# Pooling Seasonal Forecast Ensembles to Estimate Storm Tide Return Periods in Extra-Tropical Regions

Irene Benito Lazaro[1], Dirk Eilander[1], Timo Kelder[2], Philip J. Ward[1], Jeroen Aerts[3], and Sanne Muis[4]

[1]Institute for Environmental Studies, Vrije Universiteit Amsterdam
[2]Climate Adaptation Services
[3]Vrije Universiteit
[4]Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam

April 01, 2025

### Abstract

Low-pressure systems and strong winds can generate severe storm tides, leading to coastal flooding and significant economic losses. Accurate estimates of storm tide frequency and intensity are crucial for flood hazard assessments and risk reduction. However, the limited observational records pose a challenge in estimating high return periods with low uncertainty. In this study, we evaluate the potential of pooling ensembles from the SEAS5 seasonal forecast archive to generate an extensive storm tide dataset for robust return period estimates in extra-tropical regions at large spatial scale. Using SEAS5 to force the hydrodynamic model GTSM, we generate 525 synthetic years of storm tides and apply extreme value analysis to estimate 40-year and 500-year return periods. Our findings demonstrate that SEAS5 produces unbiased and independent synthetic mean sea level pressure events across major extra-tropical regions, including Europe, China, Russia, South America and Australia. In Europe, unbiased SEAS5-derived storm tide extremes along the Atlantic coast are particularly well-suited for return period analysis. The results show the benefits of using longer records to improve extreme return periods. SEAS5 not only reduces uncertainties in high return period estimates but also provides more extreme events, enhancing the reliability of extreme value distributions compared to short observational records.

### Hosted file

`Irene_Benito_Paper3_v5.docx` available at https://authorea.com/users/903251/articles/1278266-pooling-seasonal-forecast-ensembles-to-estimate-storm-tide-return-periods-in-extra-tropical-regions

# Pooling Seasonal Forecast Ensembles to Estimate Storm Tide Return Periods in Extra-Tropical Regions

**Irene Benito[1], Dirk Eilander[1,2], Timo Kelder[1,3], Philip J. Ward[1,2], Jeroen C.J.H. Aerts[1,2] and Sanne Muis[1,2]**

[1]Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam, The Netherlands

[2]Deltares, Delft, The Netherlands.

[3]Climate Adaptation Services Foundation (CAS), The Netherlands

Corresponding author: Irene Benito (i.benito.lazaro@vu.nl)

**Key points**

- Using ensemble pooling of a seasonal forecast archive with a hydrodynamic model to extend the sample size of storm tide extremes
- The ECMWF seasonal forecast archive can generate unbiased, independent storm tide events in several extra-tropical regions
- The 525 synthetic years of storm tides generated reduce uncertainty in extreme return periods at the European scale

## Abstract

Low-pressure systems and strong winds can generate severe storm tides, leading to coastal flooding and significant economic losses. Accurate estimates of storm tide frequency and intensity are crucial for flood hazard assessments and risk reduction. However, the limited observational records pose a challenge in estimating high return periods with low uncertainty. In this study, we evaluate the potential of pooling ensembles from the SEAS5 seasonal forecast archive to generate an extensive storm tide dataset for robust return period estimates in extra-tropical regions at large spatial scale. Using SEAS5 to force the hydrodynamic model GTSM, we generate 525 synthetic years of storm tides and apply extreme value analysis to estimate 40-year and 500-year return periods. Our findings demonstrate that SEAS5 produces unbiased and independent synthetic mean sea level pressure events across major extra-tropical regions, including Europe, China, Russia, South America and Australia. In Europe, unbiased SEAS5-derived storm tide extremes along the Atlantic coast are particularly well-suited for return period analysis. The results show the benefits of using longer records to improve extreme return periods. SEAS5 not only reduces uncertainties in high return period estimates but also provides more extreme events, enhancing the reliability of extreme value distributions compared to short observational records.



## Plain Language Summary

During storms, extreme sea levels can be caused by low pressure and strong winds. Understanding how often these extreme sea levels occur and how intense they are is crucial for effective coastal flood risk management. However, the probabilities of extreme sea levels beyond the available observational data are highly uncertain. In this study, we present an approach to reduce that uncertainty. Seasonal forecasts, which predict the weather and climate few months ahead, are typically used  in agriculture, water resources and energy demand applications. In our work, we explore a new way to use these forecasts by combining them with a model to simulate sea levels. We pool together multiple seasonal forecast simulations, allowing us to generate a larger dataset of plausible extreme sea levels, including events that may not have occurred yet. With more data, we can better estimate the probabilities of extreme sea levels, leading to improved coastal flood risk assessments and more effective adaptation and mitigation strategies.


## 1 Introduction

Extreme sea levels, driven by storm tides, can cause severe coastal flooding, economic losses, and threats to life, particularly in low-lying areas. Historical high-impact storm tide events include Tropical Cyclone Harvey, which caused 125 billion U.S. dollars in damages in 2017 (Sebastian et al., 2021), storm Xynthia, which hit France in 2010 resulting in 2.5 billion euros in damages (CGEDD,

2010) and the “Pasha Bulker” storm, which hit the east coast of Australia in 2007, causing 10 deaths and 1.4 billion Australian dollars in damages (Dowdy et al., 2013). Reliable estimates of storm tide frequency and intensity are essential for developing flood hazard assessments that help implement risk reduction and adaptation measures.

Large-scale coastal flood risk assessments typically rely on extreme value analysis (EVA) of observed or modelled extreme sea levels to estimate return periods beyond the observational record (Dullaart et al., 2021; Menéndez & Woodworth, 2010; Wahl et al., 2017). Storm tide observations are usually obtained from tide gauge records (Haigh et al., 2023). However, these records vary significantly in spatial coverage, are often short in duration, and may fail to record extreme storm events. To address these limitations, storm tide datasets can be derived from models driven by meteorological inputs from reanalysis data (Muis et al., 2016; Rose et al., 2024). While these datasets offer more globally consistent sea level information, reliable global observational records of key meteorological inputs (i.e. wind and pressure), validated and assimilated with satellite data, extend only about 40 years (Hersbach et al., 2019).

Applying EVA methods to such short records and extrapolating beyond observational data introduces significant uncertainties, particularly for extreme return periods (van den Brink et al., 2004). An alternative approach to traditional observation-based EVA involves using synthetic datasets, which expand the sample size and allow for more robust statistical analysis. Several methods exist to generate synthetic datasets. Fully statistical approaches, such as copulas, max-stable models, or the conditional multivariate exceedance model by Heffernan & Tawn (2004), create synthetic events based on observed or modelled sea levels (Li et al., 2023; Rashid et al., 2024). These approaches are computationally efficient and have evolved to incorporate spatiotemporal correlations. While they can expand the sample size of extreme events, they still rely on extreme value techniques, meaning they do not fully resolve the uncertainty in estimating extremes beyond observations.

Hybrid approaches use statistical methods to derive synthetic meteorological datasets, which are then used to force hydrodynamic models (e.g. Lin & Chavas, 2012). These approaches have been successfully applied in previous studies, where global synthetic datasets for tropical cyclones (Bloemendaal et al., 2020; Emanuel et al., 2006; James & Mason, 2005; Lee et al., 2018) have been used to drive storm tide models (Benito et al., 2024; Dullaart et al., 2021; Haigh et al., 2014; Marsooli et al., 2019). However, no equivalent synthetic meteorological dataset currently exists for simulating synthetic storm tides in extra-tropical regions. As a result, global estimates of return periods for storm tide levels in these regions rely on approximately 40 years of reanalysis data (Hersbach et al., 2019), leading to substantial uncertainties for extreme return periods.

Physically-based approaches offer an alternative for generating synthetic meteorological forcing datasets in extra-tropical regions. These approaches use large-ensemble climate models to drive hydrodynamic models. Although more computationally expensive than fully statistical approaches, these methods can capture the physical principles of atmospheric processes and storm tide dynamics. Howard & Williams (2021), for example, used a 483-year present-day simulation from the HadGEM3-GC3-MM climate model to generate synthetic storm surges for the UK. Other studies used pooling techniques across climate models (Meucci et al., 2020) or model ensembles to construct synthetic datasets (Breivik et al., 2014; Thompson et al., 2017). One example of such a method is the UNSEEN (UNprecedented Simulated Extremes using Ensembles) approach (Thompson et al., 2017), which has been applied, among others, to estimate the likelihood of extreme precipitation in the UK (Thompson et al., 2017) and fluvial floods in Europe (Brunner & Slater, 2022).

Large ensembles of climate variables are typically derived from climate models (Howard & Williams, 2021; Meucci et al., 2020; Thompson et al., 2017), though they often present significant biases (Eyring, Righi, et al., 2016; C. Wang et al., 2014). An alternative physics-based approach for generating synthetic extreme events is through the use of seasonal forecast archives. These forecasts simulate realistic atmospheric phenomena while capturing weather variability with their various ensemble members. One such archive is the fifth-generation European Centre for Medium-range Weather Forecasts (ECMWF)'s SEAS5, which provides 7-month forecasts initialised every month (ECMWF, 2021). Traditionally, seasonal forecasts have been used to predict weather and climate patterns over several months, supporting applications in agriculture, water resources and energy demand, among others (Dessai & Bruno Soares, 2013). However, their limited predictive skill beyond certain lead times makes their archived simulations useful for generating independent and unbiased event sets. As a result, seasonal forecast archives are well-suited for generating synthetic events that expand the sample size of extreme events (Kelder et al., 2020; Kelder, Marjoribanks, et al., 2022). Additionally, the spatiotemporal resolution of seasonal forecasts allows them to capture synoptic-scale storms, which typically drive extreme sea levels in extra-tropical regions (ECMWF, 2021). This capability makes them suitable to drive storm tide simulations.

