Article

# Trends in Europe storm surge extremes match the rate of sea-level rise

https://doi.org/10.1038/s41586-022-04426-5

Received: 23 September 2021

Accepted: 14 January 2022

Published online: 30 March 2022

Check for updates

Francisco M. Calafat[1]✉, Thomas Wahl[2,3], Michael Getachew Tadesse[2,3] & Sarah N. Sparrow[4]

Coastal communities across the world are already feeling the disastrous impacts of climate change through variations in extreme sea levels[1]. These variations reflect the combined effect of sea-level rise and changes in storm surge activity. Understanding the relative importance of these two factors in altering the likelihood of extreme events is crucial to the success of coastal adaptation measures. Existing analyses of tide gauge records[2–10] agree that sea-level rise has been a considerable driver of trends in sea-level extremes since at least 1960. However, the contribution from changes in storminess remains unclear, owing to the difficulty of inferring this contribution from sparse data and the consequent inconclusive results that have accumulated in the literature[11,12]. Here we analyse tide gauge observations using spatial Bayesian methods[13] to show that, contrary to current thought, trends in surge extremes and sea-level rise both made comparable contributions to the overall change in extreme sea levels in Europe since 1960 . We determine that the trend pattern of surge extremes reflects the contributions from a dominant north–south dipole associated with internal climate variability and a single-sign positive pattern related to anthropogenic forcing. Our results demonstrate that both external and internal influences can considerably affect the likelihood of surge extremes over periods as long as 60 years, suggesting that the current coastal planning practice of assuming stationary surge extremes[1,14] might be inadequate.

Floods resulting from extreme sea levels are among the costliest natural hazards, causing tens of billions of dollars in economic losses globally each year[1]. Without adaption, such losses are certain to worsen in the decades ahead as the sea level rises[15,16]. Cost-effective adaptation plans are key to reducing this vulnerability while also avoiding costly overprotection measures[17]. However, their success relies on robust understanding of how changes in mean climate affect the likelihood of extreme sea-level events. This effect can occur primarily through (omitting tides and waves) changes in storminess affecting the occurrences of storm surges and changes in mean sea level (MSL) raising or lowering the baseline level for storm surges. While most studies agree that sea-level rise[18,19] has made extreme sea-level events more likely across the world since at least the mid-twentieth century[2–10], the contribution from changes in surge extremes is subject to debate[11,12]. Clarifying this debate is a priority because, in most countries, current practice for assessing future coastal flood risk assumes that the probability of surge extremes is the same now as in the future[1,14] (that is, only sea-level rise is considered). Should this assumption turn out to be invalid, this could cause adaptation plans to be ineffective.

Numerical models generally predict that storm surge activity will change this century in many places as the climate warms[20–26], albeit with varying intensity depending on the study. Yet observational evidence for such changes is lacking[11,12], despite the expectation from models that surge extremes are influenced by climate change. Most observational studies emphasize that, after removal of MSL influences, trends in sea-level extremes become statistically non-significant at most data sites. This is often interpreted as evidence of no trends in surge extremes. It is, however, incorrect to conclude that a trend is absent on the basis of its statistical non-significance[27]. In studies with small sample sizes, such as those of extremes, lack of statistical significance is often because of insufficient statistical power and does not indicate that a trend is small or absent[28]. While trends in surge extremes have been reported at some data sites[6,8,9], the sites and magnitude of the trends are often inconsistent across studies.

These inconclusive, and often conflicting, results largely reflect the difficulty of estimating event probabilities from the historical sea-level record. Tide gauge observations are noisy, gappy, spatially sparse and highly heterogeneous, with complex spatiotemporal dependencies. Traditional approaches to analysing extremes are not designed to handle this complexity. Typically, event probabilities are estimated by fitting asymptotically justified distributions[29] to the observational data. Crucially, this is done on a site-by-site basis, with few extreme data at each site, leading to analyses that have low statistical power and are subject to substantial sampling error and small-scale noise. Such analyses have a low ability for detecting trends in extremes, tend to overestimate the magnitude of the trends when these are detected and may produce spurious trends (see the Supplementary Information). These difficulties pose an enormous challenge to understanding how surge extremes change in a warming climate.

[1]National Oceanography Centre, Liverpool, UK. [2]Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL, USA. [3]National Center for Integrated Coastal Research, University of Central Florida, Orlando, FL, USA. [4]Oxford e-Research Centre, Department of Engineering Science, University of Oxford, Oxford, UK.
✉e-mail: francisco.calafat@noc.ac.uk

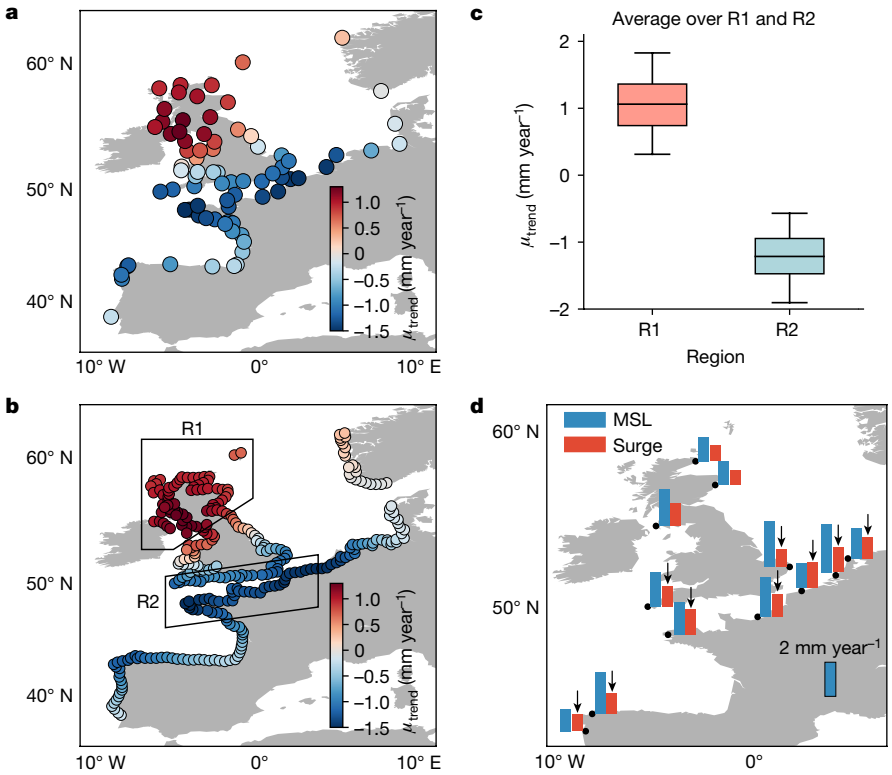



**Fig. 1 | Historical trends in storm surge extremes.** Estimates of the trend in $\mu$ during the period 1960–2018 as derived using the BHM at tide gauge sites (**a**) and gridded locations (**b**). **c**, Posterior mean (central mark), interquartile range (shaded box) and 5–95% CI (whiskers) for the $\mu$ trends averaged over the regions R1 and R2 (denoted by the black boxes in panel **b**). **d**, Comparison of trends in MSL (blue) and $\mu$ (red) at tide gauge sites for the period 1960–2018. Downward arrows indicate negative trends.

Here we provide observational evidence of significant trends in surge extremes, similar in magnitude to MSL trends, over Europe since 1960 and quantify the contributions from anthropogenic forcing and internal climate variability to these trends. We achieve this by using a Bayesian hierarchical model (BHM) that accommodates the complexity typical of tide gauge observations and leverages spatial dependencies to overcome the difficulties related to small sample sizes[13] (see Methods). By enabling sharing of information across space, our approach can separate the large-scale patterns of change from the singularities of each location and allows estimation of changes in extremes at ungauged locations. The attribution of the changes is done by extracting the fingerprint (that is, response pattern) of external forcing on surge extremes from a large ensemble of surge simulations and calibrating it to observations.

