Probabilistic reanalysis of storm surge extremes in Europe

Francisco M. Calafat[a,1] and Marta Marcos[b,c]

[a]Department of Marine Physics and Ocean Climate, National Oceanography Centre, Liverpool L3 5DA, United Kingdom; [b]Department of Oceanography and Global Change, Mediterranean Institute for Advanced Studies, Spanish National Research Council and University of the Balearic Islands (CSIC-UIB), Esporles 07190, Spain; and [c]Department of Physics, University of the Balearic Islands, Palma 07122, Spain

Extreme sea levels are a significant threat to life, property, and the environment. These threats are managed by coastal planers through the implementation of risk mitigation strategies. Central to such strategies is knowledge of extreme event probabilities. Typically, these probabilities are estimated by fitting a suitable distribution to the observed extreme data. Estimates, however, are often uncertain due to the small number of extreme events in the tide gauge record and are only available at gauged locations. This restricts our ability to implement cost-effective mitigation. A remarkable fact about sea-level extremes is the existence of spatial dependences, yet the vast majority of studies to date have analyzed extremes on a site-by-site basis. Here we demonstrate that spatial dependences can be exploited to address the limitations posed by the spatiotemporal sparseness of the observational record. We achieve this by pooling all of the tide gauge data together through a Bayesian hierarchical model that describes how the distribution of surge extremes varies in time and space. Our approach has two highly desirable advantages: 1) it enables sharing of information across data sites, with a consequent drastic reduction in estimation uncertainty; 2) it permits interpolation of both the extreme values and the extreme distribution parameters at any arbitrary ungauged location. Using our model, we produce an observation-based probabilistic reanalysis of surge extremes covering the entire Atlantic and North Sea coasts of Europe for the period 1960–2013.

sea level | extremes | storm surge | flooding | Bayesian hierarchical model

Extreme sea levels can have profound impacts on coastal areas, including significant loss of life and damage to property and to the environment. They cause billion-dollar disaster events in countries across the globe and floods related to sea level now cost the world 10s of billions of dollars each year (1). And with climate projections indicating a significant increase in the intensity and frequency of sea-level extremes by 2100 (2, 3), those numbers are bound to grow even further. To manage these threats, coastal planners use measures of extreme event likelihood to estimate risk and determine appropriate levels of protection that balance expected damage with protection costs. When risk can be accurately estimated, a well-designed risk mitigation plan can save both lives and money by reducing disaster impact while avoiding needless costly overprotection measures. In reality, however, estimates of event probabilities, which are central to risk estimation, are often subject to large uncertainty, primarily due to the sparseness of the observational record. This uncertainty can lead to a significant shortfall in the performance of risk mitigation strategies, including the premature failure of infrastructure, with disastrously expensive consequences. Reducing uncertainty in existing estimates of probabilities of sea-level extremes is therefore a priority, in order to enable more accurate risk estimation, and thus more effective mitigation.

Extreme value theory (EVT) provides the most rigorous statistical framework for the analysis of extremes and underlies the majority of existing work on sea-level extremes. Generally, there are two ways of defining extremes, both widely used: the block-maxima method, which divides the observations into consecutive nonoverlapping blocks (or periods), typically years, and selects the maximum value in each block; and the peaks-over-threshold method, which considers all of the values above a certain threshold. The central result of EVT asserts that the only possible limiting distribution of block maxima is the generalized extreme value (GEV) distribution (4, 5). Analogously, threshold excesses can only converge in distribution to the generalized Pareto distribution (6). This fundamental result means that, assuming the observations are independent and identically distributed, the distribution of extremes can be characterized simply by fitting one of those limiting distributions to the extreme data.

While simple conceptually, the application of EVT to sea-level data involves a number of challenges. First, the historical sea-level record, consisting of tide gauge measurements, provides only a small sample of extreme events, from which accurate estimation of the distribution parameters is difficult. Second, there is abundant observational evidence that the sea-level distribution is changing with time, both its mean values (7) and its tail behavior (8–11), which violates the assumption of stationarity on which classical EVT is predicated. Taking this nonstationarity into account is essential to ensure not only that EVT remains applicable, but also that risk mitigation strategies select a level of protection that matches the real risk of extremes. Last but not least, tide gauge observations are only available at a small number of sites, while

estimates of event probabilities are also needed at many other ungauged coastal locations. Dynamical models can provide simulated data with better spatial coverage, but extreme values tend to be underestimated in such models (12–15). Furthermore, existing dynamical simulations span only the most recent decades, and hence they face the same problem of a small sample of events as the observations.