The UNSEEN approach has been successfully applied together with SEAS5 to estimate the probabilities of extreme precipitation events (Kelder et al., 2020; Kelder, Marjoribanks, et al., 2022). In the context of extreme sea levels, Van den Brink et al. (2004) used ensemble pooling techniques on an earlier version of the ECMWF's seasonal forecast archive to reduce statistical uncertainty of 10,000-year surge level estimates for a coastal location in the Netherlands.

Our study investigates the potential of using seasonal forecast archives for generating extended storm tide events sets, enabling more robust large-scale return period estimates in extra-tropical regions. First, we evaluate if SEAS5 can be used for generating independent and unbiased events by validating the mean sea level pressure against ERA5. Results show that SEAS5 meets the bias and independence criteria in Europe, making it the focus of our further analysis. Second, we simulate 525 years of storm tides for Europe by forcing a regional cut-out of the Global Tide and Surge Model (GTSM) with SEAS5's re-forecast wind and pressure fields. We then evaluate biases in the simulated storm tides by comparing against ERA5 simulations with the same model setup. Third, we conduct a statistical analysis to compare storm tide return periods and uncertainties between SEAS5's 525 synthetic years and ERA5's ~40 years.

# 2 Methods

## 2.1 General approach

Figure 1 provides an overview of the methodological framework of this study, which consists of three main steps. First, in Section 2.2, we assess if SEAS5 can be used for generating unbiased and independent extreme events (Figure 1, red panel). This global-scale assessment includes a test to evaluate SEAS5's ability to represent unbiased mean sea level pressures (Section 2.2.1) and an independence test to ensure that the number of extreme events is not influenced by an initial climate state, which could introduce biases in extreme value estimates (Section 0). These tests are applied to both the re-forecast period (1981 – 2016) and the forecast period (2017 – 2023) of SEAS5, and help us identify regions and time frames where SEAS5 can be used for the generation of reliable synthetic events. Based on these findings, we focus the remainder of the analysis in Europe, where SEAS5 performs particularly well (Figure 1 blue panels).

Second, we perform storm tide modelling for Europe using 10 m wind components and mean sea level pressure fields from SEAS5's re-forecast and ERA5 to force a regional cut-out model of GTSM (Section 2.3). Third, we construct 525 synthetic years by pairing SEAS5 storm tide time series per ensemble member. We evaluate biases in the timeseries and apply two EVA approaches to compare the robustness of high-return-period storm tide estimates derived from SEAS5 and ERA5 (Section 2.4).

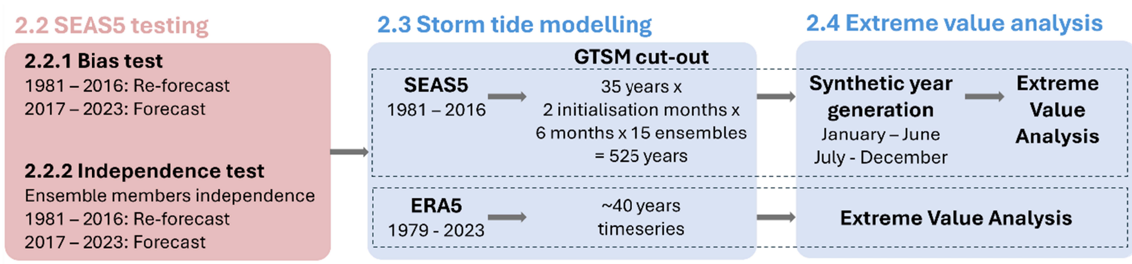



**Figure 1.** Workflow of the analysis carried out in this paper, structured in 3 parts: 2.2 SEAS5 testing at global scale (red panel), and at European scale (blue panels) 2.3 storm tide modelling and 2.4 extreme value analysis.

## 2.2 Evaluation of SEAS5 and its ability provide unbiased and independent events

We use the archive of the seasonal forecasting system SEAS5 to obtain wind and pressure meteorological forcing data necessary for simulating storm tides. SEAS5 is a global coupled model released by ECMWF in 2017, which integrates ocean, sea-ice and atmospheric components. Its atmospheric component is based on cycle 43r1 of the ECMWF Integrated Forecast System. ECMWF provides re-forecast datasets (also known as hindcasts) covering a historical period used for calibrating the forecasting system. These re-forecasts span 1981 – 2016 and include 25 ensemble members. The forecast period extends from 2017 to present and includes 51 ensemble members. Both re-forecasts and forecasts are initialised monthly, have a 7-month length, and are provided at a 6-hourly temporal resolution with a 36 km horizontal resolution (ECMWF, 2021).

While both wind speed and sea level pressure are used for storm tide modelling (Section 2.3), we focus on sea level pressure for the bias and independence tests (Section 2.2.1 and 0) because it is more stable than wind speeds. Specifically, we use the monthly minimum mean sea level pressure from SEAS5, obtained from the Climate Data Store (CDS). For the re-forecast period, we use SEAS5 data from 1981 to 2016, and for the forecast period we use data from 2017 to 2023. To reduce computational costs, this analysis is carried out at a 1-degree horizontal resolution. Although SEAS5 has a time span of 7 months, the CDS provides monthly statistics from lead times 1 to 6. Therefore, the 7-month lead time is excluded from this analysis.

### *2.2.1 Bias test*

To verify that the SEAS5 (re-)forecast datasets can provide unbiased extreme events, we conduct a bias test based on the UNSEEN approach (Kelder et al., 2020; Kelder, Marjoribanks, et al., 2022; Thompson et al., 2017). This test evaluates how well the SEAS5 forecasts represent observed climate by comparing them with ERA5 reanalysis, which we treat as the “ground truth” in this study, also obtained from CDS for the same time frames as SEAS5. To detect potential biases in SEAS5, we perform a permutation test on the re-forecast and forecast data by comparing their means with those of ERA5. At each grid cell, SEAS5 data are pooled across all lead times and ensemble members. Since the pooled SEAS5 dataset is considerably larger than the ERA5 dataset, a direct comparison would provide skewed results. To address this, we apply a bootstrapping method and resample the SEAS5 data without replacement to generate samples that match the size of ERA5 for each permutation. The permutation test provides p-values for each grid cell, allowing us to identify grid cells where the differences between the means of SEAS5 and ERA5 are statistically significant. As suggested by Wilks (2016), we apply a false discovery rate correction using the Benjamini-Hochberg procedure, ranking the p-values and applying a significance threshold of $\alpha < 0.1$. We do this permutation test for the full year and separately for each season to assess the biases more specifically in the storm seasons.

### *2.2.2 Independence test*

To test if the SEAS5 ensemble members are independent, we conduct an independence test based on Kelder et al. (2020, 2022). At the global scale, for each grid cell and each lead time, we use SEAS5 data to construct a new timeseries. These timeseries are created by concatenating monthly minimum mean sea level pressures for each month and year per ensemble member (Figure 2 panel a).

After constructing these new timeseries, we de-seasonalise the minimum monthly pressures by subtracting for each month the mean of the monthly minimum pressures. We then calculate pairwise Spearman correlations along ensemble members (Figure 2 panel b). For each grid cell and lead time, this results in a correlation matrix with 300 pairwise correlations among the 25 ensemble members in the re-forecast period, and 1,275 pairwise correlations among the 51 ensemble members in the forecast period. From these pairwise correlations, we calculate the median values of the Spearman correlations for each grid cell and lead time (Figure 2 panel c). These values are then combined to generate global gridded maps, showing the median Spearman correlations for each lead time.

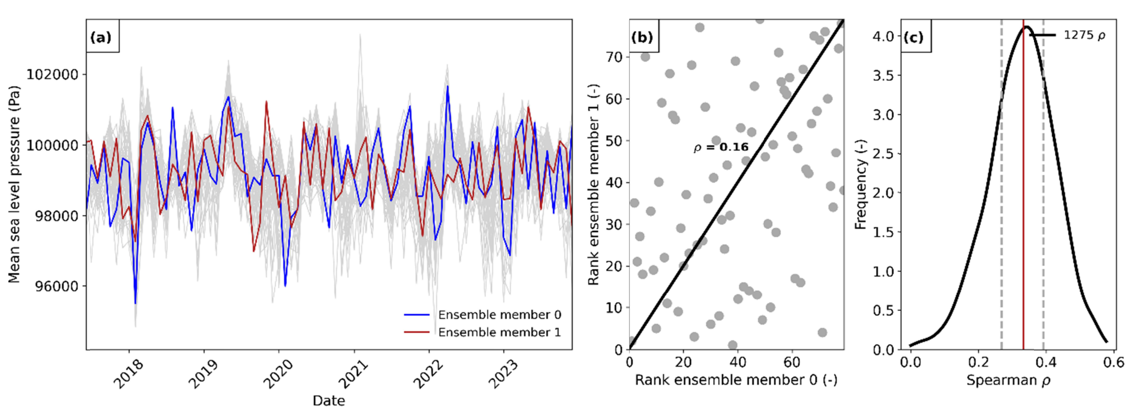


**Figure 2.** Visualisation of the workflow to calculate independence: (a) For a specific lead time and grid cell, monthly minimum pressure timeseries for the 51-member seasonal forecasts between January 2017 and December 2023. (b) For a specific lead time and grid cell, Spearman's ranks of the monthly minimum pressure for ensemble member 0 and member 1. (c) For a specific lead time and grid cell, the black line shows the probability density function of the 1275 Spearman correlations, the red line shows the median of the distribution.

