https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

Extreme high-water anomalies in southwest coastal Bangladesh: model uncertainty, cyclone-event checks and non-stationarity

Sujoy Biswas[a,*], Md. Shahjahan Ali[a], Arpon Roy[a], Nikhil Ronjon Roy[a], Juganto Roy[a]
[a] *Department of Civil Engineering, Khulna University of Engineering & Technology (KUET), Khulna 9203, Bangladesh*
**Corresponding author: Sujoy Biswas; e-mail: sujoybiswas10112002@gmail.com*

ABSTRACT
Extreme high-water estimates are needed for coastal flood-risk assessment and infrastructure design, but short, incomplete and strongly seasonal tidal-river records make them difficult to constrain. We analysed daily maximum water levels through 2025 at five Bangladesh Water Development Board stations in southwest coastal Bangladesh using a station-specific leave-one-year-out Monthly Climatological High-Water Reference (MCHR) derived from monthly means of daily maxima. Annual maxima were modelled with Normal, Gumbel-I and generalized extreme-value distributions fitted by L-moments (GEV-LM), while declustered peaks over threshold were analysed with a generalized Pareto distribution (POT-GPD). After 90 % annual completeness screening, 25–34 annual maxima remained per station. At the 100-year return period, the four stationary approaches differed by 15.6–21.5 %. POT estimates were stable across threshold and declustering choices at four stations, whereas SW244 showed an 18.75 % maximum deviation from its baseline estimate. Combined trend, change-point and likelihood evidence supported time-varying GEV location at four stations after false-discovery-rate adjustment and trend-free prewhitening sensitivity; likelihood evidence remained significant at all four, while SW243 remained stationary. Replacing the mean MCHR with a median reference changed 100-year GEV-LM and POT-GPD estimates by at most 2.79 % and 3.48 %, respectively. Cyclone-date checks showed that Aila, Amphan, Yaas and Remal coincided with annual maxima or unusually high anomalies at multiple stations. Extreme high-water estimates from short tidal-river records are most informative when model choice, sampling, non-stationarity, parameter uncertainty and data quality are reported alongside the point estimate instead of being collapsed into a single deterministic return level.

Keywords: High-water anomaly; extreme-value analysis; POT-GPD; non-stationary GEV; model uncertainty; coastal Bangladesh

1. Introduction

Coastal flooding and infrastructure damage often arise from several water-level processes acting together; a single driver rarely explains the full response. Mean sea level provides the background state, astronomical tides impose regular fluctuations and meteorological forcing can add surge. Waves, river discharge, rainfall and local hydraulics can further modify the level reached at a site. In shallow coastal waters and estuaries, these components may interact, so an extreme level cannot always be represented by a simple sum of independent contributions (Arns et al., 2020; Muis et al., 2016; Palmer et al., 2024; Zhang and Sheng, 2015). Reliable estimates of extreme high water are needed for flood-risk assessment and for the design or screening of embankments, drainage systems, jetties, roads and port infrastructure (Guan et al., 2021; Wahl et al., 2017).

Bangladesh is particularly exposed to this problem. Its low-lying deltaic setting at the northern Bay of Bengal combines tropical cyclones, a broad shallow shelf and an extensive tidal-river network, all of which favour severe coastal flooding (Bernard et al., 2022; Karim and Mimura, 2008; Lewis et al., 2014). In the southwest, elevated water levels can travel inland through tidal and estuarine channels and affect settlements, polders, embankments and transport or industrial assets. The magnitude of that response varies from place to place because cyclone properties, coastal geometry, channel configuration and local topography differ (Khan et al., 2022; Lewis et al., 2014). Observational and modelling studies also indicate that sea-level rise can increase future storm-surge and inundation hazards in Bangladesh (Jisan et al., 2018; Lee, 2013). These spatial differences favour station-specific analysis over assigning one regional extreme value to the entire southwest coast.

The statistical problem is made harder by the form and length of the available observations. Where suitable data exist, storm-surge studies usually combine high-frequency water levels with predicted astronomical tides to estimate non-tidal residuals or skew surges (Cid et al., 2018; Saint-Criq et al., 2022). Even skew surge can retain a dependence on tidal conditions in shallow or mixed tidal environments, so physically separated components still need careful interpretation (Santamaria-Aguilar and Vafeidis, 2018). In many data-limited settings, however, long records are available only as daily maxima. Extrapolation from such short records is inherently uncertain because extreme return levels extend well beyond the range directly observed (Goring et al., 2011; Wahl et al., 2017). The stations used in this analysis all have this restriction: each station provides only 25-34 annual maxima after quality and completeness screening, and a continuous predicted astronomical-tide series is not available for the entire analysis period.

Given these constraints, we use a Monthly Climatological High-Water Reference (MCHR) to represent the recurring seasonal high-water pattern at each station. Each daily maximum is expressed relative to the climatological

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

reference for its calendar month. This reduces much of the regular seasonal variation but retains unusually large positive departures. This adjustment is important because coastal water levels and storm-surge behaviour vary seasonally, and changes in the timing or amplitude of the seasonal cycle can affect when extremes occur (Barroso et al., 2025; D'Arcy et al., 2023; Palmer et al., 2024). We describe the resulting variable as a high-water anomaly because astronomical tide has not been removed, so it is not a pure meteorological surge residual. Tidal departures, tide–surge interaction, freshwater influence, rainfall, and local estuarine effects may still contribute to the anomaly (Arns et al., 2020; Saint-Criq et al., 2022).

Long-return-period extrapolation introduces a second layer of uncertainty. Even if the record is only a few decades long, 50- or 100-year levels may be required for engineering and hazard studies. Annual Maximum Series (AMS) provide a transparent block-maxima sample. Peaks-Over-Threshold (POT) sampling retains several independent high events, allowing more information from the upper tail to contribute (Coles, 2001; Davison and Smith, 1990). POT estimates, however, depend on both the threshold and the treatment of temporal dependence. A threshold that is too low can bias the tail model; one that is too high leaves too few exceedances and inflates sampling uncertainty (Bernardara et al., 2012; Caballero-Megido et al., 2018; Silva Lomba and Fraga Alves, 2020). Different extreme-value models can also fit the observed range similarly and then separate substantially at long return periods (Caruso and Marani, 2022; Fang et al., 2021; Hamdi et al., 2014). Choosing a model only because it has the smallest in-sample error can therefore understate uncertainty in the far tail.

The assumption of stationarity is also restrictive. Extreme-water behaviour can evolve as mean sea level, tides, storm climate, river conditions, morphology and other local controls change. Long-term changes in extreme sea levels have been documented in several regions and recent work shows that storm-surge seasonality and the relative contributions of mean sea level, tide and surge can also vary through time (Barroso et al., 2025; Marcos and Woodworth, 2017; Menéndez and Woodworth, 2010; Palmer et al., 2024). Under those conditions, a stationary return level blends information from different historical states and may not represent any single year equally well. Comparing stationary and non-stationary formulations becomes especially important when short records are extrapolated to long return periods.

Previous work provides strong foundations for the individual components of this problem. Bangladesh studies have examined storm-surge hazard, inundation and sea-level-rise effects (Jisan et al., 2018; Karim and Mimura, 2008; Khan et al., 2022; Lee, 2013; Lewis et al., 2014), while a broader literature has developed methods for extreme-value modelling, POT threshold choice, seasonality, dependence and non-stationarity (Caballero-Megido et al., 2018; Coles, 2001; D'Arcy et al., 2023; Hamdi et al., 2014). What remains less well documented is how these uncertainty sources interact in observational analyses of several inland tidal-river stations in southwest Bangladesh. In particular, model choice, AMS/POT sampling, threshold and declustering decisions, temporal change, parameter uncertainty and the definition of the seasonal reference have rarely been assessed together for this setting.

To address this gap, we analysed daily maximum water levels from five BWDB stations spanning the Baleswar-Haringhata, Barisal-Buriswar and Rupsa-Pasur systems. Annual maxima were evaluated with Normal Probability, Gumbel-I and GEV-LM models and POT-GPD provided a complementary threshold-based estimate. Return levels were compared at 10, 25, 50 and 100 years. The workflow also examined POT threshold and declustering sensitivity, Mann-Kendall trends, Sen slopes, Pettitt change points, stationary GEV-LM and linear-location GEV-MLE models, bootstrap and profile-likelihood uncertainty, leave-one-year-out construction of the climatological reference and correspondence with documented cyclone periods.

Three questions frame the analysis: (1) how much do probability-model choice and AMS/POT sampling affect long-return-period anomaly estimates and station ranking; (2) do documented major cyclone periods coincide with unusually high anomalies; and (3) do neighbouring tidal-river stations show similar or contrasting evidence of non-stationarity? Addressing these questions together separates observed-range fit from uncertainty in the unobserved tail and provides a stronger basis for interpreting extreme high water from short, incomplete and strongly seasonal records.

2. Data and methods

2.1 Study area and data sources

The analysis is based on five Bangladesh Water Development Board (BWDB) stations in the principal southwest tidal-river systems (Fig. 1): SW107.2 (Rayenda) in the Baleswar-Haringhata system, SW20 (Amtali) in the Barisal-Buriswar system, and SW241 (Khulna), SW243 (Chalna) and SW244 (Mongla) in the Rupsa-Pasur system. Daily maximum water levels were available mainly from 1985/1986 through 2025. Because station datums differ, absolute levels were not pooled; each anomaly was referenced to a station-specific baseline in its source datum. Table 1 summarises station location, river system, analysis period and effective AMS sample size.