Studies of extremes have become a preeminent focus for sea-level research in recent decades. Although EVT is central to most of these studies, they differ in their geographical scope, their definition of extremes, and in how they address nonstationarity. Some studies are global in scope (8, 9, 14, 16, 17), whereas many others have a regional or local focus (10, 13, 18–28); some studies adopt a block-maxima approach (e.g., ref. 9), while others use threshold excesses (e.g., ref. 23); some incorporate nonstationarity through state-space methods (e.g., ref. 9), whereas others model it using fits to running windows (e.g., ref. 10) or parametric approaches (8). Beyond these differences, all these studies share a common trait: they analyze extremes on a site-by-site basis by fitting a model separately at each tide gauge station. The limitations of this approach are readily apparent. At sites with long tide gauge records, the approach is generally able to constrain the distribution parameters reasonably well, however at sites with little data (the vast majority) estimates have large uncertainty. Crucially, the approach tells us nothing about extremes at ungauged locations.

Here, we present an approach based on a Bayesian hierarchical model (see ref. 29 for a general description of hierarchical models) that addresses the challenges posed by the sparseness of the observational record and overcomes the limitations of the traditional site-by-site analysis. Our focus here is on the surge contribution, which is the part of sea level that remains after removal of the tide and the mean sea level. While spatial hierarchical models have never been used for sea-level extremes (to our knowledge), they have been successfully applied to other types of extremes (30–35), and are emerging as the state-of-the-art tool for geostatistical modeling in general. Our approach relies on the fact that, although observations are spatially discrete, the storm surge process is continuous in space and varies smoothly with length scales similar to those of the weather regimes that give rise to the surges. This means that, although tide gauge sites might have some degree of individuality due, for example, to local bathymetric features, there will also be dependences among them. The key idea is to exploit these dependences in order to enable sharing of information across stations. Such data pooling cannot only drastically reduce estimation uncertainty, but also allows for the interpolation of the extreme surge field at unobserved locations and times.

The idea of pooling information across space has been applied before to the analysis of surge extremes, primarily through the method of regional frequency analysis (RFA) (36, 37). RFA involves, first, predefining homogeneous spatial regions, then normalizing the extreme data in each region by an index flood measure, and last fitting an extreme value distribution to the pooled normalized data. While RFA represents an improvement over the traditional single-site analysis as it allows for more precise estimates of event probabilities, it has its limitations. In particular, the specification of homogeneous regions introduces artificial boundaries that can lead to discontinuities in the extreme field, contradicting the physical expectation that the surge process is spatially continuous. In addition to this, RFA does not permit incorporation of physical information through covariates, and accounting for nonstationarity is unfeasible, or at best problematic. Notably, it is unclear how errors propagate through the various steps of the procedure, which presents an obstacle to obtaining proper uncertainty estimates. Relatedly, RFA typically ignores spatial dependence when estimating the extreme distribution parameters, which artificially narrows confidence intervals,

although a modification to correct for this has recently been proposed (38).

In contrast to RFA, our approach captures spatial dependence in both the surge annual maxima, via a max-stable process, and in the GEV parameters, through latent processes and physical covariates. A max-stable process is the infinite-dimensional generalization of the GEV distribution (39), and hence the most appropriate choice to model pointwise maxima. The use of spatial models avoids spatial discontinuities and enables us to flexibly account for site differences while also pooling information across sites. Furthermore, the hierarchical approach allows us to prescribe rich classes of spatiotemporal models. For example, our model introduces temporal variation in the GEV location parameter to account for long-term changes in extremes arising from climate change. Importantly, our approach naturally accommodates data gaps, involves a comprehensive treatment of uncertainties with rigorous error propagation, and allows estimation of both the annual maxima and the GEV parameters at any arbitrary location, either gauged or ungauged. This enables us to produce observation-based estimates of the GEV parameters and annual maxima on a grid covering the entire Atlantic and North Sea coastlines of Europe for the period 1960 through 2013 (40).

This study focuses on describing the Bayesian hierarchical model, assessing its performance through a number of evaluation metrics, and presenting the reanalysis.