## 2.3 Storm tides modelling

We simulate the storm tides in Europe using a regional cut-out of the calibrated depth-averaged hydrodynamic model Global Tide and Surge Model Version 4.1 (GTSMv4.1), which is based on Delft3D Flexible Mesh (Kernkamp et al., 2011; Muis et al., 2016; X. Wang et al., 2022). GTSMv4.1 has a variable spatial resolution, going from 25 km in the deep ocean to 2.5 km along the coasts (1.25 km in Europe). It uses bathymetric data from EMODNET for Europe (Consortium EMODnet Bathymetry, 2018), Bedmap2 (Fretwell et al., 2013) for the Arctic, and General Bathymetry Chart of the Ocean (GEBCO) 2019 (GEBCO, 2014) for the rest of the globe together with some local datasets. The regional model derived from GTSM covers a domain between latitudes 30.2N to 71.5N and longitudes 12.7W to 42.9E.

For the tidal boundary conditions, we use tidal constituents from the Finite Element Solution global ocean tidal atlas (Lyard et al., 2021). Storm tides are simulated by applying wind and pressure fields as forcing inputs to the model. Both tidal and meteorological forcing are applied together to capture the non-linear interactions between tides and storm surges. For meteorological forcing, we use 10 m wind components and mean sea level pressure fields from SEAS5 at a spatial resolution of 0.4 degrees and a temporal resolution of 6 hours, obtained from the MARS Catalogue of ECMWF (ECMWF, 2018). As further explained below, this simulation covers 525 years. In addition, to validate the SEAS5 storm tides we also simulate storm tides derived from ERA5 data. For this, we use 10 m wind components and mean sea level pressure fields at a spatial resolution of 0.25 degrees and a temporal resolution of 1 hour, obtained from the CDS. This simulation covers 44 years, spanning from 1979 to 2023.

By pairing ensemble member storm tide timeseries from initialisation months that are six months apart and disregarding the 1-month lead time (as explained below; Figure 3), it is possible to obtain 5,250 years from the re-forecast period (6 pairs of initialisation months x 25 ensemble members x 35 years), and 1,836 years from the forecast period (6 pairs of initialisation months x 51 ensemble members x 6 years). This results in timeseries with a total length of 6,936 years. To reduce computational costs while accounting for interannual variability, we conduct the storm tide simulation for one pair of initialisation months and 15 ensemble members in the re-forecast period, resulting in 525 years. We focus on the re-forecast period because its longer historical record provides a more comprehensive representation of climate variability compared to the forecast period. While any pair of initialisation months separated by six months could have been used to generate the synthetic years, we specifically execute GTSM for seasonal forecasts spanning December 1981 and June 2016, for each forecast initialised in the months of December and June, and for each of the 15 ensemble members. Each forecast spans seven months, but the first month is excluded due to dependencies between ensemble members (see Section 3.1). Therefore, we use it to only to spin up GTSM, and only the six months (lead times of 2 to 7 months) of storm tide data are used for further analysis. This results in 525 timeseries with a duration of 6 months, covering January to June, and another covering July to December.

| 1981 | 1982 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Dec** | **Jan** | **Feb** | **Mar** | **Apr** | **May** | **Jun** | **Jul** | **Aug** | **Sep** | **Oct** | **Nov** | **Dec** |
| 1 | 2 | | | | | 7 / 1 | 2 | | | | | 7 |

Initialisation month for the first half a year

Lead time 7 of the first half a year & initialisation month of the second half a year

**Figure 3.** Example of the construction of one synthetic year, disregarding the 1-month lead time and stitching the 2-month lead time to 7-month lead time of two initialisation months that differ six months.

## 2.4 Extreme value analysis

We conduct EVA using the peaks-over-threshold (POT) method, fitting a generalised pareto distribution (GPD) to SEAS5 and ERA5 storm tide timeseries. For the SEAS5 timeseries, the first step of the EVA is to construct synthetic years by pairing ensemble members. Specifically, for each ensemble member, we combine the forecast initialised in December with the forecast from the following June, resulting in a synthetic year spanning 1$^{st}$ of January to 31$^{st}$ of December. By applying this procedure to 15 ensemble members across 35 years (1981 – 2016), we obtain a total of 525 synthetic years.

Next, we derive the extremes values of SEAS5 and ERA5 by applying the POT method with a threshold at the 99.5$^{th}$ percentile, and ensuring that the events are independent by using a declustering time of 3 days between events (Wahl et al., 2017).

2.2.1In the first step of our methodology, we assess the biases in SEAS5 minimum mean sea level pressure  (Section 2.2.1). However, additional biases may arise from SEAS5's wind fields or differences in spatiotemporal resolution – SEAS5 being coarser in both spatial and temporal resolution – potentially leading to underestimations of the simulated storm tide extremes. Therefore, before fitting extreme values to statistical distributions, we test for biases in SEAS5-derived storm tides. We do this by comparing the mean of ERA5-derived POT extremes with the 95% confidence intervals of 1,000 bootstrapped samples (with replacement) of SEAS5-derived mean POT extremes. This approach allows us to identify regions in Europe where SEAS5 produces unbiased extreme storm tides relative to ERA5. Since global hydrodynamic models such as GTSM typically have an uncertainty of approximately 10 cm, we further classify biased regions based on the differences between the mean of ERA5 and the median of the SEAS5 bootstrapped samples. We define small biases where this difference is less than 10 cm. Differences exceeding 10 cm are classified as over- and underestimations.

Finally, we fit the POT extremes to the GPD using the SciPy Python package. We estimate 95% confidence intervals (CIs) through bootstrapping with 1,000 resampled samples (with replacement). Additionally, we empirically derive return periods for SEAS5 and ERA5 using the Weibull's formula (Coles, 2001) for POT extremes.

We perform the same analysis for annual maxima (AM), fitting a generalised extreme value (GEV) distribution. However, since both methods lead to similar results (see Figure 6 and S1), we present only the results for the POT-GEV approach.

# 3 Results

## 3.1 SEAS5 evaluation

First, we asses potential biases of SEAS5 by comparing the monthly minimum mean sea level pressure of SEAS5 against ERA5. Figure 4 shows the p-value results from the bias test comparing the mean values of SEAS5 and ERA5 for the re-forecast and forecast periods. Results are presented for the entire year and disaggregated by seasons, with stippling that marks where the mean differences are statistically significant.

The results for the entire year (panels a and f) show significant differences across large areas during the re-forecast period, particularly in tropical zones and parts of the Southern Hemisphere. Conversely, in the Northern Hemisphere, especially in Europe, parts of Asia and along the west coast of the U.S., the mean values from SEAS5 closely match those of ERA5, providing unbiased representations of the minimum mean sea level pressure. During the forecast period, the results show statistical significance in only a few regions, suggesting that SEAS5 is generally suitable for generating unbiased events. Differences between the re-forecast period and forecast period can be attributed to the short time span used to assess the biases for the forecast period in comparison to the re-forecast period. Furthermore, differences can occur due to differences in the initialisation between the re-forecast, that is initialised with ERA-Interim (Dee et al., 2011) and OCEAN5 (Zuo et al., 2018), and the forecast, initialised with ECMWF operational analyses (Kelder, Marjoribanks, et al., 2022).

In extra-tropical regions, the storm season occurs in the winter months, spanning September to March in the Northern Hemisphere and April to August in the Southern Hemisphere. In the Northern Hemisphere, extra-tropical regions that experience substantial storm tides include Northern Europe, the U.S., Russia and China (Dullaart et al., 2021; Priestley et al., 2020). Panel (e) shows that in the re-forecast dataset, SEAS5 does not capture ERA5 mean values well for the SON season along the eastern U.S. coast. However, it performs better along the U.S. west coast, Europe (excluding Spain's Atlantic coast), China and most of Russia. For the DJF season, SEAS5 shows no statistically significant differences with ERA5 along European and U.S. coastlines. For the forecast period, both SON and DJF seasons show minimal statistically significant differences globally.

In the Southern Hemisphere, storm tides are substantial along the Australian coast and the Atlantic coast of South America (Dullaart et al., 2021). In these regions and during the MAM season, SEAS5 accurately captures minimum mean sea level pressures, but it shows statistically significant grid cells in northern Australia. In the JJA season, SEAS5 performs well in South America but shows significant deviations in the northern and southern regions of Australia.

Second, we assess the independence of the ensemble members of the SEAS5 forecasts. Figure 5 shows the results with panels (a) – (e) representing the re-forecast period and panels (f) – (j) covering the forecast period. Each panel shows the median Spearman correlation for lead times ranging from one month to six months. In general, the Spearman correlation decreases as lead time increases. For a lead time of one month, the mean Spearman correlations are more than 0.3 in large parts of the domain, indicating dependencies between ensemble members. From a lead time of two months onwards, most regions exhibit Spearman correlation coefficients near zero, indicating independence among ensemble members. However, tropical regions in the Pacific and Indian Oceans show high correlations, likely attributable to SEAS5's forecasting skill for El Niño-Southern Oscillation (ENSO). SEAS5 has been shown to exhibit lower ensemble spread in ENSO regions (ECMWF, 2021; Johnson et al., 2019).