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

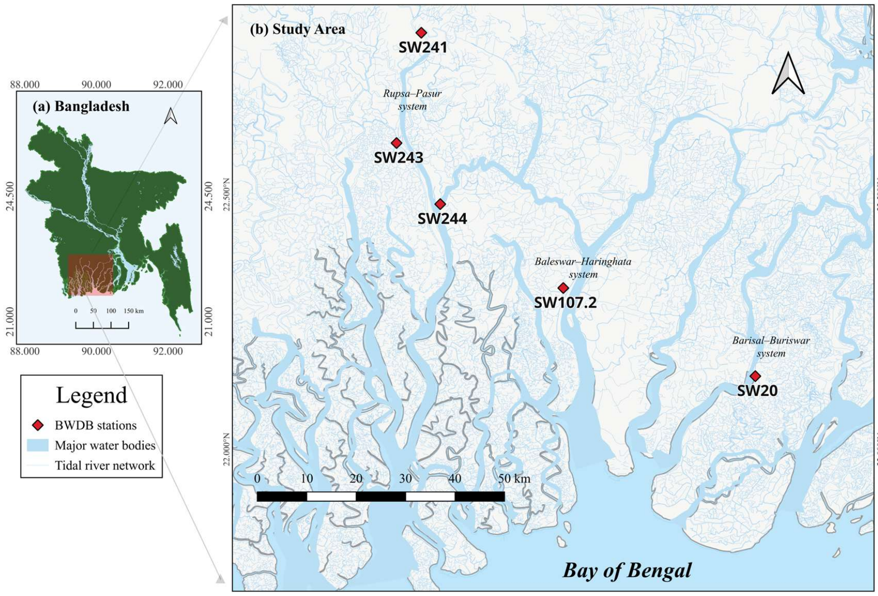

Figure 1. Map of the five analysed BWDB water-level stations in southwest coastal Bangladesh. The station network samples the Baleswar-Haringhata, Barisal-Buriswar and Rupsa-Pasur tidal-river systems. Administrative boundaries are based on Bangladesh Bureau of Statistics (BBS) data distributed through the Humanitarian Data Exchange (HDX) (United Nations Office for the Coordination of Humanitarian Affairs, 2020); river-network data are from the Local Government Engineering Department (LGED), distributed through HDX (Local Government Engineering Department (LGED), 2018); and inland-water polygons are from Global Map Bangladesh, Survey of Bangladesh/International Steering Committee for Global Mapping (ISCGM) (Survey of Bangladesh, 2016). Map prepared by the authors in QGIS.

Table 1. Station metadata and effective annual-maximum sample after 90 % completeness screening.

Station	Location	River system	Lat. (° N)	Lon. (° E)	Analysis period	AMS n
SW107.2	Rayenda	Baleswar-Haringhata	22.3161	89.8495	1985–2025	30
SW20	Amtali	Barisal-Buriswar	22.1432	90.2265	1986–2025	32
SW241	Khulna	Rupsa-Pasur	22.8169	89.5708	1985–2025	34
SW243	Chalna	Rupsa-Pasur	22.6005	89.5222	1985–2025	32
SW244	Mongla	Rupsa-Pasur	22.4807	89.6080	1985–2025	25

The records were first ordered by date and checked for duplicate entries, non-numeric values and suspicious observations. Large values were reviewed on their merits; magnitude alone was not used as a reason for removal. Completeness was calculated from the expected number of calendar days at both annual and monthly scales. The 90 % annual threshold controlled AMS eligibility, whereas the 90 % monthly threshold determined whether a year-month could enter the climatological reference. Valid observations within an accepted year were neither imputed nor discarded because another day in the same month was missing. Sensitivity to this rule was tested at 85 %, 90 % and 95 %. We also compared missingness in April–May and October–November with full-record missingness to determine whether gaps were concentrated in the main cyclone months.

SW244 required a dedicated data-quality sensitivity check. The source record reports 5.88 m on 26 June 1996, but the audited workflow contains a documented QC override of 3.39 m for that date; the raw value was retained separately and not overwritten. No independent station-history document supplied with the dataset explains the physical basis of the override. The primary QC series uses 3.39 m, with sensitivity reported for the raw 5.88 m value and for exclusion of 1995–1996.

The 18 April 1996 anomaly of 2.081 m remains unchanged. Because the mid-1990s Pettitt change point coincides with unresolved gauge/datum history, temporal change at SW244 is interpreted statistically and not attributed to a specific hydrological or climatic cause (Supplementary Table S1).

2.2 Monthly climatological high-water reference and anomaly

For every annually accepted year y and calendar month m, we first calculated the mean of all valid daily maxima in that year-month. A year-month entered the reference pool only when monthly completeness reached 90 %. For a target year y, the Monthly Climatological High-Water Reference (MCHR) for month m was then calculated from the accepted year-month means for the same calendar month after excluding the target year. By construction, the primary reference is leave-one-year-out (LOYO), which prevents the target year from contributing to its own baseline. The MCHR is a study-specific climatological reference, not an established tidal datum. It was adopted because a continuous predicted astronomical-tide series was unavailable for the full record and because the aim was to reduce the recurring seasonal high-water cycle while retaining unusual positive departures. A median-based LOYO reference was used as a robustness alternative (Supplementary Table S2).

$$A(t) = \mathrm{WL}_{\max}(t) - \mathrm{MCHR}(y, m), \tag{1}$$

Here WLmax(t) denotes the observed daily maximum on day t and MCHR(y,m) is the LOYO climatological reference for the corresponding target year and calendar month. Equation (1) defines the operational anomaly used in this study; it is not a conventional storm-surge, skew-surge or non-tidal-residual formulation.

The anomaly represents the departure of a daily maximum from the station's typical high-water condition for that calendar month. Because astronomical tide is not removed, the anomaly can retain contributions from meteorological surge, tidal variability, tide-surge interaction, freshwater discharge, rainfall, background water-level variability and local estuarine hydraulics (Arns et al., 2020; Palmer et al., 2024; Saint-Criq et al., 2022; Santamaria-Aguilar and Vafeidis, 2018). Throughout the manuscript, we refer to it as a seasonally adjusted high-water anomaly and not as a storm-surge residual or skew surge.

We tested the seasonal reference in two ways. First, the mean-based LOYO MCHR was replaced by a median-based version and the principal AMS, POT and trend results were recalculated. Second, annual and monthly completeness thresholds of 85 %, 90 % and 95 % were compared. These checks quantify sensitivity to the reference statistic and screening rule; neither is treated as an independent physical validation.

Monthly means of daily maxima, the corresponding LOYO MCHR and monthly anomalies were inspected together to assess seasonality and continuity. Year-months that failed the completeness criterion remained as gaps in the reference-input series and were not interpolated.

2.3 Extreme-value samples

We extracted two complementary extreme-value samples from the screened anomaly series. The AMS retained the single largest daily anomaly from each accepted year, following the standard block-maxima framework (Coles, 2001; Hamdi et al., 2014). POT sampling retained observations above a high threshold, so threshold selection and temporal dependence become part of the sampling decision. The baseline used a station-specific 95th-percentile threshold with 3-day run declustering after considering mean residual life, GPD shape and modified-scale stability, together with the number of retained peaks. The threshold choice reflects the usual bias-variance trade-off: thresholds that are too low can introduce model bias, while very high thresholds reduce sample size and increase variance (Bernardara et al., 2012; Caballero-Megido et al., 2018; Davison and Smith, 1990; Kang and Song, 2017; Silva Lomba and Fraga Alves, 2020). Supplementary Figs. S1-S2 provide the corresponding diagnostics.

Exceedances separated by no more than three calendar days were grouped into one cluster, from which only the largest anomaly was retained; a new cluster began once the gap exceeded three days. Sensitivity was assessed across the 94th, 95th, 96th and 97th percentile thresholds combined with 2-, 3- and 5-day run lengths, yielding 12 scenarios per station. For each station, POT sampling sensitivity is reported as the maximum absolute percentage deviation of the 100-year return level from the baseline 95th-percentile/3-day estimate: $\max|z_{100,s}-z_{100,base}|/z_{100,base}\times 100$ %. This quantity is a sensitivity measure, not a confidence interval.

2.4 Stationary extreme-value models

Four stationary approaches were used to estimate anomaly return levels. Normal Probability, Gumbel-I and GEV-LM were fitted to the same AMS; POT-GPD used the declustered threshold-exceedance sample. Normal and Gumbel-I were retained as transparent benchmarks; GEV-LM and GPD provide more flexible tail behaviour. A multi-model comparison is warranted because extrapolation from a few decades of data to 50- or 100-year return periods can be highly model dependent (Caruso and Marani, 2022; Fang et al., 2021; Goring et al., 2011; Hamdi et al., 2014). For the Normal distribution, the T-year return level followed the standard frequency-analysis form (Wan Deraman et al., 2017):

$$z_T = \bar{x} + s\,\Phi^{-1}\left(1 - \frac{1}{T}\right), \tag{2}$$

where $\bar{x}$ and s are the AMS sample mean and standard deviation, respectively and Φ^{-1} is the inverse standard-Normal distribution function. For Gumbel-I, location and scale were estimated by the method of moments. The moment relations and quantile form are standard results for the type-I extreme-value distribution (Coles, 2001; Landwehr et al., 1979):

$$\beta = \frac{s\sqrt{6}}{\pi}, \qquad \mu_G = \bar{x} - \gamma\beta, \tag{3}$$

where $\gamma = 0.5772$ is the Euler-Mascheroni constant. The corresponding T-year Gumbel-I return level was

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

$$z_T = \mu_G - \beta \ln\left[-\ln\left(1-\frac{1}{T}\right)\right]. \tag{4}$$

The stationary GEV-LM model was estimated from the AMS using probability-weighted moments and L-moments. Probability-weighted moments were formalised by Greenwood et al. (1979), their application to GEV parameter estimation was developed by Hosking et al. (1985) and the broader L-moment framework is described by Hosking (1990). With location μ, scale $\sigma > 0$ and shape ξ, the T-year return level for $\xi \neq 0$ was

$$z_T = \mu + \frac{\sigma}{\xi}\left[\left\{-\ln\left(1-\frac{1}{T}\right)\right\}^{-\xi} - 1\right], \tag{5}$$

with the Gumbel limiting form used as ξ approaches zero (Coles, 2001; Hosking et al., 1985). A negative fitted ξ implies a finite upper endpoint for the fitted statistical distribution, but we do not interpret that endpoint as a known physical maximum water level. For POT analysis, exceedances $Y = X-u$ were fitted with a generalized Pareto distribution using method-of-moments estimates consistent with Hosking and Wallis (1987) and the POT framework of Coles (2001) and Davison and Smith (1990). The annual rate of declustered peaks was $\lambda = N_p/N_\gamma$, where N_p is the number of retained peaks and N_γ is the number of accepted analysis years. Combining the GPD tail with a Poisson occurrence model gives, for $\xi \neq 0$,

$$z_T = u + \frac{\sigma_u}{\xi}\left[\left(\frac{\lambda}{-\ln\left(1-\frac{1}{T}\right)}\right)^{\xi} - 1\right], \tag{6}$$

where u is the selected threshold and σ_u is the GPD scale parameter. For $\xi = 0$, the continuous limit $z_t = u + \sigma_u \ln\{\lambda/[-\ln(1-1/T)]\}$ was used. This is the standard Poisson-GPD return-level formulation (Coles, 2001; Davison and Smith, 1990; Tebaldi et al., 2012). All stationary models were evaluated at 10-, 25-, 50- and 100-year return periods. AMS and POT estimates were treated as complementary; exact agreement was not expected because the two approaches rely on different samples, dependence assumptions and tail models.

