (grey bars, Fig. 6) are excluded from the subselection. Overall, the results demonstrate a robust dynamical

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

separation between surge-promoting westerly regimes and surge-suppressing blocked states, with generally consistent magnitudes between ERA5 and MPI-ESM-HR.

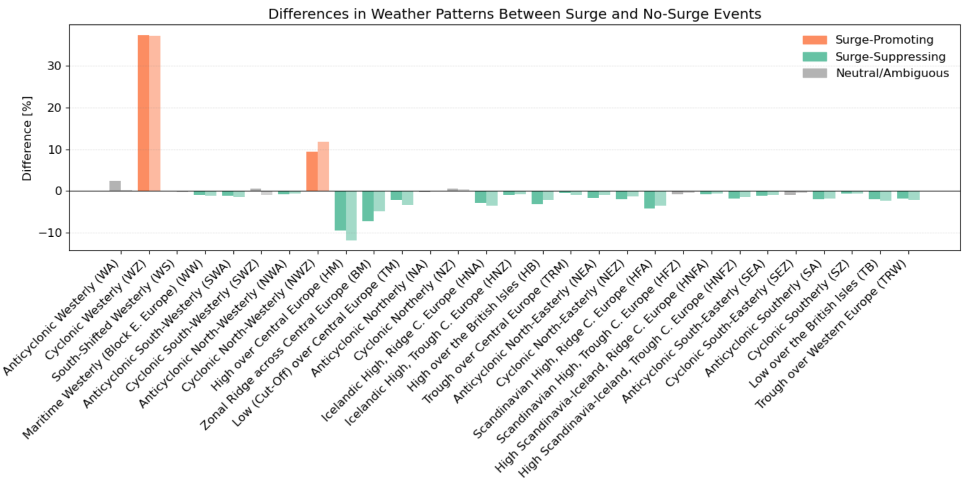

Figure 6. Percentage differences in weather-pattern frequencies between surge and no-surge events. Bars show ERA5 observations (darker colours) and MPI-ESM-HR predictions (lighter colours) for each weather pattern, labelled with full name and abbreviation. Colours indicate surge classification, orange (surge-promoting), green (surge-suppressing) and neutral (grey) indicates that either ERA5 and prediction disagree or the signal is below ± 0.5%. Frequencies were computed from daily observations and predictions over common time period considering only predictions within the first year of the model run.

3.4 Improving seasonal storm surge forecasts through ensemble subselection

Now that a clear link between large-scale weather patterns and storm surge occurrence is established, we assess whether this information can improve seasonal storm surge forecasts through weather pattern-informed ensemble subselection. Over the full hindcast period (1960–2023), weather-pattern-based ensemble subselection improves deterministic forecast skill for both storm surge metrics (Table 1). For seasonal cumulative q98 surge height, the anomaly correlation coefficient (ACC) increases from 0.21 for the full ensemble to 0.27 for the subselected ensemble, resulting in statistically significant skill (p = 0.032). However, this improvement is not accompanied by improvements in categorical verification metrics, with slightly lower accuracy (0.53 vs. 0.56), positive predictive value (PPV; 0.57 vs. 0.61), negative predictive value (NPV; 0.50 vs. 0.53), and true skill statistic (TSS; 0.07 vs. 0.13) compared with the full ensemble. RMSE is similar between the two configurations (18.80 and 18.96, respectively), while CRPS increases from 11.28 to 12.67. In contrast, the idealized ERA-matched ensemble achieves substantially higher skill, with ACC = 0.78, lower RMSE (12.66), and improved categorical performance (accuracy = 0.80, PPV = 0.82, NPV = 0.77, TSS = 0.59).

For seasonal surge-event counts, subselection results in more consistent improvements. ACC increases from 0.23 for the full ensemble to 0.31 for the subselected ensemble, with a small reduction in RMSE from 8.53 to 8.35 events (Table 1). Categorical performance also improves, with accuracy increasing from 0.55 to 0.63, PPV from 0.52 to 0.61, NPV from 0.58 to 0.64, and TSS from 0.09 to 0.24. CRPS, however, increases from 5.33 to 6.48. ERA-Sub again shows the highest performance, achieving ACC = 0.64, RMSE = 7.02 events, and improved categorical skill (accuracy = 0.66, PPV = 0.63, NPV = 0.68, TSS = 0.31).

Overall, weather-pattern-based subselection consistently increases anomaly correlation skill, indicating improved representation of interannual storm surge variability, with the largest improvements for seasonal surge-event counts. However, improvements in categorical and probabilistic metrics are less consistent, particularly for seasonal surge height. The substantially higher skill achieved by the ERA-matched ensemble demonstrates the gap between current weather-pattern-based

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

subselection and the potential skill attainable when the seasonal evolution of surge-relevant weather patterns is perfectly represented.

Table 1. Forecast verification summary for the full ensemble (Full), the weather-pattern–based subselected ensemble (Sub), and the ERA-matched subselected ensemble (ERA-Sub), using ERA5 as the reference dataset for seasonal q98 surge height and seasonal surge-event count (NDJFM anomalies) over the common time period (1961-2023). Continuous skill metrics include the Anomaly Correlation Coefficient (ACC) with statistical significance indicated at the 5% (*) and 1% (**) levels, the Root Mean Squared Error (RMSE) and the Continuous Ranked Probability Score (CRPS). The ideal values are 1 for ACC, Accuracy, PPV, NPV, and TSS, and 0 for RMSE and CRPS. Model improvements relative to the full ensemble are indicated in bold.