## Historical trends in surge extremes

We use our approach to analyse surge annual maxima from 79 tide gauges located along the Atlantic and North Sea coastlines of Europe for the period 1960–2018. We model the annual maxima as a max-stable process[13,30] (see Methods), meaning that the marginal distribution at each location follows a univariate generalized extreme value (GEV) distribution[31] with parameters $\mu$ (location), $\sigma$ (scale) and $\xi$ (shape). In our model, any effects of climate change on the probabilities of surge extremes are captured through changes in $\mu$. Hence, much of our focus here is on describing the evolution of $\mu$ over time, with emphasis on long-term changes. Inferences from our BHM take the form of samples from the posterior distribution of processes and parameters given the data. These samples enable us to make meaningful inferences with quantifiable uncertainty on any function of the variables of interest, such as spatial and temporal averages of trends in $\mu$. Our estimates are summarized by posterior means, standard deviations and 5–95% credible intervals (CIs), and are provided both at tide gauge sites and on a coastal grid. CIs are not necessarily symmetric and, hence, they are denoted by square brackets.

Trends in $\mu$ for 1960–2018 show a marked north–south dipole structure characterized by positive values along the coasts of the British Isles north of 52° N and negative values along most of the European coastlines south of that latitude from Portugal to the Netherlands (Fig. 1a, b). Positive (negative) trends in $\mu$ indicate increasing (decreasing) probability of extreme events. Hereafter, we refer to the positive and negative poles of the trend pattern as R1 and R2, respectively, as denoted in Fig. 1b by the black boxes. The standard deviations associated with the trend estimates at individual locations are shown in Extended Data Fig. 1. In each region, trends are remarkably uniform and consistent, reflecting the ability of our approach to separate the large-scale signals from the small-scale noise. The magnitude of the trends is similar in R1 and R2, with spatially averaged values of 1.1 mm year$^{-1}$ [0.3, 1.8] and −1.2 mm year$^{-1}$ [−1.9, −0.6] (Fig. 1c), respectively. The trend estimates are robust to prior choices (see the Supplementary Information). To put these trends into context, we compare them with the average MSL rise for 1960–2018 as estimated from annual tide gauge records (Fig. 1d). Trends in surge extremes are similar in magnitude to the rate of MSL rise at most sites with surge-to-MSL trend ratios ranging from 0.4 to 1.0. In R1, both contributions are positive, thus creating a compounding effect that drives event probabilities substantially higher, whereas in R2, they have opposite sign, thus largely cancelling one another out.

To illustrate the effect of the $\mu$ trends on the occurrence probabilities of surge extremes, we show how the return period corresponding to a 50-year surge event in 1960 has changed over time in both R1 and R2

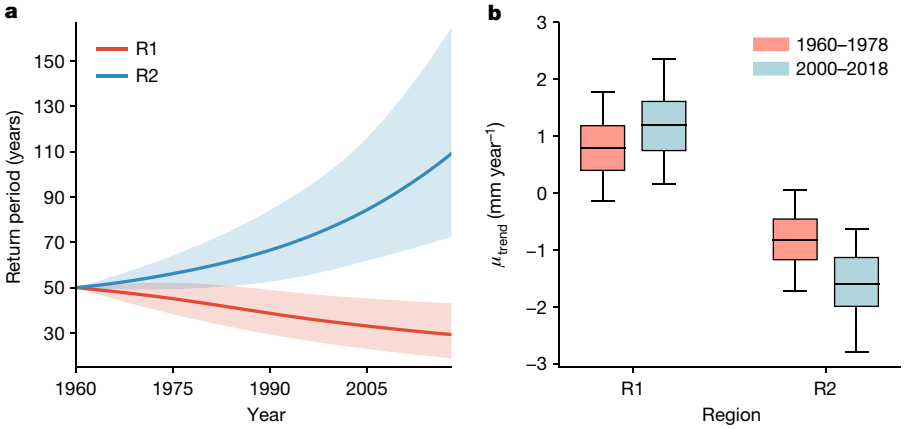



**Fig. 2 | Temporal changes in return period. a**, Changes in the return period corresponding to a 50-year surge event in 1960 averaged over the regions R1 and R2 (denoted by the black boxes in Fig. 1b). The thick lines indicate posterior means, whereas the shading represents the 5–95% CI. **b**, Posterior mean (central mark), interquartile range (shaded box) and 5–95% CI (whiskers) for the $\mu$ trends during the periods 1960–1978 (red) and 2000–2018 (blue) averaged over the regions R1 and R2.

(Fig. 2a). In R1, the return period decreased from 50 years in 1960 to 29.3 years [18.8, 43.1] in 2018, whereas in R2, it more than doubled to 109.1 years [71.8, 166.2] in 2018. In other words, extreme surge events in R1 (R2) are about twice (half) as likely to occur now as they were in 1960. The factor by which return periods change as $\mu$ changes is insensitive to the value of the return period. For example, in R1, the return period corresponding to a 1,000-year surge event in 1960 fell by a factor of 1.87 from 1,000 years in 1960 to 534.4 years [309.1, 845.5] in 2018, which is almost the same reduction factor as for the 50-year return period. We also note that the rate of change has not been constant over time (Fig. 2b), which suggests an acceleration of $\mu$. Indeed, we find it likely (probability $P = 0.69$) that the $\mu$ trend in R1 during the period 2000–2018 (1.2 mm year$^{-1}$; 0.2 to 2.4) is larger than for 1960–1978 (0.8 mm year$^{-1}$; −0.1 to 1.8). Similarly, it is likely ($P = 0.83$) that the $\mu$ trend in R2 for 2000–2018 (−1.6 mm year$^{-1}$; −2.8 to −0.6) is larger in absolute value than for 1960–1978 (−0.8 mm year$^{-1}$; −1.7 to 0.1).

## Attribution of historical trends

Observational changes in surge extremes will, in general, reflect the combined effect of external forcing and internal climate variability. Separating the individual contributions from these two factors is crucial to understanding how surge extremes might change in the future under human-induced climate change. Robust separation is, however, not possible from the single climate realization provided by the observational record. To address this challenge, we resort to a large ensemble (100 members) of historical surge simulations (see Methods), which we generate by driving a storm surge model[32,33] with the atmospheric fields from an initial condition ensemble of climate simulations[34]. We extract the fingerprint of anthropogenic forcing on surge extremes by fitting the BHM to the simulated surge annual maxima and then averaging the estimated patterns of $\mu$ trends over the ensemble members. If climate models were perfect, the simulated fingerprint would already give us the anthropogenic contribution to the observed $\mu$ trends. Models, however, are affected by biases and, thus, it is crucial that their outputs be calibrated to observations. We achieve this by incorporating the simulated fingerprint into the BHM as a covariate for $\mu$ and then calibrating it by fitting the BHM to the tide gauge observations (see Methods). The contribution from internal climate variability is defined as the portion of $\mu$ trends unexplained by anthropogenic forcing. This type of analysis falls in the realm of what is known as extreme event attribution[35] (EEA) (see Methods for a discussion on EEA approaches).

The anthropogenic fingerprint on $\mu$ trends is characterized by a single-sign positive pattern (except for Lerwick) of meridional variation, wherein trends increase from nearly zero in the north of the UK to values of up to 0.8 mm year$^{-1}$ in the English Channel and the east coast of England (Fig. 3a). The contribution from internal variability to $\mu$ trends (Fig. 3b) shows the same north–south dipole structure as the total $\mu$ trends (Fig. 1a, b), although with smaller positive values in R1 and larger negative values in R2. The relative importance of anthropogenic forcing and internal variability is further quantified by computing spatially averaged trends over R1 and R2 (Fig. 3c). In R1, anthropogenic forcing has an average contribution of 0.2 mm year$^{-1}$ [−0.3, 0.9], whereas internal variability contributes 0.9 mm year$^{-1}$ [−0.1, 1.8]. In R2, the contribution from anthropogenic forcing is larger, with an average value of 0.6 mm year$^{-1}$ [0.0, 1.7], compared with −1.7 mm year$^{-1}$ [−3.0, −0.8] for internal variability. Hence, the observational pattern of trends in surge extremes (Fig. 1a, b) is largely driven by the effects of internal climate variability, but it also contains a detectable response to anthropogenic forcing. Sensitivity analysis shows that estimates of the individual contributions are robust to the choice of priors on the calibration parameters (see the Supplementary Information).