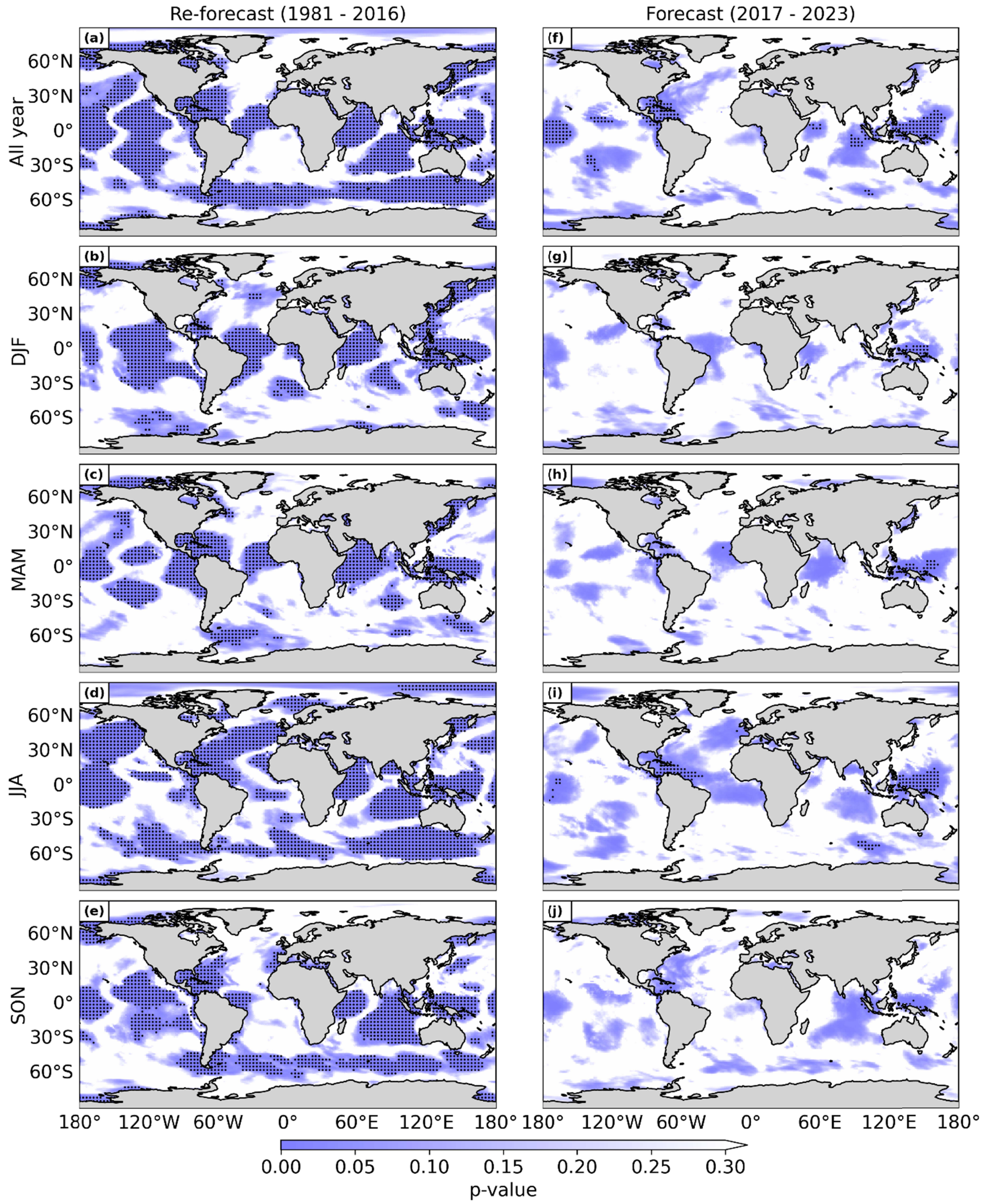
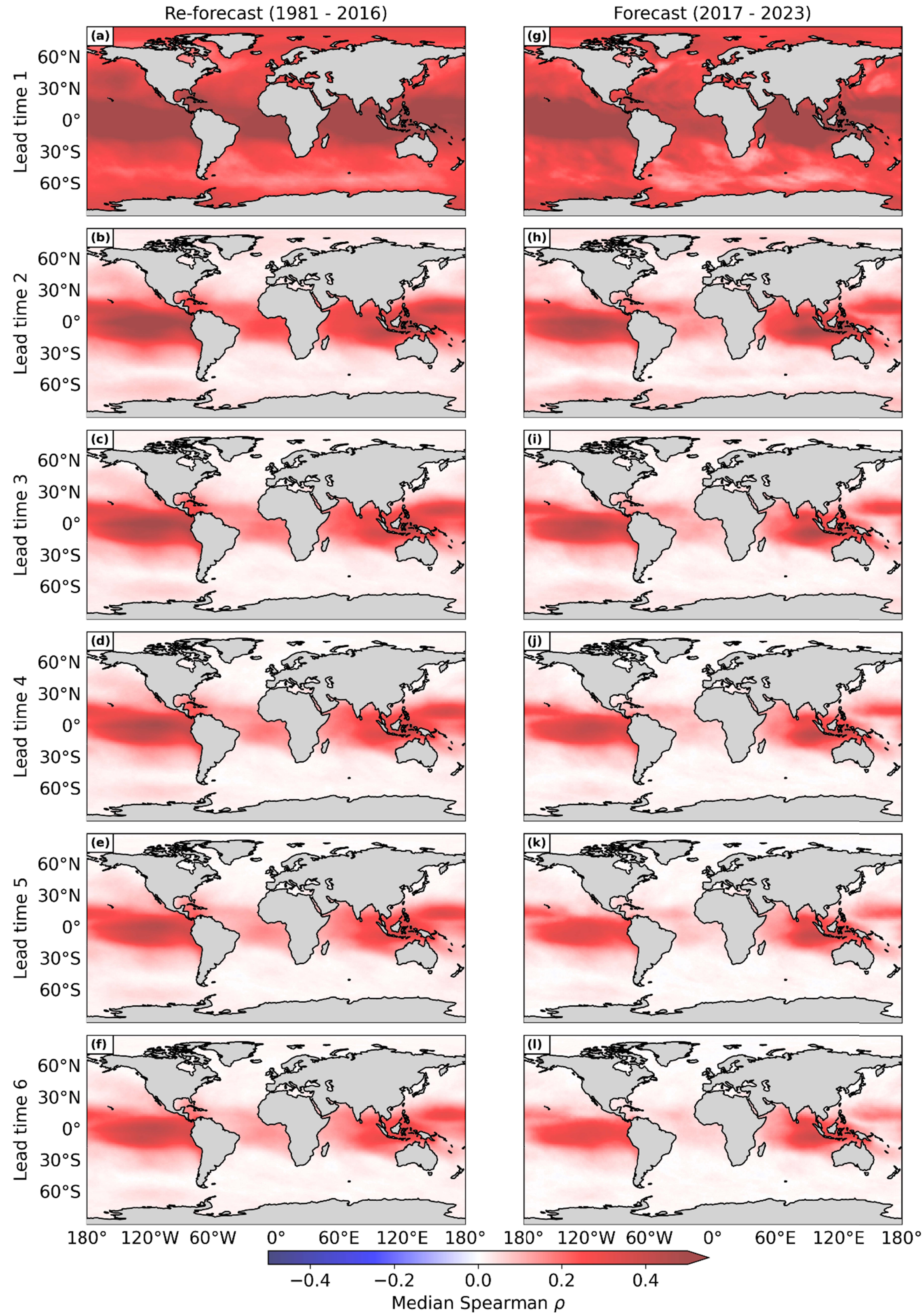



**Figure 4.** Bias test at global scale for SEAS5 compared to ERA5 for the re-forecast and forecast periods. P-value results of the mean comparison of SEAS5 and ERA5, with stippling marks in regions that show statistically significant differences with correction for false discoveries (α < 0.1). Panels (a) and (f) present the analysis for the full year, while the remaining panels show the analysis per season.


Re-forecast (1981 - 2016)
Forecast (2017 - 2023)
Lead time 1
Lead time 2
Lead time 3
Lead time 4
Lead time 5
Lead time 6
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
60°N
30°N
0°
30°S
60°S
180°
120°W
60°W
0°
60°E
120°E
180°
−0.4
−0.2
0.0
0.2
0.4
Median Spearman ρ

326

**Figure 5.** Results of the independence test of the SEAS5 monthly minimum sea level pressure. The plot shows the median Spearman correlation result of the pairwise comparison between ensemble members for each lead time month, distinguishing between the re-forecast and forecast periods.

### 3.2 Statistical analysis of storm tides

We assess whether the extreme storm tides obtained from the POT method exhibit significant biases (Figure 6 panel a). The results reveal region-specific biases in SEAS5 storm tides compared to ERA5. SEAS5 overestimates extreme storm tides in the Mediterranean and underestimates them in the Baltic region, as well as in certain locations along the southern coast of Spain and the Mediterranean. Most of these biases are small and occur in regions where extreme storm tides are low (less than 1 m + MSL). Storm tides across the rest of Europe are unbiased.

Figure 6 panel b shows the 500-year return levels (approximately corresponding to the time span of the SEAS5 synthetic dataset) derived from the SEAS5 POT-GPD fit, where the highest values are observed along the English Chanel and western UK. Figure 6 panel (c) shows the absolute differences in 500-years return levels between SEAS5 and ERA5, while panel (d) shows the relative differences. Blue shading indicates regions where SEAS5 return levels exceed those of ERA5, whereas red shading indicates regions where ERA5 return levels are higher. SEAS5 generally simulates higher return levels in northern Spain and France, the Italian Adriatic coast, the Netherlands, Ireland, and southwestern UK, with exceedances of up to 0.4 m and relative differences over 10%. Conversely, northern and western parts of the UK exhibit higher ERA5 return levels, with relative differences also exceeding 10%.

The analysis using the AM-GEV method (Figure S1 in Supporting Information), results in similar biases, return levels and differences relative to ERA5. Additionally, return periods other than 500 years, also produce results consistent with those presented here.