Variable	Ensemble	ACC (p-value)	RMSE	CRPS	Accuracy	PPV	NPV	TSS
Seasonal cumulative q98 surge height	**Full**	0.213	18.964	11.281	0.562	0.607	0.528	0.133
	Subselected	**0.268***	**18.799**	12.667	0.531	0.571	0.500	0.071
	ERA-Sub	**0.777****	**12.664**	12.182	**0.797**	**0.818**	**0.774**	**0.594**
Seasonal surge-event count	**Full**	0.229	8.533	5.332	0.547	0.516	0.576	0.092
	Subselected	**0.307****	**8.354**	6.475	**0.625**	**0.607**	**0.639**	**0.243**
	ERA-Sub	**0.643****	**7.015**	6.100	**0.656**	**0.633**	**0.676**	**0.310**

Figure 7 shows the interannual variability of seasonal cumulative q98 surge height and surge-event count, highlighting differences in predictability across ensemble subsets and time periods. The subselected ensemble reproduces interannual variability more accurately, with higher correlation to observations and reduced ensemble spread for both surge height and surge count (Fig. 7a,c). However, a systematic amplitude bias remains: MPI-ESM-HR underestimates the magnitude of observed anomalies, particularly during pronounced positive events (e.g. 1973, 1974, 1992, 2015) and negative events (e.g. 1962, 1968, 1984, 1995, 2012; Fig. 7a,c). This underestimation is present in both the full model and the subselected ensemble, with only marginal improvement in the latter, whereas ERA-Sub more frequently reproduces observed anomaly magnitudes (Fig. 7a,c).

Note that the tight co-variation of surge height and surge-event count across years (Fig. 7) reflects a strong statistical linkage between the two metrics, with both exhibiting consistently high correlations across all datasets ($r \approx 0.90$, $p < 0.001$). This relationship is partly structural, as event counts are defined by exceedances of the q98 surge-height threshold, and is therefore not fully independent.

Beyond amplitude skill, subselection also affects the directional consistency of anomalies. Weather-pattern-based subselection improves sign agreement with observed surge anomalies in several cases, with shifts in both surge height and surge count aligning with observed sign changes (e.g. 1993, 2007, 2011, 2012, 2015), but it also produces opposite shifts in some years (e.g. 1970, 1976, 1999, 2001; Fig. 7a,c). ERA-Sub generally reproduces the observed anomaly direction, although sign mismatches persist in both metrics in some years (e.g. 1970, 1973, 1977, 1989, 2021; Fig. 7a,c).

Model performance relative to climatology differs between the two variables and across ensemble subsets. For surge height, the full ensemble yields lower absolute error than the climatological reference in 34 years (53.1%). The proportion is slightly smaller for the subselected ensemble (31 years, 48.4%), but markedly higher for ERA-Sub which outperforms climatology in more than 70 percent of years (47 years; 73.4%; Fig. 7a). For surge-event count, the fraction of years with reduced absolute

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

error relative to climatology increases from 31 years (48.4%) in the full ensemble to 37 years (57.8%) after subselection, reaching 40 years (62.5%) for ERA-Sub (Fig. 7c).

Comparing model performance across the two sub-periods (1960–1991 and 1992–2023) reveals clear temporal differences across variables and ensemble configurations (Fig. 7b,d). ERA-Sub consistently exhibits the highest correlations, particularly high during the recent period, with ACC values of 0.79 for seasonal cumulative q98 surge height and 0.67 for surge-event counts, compared with 0.76 and 0.60, respectively, during 1960–1991.

Temporal differences are more pronounced for the full and subselected ensembles. During 1960–1991, both configurations show weak and non-significant skill for surge height (Full: ACC = 0.30; Sub: ACC = 0.24) and surge-event counts (Full: ACC = 0.21; Sub: ACC = 0.17), with confidence intervals extending to approximately −0.2 (Fig. 7b,d), indicating limited robustness of predictive skill during this period. In contrast, during 1992–2023, the Subselected ensemble achieves higher correlations for both surge height (ACC = 0.36) and surge-event counts (ACC = 0.54) than over the full hindcast period, with confidence intervals excluding zero, indicating statistically significant positive skill (Fig. 7b,d). Over the full period, robust positive skill is only achieved by the subselected ensemble for surge-event counts and by the ERA-Sub ensemble for both surge-event counts and surge height (Fig. 7b,d).

A similar temporal pattern is evident for categorical forecast skill, particularly for surge height. During 1992–2023, the ERA-Sub ensemble shows substantially higher accuracy, PPV, NPV, and TSS compared with 1960–1991, while the operational weather-pattern-based subselection also improves PPV and TSS (Annex, Table A1). These results indicate that forecast skill is not stationary but varies between periods, with more recent decades showing greater potential for extracting predictable information from the seasonal forecast ensemble.