The Bayesian solutions allow us to assess how much the occurrence probability of any given event has changed as a result of anthropogenic forcing. As an example, we quantify the change in the likelihood of the surge event caused by cyclone Xaver[36] along the North Sea coastlines in December 2013 attributable to anthropogenic forcing since 1960 (see Methods). For this analysis, we focus on the Lowestoft and Dunkirk tide gauges (their location is shown in Fig. 3a), as these stations exhibit the largest surge levels induced by Xaver in our study domain, with values of 2.1 m and 2.4 m, respectively. We quantify the change in occurrence probability attributable to anthropogenic forcing in terms of the fraction of attributable risk (FAR) (see Methods). We find a FAR of 0.17 [0.02, 0.41] at Lowestoft and 0.16 [0.01, 0.40] at Dunkirk (Fig. 3d). In other words, although the uncertainty in the FAR estimates is large, the posterior means (that is, the FAR values best supported by the data) indicate that anthropogenic forcing has increased the occurrence probability of such an event by a factor of 1.2 since 1960.

## Conclusions

We have addressed the continuing debate on whether the likelihood of storm surge extremes has changed during the instrumental period of sea-level observations in Europe. We find that, contrary to what past studies suggest, trends in surge extremes for the period 1960–2018 were comparable with those in MSL. The trend pattern reflects a combination of a north–south dipole associated with internal climate variability and a single-sign positive pattern related to anthropogenic

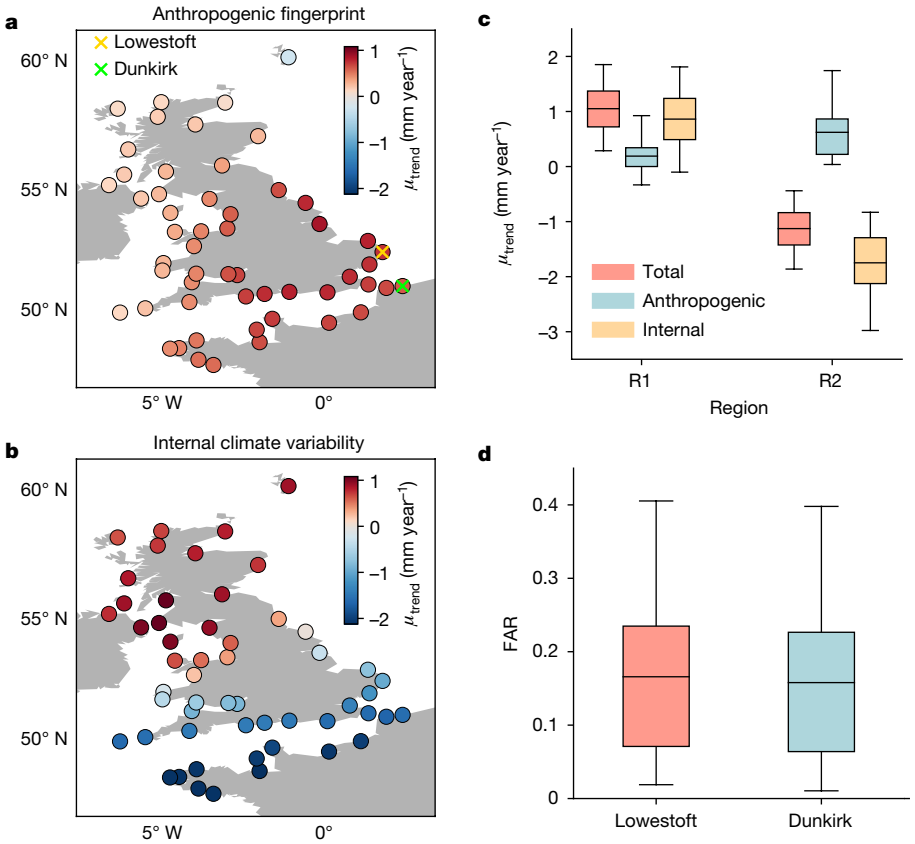



**Fig. 3 | Attribution of trends in surge extremes.** Contributions from anthropogenic forcing (**a**) and internal climate variability (**b**) to $\mu$ trends during the period 1960–2018 as estimated by fitting the BHM to the tide gauge observations. **c**, Posterior mean (central mark), interquartile range (shaded box) and 5–95% CI (whiskers) for the total $\mu$ trends (red) and the contributions of anthropogenic forcing (blue) and internal variability (orange) averaged over the regions R1 and R2 (denoted by the black boxes in Fig. 1b). **d**, The change in the occurrence probability of the surge event caused by cyclone Xaver at the Lowestoft and Dunkirk tide gauges attributable to anthropogenic forcing as measured by the fraction of attributable risk (FAR). The boxes and whiskers are defined as in **c**. The location of the two tide gauges is shown in **a**.

forcing. The anthropogenic fingerprint is consistent with a strengthening and eastward extension of the North Atlantic storm track, leading to increased storminess over the UK and central Europe as predicted by climate models in response to anthropogenic forcing[37–39]. It is also qualitatively similar to projections of changes in storm surge extremes for this century[23,24].

Our results are contrary to the current planning practice of assuming stationary storm surge extremes[1,14] and show that anthropogenic forcing might have already altered the likelihood of such extremes. Furthermore, we find that internal climate variability alone can give rise to trends in surge extremes of more than 2 mm year$^{-1}$ during periods that are long enough to have implications for coastal planning. Ignoring this non-stationarity in surge extremes can lead to the premature failure of flood defences, with catastrophic consequences. The substantial effect of internal variability on surge extremes over long time periods also means that large ensembles are needed to robustly isolate the anthropogenically forced signal, much larger than the current ensembles (<10 members) being used for the analysis of future surge extremes[23,24,40].

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04426-5.


1. Hinkel, J. et al. Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proc. Natl Acad. Sci. USA* **111**, 3292–3297 (2014).
2. Zhang, K., Douglas, B. C. & Leatherman, S. P. Twentieth-century storm activity along the U.S. East Coast. *J. Clim.* **13**, 1748–1761 (2000).
3. Woodworth, P. L. & Blackman, D. L. Evidence for systematic changes in extreme high waters since the mid-1970s. *J. Clim.* **17**, 1190–1197 (2004).
4. Marcos, M., Tsimplis, M. N. & Shaw, A. G. P. Sea level extremes in southern Europe. *J. Geophys. Res. Oceans* **114**, C01007 (2009).
5. Haigh, I., Nicholls, R. & Wells, N. Assessing changes in extreme sea levels: application to the English Channel, 1900–2006. *Cont. Shelf Res.* **30**, 1042–1055 (2010).
6. Menéndez, M. & Woodworth, P. L. Changes in extreme high water levels based on a quasi-global tide-gauge data set. *J. Geophys. Res.* **115**, C10011 (2010).
7. Weisse, R. et al. Changing extreme sea levels along European coasts. *Coastal Eng.* **87**, 4–14 (2014).
8. Wahl, T. & Chambers, D. P. Evidence for multidecadal variability in US extreme sea level records. *J. Geophys. Res. Oceans* **120**, 1527–1544 (2015).
9. Marcos, M. & 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 (2017).
10. Rohmer, J. & Le Cozannet, G. Dominance of the mean sea level in the high-percentile sea levels time evolution with respect to large-scale climate variability: a Bayesian statistical approach. *Environ. Res. Lett.* **14**, 014008 (2019).
11. Seneviratne, S. I. et al. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. et al.) 109–230 (Cambridge Univ. Press, 2012).
12. Intergovernmental Panel on Climate Change in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) (Cambridge Univ. Press, 2013).
13. Calafat, F. M. & Marcos, M. Probabilistic reanalysis of storm surge extremes in Europe. *Proc. Natl Acad. Sci. USA* **117**, 1877–1883 (2020).
14. Luxford, F. & Faulkner, D. Recommendations for future research and practice on non-stationarity in UK flooding. FRS18087/REA/R2. *Environment Agency* https://assets.publishing.service.gov.uk/media/6038f813e90e07055685020c/Recommendations_for_future_research_and_practice_on_non-stationarity_in_UK_flooding_-_report__2_.pdf (2020).