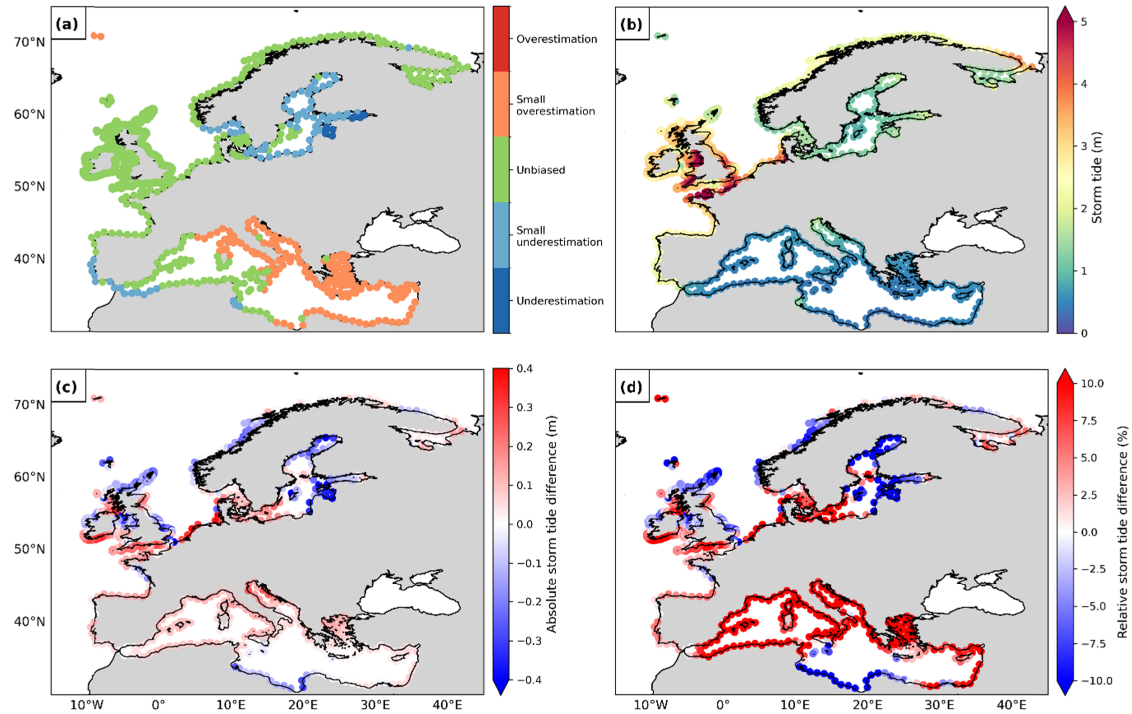



**Figure 6.** Statistical analysis of storm tides using the POT-GPD method: (a) Bias of extremes sampled with the POT method, (b) storm tide levels for the 500-year return period, (c) absolute difference between SEAS5- and ERA5-derived storm tide levels for the 500-year return period and (d) relative difference between SEAS5- and ERA5-derived storm tide levels for the 500-year return period.

### 3.3 Statistical uncertainty of storm tides

We assess the uncertainty of the storm tide return levels looking at the relative 95% CI ranges of SEAS5 (Figure 7 panels b, e) and ERA5 (Figure 7 panels a, d) for RP40 (approximately corresponding to the ERA5 time span) and RP500. These 95% CI ranges are expressed as percentages of the storm tide estimates from the POT-GPD fit. The results indicate that ERA5 exhibits the highest relative uncertainties in the Mediterranean and Baltic regions across both return periods. This is also the result of the low return values of those regions. Specifically, for RP40, the 95% CI range reaches up to 80% of the storm tide level, and more than 400% for RP500. In contrast, SEAS5 shows considerably lower uncertainties for both EVA methods, with 95% CI ranges remaining below 10%, and 20%, for RP40 and RP500, respectively.

While the largest uncertainties in ERA5's relative 95% CI ranges occur in the Mediterranean and Baltic regions, where storm tide levels are relatively low (less than 1.5 m +MSL), substantial uncertainties are also observed in regions with higher storm tides (above 4 m +MSL), such as the UK and Germany. In these regions, the use of SEAS5 significantly reduces uncertainty of the return period estimates. For example, in London and Cuxhaven, the uncertainty decreases substantially across all return periods (Figure 8). For RP40, the uncertainty in London drops from 8% (0.3 m) for ERA5 to 3% (0.1 m) with SEAS5, while for RP500, the reduction is from 19% (0.7 m) to 8% (0.3 m). Similarly, in Cuxhaven, uncertainty decreases substantially with SEAS5, with reductions from 12% (0.5 m) to 5% (0.2 m) for RP40 and from 43% (1.9 m) to 11% (0.5 m) for RP500, resulting in a reduction of uncertainty of 1.4 m.

The ratio of 95% CI ranges between ERA5 and SEAS5, defined as the 95% CI range of ERA5 divided by that of SEAS5, is notably large in several regions. In particular, areas north and west of the UK, as well as parts of the Mediterranean, show high CI ratios for both return periods. For example, in Cuxhaven (see Figure 8), the 95% CI range for ERA5 is about 4 times larger than for SEAS5 for RP500. In other regions, while the CI ratio between ERA5 and SEAS5 is not significant, the SEAS5 CIs fall outside those of ERA5. This indicates that although the uncertainty with ERA5 may not be large, its fit produces considerably different storm tide levels compared to SEAS5 (see Figure 7 panels c and f). Examples of this behaviour are also observed in Figure 8, in cities such as Brighton, Edinburgh and San Sebastian.

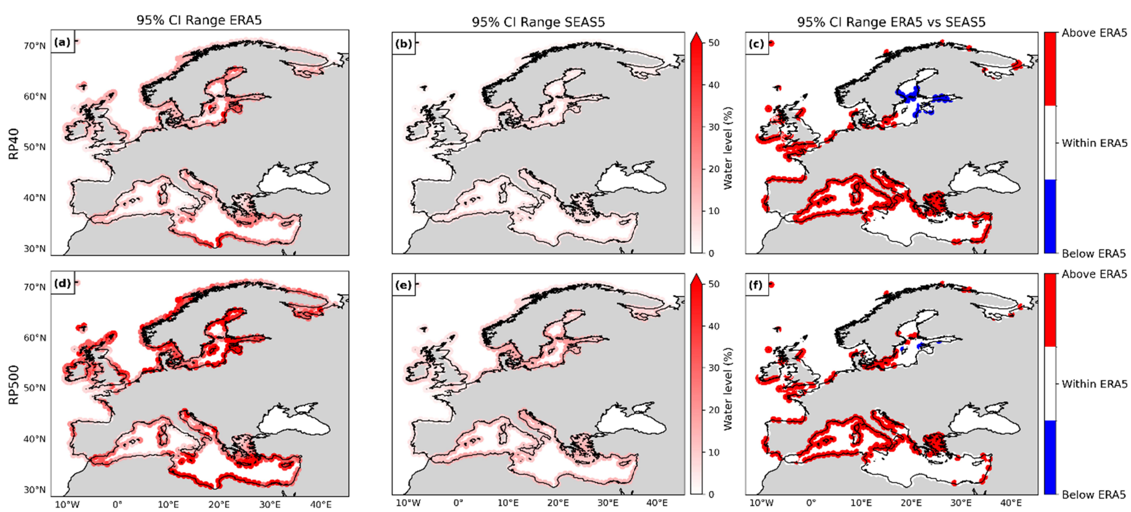



**Figure 7.** Statistical uncertainty of ERA5 and SEAS5: Relative RP40 and RP500 95% CI ranges for ERA5 (panels a, d) and SEAS5 (panels b, e), and regions where the SEAS5 95% CIs fall above (red), within (white) or below (blue) the ERA5 95% CIs (panels c, f).