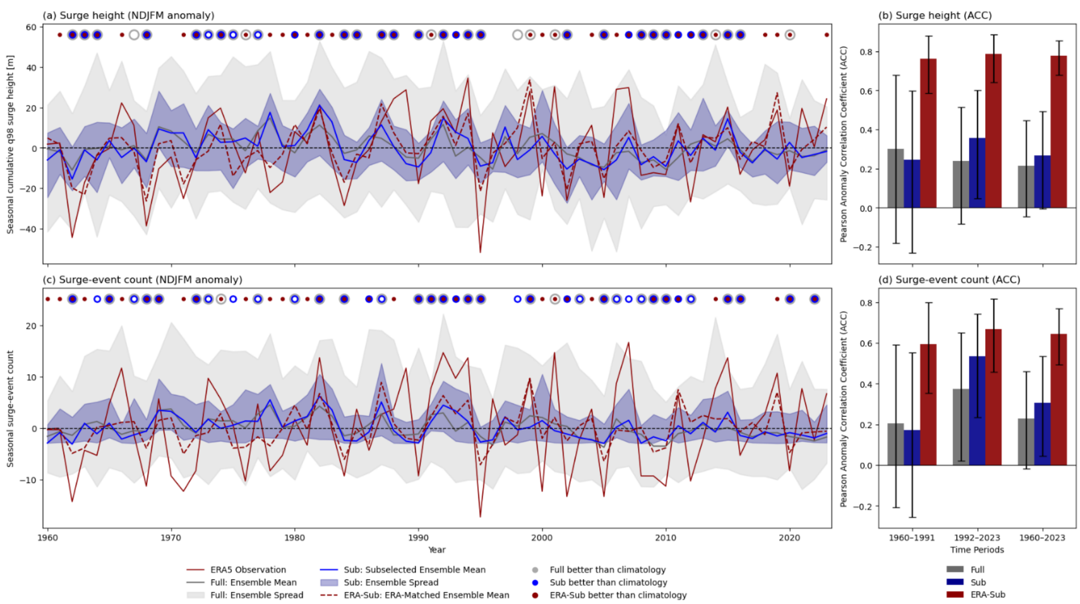

Figure 7. NDJFM seasonal storm surge anomalies and time period-dependent correlation for (a,b) cumulative q98 surge height and (b,d) surge-event count. Interannual variability in the two metrics is shown in a&c, with MPI-ESM-HR full ensemble (Full: grey; mean and spread), subselected ensemble (Sub: blue; mean and spread), and ERA-matched ensemble mean (ERA-Sub: red dashed), with ERA5 observations (red solid) for comparison. Markers at the top of each year (a,c) indicate whether a given forecast system outperforms a climatological reference in terms of absolute error, where filled markers indicate years in which the forecast error is smaller than the climatological error. The bar diagrams (b,d) report Pearson anomaly correlation coefficients (ACC) for 1960-1991, 1992-2023 and the full timeseries (1960-2023) with error bars indicating the 95% confidence interval obtained from a nonparametric bootstrap with independent resampling.

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

4 Discussion

This study demonstrates that seasonal storm surge predictability in the German Bight is closely linked to the representation and predictability of large-scale atmospheric circulation regimes. While the MPI-ESM-HR ensemble mean shows limited skill in predicting interannual variations of seasonal storm surge activity at Cuxhaven, the ensemble contains substantial predictive information, as demonstrated by the high skill of the best-performing members (Fig. 4). By identifying weather patterns associated with increased or reduced surge likelihood (Fig. 6) and applying weather pattern-informed ensemble subselection, we increase forecast skill, transforming the initially non-significant relationship between the ensemble mean and observed surge variability into statistically significant predictive skill (Fig. 7). These results indicate that seasonal storm surge predictability is not solely determined by the mean forecast state, but also depends on the ability of the forecast system to represent physically meaningful circulation regimes. Weather-pattern-based subselection therefore provides a physically interpretable approach for extracting circulation-dependent information from seasonal ensembles and translating it into improved storm surge predictions.

The seasonal forecast system represents the relevant characteristics of observed storm surge variability. MPI-ESM-HR captures the dominant seasonal cycle of storm surge activity, with both observations and simulations showing a clear winter maximum during November–March. The close agreement in the timing of enhanced surge activity confirms the suitability of this period as the main storm surge season and indicates that the model captures the large-scale atmospheric conditions responsible for enhanced wintertime storm activity in the North Sea.

The use of the 98th percentile threshold provides a physically meaningful definition of extreme surge conditions. In ERA5, the resulting threshold of approximately 1.5 m is consistent with the operational definition of storm surges used by the German Federal Maritime and Hydrographic Agency (Jensen and Müller-Navarra, 2008), supporting the relevance of this metric for identifying impact-relevant events. However, despite reproducing the overall seasonal distribution, MPI-ESM-HR systematically underestimates both the magnitude and frequency of extreme surges. The lower model-derived 98th percentile threshold, reduced seasonal event counts, and weaker anomaly amplitudes indicate a consistent underrepresentation of the upper tail of the surge distribution. Since these biases are present across the full ensemble and are not reduced through ensemble subselection, they likely reflect structural limitations rather than insufficient sampling.

Several factors may contribute to this systematic bias. Storm surges result from a complex chain involving large-scale circulation, cyclone intensity and trajectory, near-surface wind forcing, air pressure effects, and the nonlinear interaction with tides and coastal geometry (Jensen and Müller-Navarra, 2008). Any bias within this chain can influence the final coastal water-level response. In particular, insufficient representation of intense North Atlantic cyclones, storm-track variability, or wind extremes may reduce the amplitude of simulated surges. The coarse representation of coastal processes in Earth system models may further contribute to differences between atmospheric forcing and observed coastal water levels.

Nevertheless, the consistency between MPI-ESM-HR and ERA5 in reproducing the seasonal cycle and the circulation dependence of storm surge occurrence indicates that the fundamental dynamical mechanisms are represented. This distinction is important: the model exhibits a bias in the amplitude of extremes, but it does not appear to fundamentally misrepresent the atmospheric states associated with surge generation. Therefore, limitations in seasonal prediction skill are likely related less to an absence of a physical signal and more to the challenge of predicting the correct atmospheric circulation states at seasonal lead times.

Despite these limitations in representing the magnitude of extremes, **the seasonal forecast ensemble contains additional predictive information beyond that expressed by the ensemble mean.** Ensemble-mean forecasts show only weak, non-