15. Jevrejeva, S., Jackson, L. P., Grinsted, A., Lincke, D. & Marzeion, B. Flood damage costs under the sea level rise with warming of 1.5 °C and 2.0 °C. *Environ. Res. Lett.* **13**, 074014 (2018).
16. Kulp, S. A. & Strauss, B. H. New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding. *Nat. Commun.* **10**, 4844 (2019).
17. Tiggeloven, T. et al. Global-scale benefit–cost analysis of coastal flood adaptation to different flood risk drivers using structural measures. *Nat. Hazards Earth Syst. Sci.* **20**, 1025–1044 (2020).
18. Church, J. et al. in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (eds Stocker, T. F. et al.) Ch. 13, 1137–1216 (Cambridge Univ. Press, 2013).
19. Frederikse, T. et al. The causes of sea-level rise since 1900. *Nature* **584**, 393–397 (2020).
20. Marcos, M., Jordà, G., Gomis, D. & Pérez, B. Changes in storm surges in southern Europe from a regional model under climate change scenarios. *Glob. Planet. Change* **77**, 116–128 (2011).
21. Conte, D. & Lionello, P. Characteristics of large positive and negative surges in the Mediterranean Sea and their attenuation in future climate scenarios. *Glob. Planet. Change* **111**, 159–173 (2013).
22. Little, C. M. et al. Joint projections of US East Coast sea level and storm surge. *Nat. Clim. Change* **5**, 1114–1120 (2015).
23. Vousdoukas, M. I. et al. Projections of extreme storm surge levels along Europe. *Clim. Dyn.* **47**, 3171–3190 (2016).
24. Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M. & Feyen, L. Extreme sea levels on the rise along Europe's coasts. *Earths Future* **5**, 304–323 (2017).
25. Howard, T., Palmer, M. D. & Bricheno, L. M. Contributions to 21st century projections of extreme sea-level change around the UK. *Environ. Res. Commun* **1**, 095002 (2019).
26. Muis, S. et al. A high-resolution global dataset of extreme sea levels, tides, and storm surges, including future projections. *Front. Mar. Sci.* **7**, 263 (2020).
27. Greenland, S. et al. Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol* **31**, 337–350 (2016).
28. Button, K. et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
29. Coles, S. G. *An Introduction to Statistical Modelling of Extreme Values* 208 pp (Springer, 2001).
30. Schlather, M. Models for stationary max-stable random fields. *Extremes* **5**, 33–44 (2002).
31. McFadden, D. Modeling the choice of residential location. *Transp. Res. Rec.* **672**, 72–77 (1978).
32. Tadesse, M. G., Wahl, T. & Cid, A. Data-driven modeling of global storm surges. *Front. Mar. Sci.* **7**, 260 (2020).
33. Tadesse, M. G. & Wahl, T. A database of global storm surge reconstructions. *Sci. Data* **8**, 125 (2021).
34. Guillod, B. P. et al. A large set of potential past, present and future hydro-meteorological time series for the UK. *Hydrol. Earth Syst. Sci.* **22**, 611–634 (2018).
35. Stott, P. A. et al. Attribution of extreme weather and climate-related events. *Wiley Interdiscip. Rev. Clim. Change* **7**, 23–41 (2016).
36. Dangendorf, S., Arns, A., Pinto, J. G., Ludwig, P. & Jensen, J. The exceptional influence of storm 'Xaver' on design water levels in the German Bight. *Environ. Res. Lett.* **11**, 054001 (2016).
37. Zappa, G., Shaffrey, L. C., Hodges, K. I., Sansom, P. G. & Stephenson, D. B. A multimodel assessment of future projections of North Atlantic and European extratropical cyclones in the CMIP5 climate models. *J. Clim.* **26**, 5846–5862 (2013).
38. Feser, F. et al. Storminess over the North Atlantic and northwestern Europe—a review. *Q. J. R. Meteorol. Soc* **141**, 350–382 (2015).
39. Barcikowska, M. J. et al. Euro-Atlantic winter storminess and precipitation extremes under 1.5 °C vs. 2 °C warming scenarios. *Earth Syst. Dyn* **9**, 679–699 (2018).
40. Vousdoukas, M. I. et al. Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard. *Nat. Commun.* **9**, 2360 (2018).

# Methods

### Tide gauge observations

Hourly sea-level observations from 79 tide gauges in Europe were obtained from the Global Extreme Sea Level Analysis (GESLA) dataset version 2 (ref. [41]). The tide gauge data are exactly the same as used in ref. [13], with the only difference being that, for this study, we have extended many of the records by 5 years from 2014 to 2018. The new data have been downloaded from the British Oceanographic Data Centre and the Copernicus Marine Environment Monitoring Service. The procedure to extract surge annual maxima from the hourly time series is also the same as that described in ref. [13] and, thus, we defer to that study for full details. Here we provide a brief description. We note that, sometimes, the skew surge is used instead of the storm surge[9]. Here we choose to analyse the storm surge because it is the most direct measure of how sea level responds to a storm's wind and low pressure acting on the ocean. It is also what a (tideless) hydrodynamic model provides as output when forced by atmospheric conditions.

Here the surge component of sea level is defined as the part of a tide gauge record that remains after removal of the MSL and the tide. To extract the surge, we begin by identifying and correcting artificial jumps or shifts in the tide gauge records using a method[42] that is able to detect abrupt changes in the mean of a signal. Then, we remove the MSL from the shift-adjusted time series by subtracting the annual medians along with a seasonal cycle. Next, we remove the tidal component by means of harmonic analysis while also correcting for inaccuracies in the tidal predictions that can arise from tide–surge interaction and timing errors (see ref. [13] for details on how we achieve this). The residual time series are then visually inspected to identify and remove outliers. Once a data point has been flagged as suspicious by visual inspection, we scrutinize it using scientific judgement in several different ways. For example, we check if the temporal profile at the time of the suspicious point is consistent with what we would expect from a storm surge (that is, a rise in water level over several hours, as opposed to a large jump). We also perform a buddy check with data from nearby stations, whenever those data are available, and use historical daily sea-level pressure data from a reanalysis[43] to determine whether stormy weather was present at the time. After removing outliers, we extract the annual maxima, noting that years are defined as starting on 1 April (and ending on 31 March) and only years with at least 6,000 valid hourly values are considered (roughly 12% of all years with at least one observation have less than 6,000 hourly values). Setting a lower bound on the number of hourly values is necessary to preserve the asymptotic properties of the BHM. The final set of extremes comprises 2,789 annual maxima from a total of 79 tide gauge records spanning the period 1960–2018 (the locations of the tide gauges are shown in Extended Data Fig. 2).

Annual time series of MSL, which are used to estimate MSL trends, were obtained from the Permanent Service for Mean Sea Level Revised Local Reference database[44]. Only tide gauge records that have at least 75% of valid data (according to the quality flags assigned by the Permanent Service for Mean Sea Level) in the period 1960–2018 are used. MSL trends are estimated by fitting a linear model to the annual time series using ordinary least squares. We note that tide gauges measure sea level relative to the land on which they reside. Therefore, the MSL trends reported here (Fig. 1d) reflect a combination of both changes in sea surface height and vertical land motion.