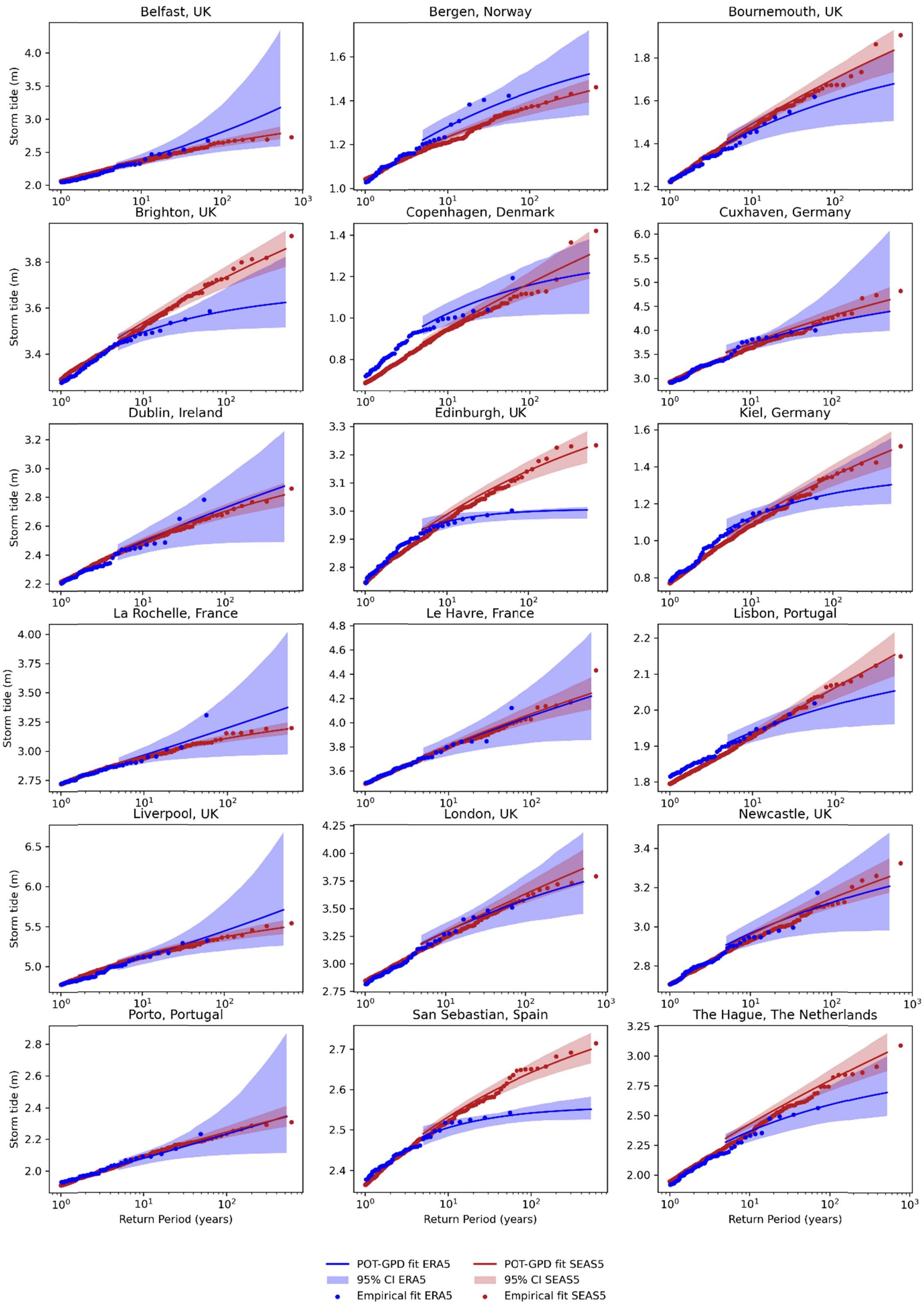



**Figure 8.** Return level curves for 18 coastal cities in the model domain. The empirical and POT-GPD-derived loss-exceedance curves, along with the 95% CIs for the POT-GPD fit are shown for ERA5 (blue) and SEAS5 (red).

## 4 Discussion

Pooling ensemble members from SEAS5 to drive GTSM and generate 525 years of synthetic storm tide timeseries has proven effective in improving the robustness of storm tide return period estimates, reducing their uncertainty, and enhancing their accuracy (Figure 7 and Figure 8). The number of synthetic years could be significantly increased by expanding the number of initialisation months and ensemble members used. However, combining re-forecast and forecast datasets is challenging, due to differences in both ensemble sizes and the bias characteristics presented in Figure 4. Re-forecasts show regional mean sea level pressure biases at the global scale, whereas forecasts appear unbiased across most of the domain in all seasons. Before merging these datasets, further analysis is required to quantify and address these differences, which can arise from differences in initialisation data used for the re-forecasts and forecasts (Kelder, Marjoribanks, et al., 2022), and the shorter observational record (six years) available for testing the bias of the forecast dataset.

Our findings demonstrate that the SEAS5 seasonal forecast archive can be used to generate unbiased (Figure 4) and independent (Figure 5) synthetic storm tide events for most extra-tropical regions. In the Northern Hemisphere, particularly along the U.S. West Coast, Europe, China and Russia, as well as in the Southern Hemisphere, including South America and certain seasons of Australia, mean sea level pressure estimates from SEAS5 show the potential to be used to drive hydrodynamic models. While this is the first attempt to systematically assess SEAS5's suitability for extreme mean sea level pressure conditions at a global scale, previous studies have successfully applied ensemble pooling to SEAS5 or its predecessors to study extremes in precipitation, temperature, wind and storm surges, in specific regions in Europe. However, certain variables required bias correction to generate reliable extremes (van den Brink et al., 2004; Hillier & Dixon, 2020; Kelder et al., 2020; Kelder, Marjoribanks, et al., 2022; Van Den Brink et al., 2005; Walz & Leckebusch, 2019).

Our application of SEAS5 to force GTSM in Europe successfully simulates synthetic storm tides with good performance across most regions, except in parts of the Mediterranean and Baltic Seas, where slight biases remain (Figure 6). The regional biases found in this analysis either in the mean sea level pressure or in the storm tide extremes can be attributed to differences in the spatiotemporal resolution between ERA5 and SEAS5 (Agulles et al., 2024), and due to the data assimilation applied in ERA5, which constrains it more closely to observations. The biases in storm tides cannot be attributed to the GTSM hydrodynamic model, as the same GTSM cut-out model setup was used for both ERA5 and SEAS5 meteorological forcing. In our current analysis, these regional biases have not yet been corrected. However, we suggest that in future studies, comprehensive bias assessment should be conducted and bias correction techniques applied where necessary. While significant progress has been made in bias correction for climate variables, correcting biases in extremes beyond the observational timeframe remains a challenge (Berg et al., 2024; Kelder, Wanders, et al., 2022), and future research should focus on the development of bias correction methods to mitigate storm tide biases (Agulles et al., 2024). Additionally, further investigation is needed to evaluate the physical credibility of the SEAS5-derived extremes to assess the drivers and plausibility of the simulated extremes that occur beyond the observational record (Kelder, Wanders, et al., 2022).

While the re-forecast period allows us to capture the maximum possible climate variability, the SEAS5 time span (1981 to present) also captures key tidal cycles, such as the 8.5 year cycle of lunar perigee or the 18.61-year lunar nodal cycle (Haigh et al., 2011). However, its representation of multidecadal climate variability remains limited to the past 40 years, restricting its ability to estimate probabilities under future conditions. Nevertheless, the same methodology used in this study could be applied to large-ensemble climate simulations (Eyring, Bony, et al., 2016; Ishii & Mori, 2020), providing the possibility to extend return period estimates beyond present-day conditions.

Considering the full time span of SEAS5 and its complete ensemble set, up to 6,936 years of synthetic storm tides could be generated, with the dataset continuously expanding for each initialisation month. This extended dataset offers substantial opportunities for various applications, including: (1) reassessing current design standards for coastal flood defences (Van Den Brink et al., 2005), (2) improving the understanding of historical return periods for extreme events, (3) identifying plausible yet unprecedented storm tide events (Horsburgh et al., 2021), and (4) enabling robust compound flood risk modelling when combining the storm tides with other variables present in SEAS5 forecasts, such as precipitation or significant wave heights.

While this dataset includes all storms occurring in extra-tropical regions, a significant advancement would be the creation of a global dataset that combines storm tide time series from both extra-tropical and tropical regions. Although the approach presented here could be applied to both tropical regions and extra-tropical regions, the resolution of SEAS5 is insufficient to accurately resolve tropical cyclones (Hodges et al., 2017; Murakami, 2014; Thomas et al., 2021). To address this, a hybrid approach using, for example, synthetic tropical cyclone tracks (Bloemendaal et al., 2020; Emanuel et al., 2006; James & Mason, 2005; Lee et al., 2018) in combination with pooled ensembles from seasonal forecast archives to drive hydrodynamic models could be a solution. However, tropical cyclones should be removed from the seasonal forecast archives to ensure consistency in the modelling process.

Physically-based approaches, such as the one presented here, use long-record meteorological datasets to drive hydrodynamic models, offering a more physically consistent representation of storm tides compared to fully statistical methods applied to observed sea levels. However, their high computational cost may limit their feasibility at large spatial scales. Future work could explore hybrid methods for generating synthetic storm tide events and their ability to produce robust and physically sound return periods for extreme events. These methods could include weather generators that stochastically simulate meteorological forcing while still accounting for the physics of storm tides through hydrodynamic models (Z. Wang et al., 2025).

# 5 Conclusions

Ensemble pooling of seasonal forecasts has proven to be a useful technique to estimate the probabilities of extreme events across multiple variables. This approach is effective when the events obtained from those variables are independent and unbiased. Our results show that SEAS5 helps reduce uncertainties in Europe when used to generate large datasets of storm tides for estimating extreme return periods. These results are mostly unbiased, with the exception of the Baltic and Mediterranean seas. In other extra-tropical regions, such as the U.S. West Coast, China and Russia in the Northern Hemisphere and South America in the Southern Hemisphere, SEAS5 ensemble pooling, shows a great potential for estimating robust storm tide return periods when combined with a hydrodynamic model.

Traditional large-scale return period estimates remain highly uncertain, especially for extreme events, which often have the greatest impacts. Pooling SEAS5 ensembles offers an alternative approach to improve the robustness of return period estimates for the present climate. Establishing reliable return period estimates under current conditions is essential, as it provides a solid foundation for assessing future climate risks, which are significantly more uncertain. The methods used in this study, combining ensemble pooling with hydrodynamic modelling, can be applied to large-ensemble climate models. This would allow for more robust future return period estimates, leading to enhanced coastal flood risk assessments at large scale and ultimately supporting more effective adaptation and mitigation strategies.

## Author contributions

I.B.: Conceptualisation, Investigation, Methodology, Modelling, Visualisation, Analysis, Writing – Original Draft. D.E.: Conceptualisation, Investigation, Methodology, Writing – Review & Editing, Supervision. T.K.: Conceptualisation, Investigation, Methodology, Writing – Review & Editing. P.J.W.: Conceptualisation, Investigation, Methodology, Writing – Review & Editing, Supervision. J.C.J.H.A.: Conceptualisation, Investigation, Methodology, Writing – Review & Editing, Supervision. S.M.: Conceptualisation, Investigation, Methodology, Writing – Review & Editing, Supervision.