Formulation of the Bayesian Hierarchical Model

A Bayesian hierarchical model is a full probability model that makes inferences from data about unobserved quantities and is expressed, by Bayes' rule, as a product of conditional distributions or submodels (29, 41): 1) an observation model (called the likelihood) that links the observed data to the spatiotemporal processes; 2) a process model that describes the dynamics of the processes; and 3) a parameter model that models the uncertainty in the parameters and incorporates our prior knowledge about the data and the processes. Bayesian inference relies on evaluating the joint distribution of processes and parameters conditioned on the observed data (called the posterior distribution), and accounts for uncertainty in the observations, processes, and parameters.

In modeling spatial dependence of extremes, it is important to distinguish between two types of dependence (30), which here we refer to as residual and climatological. The former occurs when multiple locations are affected by the same events, while the latter implies locations with similar storminess but not necessarily cooccurrence of events. In other words, residual dependence implies dependence among annual maxima whereas climatological dependence reflects spatial association among the GEV parameters. Here, residual dependence is captured via a max-stable process, while dependence among the GEV parameters is described via latent processes with random effects and bathymetric covariates. The implementation of the max-stable process is the same as described by ref. 32 and we defer to that study for full details.

In our model, the GEV location parameter is allowed to vary in both time and space, whereas the scale parameter varies only in space. The shape parameter is assumed to be constant over the entire domain. This latter assumption is justified by exploratory analysis based on individual GEV fits to the observed annual maxima (*SI Appendix*). Inference in our model is performed using Markov chain Monte Carlo (MCMC) sampling as described in *Materials and Methods*. In the following, we describe in detail the observation and process layers of the hierarchical model. The parameter layer is described in *SI Appendix*.

Observation Layer

Let $Y_t(\mathbf{s})$, $t = 1, \ldots, T$, be the annual maximum surge for year t and at location $\mathbf{s}$, and let $\mathbf{s}_1, \ldots, \mathbf{s}_n$ denote the locations of the tide gauge stations. We assume that the surge process

$Y_t(\mathbf{s})$ is max-stable, and hence its marginal distributions are GEV($\mu_t(\mathbf{s}),\sigma(\mathbf{s}),\xi$), where $\mu_t(\mathbf{s})$ is the location parameter, $\sigma(\mathbf{s})$ is the scale parameter, and ξ is the shape parameter. Following ref. 32, the likelihood can be written as:

$$Y_t(\mathbf{s}_i)\Big|\theta_t(\mathbf{s}_i),\mu_t(\mathbf{s}_i),\sigma(\mathbf{s}_i),\xi,\alpha \overset{\text{ind}}{\sim} \text{GEV}\big(\mu_t^*(\mathbf{s}_i),\sigma_t^*(\mathbf{s}_i),\alpha\xi\big), \quad \textbf{[1]}$$

$$\mu_t^*(\mathbf{s}) = \mu_t(\mathbf{s}) + \frac{\sigma(\mathbf{s})}{\xi}\big(\theta_t(\mathbf{s})^\xi - 1\big), \quad \textbf{[2]}$$

$$\sigma_t^*(\mathbf{s}) = \alpha\sigma(\mathbf{s})\theta_t(\mathbf{s})^\xi, \quad \textbf{[3]}$$

where $\theta_t(\mathbf{s})$ is a spatial process capturing residual dependence and $\alpha \in (0,1)$ is a parameter that controls the relative contribution of small-scale errors.