### Bayesian hierarchical model for extremes

The BHM used here was recently developed and thoroughly tested in ref. [13] and, thus, we defer to that study for full details of the model. Here we provide a conceptual overview of the model and summarize a few modifications to the original model, as well as the adjustments needed for separating the contributions from anthropogenic forcing and internal variability.

Our approach relies on the fact that extreme events at locations that are geographically close tend to show similar occurrence probabilities and, if the locations are close enough to be affected by the same events, they also show temporally correlated values. The key idea is to exploit these spatial dependencies to enable sharing of information across data sites. Rather than analysing extremes on a site-by-site basis, as most studies so far have done, we analyse them by pooling data from all locations together by means of a BHM that describes how the distribution of extremes varies in time and space. This means that inferences at one location are informed not only by observations at that location but also by observations at all other locations. This leads to several key advantages over traditional approaches: (1) it can reduce estimation uncertainty by a factor of more than two[13]; (2) it enables us to separate the large-scale signals of interest from small-scale noise, as well as the long-term changes from sampling variability; (3) it allows separation of the contributions from external and internal influences; and (4) it permits estimation of changes in extremes at ungauged locations (see ref. [13] for an assessment of the skill of the BHM at prediction sites).

Spatial dependence in surge annual maxima (called residual dependence) is described by means of a max-stable model[13,30] (the infinite-dimension generalization of the GEV). This means that, at each location, the marginal distribution of the surge process is a GEV. The spatial residual process is expressed as a linear combination of scaled Gaussian kernels defined on a regular grid of spatial knots. The residual process converges to a max-stable process asymptotically as the number of knots grows infinitely large[45]. In ref. [13], we used a 2° × 2° grid of knots, but subsequent analysis suggests that such grid spacing may be too coarse. Hence, here, we increase the grid resolution to 0.5° × 0.5°. This comes at substantially increased computational cost. To make the model more computationally tractable, we build the residual process using only the nearest 30 knots to each tide gauge station. This is equivalent to setting the kernels to zero for distances larger than roughly one length scale, a modification that preserves the max-stability properties of the model[46]. The knot grid is shown in Extended Data Fig. 2.

Spatial dependence among the marginal GEV parameters is captured through latent Gaussian processes and bathymetric covariates. In particular, the GEV location parameter $\mu$ is modelled as a time-varying spatial process, the scale parameter $\sigma$ is assumed to follow a time-invariant spatial process, and the shape parameter $\xi$ is treated as constant in both time and space (see ref. [13] for justification of this assumption). Fitting a BHM with a time-varying $\sigma$ to the tide gauge data shows no evidence of trends in that parameter, hence we keep it constant in time. In ref. [13], we modelled $\mu$ as a spatiotemporal integrated random walk (see equations (8) and (9) in ref. [13]). This is the formulation that we use here when estimating historical trends (for example, Figs. 1 and 2), as well as when extracting the fingerprint of external forcing from the ensemble of surge simulations (as described later). Note that this formulation (hereafter referred to as BHM1) aims to estimate total trends in $\mu$ without distinguishing between external and internal influences.

To quantify the individual contributions from external forcing and internal climate variability, a different formulation is required (hereafter referred to as BHM2). Specifically, we assume that $\mu$ can be decomposed into a linear trend associated with external forcing ($\mu^{\text{ext}}$) plus a spatiotemporal integrated random walk that captures the effects of internal climate variability ($\mu^{\text{int}}$). The fingerprint of external forcing extracted from the ensemble of surge simulations, $x_{\text{ext}}$, is incorporated into the BHM as a covariate for $\mu^{\text{ext}}$ with multiplicative ($\beta_{\text{ext}}$) and additive ($\delta_{\text{ext}}$) biases that account for errors in the model-simulated fingerprint ($x_{\text{ext}}$). With this, the model for $\mu$ can be written as:

$$\mu_t(\mathbf{s}) = \mu_t^{\text{ext}}(\mathbf{s}) + \mu_t^{\text{int}}(\mathbf{s}) \quad (1)$$

$$\mu_t^{\text{ext}}(\mathbf{s}) = \mu_{t-1}^{\text{ext}}(\mathbf{s}) + \beta_{\text{ext}}(x_{\text{ext}}(\mathbf{s}) + \delta_{\text{ext}}) \quad (2)$$

$$\mu_t^{\text{int}}(\mathbf{s}) = \mu_{t-1}^{\text{int}}(\mathbf{s}) + \mu_{\text{trend},t-1}^{\text{int}}(\mathbf{s}) \quad (3)$$

$$\mu_{\text{trend},t}^{\text{int}}(\mathbf{s}) = \mu_{\text{trend},t-1}^{\text{int}}(\mathbf{s}) + \omega_t^{\text{int}}(\mathbf{s}) \quad (4)$$

$$\omega_t^{\text{int}}(\mathbf{s}) \sim \text{GP}(0, c_{\text{int}}(\mathbf{s}, \mathbf{s}'; \gamma_\mu, \rho_\mu)) \quad (5)$$

in which $t$ and $\mathbf{s}$ denote year and location, respectively, GP($m$, $c(\cdot,\cdot)$) denotes a Gaussian process with mean $m$ and covariance function $c$, and $\gamma_\mu$ and $\rho_\mu$ are, respectively, the standard deviation and length scale defining the covariance function $c_{\text{int}}$. The procedure to extract $x_{\text{ext}}$ from the ensemble of surge simulations is described later in Methods.

For the additive bias, we assume a normal distribution: $\delta_{\text{ext}} \sim \mathcal{N}(0,0.2)$. A standard deviation of 0.2 corresponds to 0.2 mm year$^{-1}$. For the amplitude of the fingerprint, $\beta_{\text{ext}}$, we assume a moderately informative gamma prior distribution with mode at 1: $\beta_{\text{ext}} \sim$ gamma(1.5, 0.2), in which the gamma distribution is parameterized in terms of shape and rate parameters. From $x_{\text{ext}}$, we expect the trends induced by external forcing to be relatively small compared with those from internal variability. This means that the likelihood for $\beta_{\text{ext}}$ is probably only weakly identifiable. In such cases, providing sufficient prior information is crucial to regularizing the likelihood and ensuring reasonable estimates. If the magnitude of the model-simulated fingerprint was consistent with the observations, then $\beta_{\text{ext}}$ would be equal to 1. This is the justification for choosing a prior with mode at 1. The chosen prior, however, has a standard deviation of 2.4 and, so, it is diffuse enough to allow the posterior to concentrate on smaller or larger values if the likelihood favours such values. Additionally, $\beta_{\text{ext}}$ is constrained to be positive to respect the sign of $x_{\text{ext}}$, but the chosen prior allows $\beta_{\text{ext}}$ to be as close to zero as required by the likelihood. Comparing the posterior and prior distributions for $\beta_{\text{ext}}$ shows that the former concentrates in a narrower region, indicating that the observations are sufficiently informative to constrain the value of $\beta_{\text{ext}}$ (Extended Data Fig. 3).

The estimated value of all scalar parameters in BHM1 as well as the ascribed prior distributions are summarized in Extended Data Table 1. Additionally, we have conducted a suite of experiments to assess: (1) the ability of the BHM to estimate trends in $\mu$; (2) the resolvability of the anthropogenic signal; and (3) the sensitivity of the Bayesian solutions to prior choices. The results of these experiments are discussed in the Supplementary Information.

The return period for an event of height $h$ is calculated as the inverse of $1 - \text{F}(h; \mu, \sigma, \xi)$, in which F is the cumulative distribution function of the GEV distribution.

### Inference in the BHM

To fit the BHM, we use the no-U-turn sampler (NUTS), as implemented by the Stan probabilistic programming language[47]. We run the sampler with four chains of 1,425 iterations each (warm-up = 800), for a total of 2,500 post-warm-up draws. Our fits did not show any divergent transitions and none of the iterations saturated the maximum tree depth, indicating that the sampler is able to explore the posterior distribution adequately. Convergence and mixing diagnostics for the model parameters, such as the potential scale reduction statistic and the effective sample size per iteration, are provided in Extended Data Table 1.