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

significant correlations with observed interannual variability of seasonal storm surge activity, which could suggest that seasonal storm surge variability in the German Bight is largely unpredictable. However, the ensemble spread and best-member analysis reveal that some ensemble members reproduce the observed year-to-year variability much more closely, indicating that the forecast system contains predictive information that is not fully captured by the ensemble mean. This difference highlights the challenge of extracting predictable signals from seasonal forecast ensembles. Ensemble averaging improves reliability but can also reduce signals associated with specific atmospheric circulation states. This is particularly relevant for storm surges, which are strongly influenced by episodic, synoptic-scale weather patterns that can generate surge-favourable or surge-suppressing atmospheric forcing, rather than by gradual variations in the mean atmospheric state.
Previous studies have demonstrated that physically informed ensemble subselection can exploit such circulation-dependent information. Dalelane et al. (2020) showed that selecting ensemble members according to large-scale circulation modes improved seasonal forecast skill, while Heinrich-Mertsching et al. (2023) demonstrated additional skill from teleconnection-based subselection when the relevant physical pathway is active. Our results extend these findings to coastal storm surges, showing that conditioning the ensemble on circulation regimes strongly associated with surge occurrence can improve deterministic seasonal predictions. However, this additional deterministic skill does not necessarily translate into improved probabilistic forecasts. CRPS remains largely unchanged for seasonal cumulative surge height and even deteriorates for seasonal surge-event counts, indicating that the improvement in deterministic skill is not accompanied by a consistent improvement in the representation of the full forecast distribution. Further improvements in seasonal storm surge prediction will therefore require not only better prediction of large-scale circulation regimes but also improved representation of uncertainty in the translation from atmospheric forcing to coastal surge metrics.

The ability to extract additional predictive information through weather-pattern-based ensemble subselection is explained by the **strong dependence of storm surge occurrence on large-scale atmospheric circulation regimes**. This strong role of atmospheric forcing is consistent with Dangendorf et al. (2013) who found that meteorological forcing, particularly zonal wind stress, explains up to 90% of winter sea-level variability at Cuxhaven. Cyclonic Westerly (WZ) and North-Westerly (NWZ) patterns favour surge occurrence, while blocked regimes suppress activity, consistent with previous studies identifying cyclonic, north-westerly, and westerly circulation types as dominant drivers of North Sea storm surges (Gerber et al., 2016; Meyer and Gaslikova, 2024; Montreuil et al., 2016). The agreement between ERA5 and MPI-ESM-HR in representing these circulation–surge relationships is particularly important. Although MPI-ESM-HR underestimates the magnitude and frequency of extreme surges, it reproduces the dominant atmospheric regimes associated with surge occurrence.

Although weather-pattern-informed ensemble subselection significantly improves deterministic forecast skill, the substantially higher skill achieved by the ERA-matched ensemble indicates that **considerable predictive potential remains unrealised**. Because the ERA-matched ensemble assumes perfect knowledge of weather patterns, it provides an estimate of the upper bound of forecast skill attainable through ensemble subselection. The large gap between operational and ERA-matched subselection therefore suggests that the main limitation is not the weather pattern–storm surge relationship itself, but the ability of seasonal forecast systems to predict surge-relevant circulation regimes. This interpretation is consistent with previous work showing that the predictability of large-scale atmospheric circulation remains a major constraint on seasonal climate prediction (Robertson et al., 2020; Vitart and Robertson, 2018). Studies of European windstorms further demonstrate that seasonal forecast skill depends strongly on the ability of forecast systems to represent and predict large-scale MSLP circulation patterns (Degenhardt et al., 2024). Recent work further highlights that coastal hazard predictability emerges from interactions between large-scale climate variability and regional processes linking atmospheric circulation to coastal impacts (Boucharel et al., 2026). Similar circulation-based approaches have improved coastal hazard prediction at shorter lead times by linking atmospheric weather patterns to coastal impacts (Costa et al., 2020; Neal et al., 2018). Our results extend this concept to

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

seasonal forecasting by demonstrating that large-scale circulation regimes can be used to extract additional predictive information from climate model ensembles. Future improvements in seasonal storm surge forecasting are therefore likely to depend primarily on advances in predicting North Atlantic and European circulation rather than on modifications of the storm surge model itself. Preliminary analyses further indicate that predictive skill differs substantially among individual weather patterns, suggesting that some circulation regimes are inherently more predictable than others. Identifying these regimes and their associated sources of predictability represents an important direction for future research.

Forecast skill was generally higher during the more recent period (1992–2023), particularly for surge-event counts. Several factors may contribute to this temporal variation. One possible explanation is the improved quality and coverage of observations used to constrain atmospheric reanalyses and initialise seasonal forecasts. The expansion of satellite observations since the late twentieth century, together with advances in data assimilation, has substantially improved the representation of the large-scale atmospheric state in reanalysis products such as ERA5 (Hersbach et al., 2020). A more accurate initial conditions may therefore improve the representation and prediction of large-scale circulation features, including MSLP- and z500-based weather patterns, and thereby contribute to improved predictions of associated storm surge variability.
However, seasonal predictability is also influenced by variability in the predictability of large-scale circulation itself and possible changes in the occurrence or persistence of surge-relevant weather regimes. Distinguishing between these mechanisms is beyond the scope of this study but remains an important direction for future research.
Interestingly, the ERA-matched ensemble shows relatively stable skill across both time periods (Fig. 7), suggesting that the physical relationship between circulation regimes and storm surge variability remains robust over time. The temporal changes in operational forecast skill therefore appear to primarily reflect variations in the ability of the forecast system to predict relevant circulation states rather than changes in the underlying circulation–surge relationship.

The **variability in forecast performance between years** further suggests that storm surge predictability is conditional rather than uniform. While some seasons exhibit little forecast skill, others are reproduced remarkably well by both the circulation-informed and ERA-matched ensembles. Such periods may represent "windows of opportunity" during which enhanced predictability of the large-scale atmospheric circulation translates into increased predictability of coastal impacts (Mariotti et al., 2020). Recent studies similarly demonstrate that coastal flood predictability can emerge from interactions between large-scale climate modes and regional atmospheric conditions, highlighting that seasonal predictability is often conditional on the presence of specific large-scale circulation configurations rather than being constant throughout the year (Boucharel et al., 2026). Determining whether these high-skill periods are associated with particularly persistent circulation regimes, strong large-scale climate drivers such as the North Atlantic Oscillation, or unusually active storm seasons represents an important avenue for future research.