2.5 Diagnostics, trend tests and non-stationary GEV-MLE

Normal, Gumbel-I and GEV-LM were evaluated on the same AMS sample to keep the observed-range fit comparison on a common sample. Ordered annual maxima were assigned Gringorten non-exceedance probabilities and empirical return periods (Gringorten, 1963):

$$F_i = \frac{i-0.44}{n+0.12}, \qquad T_i = \frac{1}{1-F_i}. \tag{7}$$

Theoretical Normal, Gumbel-I and GEV-LM quantiles were calculated at the same empirical probabilities and RMSE and MAE were used as descriptive measures of in-sample agreement. POT-GPD was assessed separately with GPD Q-Q R^2 and Q-Q RMSE because it uses a different exceedance sample. At each return period, multi-model/method spread was calculated as (maximum - minimum)/mean × 100 %. This study-specific quantity combines model-form differences, parameter-estimation differences and, for POT-GPD, a different extreme-value sample; we use it as a descriptive sensitivity measure, not as a formal statistical test (Fang et al., 2021; Hamdi et al., 2014).

Temporal structure in AMS was examined with the Mann-Kendall (MK) trend test, Sen median slope and Pettitt single change-point test at $\alpha=0.05$ (Mann, 1945; Pettitt, 1979; Sen, 1968). Because the analysis involved five stations, Benjamini-Hochberg false-discovery-rate (BH-FDR) adjusted q-values were reported separately for MK, Pettitt and likelihood-ratio tests (Benjamini and Hochberg, 1995). Lag-1 dependence was checked and trend-free prewhitening was used as an MK sensitivity analysis where serial structure could affect inference (Yue et al., 2002). Pettitt results were treated as evidence of an abrupt statistical change, not as proof of physical causation or as a reason to impose a step model. With only 25–34 annual maxima per station, the non-stationary GEV was intentionally parsimonious: M0 retained constant location, scale and shape, whereas M1 allowed only the location parameter to vary linearly with centred time.

$$\mu(t) = \mu_0 + \mu_1\left(t - t_{ref}\right). \tag{8}$$

Here t is calendar year and tref is the mean calendar year of the accepted AMS observations at that station; centring changes the intercept but not μ_1 or the fitted likelihood. AICc was calculated for M0 and M1 (Akaike, 1974; Hurvich and Tsai, 1989) and the nested models were compared with a one-degree-of-freedom likelihood-ratio test. M1 was selected as primary when the LR result remained significant after BH-FDR adjustment and the profile-likelihood interval for μ_1 excluded zero. Conditional T-year return levels were evaluated at the 2025 covariate value to show the effect of stationarity without extending the fitted trend beyond the observation period. These conditional quantiles should not be read as forecasts of an event expected once in the next T years while the distribution continues to evolve.

2.6 Uncertainty and reproducibility

Uncertainty in return levels was quantified with 1,000 nonparametric bootstrap resamples for M0, 1,000 parametric bootstrap samples from the fitted M1 and 2,000 nonparametric resamples of retained POT excesses. Percentile 95 % intervals were used (Davison and Hinkley, 1997). Reproducibility settings included base random seed 20260905,

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

station-specific deterministic substreams, bounded maximum-likelihood optimisation ($-0.8 \leq \xi \leq 0.8$) and storage of all bootstrap replicates and profile-likelihood evaluations. No final bootstrap replicate was missing. The primary POT bootstrap conditions on the observed peak count; a separate Poisson-count sensitivity bootstrap propagates event-frequency uncertainty. Profile likelihood was used independently for μ_1. Parameters, likelihoods, AICc, LR statistics, return levels and uncertainty summaries are supplied in the reviewer summary workbook and derived tables and the included verification script checks the reported core statistics.

2.7 External cyclone-event check of high-water anomalies

Five pre-specified Bay of Bengal cyclone dates were used for an external event-date check: Sidr (15 November 2007), Aila (25 May 2009), Amphan (20 May 2020), Yaas (26 May 2021) and Remal (26 May 2024) (India Meteorological Department, 2010, 2020, 2021, 2024; Lewis et al., 2014). For each station, the largest anomaly within D-1 to D+1 was compared with both the accepted AMS and the station-specific daily-anomaly distribution. An annual-maximum match was recorded only when that year's accepted maximum fell inside the event window. A D-2 to D+2 window was used as a timing-sensitivity check; missing observations were recorded as NA; no reconstruction was attempted. The exercise measures correspondence with externally specified cyclone dates; it is neither an independent surge dataset nor a reconstruction of meteorological storm surge. Figure 2 summarises the complete analytical sequence.

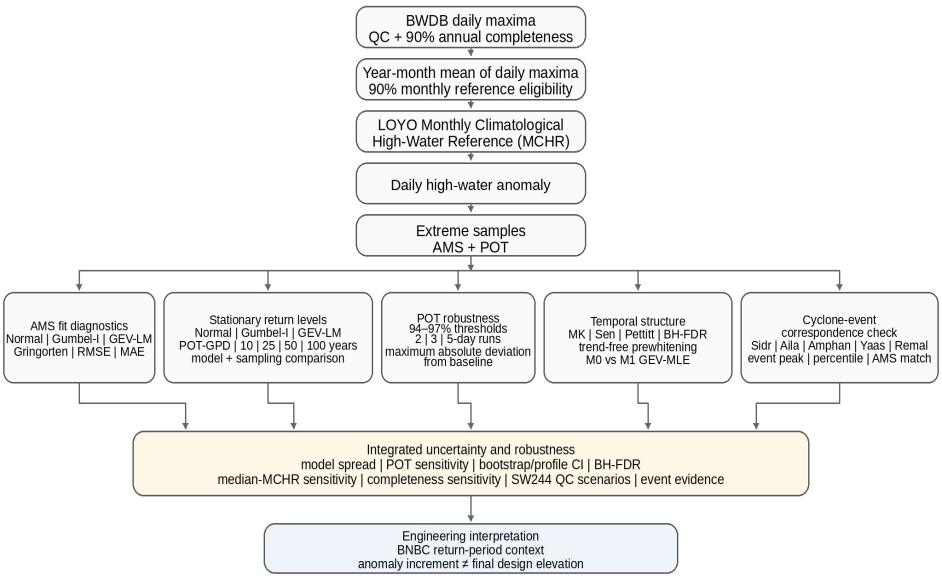

Figure 2. Analytical workflow from BWDB screening and MCHR anomaly construction to extreme-value modelling, robustness tests, external cyclone-event checks and engineering interpretation.

3. Results

3.1 Data screening and extreme samples

After completeness screening, the AMS contained 30, 32, 34, 32 and 25 annual maxima at SW107.2, SW20, SW241, SW243 and SW244, respectively (Table 1). The baseline POT analysis retained between 106 and 169 independent peaks, including 123 at SW107.2. Station-specific thresholds ranged from 0.616 m at SW20 to 0.920 m at SW243. Because only 25–34 annual maxima remain at each station, the 50- and 100-year return levels extend beyond the length of the observed AMS record.

Missingness in the main cyclone months (April–May and October–November) closely tracked missingness across the full record at every station, with differences of less than one percentage point. This check cannot recover events that were never observed, but it does not indicate a disproportionate concentration of gaps during the principal cyclone seasons.

Stationary GEV-LM shape estimates were negative at all stations, from $\xi = -0.066$ to -0.261, while baseline GPD shape estimates ranged from -0.139 at SW241 to $+0.122$ at SW107.2 (Table 2). The two sets of shape parameters should not be compared directly because AMS and POT rely on different samples and estimators. POT-GPD Q-Q R^2 values ranged from 0.956 at SW244 to 0.994 at SW20, with the weakest upper-tail agreement at SW244.

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

Table 2. Key stationary GEV-LM and baseline POT-GPD diagnostics.

Station	AMS n	GEV-LM ξ	GEV-LM RMSE (m)	POT peaks	u (m)	GPD ξ	GPD Q-Q R^2
SW107.2	30	-0.066	0.065	123	0.662	0.122	0.989
SW20	32	-0.116	0.048	166	0.616	-0.008	0.994
SW241	34	-0.202	0.072	169	0.636	-0.139	0.988
SW243	32	-0.261	0.078	106	0.920	-0.055	0.987
SW244	25	-0.185	0.112	111	0.678	0.076	0.956

3.2 Seasonal variation in high water and anomaly patterns

The monthly series at all five stations show a clear recurring seasonal high-water cycle (Fig. 3). The station-specific LOYO MCHR follows the seasonal envelope of the monthly mean of daily maxima, while episodic departures remain visible. Positive departures vary in both timing and magnitude among stations, including within the Rupsa-Pasur system. Months that failed the reference-input completeness criterion remain as gaps and were not interpolated.