### Ensemble of climate and surge simulations

To obtain the ensemble of storm surge simulations, we use atmospheric fields (that is, wind speed and sea-level pressure) from an ensemble of climate simulations to force a statistical storm surge model as described below. The reason for this two-step procedure is that the models used for the climate simulations do not provide sea level as an output.

The large ensemble of climate data used in this analysis was produced by means of public distributed computing on the climate*prediction*.net platform using the weather@home European 25 km (EU25) model[48]. The weather@home model consists of a high-resolution regional climate model, which is one-way nested inside a driving coarse-resolution global circulation model. The driving global circulation model used is the Hadley Centre Atmospheric Model version 3 (HadAM3P) run with a horizontal resolution of 1.875° × 1.25° in longitude and latitude, respectively, and 19 vertical levels. The nested regional climate model is the Hadley Centre Regional Model version 3 (HadRM3P or PRECIS model), which is run at 25 km over the European domain. Performance and bias analysis for the weather@home EU25 model setup has been performed previously[48]. The data used in this analysis have also been the subject of hydrometeorological studies in Europe[34] and a detailed description of the experimental design is provided therein. To summarize, this large ensemble comprises of a set of 100 time series over a baseline period of 1900–2006, driven with historical forcing conditions. Sea surface temperatures and sea ice concentrations are taken from the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST) version 2 (ref. [49]). All other forcings — that is, well-mixed greenhouse gas concentrations, ozone, sulfur dioxide emissions, volcanic activity and solar forcing — were as provided by the Met Office and are consistent with the Coupled Model Intercomparison Project Phase 5 (CMIP5) historical experiment[34,50]. Each year was run independently and subsequently stitched together, to ensure continuity in the slowly evolving variability[34].

The models used in weather@home are atmosphere-only and, thus, they do not provide storm surges as an output. To obtain ensembles of surges, we use the atmospheric fields from weather@home to drive a statistical storm surge model. In particular, we simulate daily maximum storm surge values using the same modelling approach as that described in refs. [32,33], but with some simplifications for efficiency. We use daily mean sea-level pressure and daily mean wind speed from grid points within a 2° × 2° box around the tide gauge stations as predictors. This leads to a total of 190 predictor time series; no lagged predictors are considered, only the ones that occurred on the same day as the surge. Principal component analysis is applied to reduce the number of predictors and the ones that explain 95% of the variability are used with multiple linear regression models to infer the daily maximum surge values at 54 tide gauges. In a first step, the regression coefficients are estimated by fitting the surge model to tide gauge observations using predictors from the ERA5 reanalysis[51] for the period 1979–2018. Then, the estimated regression coefficients are applied to the atmospheric predictors from the ensemble of climate simulations to obtain 100 time series of daily maximum surge values at each tide gauge station spanning the period 1960–2006. The number of tide gauges used in the attribution analysis is less than 79 because the weather@home models only cover a subdomain of the European region centred on the UK. Fortunately, this subdomain is centred on the north–south dipole axis associated with the pattern of $\mu$ trends (see Fig. 1), thus giving us a nearly complete view of the pattern. A validation of the storm surge model is presented in the Supplementary Information.

### Estimation of the response pattern

First, we note that the climate simulations used here are not only forced by anthropogenic forcing but also by other external climate drivers, such as volcanic and solar activity. Hence, when we refer to the anthropogenic fingerprint on surge extremes, this assumes that the period over which the trends are computed (1960–2018) is long enough to cancel the effects of volcanism and that the influence of solar activity is minimal. To estimate the pattern of response to external forcing, we fit BHM1 to the annual maxima from the ensemble of surge simulations. Note that this involves 100 separate fits, one for each of the ensemble members. To make the problem computationally tractable, we modify the BHM by assuming that annual maxima are uncorrelated across data sites (that is, no residual dependence). This assumption greatly simplifies the BHM and is equivalent to setting the residual spatial

process equal to 1 (ref. [13]). Note that this version of the BHM still captures spatial dependence in the GEV parameters, but assumes that the annual maxima are independent. The location parameter $\mu$ is allowed to vary in time but only linearly. This simplification is necessary to reduce confounding between $\mu$ trends and sampling variability when not accounting for residual dependence. Exploratory analysis shows that BHMs with and without residual dependence yield patterns of trends that are similar in spatial structure, although the latter tends to produce larger values, owing to an influence of sampling variability on the trends. This is not an issue, as the amplitude of the response pattern is inferred from the tide gauge observations.

Fitting BHM1 to the ensemble of annual maxima produces 100 patterns of $\mu$ trends. The pattern of response to external forcing, $x_{ext}$, is computed as the mean of the 100 patterns of $\mu$ trends. This is then incorporated into BHM2 for calibration as described earlier in Methods. Note that $x_{ext}$ is computed using simulated data during the period 1960–2006, while the observations span the period 1960–2018, so we are assuming that the response pattern is similar in both periods. To test the robustness of the response pattern, we perform an experiment in which we randomly exclude several ensemble members and then compute the response pattern by averaging only over the retained ensemble members. We start by excluding only one member, then we exclude two members and so on up to 80 excluded members. We repeat this calculation 100 times, each time excluding a different set of members selected randomly. The resulting response patterns are then compared with the pattern derived from all 100 ensemble members. This experiment provides an indication of whether 100 ensemble members is sufficient to average out the influence of internal variability and isolate the response to external forcing. We find that using 80 members produces a response pattern that is almost identical to that derived from all the ensemble members (average spatial correlation of 0.95). Using 50 members still produces a good estimate of the response pattern, with an average spatial correlation of 0.82. However, correlations start to decrease rapidly at this point and are as low as 0.5 when using only 20 ensemble members.

### Extreme event attribution

The BHM solutions allow us to compute the anthropogenic contribution to changes in the occurrence probability of an event between any two years in the period 1960–2018. This type of assessment is referred to in the literature as EEA[35]. Past studies have followed two different approaches to EEA. The first approach compares the occurrence probability of an event in a factual world, in which climate conditions are similar to those observed at the time of the event, with its probability in a counterfactual world without human influences. Both the factual and counterfactual worlds are simulated by means of numerical models. This approach provides an estimate of the total anthropogenic contribution. The second approach uses transient model runs spanning a particular time period. If the model runs go back to pre-industrial times, then this second approach also gives the total anthropogenic contribution, otherwise, it gives the contribution since the starting year of the runs. In deciding whether to assess the total or partial anthropogenic contribution, it is important to note that models are subject to biases and, thus, they need to be calibrated to observations. While this is possible for simulations of the recent past climate, calibration of pre-industrial simulations is challenging because of the lack of observations. Therefore, a trade-off between reliable calibration and partial attribution needs to be made. Here we adopt the second approach and use simulations starting in 1960. This means that our approach only gives the anthropogenic contribution to changes in extremes since 1960, but it has the benefit of using a sufficient amount of observations for calibration. To our knowledge, our study presents the first EEA analysis conducted using a max-stable model. The advantage of using such a model is that it enables us to pool all the observations together in a mathematically consistent way, leading to a more robust calibration.

Changes in the occurrence probability attributable to anthropogenic forcing since 1960 are assessed in terms of the FAR[35], which is calculated as:

$$\mathrm{FAR} = 1 - \frac{p_{1960}}{p_{\mathrm{event}}} \tag{6}$$

in which $p_{1960}$ is the occurrence probability of the event in 1960 and $p_{\mathrm{event}}$ is its probability at the time (year) when the actual event occurred if only changes because of anthropogenic forcing are considered. For example, if the event occurred in 2013, then $p_{\mathrm{event}}$ would be computed as:

$$p_{\mathrm{event}} = 1 - \mathrm{F}\left(h; \mu_{1960} + \Delta\mu_{2013}, \sigma, \xi\right) \tag{7}$$

in which $\mu_{1960}$ is the value of $\mu$ in 1960 and $\Delta\mu_{2013}$ is the change in $\mu$ owing to anthropogenic forcing between 1960 and 2013.