Predictability also differs between the two storm surge metrics. Seasonal cumulative q98 surge height generally achieves higher potential skill than surge-event counts, as demonstrated by the ERA-matched ensemble, likely because seasonal averaging of surge magnitude reduces sensitivity to the exact timing of individual storms and more clearly reflects large-scale circulation variability. In contrast, seasonal event counts depend on the number of individual threshold exceedances and are therefore more sensitive to small errors in the amplitude of daily surge events. Nevertheless, during the more recent period (1992–2023), surge-event counts achieve comparable or even higher correlation skill than surge height, indicating that the predictability of different storm surge metrics is itself time dependent. The smaller ensemble spread of surge-event counts should also not be interpreted as greater forecast confidence, as it largely reflects the bounded nature of this discrete variable rather than lower forecast uncertainty.

Several **limitations** remain. Seasonal forecasts cannot predict the timing and magnitude of individual storm surge events; instead, they provide information on whether atmospheric conditions are more or less favourable for enhanced surge activity. Individual extreme events may therefore be poorly represented by seasonal metrics if they occur within otherwise normal seasons. For example, exceptional surge events such as the February 1962 North Sea storm surge contribute only modestly to seasonal anomalies if the remainder of the season is near normal. Seasonal forecasts should therefore be interpreted as predictions of the background hazard state rather than forecasts of individual coastal disasters. Moreover, weather-pattern classifications represent only one component of the physical pathway leading to storm surges. Although they successfully capture the large-scale atmospheric background state, they do not explicitly account for storm intensity, cyclone track, effective wind forcing over the German Bight (Müller-Navarra and Giese, 1999), or tide–surge interactions, all of which influence the resulting coastal water levels (Jensen and Müller-Navarra, 2008; Weisse et al., 2012). This reflects a broader challenge in coastal hazard prediction, where extreme events often arise from the compound interaction of multiple physical components rather than from a single controlling factor. Recent work has therefore emphasized the need to identify and disentangle the individual contributions of atmospheric forcing, ocean dynamics, and coastal processes to improve understanding and risk management of extreme coastal events (Achterberg et al., 2025). Future circulation-based forecasting approaches may therefore benefit from incorporating additional physically relevant predictors, such as effective wind forcing and storm-track characteristics, while continuously reassessing the stability of circulation–impact relationships under changing climate conditions.

An additional limitation is that the circulation–surge relationships identified here are derived from the historical climate period and implicitly assume that these relationships remain sufficiently stationary over time. However, future sea-level rise and morphological changes in the Wadden Sea may modify coastal hydrodynamics and alter the translation of atmospheric forcing into coastal water levels, potentially affecting the transferability of historically derived relationships. In the German Bight, sea-level rise has been shown to cause nonlinear changes in extreme water levels, particularly in shallow Wadden Sea regions, through modifications of tidal propagation and shallow-water processes (Arns et al., 2015).

Moreover, future changes in storm activity may alter the frequency and characteristics of surge-relevant circulation states. CMIP6-based assessments suggest that future Northeast Atlantic and German Bight storm activity may involve changes in both circulation frequency and storm intensity, including an increased occurrence of westerly and north-westerly flow regimes but potentially reduced overall storm activity due to weaker wind speeds (Krieger and Weisse, 2026). While these effects are beyond the scope of this study, future coastal risk assessments will need to integrate changes in mean sea level, coastal morphology, storm activity, and atmospheric circulation to fully capture future changes in storm surge impacts. Future developments should therefore not only focus on improving the prediction of present-day circulation regimes but also on understanding how these regimes and their associated coastal impacts may change under future climate conditions.

5 Conclusions

This study demonstrates that large-scale atmospheric circulation regimes provide a key source of seasonal storm surge predictability in the German Bight. By exploiting circulation-dependent information within seasonal forecast ensembles, weather-pattern-based subselection improves deterministic predictions of seasonal storm surge activity beyond the ensemble mean. These results highlight the potential of circulation-based approaches to extend coastal hazard forecasting towards seasonal timescales, while also demonstrating that further progress depends on improving the prediction of surge-relevant atmospheric regimes and the representation of their translation into coastal impacts. Weather-pattern-based ensemble conditioning therefore provides a physically interpretable framework for extracting additional predictive information from seasonal forecast ensembles and represents a promising pathway towards improved seasonal forecasting of coastal storm surge hazards

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

Appendices

Table A1. Summary of forecast verification for the full MPI-ESM-HR ensemble (Full), the weather-pattern–based subselected ensemble (Sub), and the ERA-matched subselected ensemble (ERA-Sub), using ERA5 as the reference dataset for seasonal q98 surge height and seasonal surge-event count (NDJFM anomalies) comparing 1960-1991 to 1992-2023. Continuous skill metrics include the Anomaly Pearson Correlation Coefficient (ACC), the Root Mean Squared Error (RMSE) and the Continuous Ranked Probability Score (CRPS). Categorical (binary) skill metrics: Accuracy, Positive Predicted Value (PPV), Negative Predictive Value (NPV) and the True Skill Statistic (TSS). The 95% confidence interval obtained from a nonparametric bootstrap with independent resampling, in indicated in parenthesis.