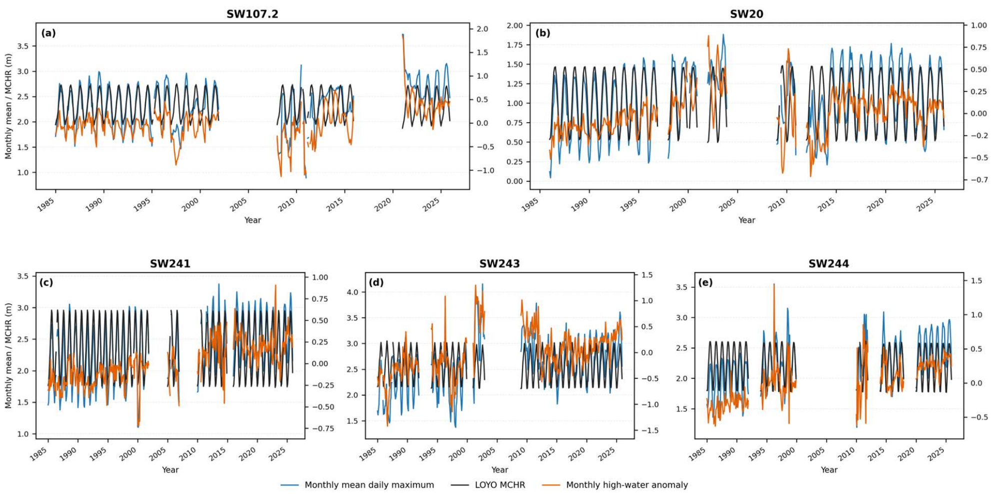

Figure 3. Monthly mean of daily maximum water level (blue), leave-one-year-out Monthly Climatological High-Water Reference (LOYO MCHR; black), and monthly high-water anomaly (orange) for the five stations. Gaps mark months that failed the reference-input completeness criterion. Station-specific secondary axes are used only for visual clarity; all statistical analyses use the original daily anomaly values in metres.

Figure 3 provides a visual check of MCHR behaviour. It follows the recurring monthly high-water pattern without moving unusual observations toward a common absolute level, so episodic positive departures remain available for extreme-value analysis. The resulting anomaly is a relative measure of unusually high water for a station and calendar month, not an estimate of meteorological storm surge.

That distinction is important for these data. A complete predicted astronomical-tide series is unavailable, so the MCHR does not separate tide from the observed water level. Individual anomalies may contain contributions from meteorological forcing, unusual tidal conditions, tide–surge interaction, freshwater discharge, rainfall, background water-level variability and local estuarine effects. The variable is therefore used throughout as a seasonally adjusted high-water indicator, not as a skew surge or non-tidal residual.

The secondary axes in Fig. 3 are used only to make the temporal anomaly patterns easier to see; they do not rescale any values used in the statistical analysis. AMS, POT, trend and return-level calculations all use the original daily anomalies in metres. Figure 3 is best read as evidence that the seasonal reference behaves as intended and that unusual high-water departures remain distinguishable despite strong seasonality and incomplete records.

3.3 Empirical AMS fits and model-performance diagnostics

Empirical annual maxima were compared with fitted Normal, Gumbel-I and GEV-LM distributions at common Gringorten plotting positions (Fig. 4). All three AMS models represented the centre of the observed distributions reasonably well. Separation between fitted curves became more visible in the upper tail, where a small number of large observations exert greater leverage on the fitted shape. The divergence was clearest at SW107.2 and SW244.

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

SW20 and SW241 showed closer agreement across most of the observed probability range and SW243 displayed broadly similar behaviour among the three fits. Because Fig. 4 is based on empirical plotting positions, it is used as an observed-range goodness-of-fit diagnostic rather than as proof that any one model extrapolates most reliably to long return periods.

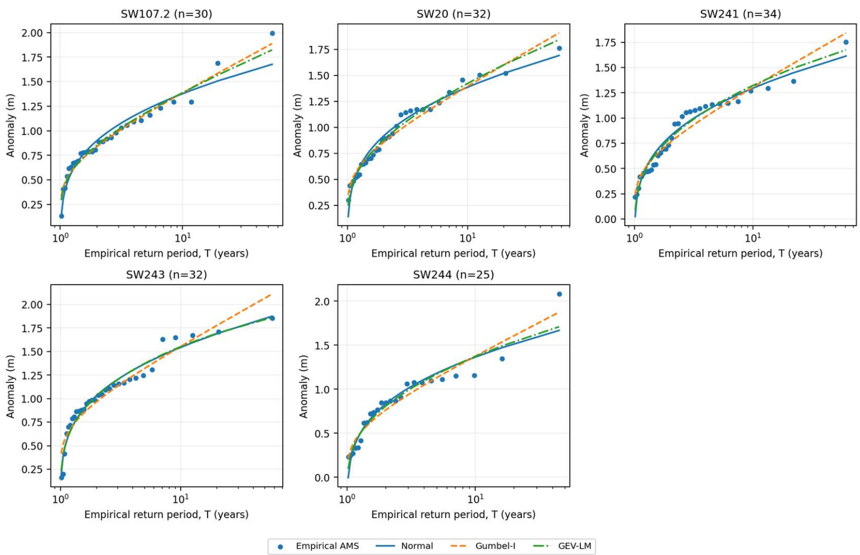

Figure 4. Empirical annual maxima and fitted Normal, Gumbel-I and GEV-LM distributions at common Gringorten return periods. Points denote accepted annual maxima; fitted curves assess agreement over the observed probability range and n is the accepted AMS size.

The numerical diagnostics confirmed that model performance varies by station (Fig. 5a, b). Gumbel-I had the lowest RMSE at SW107.2 and SW244, GEV-LM at SW20 and SW241 and Normal at SW243. Across all five stations, GEV-LM gave the lowest average RMSE and MAE (75.1 and 58.9 mm), compared with 85.6 and 66.5 mm for Normal and 85.9 and 67.6 mm for Gumbel-I. Across the network, GEV-LM was the most consistently competitive AMS model, although no distribution performed best everywhere.

POT-GPD Q-Q agreement was strong (R^2=0.956–0.994), yet good agreement within the observed exceedances did not remove uncertainty in the far tail. At the 100-year return period, the four stationary approaches differed by 15.6–21.5 %, with the largest spread at SW241. SW244 combined the weakest POT Q-Q fit with a notable inter-model spread, illustrating that an adequate in-sample fit and a stable extrapolation are not the same thing.

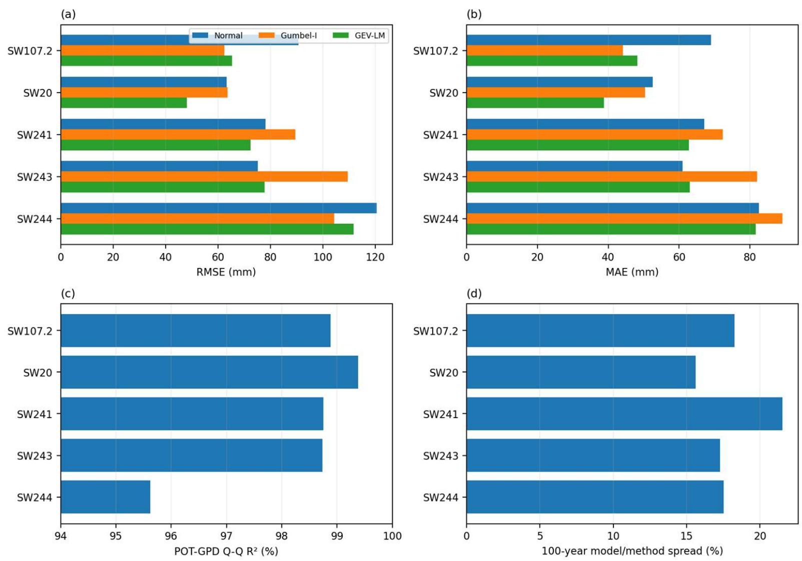

Figure 5. Station-wise model-performance and tail diagnostics: (a) common-sample AMS RMSE, (b) common-sample AMS MAE, (c) POT-GPD Q-Q R^2 and (d) 100-year multi-model/method spread. Normal, Gumbel-I and GEV-LM are compared on the same AMS sample; POT-GPD is assessed separately because it uses the declustered threshold-exceedance sample.

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

3.4 Stationary return levels and multi-model/method spread

At short return periods, the stationary return-level curves remain fairly close; they separate progressively as the return period increases (Fig. 6). This pattern is consistent with the far-tail disagreement in Fig. 5d. Table 3 further shows that the way the models diverge is not uniform from one station to another.

At 100 years, Gumbel-I gave the highest stationary estimate at SW20, SW241, SW243 and SW244, while POT-GPD was highest at SW107.2. For SW107.2, the 100-year estimates were 1.767 m (Normal), 2.069 m (Gumbel-I), 1.969 m (GEV-LM) and 2.129 m (POT-GPD), corresponding to an 18.3 % spread. Thus the model giving the largest estimate changes with location rather than following a single network-wide pattern.

SW244 illustrates this dependence most clearly. Its retained anomaly of 2.081 m on 18 April 1996 is higher than the fitted 100-year Normal (1.799 m), GEV-LM (1.850 m) and POT-GPD (1.906 m) estimates, but lower than the Gumbel-I estimate of 2.136 m. That observation does not invalidate the models: a 100-year return level is a probabilistic quantile, not a physical ceiling. The contrast shows how much an individual extreme can influence tail inference when long return periods are estimated from a short record.

Table 3. 100-year stationary high-water anomaly return levels (m) and multi-model/method spread.