## Data availability

The high-frequency tide gauge data used in this study for the period 1960–2013 are available from the Global Extreme Sea Level Analysis project (https://www.gesla.org/), whereas data for the period 2014–2018 are from the British Oceanographic Data Centre (https://www.bodc.ac.uk/data/hosted_data_systems/sea_level/uk_tide_gauge_network/) and the Copernicus Marine Environment Monitoring Service (https://resources.marine.copernicus.eu/product-detail/INSITU_GLO_NRT_OBSERVATIONS_013_030/). The ensemble of climate simulations is available from https://doi.org/10.5285/0cea8d7aca57427fae92241348ae9b03 (baseline folder). The observed annual maxima from tide gauge records, the ensemble of surge simulations, as well as the Bayesian solutions from BHM1 and BHM2 have been deposited in Zenodo (https://doi.org/10.5281/zenodo.5749736).

## Code availability

The code that implements the BHM is available at Zenodo (https://doi.org/10.5281/zenodo.5035438).


41. Woodworth, P. L. et al. Towards a global higher-frequency sea level dataset. *Geosci. Data J.* **3**, 50–59 (2017).
42. Killick, R., Fearnhead, P. & Eckley, I. A. Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* **107**, 1590–1598 (2012).
43. Kalnay, et al. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–470 (1996).
44. Holgate, S. J. et al. New data systems and products at the Permanent Service for Mean Sea Level. *J. Coast. Res.* **29**, 493–504 (2013).
45. Reich, B. J. & Shaby, B. A. A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat* **6**, 1430–1451 (2012).
46. Stephenson, A. G., Shaby, B. A., Reich, B. J. & Sullivan, A. L. Estimating spatially varying severity thresholds of a forest fire danger rating system using max-stable extreme-event modeling. *J. Appl. Meteorol. Climatol.* **54**, 395–407 (2015).
47. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
48. Guillod, B. P. et al. weather@home 2: validation of an improved global–regional climate modelling system. *Geosci. Model Dev.* **10**, 1849–1872 (2017).
49. Rayner, N. A. et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos.* **108**, 4407 (2003).
50. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
51. Hersbach, et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).


**Acknowledgements** We acknowledge the GESLA project for assembling and making the tide gauge data available. F.M.C. was supported by the Natural Environment Research Council (NERC) National Capability funding. M.G.T. and T.W. were supported by the National Aeronautics and Space Administration (NASA) under the New (Early Career) Investigator Program (NIP) in Earth Science (grant number 80NSSC18K0743) and the NASA Sea Level

Science Team (grant number 80NSSC20K1241). T.W. also acknowledges support from the National Science Foundation (under grant ICER-1854896). We acknowledge conversations with M. Marcos and also thank her for providing the tide gauge data (from the British Oceanographic Data Centre and the Copernicus Marine Environment Monitoring Service) for the period 2014–2018.

**Author contributions** F.M.C. conceived and designed the study, with input from all authors. T.W. and M.G.T. produced the ensemble of surge simulations. S.N.S. provided the ensemble of climate simulations. F.M.C. performed the analyses and wrote the manuscript, with contributions from all authors.

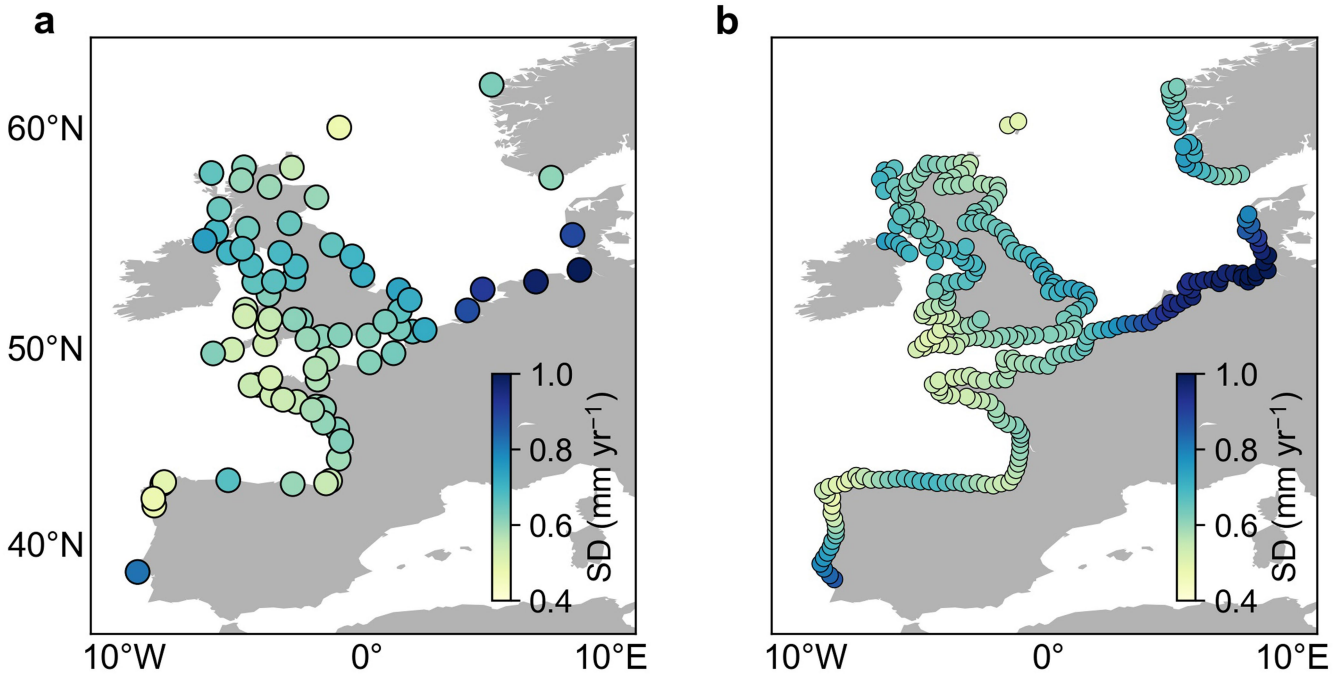
**Competing interests** The authors declare no competing interests.

**Additional information**

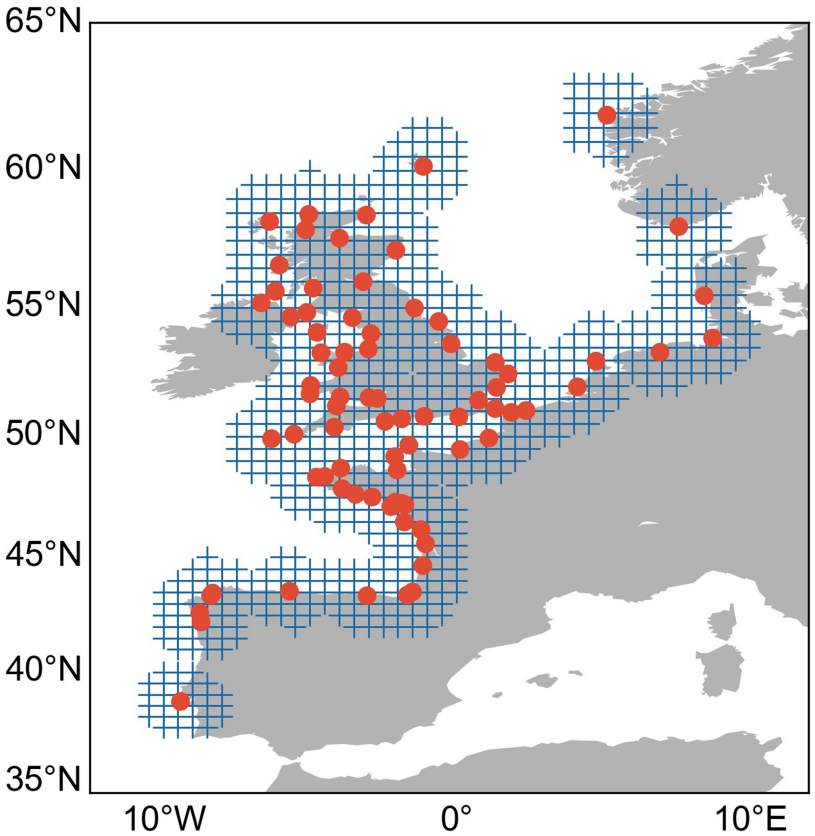
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04426-5.