Variable	Time Period	Predictor	ACC	RMSE	CRPS	Accuracy	PPV	NPV	TSS
Seasonal cumulative q98 surge height	1960-1991	Full	0.302 (-0.183, 0.677)	17.509 (13.371, 21.098)	10.214 (7.748, 12.854)	0.594 (0.438, 0.750)	0.556 (0.316, 0.786)	0.643 (0.375, 0.889)	0.196 (-0.150, 0.528)
		Sub	0.244 (-0.230, 0.599)	18.070 (13.719, 21.913)	11.812 (8.582, 15.293)	0.531 (0.375, 0.688)	0.500 (0.263, 0.733)	0.571 (0.300, 0.833)	0.071 (-0.283, 0.425)
		ERA-Sub	0.762 (0.586, 0.879)	11.868 (9.044, 14.719)	11.413 (8.284, 14.890)	0.750 (0.594, 0.875)	0.769 (0.500, 1.000)	0.737 (0.526, 0.923)	0.490 (0.174, 0.769)
	1992-2023	Full	0.239 (-0.084, 0.515)	20.315 (16.145, 24.428)	12.348 (9.499, 15.599)	0.531 (0.344, 0.719)	0.700 (0.375, 1.000)	0.455 (0.250, 0.667)	0.138 (-0.196, 0.455)
		Sub	0.356 (0.046, 0.602)	19.500 (15.472, 23.352)	13.521 (10.061, 17.335)	0.531 (0.344, 0.719)	0.700 (0.375, 1.000)	0.455 (0.250, 0.667)	0.138 (-0.193, 0.455)
		ERA-Sub	0.786 (0.640, 0.887)	13.412 (10.217, 16.454)	12.951 (9.600, 16.689)	0.844 (0.719, 0.969)	0.850 (0.682, 1.000)	0.833 (0.583, 1.000)	0.664 (0.365, 0.917)
Seasonal surge-event count	1960-1991	Full	0.205 (-0.207, 0.590)	7.590 (6.085, 8.861)	4.540 (3.489, 5.605)	0.562 (0.375, 0.750)	0.500 (0.280, 0.727)	0.667 (0.375, 0.917)	0.159 (-0.200, 0.500)
		Sub	0.172 (-0.255, 0.552)	7.672 (6.097, 9.093)	5.550 (4.189, 6.972)	0.594 (0.438, 0.781)	0.529 (0.286, 0.765)	0.667 (0.412, 0.909)	0.198 (-0.153, 0.550)
		ERA-Sub	0.595 (0.353, 0.800)	6.213 (5.027, 7.372)	5.199 (3.918, 6.511)	0.719 (0.562, 0.875)	0.692 (0.429, 0.929)	0.737 (0.526, 0.929)	0.421 (0.095, 0.733)
	1992-2023	Full	0.373 (0.022, 0.649)	9.381 (7.914, 10.738)	6.123 (4.899, 7.375)	0.531 (0.375, 0.719)	0.545 (0.222, 0.857)	0.524 (0.308, 0.737)	0.062 (-0.264, 0.405)
		Sub	0.535 (0.234, 0.741)	8.984 (7.688, 10.245)	7.400 (6.018, 8.819)	0.656 (0.500, 0.812)	0.727 (0.429, 1.000)	0.619 (0.400, 0.833)	0.312 (-0.020, 0.627)
		ERA-Sub	0.668 (0.456, 0.817)	7.735 (6.336, 9.112)	7.000 (5.583, 8.461)	0.594 (0.438, 0.750)	0.588 (0.353, 0.812)	0.600 (0.333, 0.850)	0.188 (-0.167, 0.517)

Data availability

ERA5 reanalysis products that were used to support this study are available from the Copernicus Data Store:

MSLP: Copernicus Climate Change Service (2025): ERA5 hourly time-series data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). URL: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels-timeseries

z500: Copernicus Climate Change Service, Climate Data Store, (2023): ERA5 hourly data on pressure levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.bd0915c6

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

The retrospective forecasts from the Max Planck Institute Earth System Model in high resolution (MPI-ESM-HR) can be accessed via:

Krieger, Daniel; Baehr, Johanna; Borchert, Leonard; Brune, Sebastian; Sylla, Adama; Uddin, Mohammad Basir (2024). *MPI-ESM-HR 1.2.02 extended seasonal prediction hourly atmospheric surface and daily oceanic output*. DOKU at DKRZ. https://hdl.handle.net/21.14106/4e9ddccad5e9165aa3927b6cf96116eee3c3dbe7

Residual hourly surge heights at Cuxhaven are available from:

Krieger, Daniel (2024). Residual Surge Heights at Cuxhaven (1940-2021), Esbjerg (1950-2019), and Delfzijl (1971-2017) DOI: 10.5281/zenodo.14169268. https://zenodo.org/api/records/14169269/files-archive

The Convolutional Neural Network for Meteorological Pattern Recognition (CNMN) used for classifying weather patterns was obtained from:

Heinrich, Philipp (2024). Convolutional Neural Meteorology Network (CNMN) (Version v3). Zenodo. https://doi.org/10.5281/zenodo.10959809

Author contributions

AKM, DK and LB designed the presented study. DK provided the Neural Network for translating mean sea level pressure anomalies into surge heights. AKM analysed the data and created the figures. All authors discussed the results. AKM took the lead in writing the manuscript with input from all co-authors.

Competing interests

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

Acknowledgements

We thank Sebastian Brune for running and providing us with the MPI-ESM-HR simulations. We also thank the German Climate Computing Center (Deutsches Klimarechenzentrum, DKRZ) for providing computing resources. ChatGPT (OpenAI) was used during manuscript preparation to assist with language editing, restructuring, and improving the clarity and conciseness of the text. It was also used to support coding assistance and troubleshooting. All scientific interpretations, methodological decisions, analyses, and final text were reviewed and verified by the authors, who remain fully responsible for the content of the manuscript.

Financial support

AKM, DK and LB were supported by the Federal Ministry of Research, Technology and Space (BMFTR) through the project METAscales (FKZ 03F0955J) within the German Marine Research Alliance (DAM) mission mareXtreme.

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

References

Achterberg, E. P., Rackelmann, F., Ramesh, A., O'Connor, J., García-Oliva, O., Jayachandran, S., Wichert, V., Li, X., Purkiani, K., Needham, D. M., Bussmann, I., Fischer, P., Kopf, A., Lemmen, C., Scharsack, J. P., Siebert, C., Kamjunke, N., Sokolova, I. M., and Brix, H.: Identifying and addressing the components of extreme physical-oceanographical events for improved risk management in coastal systems, Front. Mar. Sci., 12, 1681652, https://doi.org/10.3389/fmars.2025.1681652, 2025.