## Acknowledgements

This work was carried out in the EU-ERC COASTMOVE project nr 884442 and the MYRIAD-EU Grant Agreement nr 101003276. The authors would like to thank the SURF Cooperative for the support in using the Dutch national e-849 infrastructure under grant no. EINF-2224 and EINF-5779.

## Data availability statement

The datasets compiled and/or analysed during the current study will be available on Zenodo upon the acceptance of the paper.

## Software availability statement

The underlying code for this study will be made available on Github upon the acceptance of the paper.

## References

Agulles, M., Marcos, M., Amores, A., & Toomey, T. (2024). Storm surge modelling along European coastlines: The effect of the spatio-temporal resolution of the atmospheric forcing. *Ocean Modelling*, *192*, 102432. https://doi.org/10.1016/j.ocemod.2024.102432

Benito, I., Aerts, J. C. J. H., Eilander, D., Ward, P. J., & Muis, S. (2024). Stochastic coastal flood risk modelling for the east coast of Africa. *Npj Natural Hazards*, *1*(1), 10. https://doi.org/10.1038/s44304-024-00010-1

Berg, P., Bosshard, T., Bozhinova, D., Bärring, L., Löw, J., Nilsson, C., et al. (2024). Robust handling of extremes in quantile mapping – "Murder your darlings." *Geoscientific Model Development*, *17*(22), 8173–8179. https://doi.org/10.5194/gmd-17-8173-2024

Bloemendaal, N., Haigh, I. D., de Moel, H., Muis, S., Haarsma, R. J., & Aerts, J. C. J. H. (2020). Generation of a global synthetic tropical cyclone hazard dataset using STORM. *Scientific Data*, *7*(1), 1–12. https://doi.org/10.1038/s41597-020-0381-2

Breivik, Ø., Aarnes, O. J., Abdalla, S., Bidlot, J.-R., & Janssen, P. A. E. M. (2014). Wind and wave extremes over the world oceans from very large ensembles. *Geophysical Research Letters*, *41*(14), 5122–5131. https://doi.org/10.1002/2014GL060997 van den Brink, H. W., Können, G. P., Opsteegh, J. D., van Oldenborgh, G. J., & Burgers, G. (2004). Improving 104-year surge level estimates using data of the ECMWF seasonal prediction system. *Geophysical Research Letters*, *31*(17), 1–4. https://doi.org/10.1029/2004GL020610

Brunner, M. I., & Slater, L. J. (2022). Extreme floods in Europe: going beyond observations using reforecast ensemble pooling. *Hydrology and Earth System Sciences*, *26*(2), 469–482. https://doi.org/10.5194/hess-26-469-2022

CGEDD. (2010). *Tempete Xynthia: Retour d'experience, evaluation et propositions d'action* (p. 192).

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

Consortium EMODnet Bathymetry. (2018). EMODnet Digital Bathymetry (DTM). Retrieved June 21, 2022, from https://sextant.ifremer.fr/record/18ff0d48-b203-4a65-94a9-5fd8b0ec35f6/

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Dessai, S., & Bruno Soares, M. (2013). *Literature review of the use of seasonal-to-decadal (S2D) predictions across all sectors. Deliverable report 12.1 of the EUPORIAS project*. Retrieved from http://euporias-test2.wdfiles.com/local--files/events-meetings/D12.1.pdf

Dowdy, A. J., Mills, G. A., Timbal, B., & Wang, Y. (2013). Changes in the Risk of Extratropical Cyclones in Eastern Australia. https://doi.org/10.1175/JCLI-D-12-00192.1

Dullaart, J. C. M., Muis, S., Bloemendaal, N., Chertova, M. V., Couasnon, A., & Aerts, J. C. J. H. (2021). Accounting for tropical cyclones more than doubles the global population exposed to low-probability coastal flooding. *Communications Earth & Environment*, *2*(1), 1–11. https://doi.org/10.1038/s43247-021-00204-9

ECMWF. (2018). MARS - the ECMWF meteorological archive [Data set]. *Software and computing services*. ECMWF. Retrieved from https://www.ecmwf.int/node/18124

ECMWF. (2021). *SEAS5 user guide* (pp. 1–43). ECMWF. Retrieved from https://www.ecmwf.int/sites/default/files/medialibrary/2017-10/System5_guide.pdf

Emanuel, K., Ravela, S., Vivant, E., & Risi, C. (2006). A Statistical Deterministic Approach to Hurricane Risk Assessment. https://doi.org/10.1175/BAMS-87-3-299

Eyring, V., Righi, M., Lauer, A., Evaldsson, M., Wenzel, S., Jones, C., et al. (2016). ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development*, *9*(5), 1747–1802. https://doi.org/10.5194/gmd-9-1747-2016

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J. L., Barrand, N. E., Bell, R., et al. (2013). Bedmap2: improved ice bed, surface and thickness datasets for Antarctica. *The Cryosphere*, *7*(1), 375–393. https://doi.org/10.5194/tc-7-375-2013

GEBCO. (2014). General Bathymetric Chart of the Oceans (GEBCO) 2014 Grid. Retrieved June 21, 2022, from https://www.gebco.net/

Haigh, I. D., Eliot, M., & Pattiaratchi, C. (2011). Global influences of the 18.61 year nodal cycle and 8.85 year cycle of lunar perigee on high tidal levels. *Journal of Geophysical Research: Oceans*, *116*(C6). https://doi.org/10.1029/2010JC006645

Haigh, I. D., MacPherson, L. R., Mason, M. S., Wijeratne, E. M. S., Pattiaratchi, C. B., Crompton, R. P., & George, S. (2014). Estimating present day extreme water level exceedance probabilities around the coastline of Australia: tropical cyclone-induced storm surges. *Climate Dynamics*, *42*(1), 139–157. https://doi.org/10.1007/s00382-012-1653-0

Haigh, I. D., Marcos, M., Talke, S. A., Woodworth, P. L., Hunter, J. R., Hague, B. S., et al. (2023). GESLA Version 3: A major update to the global higher-frequency sea-level dataset. *Geoscience Data Journal*, *10*(3), 293–314. https://doi.org/10.1002/gdj3.174

Heffernan, J. E., & Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 497–546. https://doi.org/10.1111/j.1467-9868.2004.02050.x

Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Sabater, J. M., Nicolas, J., et al. (2019). Global reanalysis: goodbye ERA-Interim, hello ERA5. *ECMWF Newsletter*, (159), 17–24. https://doi.org/10.21957/vf291hehd7

Hillier, J. K., & Dixon, R. S. (2020). Seasonal impact-based mapping of compound hazards. *Environmental Research Letters*, *15*(11), 114013. https://doi.org/10.1088/1748-9326/abbc3d

Hodges, K., Cobb, A., & Vidale, P. L. (2017). How Well Are Tropical Cyclones Represented in Reanalysis Datasets? https://doi.org/10.1175/JCLI-D-16-0557.1

Horsburgh, K., Haigh, I. D., Williams, J., De Dominicis, M., Wolf, J., Inayatillah, A., & Byrne, D. (2021). “Grey swan” storm surges pose a greater coastal flood hazard than climate change. *Ocean Dynamics*, *71*(6–7), 715–730. https://doi.org/10.1007/s10236-021-01453-0

Howard, T., & Williams, S. D. P. (2021). Towards using state-of-the-art climate models to help constrain estimates of unprecedented UK storm surges. *Natural Hazards and Earth System Sciences*, 1–34. https://doi.org/10.5194/nhess-2021-184

Ishii, M., & Mori, N. (2020). d4PDF: large-ensemble and high-resolution climate simulations for global warming risk assessment. *Progress in Earth and Planetary Science*, *7*(1), 58. https://doi.org/10.1186/s40645-020-00367-7

James, M. K., & Mason, L. B. (2005). Synthetic Tropical Cyclone Database. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *131*(4), 181–192. https://doi.org/10.1061/(ASCE)0733-950X(2005)131:4(181)

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2019). SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, *12*(3), 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019

Kelder, T., Müller, M., Slater, L. J., Marjoribanks, T. I., Wilby, R. L., Prudhomme, C., et al. (2020). Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes. *Npj Climate and Atmospheric Science*, *3*(1), 1–13. https://doi.org/10.1038/s41612-020-00149-4

Kelder, T., Marjoribanks, T. I., Slater, L. J., Prudhomme, C., Wilby, R. L., Wagemann, J., & Dunstone, N. (2022). An open workflow to gain insights about low-likelihood high-impact weather events from initialized predictions. *Meteorological Applications*, *29*(3), e2065. https://doi.org/10.1002/met.2065

Kelder, T., Wanders, N., Van Der Wiel, K., Marjoribanks, T. I., Slater, L. J., Wilby, R. L., & Prudhomme, C. (2022). Interpreting extreme climate impacts from large ensemble simulations—are they unseen or unrealistic? *Environmental Research Letters*, *17*(4), 044052. https://doi.org/10.1088/1748-9326/ac5cf4

Kernkamp, H. W. J., Van Dam, A., Stelling, G. S., & De Goede, E. D. (2011). Efficient scheme for the shallow water equations on unstructured grids with application to the Continental Shelf. *Ocean Dynamics*, *61*, 1175–1188. https://doi.org/10.1007/s10236-011-0423-6

Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2018). An Environmentally Forced Tropical Cyclone Hazard Model. *Journal of Advances in Modeling Earth Systems*, *10*(1), 223–241. https://doi.org/10.1002/2017MS001186