Station	T (yr)	Normal	Gumbel-I	GEV-LM	POT-GPD	Spread (%)
SW107.2	100	1.767	2.069	1.969	2.129	18.3
SW20	100	1.771	2.069	1.957	1.836	15.6
SW241	100	1.686	1.988	1.750	1.609	21.5
SW243	100	1.962	2.285	1.932	2.002	17.3
SW244	100	1.799	2.136	1.850	1.906	17.5

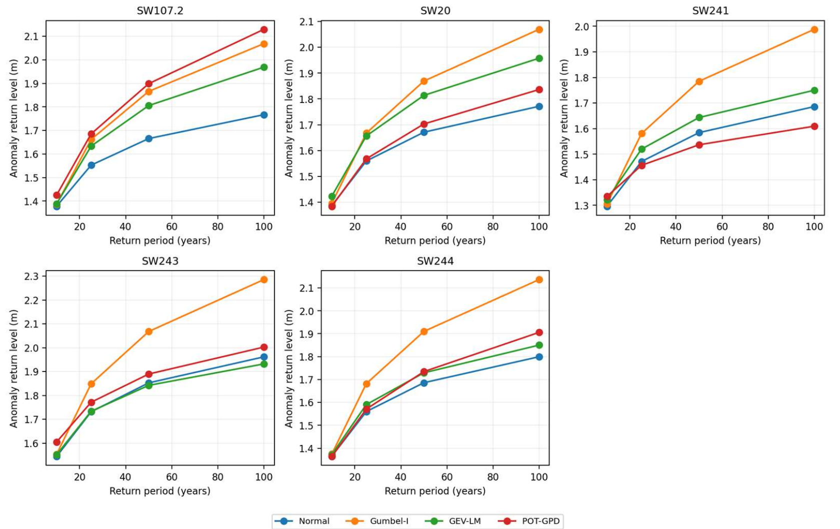

Figure 6. Station-specific stationary high-water anomaly estimates from different models across 10-, 25-, 50-, and 100-year return periods.

3.5 POT sensitivity to threshold and declustering choices

Across the 12 threshold/run combinations, the largest absolute departure from the baseline 100-year POT-GPD estimate was 2.71 % at SW107.2, 4.35 % at SW20, 5.30 % at SW241 and 4.98 % at SW243. SW244 behaved differently: its estimates ranged from 1.765 to 2.263 m, giving a maximum absolute deviation of 18.75 % from the 1.906 m baseline.

For most stations, the choice of POT threshold and declustering rule had only a modest effect, but SW244 was much more sensitive. The contrast is particularly clear within the Rupsa-Pasur system, where SW241 and SW243 remained comparatively stable while nearby SW244 did not. This result argues against applying one POT sampling assumption uniformly across neighbouring tidal-river stations and supports station-specific sensitivity checks when the upper tail is extrapolated. The baseline estimates and corresponding sensitivity ranges are summarised in Fig. 7.

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

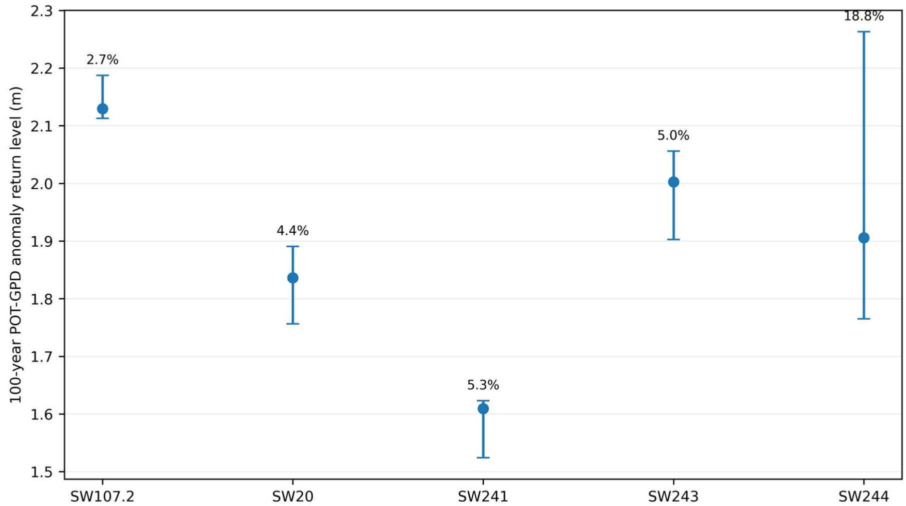

Figure 7. Baseline 100-year POT-GPD estimate and the range obtained from 94th–97th percentile thresholds and 2-, 3- and 5-day run lengths.

3.6 Non-stationarity, uncertainty and robustness

The primary MK tests indicated increasing trends at SW107.2 ($p=0.0457$), SW20 ($p=0.0010$), SW241 ($p<0.0001$) and SW244 ($p=0.0006$), with Sen slopes of 9.00, 18.06, 24.70 and 17.57 mm/year, respectively. SW243 was not significant ($p=0.2118$; 6.46 mm/year). After BH-FDR adjustment, the MK result at SW107.2 became marginal ($q=0.057$), while the other three increasing trends remained significant. Pettitt q-values stayed below 0.05 at SW107.2, SW241 and SW244; SW20 was borderline ($q=0.051$). Trend-free prewhitening reproduced the same four-station versus SW243 pattern, suggesting that the main conclusion was not an artefact of lag-1 dependence.

Likelihood-based GEV-MLE comparison provided the clearest network-level evidence. M1 LR p-values were 0.0288, 0.00807, 4.88×10^{-13} and 5.76×10^{-6} for SW107.2, SW20, SW241 and SW244, with BH-FDR q-values of 0.0360, 0.0135, 2.44×10^{-12} and 1.44×10^{-5}; the profile-likelihood intervals for μ_1 excluded zero. SW243 remained unsupported ($p=q=0.0644$; profile interval includes zero). M1 is retained as the primary model at four stations and M0 at SW243, although the strength and form of evidence are not identical among sites. The complete station-wise trend, change-point, likelihood-ratio, and model-selection results are summarised in Table 4.

Table 4. Trend, change-point and GEV-MLE likelihood-based stationarity assessment.

Station	MK p [BH q]	Sen slope mm/yr [95 % CI]	Pettitt year p [BH q]	LR p [BH q]	M1 μ_1 mm yr^{-1} [profile 95 % CI]	Selected model
SW107.2	0.0457 [0.0571]	9.00 [0.27, 18.88]	2008 0.0262 [0.0436]	0.0288 [0.0360]	11.56 [1.36, 22.46]	M1 primary
SW20	0.0010 [0.0017]	18.06 [8.42, 28.31]	1993 0.0409 [0.0512]	0.00807 [0.0135]	12.71 [3.61, 21.81]	M1 primary
SW241	<0.0001 [<0.0001]	24.70 [19.45, 30.07]	2001 <0.0001 [<0.0001]	4.88×10^{-13} [2.44×10^{-12}]	23.97 [19.87, 29.27]	M1 primary
SW243	0.2118 [0.2118]	6.46 [-3.61, 16.94]	1998 0.1551 [0.1551]	0.0644 [0.0644]	11.14 [-0.66, 20.74]	M0 primary; M1 sensitivity
SW244	0.0006 [0.0015]	17.57 [9.83, 25.66]	1995 0.0057 [0.0142]	5.76×10^{-6} [1.44×10^{-5}]	16.56 [11.86, 21.96]	M1 primary

The practical effect of non-stationarity was not uniform across the network (Fig. 8). Conditional on the 2025 covariate value, M0 and M1 100-year estimates were 1.913 and 2.010 m at SW107.2, 1.890 and 1.894 m at SW20, 1.753 and 1.781 m at SW241 and 2.005 and 2.499 m at SW244. For SW243, M1 is shown only as a sensitivity result because M0 remains the selected model. Yet evidence of temporal change does not translate into the same engineering-scale shift in return level at every station.

Across the five stations, trend detection and engineering effect are separate questions. Both the statistical support for non-stationarity and the size of its influence on return levels vary markedly from site to site.

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

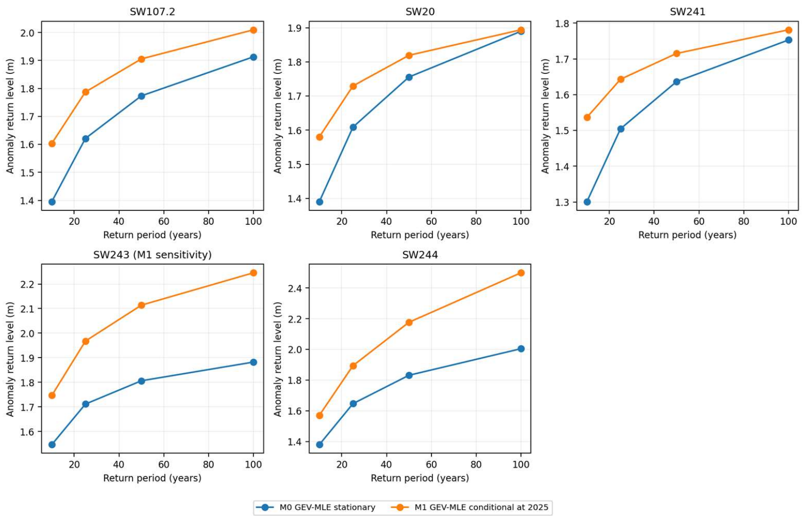

Figure 8. M0 stationary and M1 conditional GEV-MLE return levels evaluated at the 2025 covariate value. SW243 M1 is shown as sensitivity only because M0 remains selected.

Uncertainty intervals also differed substantially among stations and modelling approaches (Fig. 9). Most bootstrap intervals overlapped and they widened as the return period increased. SW244 remained the least constrained station, particularly for the conditional M1 estimate, showing that parameter uncertainty can be as important as differences among point estimates.

The influence of non-stationarity was likewise uneven. M1-2025 and M0 GEV-MLE produced nearly identical 100-year estimates at SW20 and SW241, a larger separation at SW107.2 and the largest difference at SW244. At SW243, the higher M1 estimate is retained only as a sensitivity result because neither the likelihood-ratio test nor the profile interval supports selection of the non-stationary model. POT-GPD estimates generally fell within the uncertainty ranges of the GEV-based estimates, indicating that several differences among point estimates are small relative to their associated uncertainty.

The M0 values in Fig. 9 are maximum-likelihood GEV estimates and differ from the GEV-LM values used in the four-method stationary comparison. At 100 years, the difference between MLE and L-moment estimates ranged from 0.15 % at SW241 to 8.34 % at SW244 (Supplementary Table S6). Estimation method is another contributor to the overall uncertainty budget.