**Correspondence and requests for materials** should be addressed to Francisco M. Calafat.

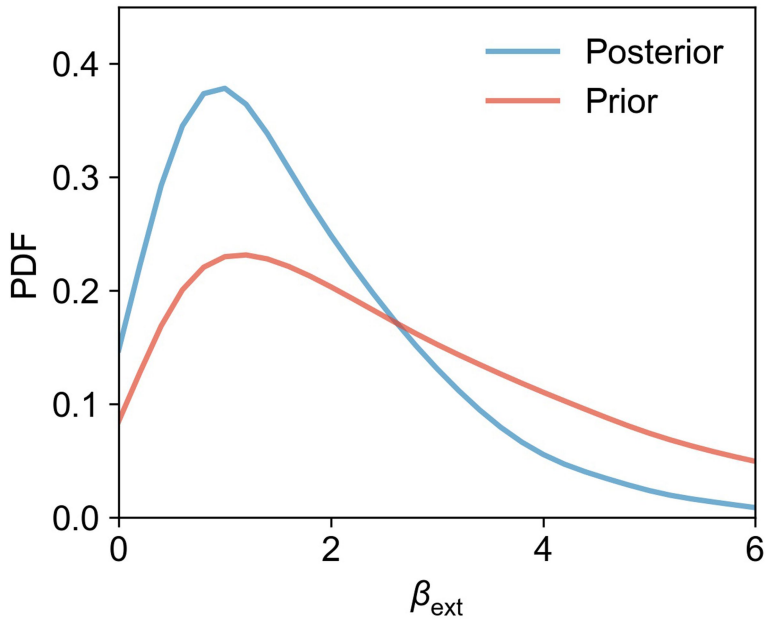
**Peer review information** *Nature* thanks Nadia Bloemendaal, Jérémy Rohmer and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.




**Extended Data Fig. 1 | Uncertainty of estimated $\mu$ trends at individual locations.** Posterior standard deviations for the $\mu$ trends at tide gauge sites (**a**) and gridded locations (**b**). These standard deviations correspond to the $\mu$ trends shown in Fig. 1a, b.




**Extended Data Fig. 2 | Tide gauge stations and spatial knots.** Location of the tide gauge stations used in the analysis of extremes (red circles), along with the spatial knots used to construct the spatial residual process in the BHM (blue crosses).




**Extended Data Fig. 3 | Amplitude of the anthropogenic fingerprint.** Posterior (blue) and prior (red) distributions for the amplitude of the anthropogenic fingerprint ($\beta_{ext}$). The posterior has been estimated by fitting BHM2 to the tide gauge observations.

**Extended Data Table 1 | Scalar parameters of the BHM and prior distributions**

| Parameter | Units | Description | Mean | 5th | 95th | Prior distribution | *R*-hat | $n_{\mathrm{eff}}/it$ |
|---|---|---|---|---|---|---|---|---|
| $\xi$ | - | GEV shape parameter | -0.06 | -0.08 | -0.04 | $\mathcal{N}(-0.05, s_\xi)$ | 1.00 | 0.38 |
| $s_\xi$ | - | Stand. deviation for $\xi$ | 0.03 | 0.00 | 0.09 | half-$\mathcal{N}(0,0.06)$ | 1.00 | 0.95 |
| $\beta_{\mu,\mathrm{width}}$ ($\times 10^3$) | m km$^{-1}$ | Regression Coeff. for $\mu$ Shelf width | 1.34 | 1.13 | 1.56 | $\mathcal{N}(0,5)$ | 1.02 | 0.08 |
| $\beta_{\mu,\mathrm{interc}}$ | m | Regression Coeff. for $\mu$ Intercept | 0.39 | 0.29 | 0.48 | $\mathcal{N}(0,5)$ | 1.03 | 0.07 |
| $\beta_{\sigma,\mathrm{width}}$ | - | Regression Coeff. for $\log\sigma$ Shelf width | 1.67 | 1.25 | 2.12 | $\mathcal{N}(0,5)$ | 1.01 | 0.29 |
| $\beta_{\sigma,\mathrm{interc}}$ | - | Regression Coeff. for $\log\sigma$ Intercept | -2.89 | -3.13 | -2.67 | $\mathcal{N}(0,5)$ | 1.01 | 0.29 |
| $\gamma_\mu$ | mm yr$^{-2}$ | Stand. deviation of GP $\mu_{\mathrm{trend}}$ | 0.20 | 0.02 | 0.45 | half-$\mathcal{N}(0,1)$ | 1.02 | 0.14 |
| $\gamma_{\mu 0}$ | m | Stand. deviation of GP Initial state of $\mu$ | 0.23 | 0.20 | 0.28 | half-$\mathcal{N}(0,1)$ | 1.02 | 0.13 |
| $\gamma_{\mu 00}$ | mm yr$^{-1}$ | Stand. deviation of GP Initial state of $\mu_{\mathrm{trend}}$ | 0.82 | 0.06 | 1.79 | half-$\mathcal{N}(0,1)$ | 1.01 | 0.15 |
| $\gamma_\sigma$ | - | Stand. deviation of GP $\log\sigma$ | 0.36 | 0.28 | 0.47 | half-$\mathcal{N}(0,1)$ | 1.01 | 0.19 |
| $\rho_\mu$ ($\times 10^{-3}$) | km | Length scale of GP $\mu_{\mathrm{trend}}$ | 0.52 | 0.15 | 0.98 | half-$\mathcal{N}(0,0.5)$ | 1.00 | 0.49 |
| $\rho_{\mu 0}$ ($\times 10^{-3}$) | km | Length scale of GP Initial state of $\mu$ | 0.05 | 0.03 | 0.07 | half-$\mathcal{N}(0,0.5)$ | 1.02 | 0.05 |
| $\rho_{\mu 00}$ ($\times 10^{-3}$) | km | Length scale of GP Initial state of $\mu_{\mathrm{trend}}$ | 0.50 | 0.13 | 0.99 | half-$\mathcal{N}(0,0.5)$ | 1.00 | 0.40 |
| $\rho_\sigma$ ($\times 10^{-3}$) | km | Length scale of GP $\log\sigma$ | 0.12 | 0.08 | 0.18 | half-$\mathcal{N}(0,0.5)$ | 1.03 | 0.11 |
| $\tau$ ($\times 10^{-3}$) | km | Length scale of Gaussian Kernels (residual process) | 0.15 | 0.12 | 0.21 | half-$\mathcal{N}(0,0.2)$ | 1.06 | 0.03 |
| $\alpha$ | - | Degree of residual dependence | 0.55 | - | - | - | - | - |

Posterior distribution mean and 5–95% CI for the scalar parameters of the BHM, along with the prior distribution ascribed to each parameter. The potential scale reduction statistic (*R*-hat) and the effective sample size per iteration ($n_{\mathrm{eff}}$/it) are also shown. In general, *R*-hat should be close to 1 at convergence, whereas $n_{\mathrm{eff}}$/it > 0.003 indicates low autocorrelation. A high-resolution knot grid leads to highly correlated samples for $\alpha$. To address this, we treat $\alpha$ as fixed and set its value equal to the posterior mean from a fit using a coarser 1°×1° grid of knots. We refer the reader to ref.[13] for a detailed description of the parameters and justification for prior choices.