Apolola, A., Ward, P. J., Tiggeloven, T., Á. Antolínez, J. A., Jäger, W., and Muis, S.: Global Analysis of Storm Surge Seasonality, J. Geophys. Res. Oceans, 130, e2025JC022841, https://doi.org/10.1029/2025JC022841, 2025.

Arns, A., Wahl, T., Dangendorf, S., and Jensen, J.: The impact of sea level rise on storm surge water levels in the northern part of the German Bight, Coast. Eng., 96, 118–131, https://doi.org/10.1016/j.coastaleng.2014.12.002, 2015.

Boucharel, J., Almar, R., Jin, F.-F., Zhao, S., Stuecker, M. F., and Dewitte, B.: Climate mode interactions amplify coastal flood risks and their seasonal predictability, Nat. Geosci., 19, 317–324, https://doi.org/10.1038/s41561-025-01903-0, 2026.

Brune, S. and Baehr, J.: Preserving the coupled atmosphere–ocean feedback in initializations of decadal climate predictions, WIREs Clim. Change, 11, 1–19, https://doi.org/10.1002/wcc.637, 2020.

Brune, S., Nerger, L., and Baehr, J.: Assimilation of oceanic observations in a global coupled Earth system model with the SEIK filter, Ocean Model., 96, 254–264, https://doi.org/10.1016/j.ocemod.2015.09.011, 2015.

Caldwell, P. C., Merrifield, M. A., and Thompson, P. R.: Sea level measured by tide gauges from global oceans — the Joint Archive for Sea Level holdings (NCEI Accession 0019568) (5.5), https://doi.org/10.7289/V5V40S7W, 2015.

Chatzistratis, D., Velegrakis, A. F., Chalazas, T., Alves, B., Schiavon, E., Monioudi, I. N., and Armaroli, C.: European Early Warning Systems for coastal floods: User needs, system availability and pertinent policy and legislation, Int. J. Disaster Risk Reduct., 126, 105602, https://doi.org/10.1016/j.ijdrr.2025.105602, 2025.

Codiga, D. L.: Unified tidal analysis and prediction using the UTide Matlab functions, (Tech. Rep. No. URI/GSO Technical Report 201101), Graduate School of Oceanography, University of Rhode Island, www.po.gso.uri.edu/pub/downloads/codiga/pubs/2011Codiga-UTide-Report.pdf, 2011.

Costa, W., Idier, D., Rohmer, J., Menendez, M., and Camus, P.: Statistical Prediction of Extreme Storm Surges Based on a Fully Supervised Weather-Type Downscaling Model, J. Mar. Sci. Eng., 8, 1028, https://doi.org/10.3390/jmse8121028, 2020.

Dalelane, C., Dobrynin, M., and Fröhlich, K.: Seasonal Forecasts of Winter Temperature Improved by Higher-Order Modes of Mean Sea Level Pressure Variability in the North Atlantic Sector, Geophys. Res. Lett., 47, e2020GL088717, https://doi.org/10.1029/2020GL088717, 2020.

Dangendorf, S., Mudersbach, C., Wahl, T., and Jensen, J.: Characteristics of intra-, inter-annual and decadal sea-level variability and the role of meteorological forcing: the long record of Cuxhaven, Ocean Dyn., 63, 209–224, https://doi.org/10.1007/s10236-013-0598-0, 2013.

Degenhardt, L., Leckebusch, G. C., and Scaife, A. A.: Understanding winter windstorm predictability over Europe, Weather Clim. Dyn., 5, 587–607, https://doi.org/10.5194/wcd-5-587-2024, 2024.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fröhlich, K., Dobrynin, M., Isensee, K., Gessner, C., Paxian, A., Pohlmann, H., Haak, H., Brune, S., Früh, B., and Baehr, J.: The German Climate Forecast System: GCFS, J. Adv. Model. Earth Syst., 13, e2020MS002101, https://doi.org/10.1029/2020MS002101, 2021.

Gerber, M., Ganske, A., Müller-Navarra, S., and Rosenhagen, G.: Categorisation of Meteorological Conditions for Storm Tide Episodes in the German Bight, Meteorol. Z., 25, 447–462, https://doi.org/10.1127/metz/2016/0660, 2016.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res. Oceans, 118, 6704–6716, https://doi.org/10.1002/2013JC009067, 2013.

Haigh, I. D., Marcos, M., Talke, S. A., Woodworth, P. L., Hunter, J. R., Hague, B. S., Arns, A., Bradshaw, E., and Thompson, P.: GESLA Version 3: A major update to the global higher-frequency sea-level dataset, Geosci. Data J., 10, 293–314, https://doi.org/10.1002/gdj3.174, 2023.

Harter, L., Pineau-Guillou, L., and Chapron, B.: Underestimation of extremes in sea level surge reconstruction, Sci. Rep., 14, 14875, https://doi.org/10.1038/s41598-024-65718-6, 2024.

Heinrich, P.: Convolutional Neural Meteorology Network (CNMN). (Version v3), Zenodo [code], https://doi.org/10.5281/zenodo.10959809, 2024.

Heinrich, P., Hagemann, S., Weisse, R., Schrum, C., Daewel, U., and Gaslikova, L.: Compound flood events: analysing the joint occurrence of extreme river discharge events and storm surges in northern and central Europe, Nat. Hazards Earth Syst. Sci., 23, 1967–1985, https://doi.org/10.5194/nhess-23-1967-2023, 2023.

Heinrich, P., Hagemann, S., and Weisse, R.: Automated classification of atmospheric circulation types for compound flood risk assessment: CMIP6 model analysis utilising a deep learning ensemble, Environ. Res. Lett., 20, 074018, https://doi.org/10.1088/1748-9326/adddcb, 2025.