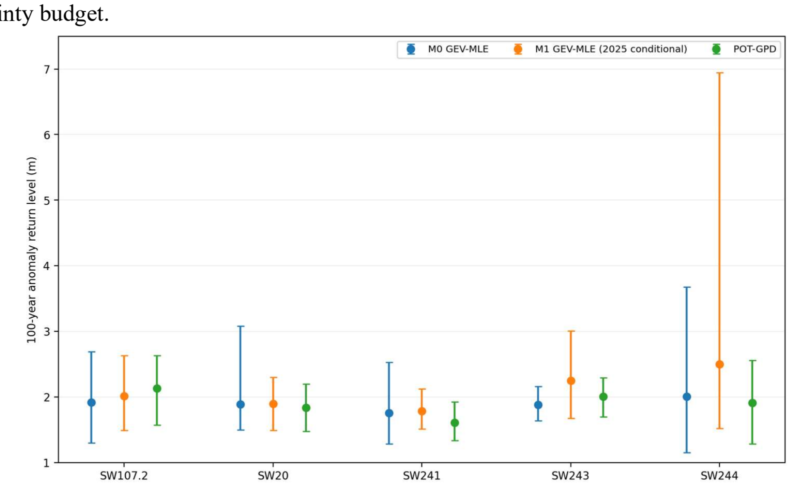

Figure 9. M0 GEV-MLE, M1 conditional-at-2025 GEV-MLE and POT-GPD 100-year estimates with 95 % intervals. POT-GPD bars show the bootstrap sensitivity that also propagates peak-count uncertainty; SW243 M1 is sensitivity only.

3.7 External cyclone-event check of high-water anomalies

Cyclone-related high-water responses differed across the station network (Fig. 10). Sidr could not be assessed because observations were unavailable within its event window. Aila could be evaluated at SW107.2 and SW243, where the 1.710 m anomaly at SW243 coincided with the 2009 annual maximum. Amphan was evaluable at four stations and matched the annual maximum at SW20 and SW243.

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

Yaas was evaluable at all five stations and produced anomalies at or above the 99th percentile at four of them; it also matched the annual maximum at SW20 and SW243. Remal produced the broadest network response, matching the annual maximum at SW107.2, SW20, SW241 and SW244 and reaching at least the 95th percentile at every station. Aila, Amphan, Yaas and Remal provide event-scale evidence that the anomaly responds to documented episodes of high water, while also showing that the response is spatially heterogeneous.

Expanding the event window from D−1/D+1 to D−2/D+2 did not alter any evaluable peak or annual-maximum match. Even so, this analysis should be read as correspondence with externally specified cyclone dates, not as independent validation of surge magnitude, because the anomaly is derived from the same BWDB water-level observations used elsewhere in the study.

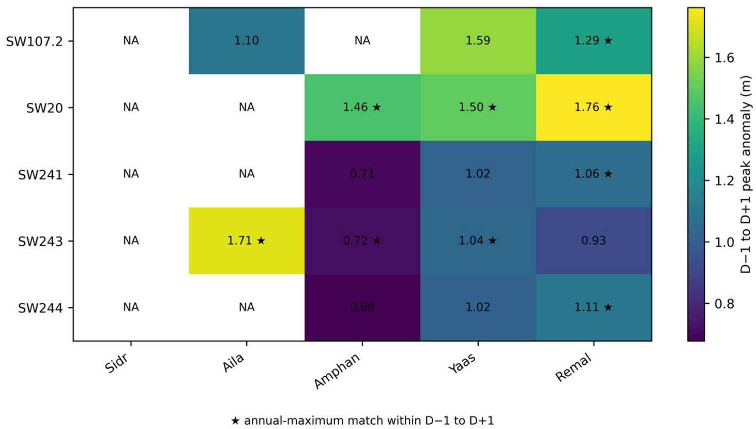

Figure 10. External cyclone event-window peak anomalies (D−1 to D+1). Stars mark station-event pairs where the accepted annual maximum occurred within the event window; NA denotes unavailable observations. The D−2 to D+2 timing check produced the same peak for every evaluable pair.

4. Discussion

4.1 Model performance, extrapolation uncertainty and non-stationarity

The first result is that observed-range fit and far-tail inference are not interchangeable. All three AMS distributions represented the accepted maxima reasonably well, but their ranking varied among stations and the 100-year estimates differed more than the in-sample error differences might suggest. GEV-LM was the most competitive model on average, although its advantage was not consistent across all stations. For engineering interpretation, the absence of a universally dominant fit is more informative than declaring a single network-wide winner.

Hamdi et al. (2014) provide a useful methodological benchmark rather than a numerical one. In French Atlantic storm-surge records, retaining more upper-tail observations through POT or r-largest sampling reduced some parameter and return-level uncertainty, while influential extremes still affected the right tail. The same tension appears here: POT-GPD described the retained exceedances well, yet SW244 had the weakest Q-Q agreement and the greatest sensitivity to threshold and declustering choices. Using more tail observations therefore did not remove the need to examine sampling assumptions; in this dataset, AMS and POT are better treated as complementary evidence than as a simple hierarchy of reliability.

Fang et al. (2021) reached a related conclusion from 13 tide gauges along coastal China, where extreme-level estimates depended on both the sampling strategy and the fitted distribution. Their use of block maxima, POT sampling and a three-day declustering rule provides a useful structural comparison with the present analysis. Our 15.6–21.5 % spread among stationary 100-year estimates reflects the same problem in a different setting. The magnitudes should not be equated because Fang et al. analysed coastal extreme sea levels after accounting for mean-level variability, whereas this study uses station-specific anomalies from inland tidal-river records. What carries across the two studies is that acceptable observed-range fit can still conceal substantial divergence after extrapolation.

Record length reinforces this uncertainty. Each station contributes only 25–34 accepted annual maxima even though 50- and 100-year quantiles are required. Goring et al. (2011) showed in a tide-dominated New Zealand setting that short-record estimates can shift markedly depending on whether major events happen to fall within the available period. Our records are different in setting, but they are similarly information-limited in the far tail. The widening separation of the fitted curves at longer return periods is therefore interpreted mainly as a consequence of limited tail information, rather than simply as disagreement among formulas.

Non-stationarity adds another source of uncertainty. The combined MK, Pettitt and likelihood results support temporal change at four stations, although the multiplicity-adjusted evidence is not equally strong for every diagnostic. SW107.2 is marginal after BH adjustment of the MK test but remains supported by Pettitt and LR evidence; SW20 is borderline after BH adjustment of the Pettitt test but remains strongly supported by MK and LR results. Reading the diagnostics together is therefore more defensible than allowing any single p-value to determine the interpretation.

The primary climatological reference is LOYO by construction, so a target year does not contribute to its own calendar-month MCHR. Robustness was examined through alternative aggregation and completeness rules rather than by repeating another nominal LOYO step. These checks assess the sensitivity of the anomaly definition while preserving the safeguard against self-influence.

The reference choice was generally robust. Replacing the mean-based LOYO MCHR with a median-based version changed 100-year GEV-LM estimates by −0.24 % to +2.79 % and POT-GPD estimates by −3.48 % to −1.31 %. The 85 % completeness setting produced the same accepted AMS and key return levels as the 90 % setting. A 95 % threshold removed additional years and had its largest effect at SW243 (GEV-LM −19.6 %; POT-GPD −14.9 %). That sensitivity is retained as an explicit limitation rather than obscured by the primary 90 % choice.

4.2 Spatial variability, estuarine position and regional context

The five stations span three tidal-river systems rather than forming a single coast-to-inland transect. SW107.2 belongs to the Baleswar-Haringhata system, SW20 to the Barisal-Buriswar system, and SW241, SW243 and SW244 to the Rupsa-Pasur system. Relative inland position can be interpreted most defensibly within the Rupsa-Pasur group; comparing coastal distance across all five stations would mix different hydraulic systems.

Within the Rupsa-Pasur system, the long-return-period estimates do not decrease monotonically with inland position. The more inland SW241 station produced the lowest 100-year stationary estimates of the three stations under all four approaches: 1.686 m for Normal, 1.988 m for Gumbel-I, 1.750 m for GEV-LM and 1.609 m for POT-GPD. This is compatible with some inland attenuation of unusually high-water departures. However, the intermediate SW243 station exceeded the more seaward SW244 under all four stationary approaches; for example, Gumbel-I gave 2.285 versus 2.136 m and POT-GPD 2.002 versus 1.906 m. Distance from the open coast alone therefore cannot explain the observed spatial ordering.

The hydrodynamic results of Khan et al. (2022) help interpret, rather than validate, the non-monotonic ordering within the Rupsa-Pasur system. Their synthetic-cyclone simulations found relatively little variation in return-period water levels over roughly the first 25 km of the Pussur estuary, with greater variation farther upstream. Our observational anomaly is a different quantity and cannot be compared numerically with those simulations, but both studies suggest that distance from the estuary mouth alone is an incomplete descriptor of extreme high water. Local estuarine controls are likely to matter.

Local channel geometry, bathymetry, convergence, friction, tidal amplification or damping, freshwater discharge, embankments and morphology could all contribute to the non-monotonic pattern. The trend and POT results also indicate a lack of a simple coastal-distance explanation. Sen slopes at SW241, SW243 and SW244 are 24.70, 6.46 and 17.57 mm/year, while maximum POT sensitivities are 5.30 %, 4.98 % and 18.75 %, respectively. SW241 and SW244 support M1, whereas SW243 retains M0. Neither temporal change nor POT sensitivity follows a simple downstream-to-upstream sequence.

This is also consistent with cyclone-sensitivity studies for the Bay of Bengal. Lewis et al. (2014) showed that variations in cyclone intensity and track can materially alter simulated peak surge and inland inundation. For the present station network, the implication is qualitative: neighbouring gauges exposed to the same named cyclone need not experience comparable anomaly magnitudes because storm characteristics interact with tidal phase and local hydraulic setting.

SW107.2 and SW20 should be considered separately from the Rupsa-Pasur coast-to-inland sequence. The Baleswar-Haringhata and Barisal-Buriswar systems have different channel configurations and exposure pathways, so contrasts with the Rupsa-Pasur stations cannot be reduced to distance from a single coastline. A station-specific interpretation is therefore preferable to transferring one distribution or return level across the entire southwest region.