Li, H., Haer, T., Couasnon, A., Enríquez, A. R., Muis, S., & Ward, P. J. (2023). A spatially-dependent synthetic global dataset of extreme sea level events. *Weather and Climate Extremes*, *41*, 100596. https://doi.org/10.1016/j.wace.2023.100596

Lin, N., & Chavas, D. (2012). On hurricane parametric wind and applications in storm surge modeling. *Journal of Geophysical Research Atmospheres*, *117*(9), 1–19. https://doi.org/10.1029/2011JD017126

Lyard, F. H., Allain, D. J., Cancet, M., Carrère, L., & Picot, N. (2021). FES2014 global ocean tide atlas: design and performance. *Ocean Science*, *17*(3), 615–649. https://doi.org/10.5194/os-17-615-2021

Marsooli, R., Lin, N., Emanuel, K., & Feng, K. (2019). Climate change exacerbates hurricane flood hazards along US Atlantic and Gulf Coasts in spatially varying patterns. *Nature Communications*, *10*(1), 3785. https://doi.org/10.1038/s41467-019-11755-z

Menéndez, M., & Woodworth, P. L. (2010). Changes in extreme high water levels based on a quasi-global tide-gauge data set. *Journal of Geophysical Research: Oceans*, *115*(C10). https://doi.org/10.1029/2009JC005997

Meucci, A., Young, I. R., Hemer, M., Kirezci, E., & Ranasinghe, R. (2020). Projected 21st century changes in extreme wind-wave events. *Science Advances*, *6*(24), eaaz7295. https://doi.org/10.1126/sciadv.aaz7295

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

Murakami, H. (2014). Tropical cyclones in reanalysis data sets. *Geophysical Research Letters*, *41*(6), 2133–2141. https://doi.org/10.1002/2014GL059519

Priestley, M. D. K., Ackerley, D., Catto, J. L., Hodges, K. I., McDonald, R. E., & Lee, R. W. (2020). An Overview of the Extratropical Storm Tracks in CMIP6 Historical Simulations. https://doi.org/10.1175/JCLI-D-19-0928.1

Rashid, M. M., Moftakhari, H., & Moradkhani, H. (2024). Stochastic simulation of storm surge extremes along the contiguous United States coastlines using the max-stable process. *Communications Earth & Environment*, *5*(1), 1–10. https://doi.org/10.1038/s43247-024-01206-z

Rose, L., Widlansky, M. J., Feng, X., Thompson, P., Asher, T. G., Dusek, G., et al. (2024). Assessment of water levels from 43 years of NOAA's Coastal Ocean Reanalysis (CORA) for the Gulf of Mexico and East Coasts. *Frontiers in Marine Science*, *11*. https://doi.org/10.3389/fmars.2024.1381228

Sebastian, A., Bader, D. J., Nederhoff, C. M., Leijnse, T. W. B., Bricker, J. D., & Aarninkhof, S. G. J. (2021). Hindcast of pluvial, fluvial, and coastal flood damage in Houston, Texas during Hurricane Harvey (2017) using SFINCS. *Natural Hazards*, (2017). https://doi.org/10.1007/s11069-021-04922-3

Thomas, S. R., Nicolau, S., Martínez-Alvarado, O., Drew, D. J., & Bloomfield, H. C. (2021). How well do atmospheric reanalyses reproduce observed winds in coastal regions of Mexico? *Meteorological Applications*, *28*(5), e2023. https://doi.org/10.1002/met.2023

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S., & Belcher, S. E. (2017). High risk of unprecedented UK rainfall in the current climate. *Nature Communications*, *8*(1), 107. https://doi.org/10.1038/s41467-017-00275-3

Van Den Brink, H. W., Können, G. P., Opsteegh, J. D., Van Oldenborgh, G. J., & Burgers, G. (2005). Estimating return periods of extreme events from ECMWF seasonal forecast ensembles. *International Journal of Climatology*, *25*(10), 1345–1354. https://doi.org/10.1002/joc.1155

Wahl, T., Haigh, I. D., Nicholls, R. J., Arns, A., Dangendorf, S., Hinkel, J., & Slangen, A. B. A. (2017). Understanding extreme sea levels for broad-scale coastal impact and adaptation analysis. *Nature Communications*, *8*(May), 1–12. https://doi.org/10.1038/ncomms16075

Walz, M. A., & Leckebusch, G. C. (2019). Loss potentials based on an ensemble forecast: How likely are winter windstorm losses similar to 1990? *Atmospheric Science Letters*, *20*(4), e891. https://doi.org/10.1002/asl.891

Wang, C., Zhang, L., Lee, S.-K., Wu, L., & Mechoso, C. R. (2014). A global perspective on CMIP5 climate model biases. *Nature Climate Change*, *4*(3), 201–205. https://doi.org/10.1038/nclimate2118

Wang, X., Verlaan, M., Veenstra, J., & Lin, H. X. (2022). Data-assimilation-based parameter estimation of bathymetry and bottom friction coefficient to improve coastal accuracy in a global tide model. *Ocean Science*, *18*(3), 881–904. https://doi.org/10.5194/os-18-881-2022

Wang, Z., Leung, M., Mukhopadhyay, S., Sunkara, S. V., Steinschneider, S., Herman, J., et al. (2025). A hybrid statistical–dynamical framework for compound coastal flooding analysis. *Environmental Research Letters*, *20*(1), 014005. https://doi.org/10.1088/1748-9326/ad96ce

Wilks, D. S. (2016). “The Stippling Shows Statistically Significant Grid Points”: How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It. https://doi.org/10.1175/BAMS-D-15-00267.1

Zuo, H., Alonso-Balmaseda, M., Mogensen, K., & Tietsche, S. (2018, 201808). OCEAN5: The ECMWF Ocean Reanalysis System and its Real-Time analysis component [text]. Retrieved February 27, 2025, from https://www.ecmwf.int/en/elibrary/80763-ocean5-ecmwf-ocean-reanalysis-system-and-its-real-time-analysis-component

*JGR: Oceans*

Supporting Information for

# Pooling Seasonal Forecast Ensembles to Estimate Storm Tide Return Periods in Extra-tropical Regions

**Irene Benito[1], Dirk Eilander[1,2], Timo Kelder[1,3], Philip J. Ward[1,2], Jeroen C.J.H. Aerts[1,2] and Sanne Muis[1,2]**

[1]Institute for Environmental Studies (IVM), VU University Amsterdam, The Netherlands

[2]Deltares, Delft, The Netherlands.

[3]Climate Adaptation Services Foundation (CAS), The Netherlands

## Contents of this file

## Introduction

Results of the alternative EVA method AM-GEV used to analyse the storm tide results.

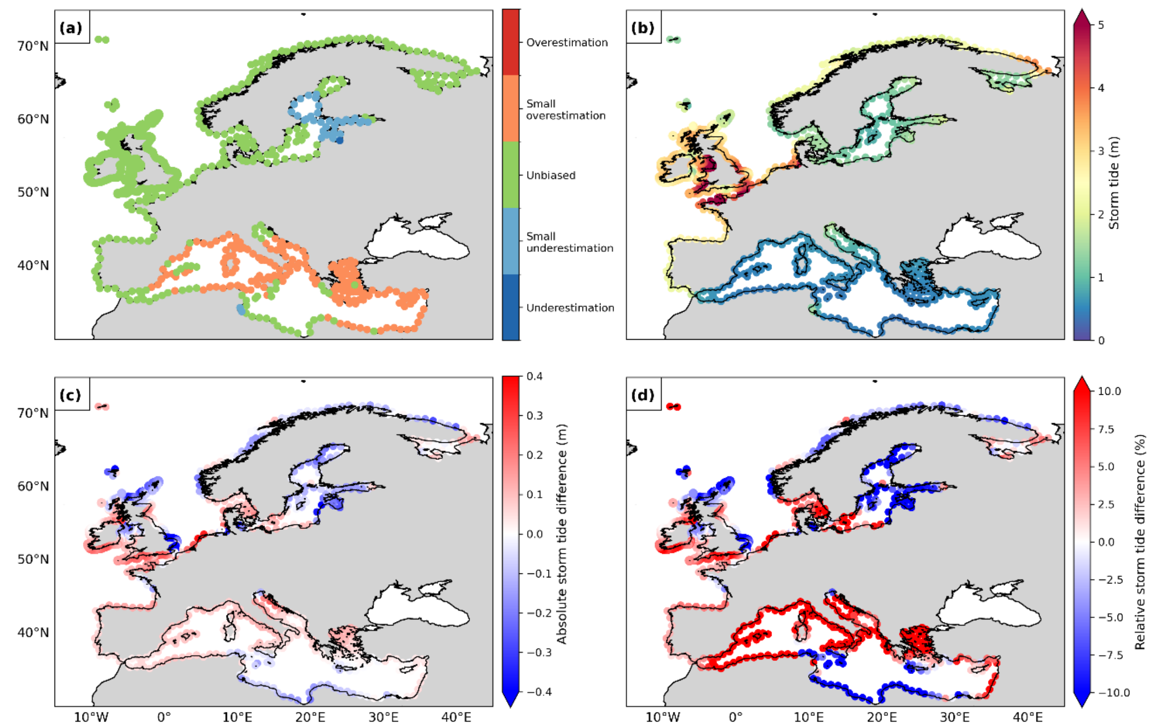



**Figure S1.** Statistical analysis of storm tides using the AM-GEV method: (a) Bias test results for AM, (b) AM-GEV derived storm tide levels for the 500-year return period, (c) absolute difference between AM-GEV SEAS5- and ERA5-derived storm tide levels for the 500-year return period and (d) relative difference between AM-GEV SEAS5- and ERA5-derived storm tide levels for the 500-year return period.