Heinrich-Mertsching, C., Sørland, S. L., Gudoshava, M., Koech, E., Bahaga, T. K., and Sobolowski, S. P.: Subselection of seasonal ensemble precipitation predictions for East Africa, Q. J. R. Meteorol. Soc., 149, 2634–2653, https://doi.org/10.1002/qj.4525, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., De Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hess, P. and Brezowsky, H.: Katalog der Großwetterlagen Europas, Deutscher Wetterdienst, 1969.

Jensen, J. and Müller-Navarra, S. H.: Storm Surges on the German Coast, 74 ICCE, 92–124, 2008.

Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification: A Practitioner's Guide in Atmospheric Science, edited by: Jolliffe, I. T. and Stephenson, D. B., Wiley-Blackwell, West Sussex, 274 pp., 2012.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and Storch, J. S.: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model, J. Adv. Model. Earth Syst., 5, 422–446, https://doi.org/10.1002/jame.20023, 2013.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, https://doi.org/10.48550/arXiv.1412.6980, 30 January 2017., 2017.

Krieger, D. and Weisse, R.: CMIP6 multi-model assessment of Northeast Atlantic and German Bight storm activity, Earth Syst. Dyn., 17, 1–21, https://doi.org/10.5194/esd-17-1-2026, 2026.

Krieger, D., Brune, S., Baehr, J., and Weisse, R.: Improving seasonal predictions of German Bight storm activity, Nat. Hazards Earth Syst. Sci., 24, 1539–1554, https://doi.org/10.5194/nhess-24-1539-2024, 2024.

Krieger, D., Weisse, R., Baehr, J., and Borchert, L. F.: Machine Learning-Driven Skillful Decadal Predictions of German Bight Storm Surges, Geophys. Res. Lett., 52, e2024GL111558, https://doi.org/10.1029/2024GL111558, 2025.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M. A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., and Pedersen, L. T.: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, The Cryosphere, 13, 49–78, https://doi.org/10.5194/tc-13-49-2019, 2019.

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., Dirmeyer, P. A., Ferranti, L., Johnson, N. C., Jones, J., Kirtman, B. P., Lang, A. L., Molod, A., Newman, M., Robertson, A. W., Schubert, S., Waliser, D. E., and Albers, J.: Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond, Bull. Am. Meteorol. Soc., 101, E608–E625, https://doi.org/10.1175/BAMS-D-18-0326.1, 2020.

Meyer, E. M. I. and Gaslikova, L.: Investigation of historical severe storms and storm tides in the German Bight with century reanalysis data, Nat. Hazards Earth Syst. Sci., 24, 481–499, https://doi.org/10.5194/nhess-24-481-2024, 2024.

https://doi.org/10.5194/egusphere-2026-5065
Preprint. Discussion started: 2 September 2026

Montreuil, A.-L., Elyahyioui, J., and Chen, M.: Effect of Large-Scale Atmospheric Circulation and Wind on Storm Surge Occurrence, J. Coast. Res., 75, 755–759, https://doi.org/10.2112/SI75-152.1, 2016.

Müller-Navarra, S. H. and Giese, H.: Improvements of an empirical model to forecast wind surge in the German Bight, Dtsch. Hydrogr. Z., 51, 385–405, https://doi.org/10.1007/BF02764162, 1999.

Neal, R., Dankers, R., Saulter, A., Lane, A., Millard, J., Robbins, G., and Price, D.: Use of probabilistic medium- to long-range weather-pattern forecasts for identifying periods with an increased likelihood of coastal flooding around the UK, Meteorol. Appl., 25, 534–547, https://doi.org/10.1002/met.1719, 2018.

Nerger, L. and Hiller, W.: Software for ensemble-based data assimilation systems—Implementation strategies and scalability, Comput. Geosci., 55, 110–118, https://doi.org/10.1016/j.cageo.2012.03.026, 2013.

Robertson, A. W., Vitart, F., and Camargo, S. J.: Subseasonal to Seasonal Prediction of Weather to Climate with Application to Tropical Cyclones, J. Geophys. Res. Atmos., 125, e2018JD029375, https://doi.org/10.1029/2018JD029375, 2020.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.: Atmospheric component of the MPI-M Earth System Model: ECHAM6, J. Adv. Model. Earth Syst., 5, 146–172, https://doi.org/10.1002/jame.20015, 2013.

Sylla, A., Brune, S., Losada, T., and Baehr, J.: Mechanisms and seasonal predictability of sea surface temperature in the Senegalo-Mauritanian Upwelling System, Clim. Dyn., 64, 159, https://doi.org/10.1007/s00382-026-08121-3, 2026.

Tiggeloven, T., Couasnon, A., Van Straaten, C., Muis, S., and Ward, P. J.: Exploring deep learning capabilities for surge predictions in coastal areas, Sci. Rep., 11, 17224, https://doi.org/10.1038/s41598-021-96674-0, 2021.

Vitart, F. and Robertson, A. W.: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events, Npj Clim. Atmos. Sci., 1, 3, https://doi.org/10.1038/s41612-018-0013-0, 2018.

Wani, M. A., Bhat, F. A., Afzal, S., and Khan, A. I.: Advances in Deep Learning, Springer Singapore, Singapore, https://doi.org/10.1007/978-981-13-6794-6, 2020.

Weisse, R., Von Storch, H., Niemeyer, H. D., and Knaack, H.: Changing North Sea storm surge climate: An increasing hazard?, Ocean Coast. Manag., 68, 58–68, https://doi.org/10.1016/j.ocecoaman.2011.09.005, 2012.

Wilks, D. S.: Forecast Verification, in: International Geophysics, vol. 100, Elsevier, 301–394, https://doi.org/10.1016/B978-0-12-385022-5.00008-7, 2011.

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