The published Bangladesh studies are most useful here as scale context, not as validation. Lee (2013) estimated a 100-year meteorological storm surge of 1.75 m at Hiron Point and a corresponding extreme level of 2.09 m after adding the estimated 2050 sea-level-rise contribution. The stationary 100-year anomalies in this study span approximately 1.61–2.29 m, but the variables are fundamentally different. Lee separated tidal motion from high-frequency observations, whereas MCHR anomalies can retain tidal, freshwater, rainfall and local estuarine influences. Similar magnitudes should not be interpreted as equivalent quantities.

Jisan et al. (2018) reinforce the importance of the background state. In their Sidr simulations for Barisal, sea-level-rise scenarios of 0.26 m and 0.54 m increased the modelled surge level by approximately 14 % and 29 %, respectively. That result supports the broader point that a changing background level can amplify cyclone-related high water, but it does not identify the cause of the trends detected here. Attribution would require independent evidence on sea level, subsidence, freshwater discharge and morphology.

The cyclone-date check adds a different kind of evidence. Aila, Amphan, Yaas and Remal aligned with annual maxima or high-percentile anomalies at several stations, but not consistently across the network. This supports the

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

sensitivity of the anomaly to major high-water episodes while also showing that event response is spatially heterogeneous. Because the event check and the extreme-value analysis use the same BWDB water-level observations, the comparison is best interpreted as correspondence rather than independent validation.

4.3 Engineering implications and limitations

For engineering use, the estimated return levels should be treated as increments above a station-specific seasonal reference, not as final design water elevations. A defensible design level would require a compatible absolute datum-based baseline together with allowances for relative sea-level rise or subsidence, waves, freshwater and rainfall effects, hydrodynamic loading, statistical uncertainty and freeboard. These components should not automatically be added as if they were independent, because interactions among mean level, tide and meteorological forcing can change the resulting extreme water level.

BNBC 2020 provides a national return-period context: for flood and surge loading, essential and hazardous facilities use 100-year values, whereas other structures use 50-year values (Government of the People's Republic of Bangladesh, 2021). The anomaly estimates reported here are useful for comparative engineering screening, but they are not BNBC absolute design-water or surge elevations.

SW244 warrants the greatest caution. It has the shortest AMS, the largest retained anomaly, the weakest POT Q-Q agreement, the greatest sensitivity to POT sampling, wide uncertainty intervals and unresolved gauge/datum history. Replacing the documented 3.39 m QC value on 26 June 1996 with the raw 5.88 m observation raises the 100-year stationary estimates substantially; excluding 1995–1996 lowers them, although the increasing trend and M1 likelihood evidence remain. The magnitude inferred at SW244 depends strongly on data-history treatment even though evidence of temporal change persists (Supplementary Table S1).

Several limitations warrant emphasis. Long-period estimates are extrapolative because only 25–34 annual maxima are available. Good observed-range fit does not guarantee stability in the far tail, POT results depend on threshold, declustering and peak-rate assumptions, and the non-stationary model allows only a linear change in GEV location. Pettitt change points may reflect abrupt, nonlinear or undocumented station-history effects and are used diagnostically. The 95 % completeness sensitivity at SW243 and the QC scenarios at SW244 show that screening and station history can matter as much as distribution choice. Most importantly, the analysed variable is a seasonally adjusted high-water anomaly rather than a conventional skew surge or non-tidal residual; its temporal trend should not be interpreted directly as a trend in mean sea level or meteorological surge.

Model choice, record length, POT sampling, temporal non-stationarity and local estuarine setting all influence the estimation of extreme high water. For short tidal-river records, engineering interpretation is stronger when these uncertainty sources are reported openly instead of being compressed into one deterministic return level.

5. Conclusions

Across five southwest tidal-river stations, the analysis shows that uncertainty in extreme high water arises from several linked sources rather than from model choice alone. The LOYO MCHR provides a station- and month-specific seasonal reference for unusual daily maxima, while AMS, POT, non-stationary GEV, bootstrap/profile uncertainty, data-quality sensitivity and cyclone-date checks test how stable the resulting tail estimates are. The analysed quantity remains a high-water anomaly rather than a reconstructed storm-surge residual.

No stationary AMS distribution dominated the network. GEV-LM gave the best average observed-range fit, yet the four stationary approaches diverged by 15.6-21.5 % at the 100-year return period. POT estimates were comparatively stable at four stations but much more sensitive at SW244, where the maximum deviation reached 18.75 %. At several sites, interval width was comparable with or larger than the separation among point estimates, so model ranking alone gives an incomplete picture of uncertainty.

The stations did not show the same pattern of temporal change. Combined trend, change-point and likelihood evidence supported a time-varying GEV location at SW107.2, SW20, SW241 and SW244, whereas SW243 remained stationary in the primary analysis. The practical effect of M1 was small at SW20 and SW241 but pronounced at SW244. Sensitivity to the median MCHR was limited (at most 2.79 % for GEV-LM and 3.48 % for POT-GPD), whereas the stricter 95 % completeness rule at SW243 and the alternative QC treatment at SW244 produced much larger changes.

Cyclone correspondence placed those statistical results in event context. Aila, Amphan, Yaas and particularly Remal aligned with annual maxima or high anomaly percentiles at multiple stations. The non-uniform response of the nearby Rupsa-Pasur stations, including SW241, SW243 and SW244, shows that shared cyclone exposure does not imply common upper-tail behaviour or a common stationarity pattern.

For engineering screening, the return levels should be treated as anomaly increments above a station-specific seasonal reference rather than as complete design elevations. A design water level would still require a suitable datum-

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

based baseline together with appropriate physical and safety allowances. For design interpretation, these uncertainty sources should be considered together instead of relying on a single extrapolated return level.

Supplementary information

Supplementary material accompanies this article and includes Tables S1-S9, Figures S1-S2 and a reproducibility/provenance note. It reports SW244 QC scenarios, mean-versus-median LOYO-MCHR sensitivity, multiplicity and serial-dependence diagnostics, complete M0/M1 likelihood results, completeness-screening sensitivity, GEV estimator sensitivity, POT threshold/declustering sensitivity, cyclone-event correspondence and 100-year uncertainty summaries.

Code availability

A verification script and derived result tables are provided in the Supplement. The script checks the reported outputs against the distributed derived results; it does not reconstruct the full analysis from restricted raw BWDB observations.

Data availability

The water-level observations analysed here were obtained from the Bangladesh Water Development Board (BWDB) and remain subject to the provider's access and reuse conditions. Processed year-month means of daily maxima, LOYO MCHR and anomaly series, AMS and POT datasets and cyclone-event correspondence summaries are available from the corresponding author on reasonable request, subject to BWDB data-sharing restrictions. Raw BWDB station records should not be redistributed without the provider's permission.

Author contribution

Sujoy Biswas: Conceptualization, Methodology, Software, Formal analysis, Data curation, Validation, Visualization, Writing - original draft. Md. Shahjahan Ali: Conceptualization, Supervision, Writing - review & editing. Arpon Roy, Nikhil Ronjon Roy and Juganto Roy: Writing - review & editing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors acknowledge the Bangladesh Water Development Board for providing the historical water-level records used in this study and the Department of Civil Engineering, Khulna University of Engineering & Technology, for academic support.

During the preparation of this manuscript, the authors used OpenAI ChatGPT to assist with language refinement, clarity and organization. All final figures and tables were prepared and finalized by the authors using conventional software. The underlying data, data processing, model development, statistical analyses, numerical results, references, scientific interpretations, and final manuscript content were reviewed and verified by the authors. The authors retain full responsibility for the accuracy, integrity, and originality of the work.

Financial support

Not applicable.

References

Akaike, H.: A new look at the statistical model identification, IEEE Trans. Automat. Contr., 19, 716–723, https://doi.org/10.1109/TAC.1974.1100705, 1974.

Arns, A., Wahl, T., Wolff, C., Vafeidis, A. T., Haigh, I. D., Woodworth, P., Niehüser, S., and Jensen, J.: Non-linear interaction modulates global extreme sea levels, coastal flood exposure, and impacts, Nat. Commun., 11, 1918, https://doi.org/10.1038/s41467-020-15752-5, 2020.

Barroso, A., Wahl, T., Enriquez, A., Morim, J., and Dangendorf, S.: Changes in the Seasonal Cycle of Storm Surge Along the Global Coastline, J. Geophys. Res. Oceans, 130, e2024JC022187, https://doi.org/10.1029/2024JC022187, 2025.

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, J. R. Stat. Soc. Series B Stat. Methodol., 57, 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x, 1995.

Bernard, A., Long, N., Becker, M., Khan, J., and Fanchette, S.: Bangladesh's vulnerability to cyclonic coastal flooding, Natural Hazards and Earth System Sciences, 22, 729–751, https://doi.org/10.5194/nhess-22-729-2022, 2022.

Bernardara, P., Mazas, F., Weiss, J., Andreewsky, M., Kergadallan, X., Benoît, M., and Hamm, L.: On the Two-Step Threshold Selection for Over-Threshold Modelling, Coastal Engineering Proceedings, 1, management.42, https://doi.org/10.9753/icce.v33.management.42, 2012.

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

Caballero-Megido, C., Hillier, J., Wyncoll, D., Bosher, L., and Gouldby, B.: Technical note: comparison of methods for threshold selection for extreme sea levels, J. Flood Risk Manag., 11, 127–140, https://doi.org/10.1111/jfr3.12296, 2018.

Caruso, M. F. and Marani, M.: Extreme-coastal-water-level estimation and projection: a comparison of statistical methods, Natural Hazards and Earth System Sciences, 22, 1109–1128, https://doi.org/10.5194/nhess-22-1109-2022, 2022.

Cid, A., Wahl, T., Chambers, D. P., and Muis, S.: Storm Surge Reconstruction and Return Water Level Estimation in Southeast Asia for the 20th Century, J. Geophys. Res. Oceans, 123, 437–451, https://doi.org/10.1002/2017JC013143, 2018.

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

D'Arcy, E., Tawn, J. A., Joly, A., and Sifnioti, D. E.: Accounting for Seasonality in Extreme Sea Level Estimation, Ann. Appl. Stat., 17, 3500–3525, https://doi.org/10.1214/23-AOAS1773, 2023.

Davison, A. C. and Hinkley, D. V.: Bootstrap Methods and their Application, Cambridge University Press, https://doi.org/10.1017/CBO9780511802843, 1997.

Davison, A. C. and Smith, R. L.: Models for Exceedances Over High Thresholds, J. R. Stat. Soc. Series B Stat. Methodol., 52, 393–425, https://doi.org/10.1111/j.2517-6161.1990.tb01796.x, 1990.

Fang, J., Wahl, T., Zhang, Q., Muis, S., Hu, P., Fang, J., Du, S., Dou, T., and Shi, P.: Extreme sea levels along coastal China: uncertainties and implications, Stochastic Environmental Research and Risk Assessment, 35, 405–418, https://doi.org/10.1007/s00477-020-01964-0, 2021.

Goring, D. G., Stephens, S. A., Bell, R. G., and Pearson, C. P.: Estimation of Extreme Sea Levels in a Tide-Dominated Environment Using Short Data Records, J. Waterw. Port Coast. Ocean Eng., 137, 150–159, https://doi.org/10.1061/(asce)ww.1943-5460.0000071, 2011.

Government of the People's Republic of Bangladesh: Bangladesh National Building Code (BNBC) 2020, Dhaka, https://www.dpp.gov.bd/bgpress/index.php/document/extraordinary_gazettes_monthly/2021-02-01 (last access: 12 September 2026), 2021.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R.: Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form, Water Resour. Res., 15, 1049–1054, https://doi.org/10.1029/WR015i005p01049, 1979.

Gringorten, I. I.: A plotting rule for extreme probability paper, J. Geophys. Res., 68, 813–814, https://doi.org/10.1029/JZ068i003p00813, 1963.

Guan, N., Chen, H., Xu, Y., Luan, Y., and Tan, Z.: Design water level calculation and analysis of Chittagong power plant project in Bangladesh, IOP Conf. Ser. Earth Environ. Sci., 621, 012152, https://doi.org/10.1088/1755-1315/621/1/012152, 2021.

Hamdi, Y., Bardet, L., Duluc, C. M., and Rebour, V.: Extreme storm surges: A comparative study of frequency analysis approaches, Natural Hazards and Earth System Sciences, 14, 2053–2067, https://doi.org/10.5194/nhess-14-2053-2014, 2014.

Hosking, J. R. M.: L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics, Journal of the Royal Statistical Society: Series B (Methodological), 52, 105–124, https://doi.org/10.1111/j.2517-6161.1990.tb01775.x, 1990.

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

Hosking, J. R. M., Wallis, J. R., and Wood, E. F.: Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments, Technometrics, 27, 251–261, https://doi.org/10.1080/00401706.1985.10488049, 1985.

Hurvich, C. M. and Tsai, C.-L.: Regression and time series model selection in small samples, Biometrika, 76, 297–307, https://doi.org/10.1093/biomet/76.2.297, 1989.

India Meteorological Department: Cyclonic Disturbances over the North Indian Ocean during 2009: A Report, New Delhi, https://rsmcnewdelhi.imd.gov.in/uploads/report/27/27_4e34f3_rsmc-2009.pdf (last access: 13 September 2026), 2010.

India Meteorological Department: Super Cyclonic Storm AMPHAN over the southeast Bay of Bengal during 16-21 May 2020: Preliminary Report, New Delhi, http://www.rsmcnewdelhi.imd.gov.in/images/pdf/publications/preliminary-report/amphan.pdf (last access: 13 September 2026), 2020.

India Meteorological Department: Very Severe Cyclonic Storm YAAS over the Bay of Bengal during 23-27 May 2021: Preliminary Report, New Delhi, https://rsmcnewdelhi.imd.gov.in/archive-report.php?internal_menu=MjY%3D&year=MjAyMQ%3D%3D (last access: 13 September 2026), 2021.

India Meteorological Department: Severe Cyclonic Storm REMAL over the Bay of Bengal (24th-28th May, 2024): A Report, New Delhi, https://rsmcnewdelhi.imd.gov.in/download.php?path=uploads%2Freport%2F26%2F26_4a868b_Severe+Cyclonic+Storm+REMAL-A+Brief+Report_03June%28Approved%29.pdf (last access: 13 September 2026), 2024.

https://doi.org/10.5194/egusphere-2026-5598
Preprint. Discussion started: 29 September 2026

Jisan, M. A., Bao, S., and Pietrafesa, L. J.: Ensemble projection of the sea level rise impact on storm surge and inundation at the coast of Bangladesh, Natural Hazards and Earth System Sciences, 18, 351–364, https://doi.org/10.5194/nhess-18-351-2018, 2018.

Kang, S. and Song, J.: Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework, J. Korean Stat. Soc., 46, 487–501, https://doi.org/10.1016/j.jkss.2017.02.003, 2017.

Karim, M. F. and Mimura, N.: Impacts of climate change and sea-level rise on cyclonic storm surge floods in Bangladesh, Global Environmental Change, 18, 490–500, https://doi.org/10.1016/j.gloenvcha.2008.05.002, 2008.

Khan, M. J. U., Durand, F., Emanuel, K., Krien, Y., Testut, L., and Islam, A. K. M. S.: Storm surge hazard over Bengal delta: a probabilistic-deterministic modelling approach, Natural Hazards and Earth System Sciences, 22, 2359–2379, https://doi.org/10.5194/nhess-22-2359-2022, 2022.

Landwehr, J. M., Matalas, N. C., and Wallis, J. R.: Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles, Water Resour. Res., 15, 1055–1064, https://doi.org/10.1029/WR015i005p01055, 1979.

Lee, H. S.: Estimation of extreme sea levels along the Bangladesh coast due to storm surge and sea level rise using EEMD and EVA, J. Geophys. Res. Oceans, 118, 4273–4285, https://doi.org/10.1002/jgrc.20310, 2013.

Lewis, M., Horsburgh, K., and Bates, P.: Bay of Bengal cyclone extreme water level estimate uncertainty, Natural Hazards, 72, 983–996, https://doi.org/10.1007/s11069-014-1046-2, 2014.

Local Government Engineering Department (LGED): Bangladesh - Rivers, https://data.humdata.org/dataset/bangladesh-water-courses, last access: 13 September 2026, 2018.

Mann, H. B.: Nonparametric Tests Against Trend, Econometrica, 13, 245–259, https://doi.org/10.2307/1907187, 1945.

Marcos, M. and Woodworth, P. L.: Spatiotemporal changes in extreme sea levels along the coasts of the North Atlantic and the Gulf of Mexico, J. Geophys. Res. Oceans, 122, 7031–7048, https://doi.org/10.1002/2017JC013065, 2017.

Menéndez, M. and Woodworth, P. L.: Changes in extreme high water levels based on a quasi-global tide-gauge data set, J. Geophys. Res. Oceans, 115, C10011, https://doi.org/10.1029/2009JC005997, 2010.

Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H., and Ward, P. J.: A global reanalysis of storm surges and extreme sea levels, Nat. Commun., 7, 11969, https://doi.org/10.1038/ncomms11969, 2016.

Palmer, K., Watson, C. S., Power, H. E., and Hunter, J. R.: Quantifying the Mean Sea Level, Tide, and Surge Contributions to Changing Coastal High Water Levels, J. Geophys. Res. Oceans, 129, e2023JC020737, https://doi.org/10.1029/2023JC020737, 2024.

Pettitt, A. N.: A Non-Parametric Approach to the Change-Point Problem, Appl. Stat., 28, 126–135, https://doi.org/10.2307/2346729, 1979.

Saint-Criq, L., Gaume, E., Hamdi, Y., and Ouarda, T. B. M. J.: Extreme Sea Level Estimation Combining Systematic Observed Skew Surges and Historical Record Sea Levels, Water Resour. Res., 58, e2021WR030873, https://doi.org/10.1029/2021WR030873, 2022.

Santamaria-Aguilar, S. and Vafeidis, A. T.: Are Extreme Skew Surges Independent of High Water Levels in a Mixed Semidiurnal Tidal Regime?, J. Geophys. Res. Oceans, 123, 8877–8886, https://doi.org/10.1029/2018JC014282, 2018.

Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall's Tau, J. Am. Stat. Assoc., 63, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934, 1968.

Silva Lomba, J. and Fraga Alves, M. I.: L-moments for automatic threshold selection in extreme value analysis, Stochastic Environmental Research and Risk Assessment, 34, 465–491, https://doi.org/10.1007/s00477-020-01789-x, 2020.

Survey of Bangladesh: Inland Waters, Bangladesh, 2016, https://purl.stanford.edu/js773kw6163, last access: 13 September 2026, 2016.

Tebaldi, C., Strauss, B. H., and Zervas, C. E.: Modelling sea level rise impacts on storm surges along US coasts, Environmental Research Letters, 7, 014032, https://doi.org/10.1088/1748-9326/7/1/014032, 2012.

United Nations Office for the Coordination of Humanitarian Affairs: Bangladesh - Subnational Administrative Boundaries, https://data.humdata.org/dataset/cod-ab-bgd, last access: 13 September 2026, 2020.

Wahl, T., Haigh, I. D., Nicholls, R. J., Arns, A., Dangendorf, S., Hinkel, J., and Slangen, A. B. A.: Understanding extreme sea levels for broad-scale coastal impact and adaptation analysis, Nat. Commun., 8, 16075, https://doi.org/10.1038/ncomms16075, 2017.

Wan Deraman, W. H. A., Abd Mutalib, N. J., and Mukhtar, N. Z.: Determination of return period for flood frequency analysis using normal and related distributions, J. Phys. Conf. Ser., 890, 012162, https://doi.org/10.1088/1742-6596/890/1/012162, 2017.

Yue, S., Pilon, P., Phinney, B., and Cavadias, G.: The influence of autocorrelation on the ability to detect trend in hydrological series, Hydrol. Process., 16, 1807–1829, https://doi.org/10.1002/hyp.1095, 2002.

Zhang, H. and Sheng, J.: Examination of extreme sea levels due to storm surges and tides over the northwest Pacific Ocean, Cont. Shelf Res., 93, 81–97, https://doi.org/10.1016/j.csr.2014.12.001, 2015.