At this point, it is instructive to make a few observations on the three equations above. First, Eq. **1** means that the annual maxima are independent, conditioned on $\theta_t(\mathbf{s})$, $\mu_t(\mathbf{s})$, $\sigma(\mathbf{s})$, ξ, and α, and are modeled as 1-dimensional GEVs. Second, climatological dependence is captured by specifying spatial models for $\mu_t(\mathbf{s})$ and $\sigma(\mathbf{s})$, whereas residual dependence is modeled by introducing random effects in the GEV parameters through $\theta_t(\mathbf{s})$, as represented by $\mu_t^*(\mathbf{s})$ and $\sigma_t^*(\mathbf{s})$. Hence, it is $\theta_t(\mathbf{s})$ that confers the model its ability to make predictions of the annual maxima at ungauged locations. To illustrate how $\theta_t(\mathbf{s})$ induces residual dependence, it is helpful to consider the cases where $\alpha \to 0$ and $\alpha \to 1$. In the first case, $\sigma_t^*(\mathbf{s}) \to 0$ and $Y_t(\mathbf{s})$ convergences in distribution to GEV($\mu_t^*(\mathbf{s}),0,0$), which implies $Y_t(\mathbf{s}) \approx \mu_t^*(\mathbf{s})$, and thus the surge process becomes a spatial process with strong residual dependence. Conversely, when $\alpha \to 1$, $\theta_t(\mathbf{s}) \to 1$ and $Y_t(\mathbf{s})$ convergences to GEV($\mu_t(\mathbf{s}),\sigma(\mathbf{s}),\xi$), and thus there is no residual dependence. Next, we describe how the processes $\theta_t(\mathbf{s})$, $\mu_t(\mathbf{s})$, $\sigma(\mathbf{s})$ are modeled.

Process Layer

The spatial residual process $\theta_t(\mathbf{s})$ is expressed as

$$\theta_t(\mathbf{s}) = \left(\sum_{l=1}^{L} A_{t,l} w_l(\mathbf{s})^{1/\alpha}\right)^\alpha, \quad \textbf{[4]}$$

where $w_l(\mathbf{s}) \geq 0$ are kernel functions and $A_{t,l}$ are their coefficients. The kernel functions are taken to be scaled Gaussian functions (other kernels are possible):

$$w_l(\mathbf{s}) = \frac{K(\mathbf{s}|\mathbf{v}_l,\tau)}{\sum_{j=1}^{L} K(\mathbf{s}|\mathbf{v}_j,\tau)}, \quad \textbf{[5]}$$

$$K(\mathbf{s}|\mathbf{v}_l,\tau) = \frac{1}{2\pi\tau^2}\exp\left(-\frac{1}{2\tau^2}(\mathbf{s}-\mathbf{v}_l)^T(\mathbf{s}-\mathbf{v}_l)\right), \quad \textbf{[6]}$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_L$ are spatial knots and τ is the characteristic length scale of the residual dependence. The scaling ensures that the kernels sum to 1 at each location, which is required to preserve the max-stability properties of the model (see ref. 32). The location of the spatial knots is shown in *SI Appendix*, Fig. S1*A*.

The coefficients of the kernel functions are assumed to follow a positive stable distribution to ensure max-stability (32):

$$A_{t,l} \overset{\text{iid}}{\sim} \text{PS}(\alpha). \quad \textbf{[7]}$$

The location parameter $\mu_t(\mathbf{s})$ is assumed to vary smoothly with time (i.e., we aim to capture long-term changes such as nonlinear trends as opposed to shorter-term variations which are likely

unresolvable), thus we model it as a spatiotemporal integrated random walk of the form

$$\mu_t(\mathbf{s}) = \mu_{t-1}(\mathbf{s}) + \mu_{trend,t-1}(\mathbf{s}), \quad \textbf{[8]}$$

$$\mu_{trend,t}(\mathbf{s}) = \mu_{trend,t-1}(\mathbf{s}) + \omega_t(\mathbf{s}), \quad \textbf{[9]}$$

where $\omega_t(\mathbf{s})$ is a zero-mean Gaussian process (42) $\omega_t(\mathbf{s}) \sim \text{GP}(0, c(\mathbf{s}, \mathbf{s}'; \gamma_\mu, \rho_\mu))$, with $c(\cdot, \cdot)$ being a covariance function and γ_μ and ρ_μ denoting, respectively, the SD and length scale defining the covariance function. The initial states of the location parameter and its trend are modeled as

$$\mu_{t=0}(\mathbf{s}) \sim \text{GP}\left(\mathbf{x}^T(\mathbf{s})\boldsymbol{\beta}_\mu, c\left(\mathbf{s}, \mathbf{s}'; \gamma_{\mu_0}, \rho_{\mu_0}\right)\right), \quad \textbf{[10]}$$

$$\mu_{trend,t=0}(\mathbf{s}) \sim \text{GP}\left(0, c\left(\mathbf{s}, \mathbf{s}'; \gamma_{\mu_{00}}, \rho_{\mu_{00}}\right)\right), \quad \textbf{[11]}$$

where $\mathbf{x}(\mathbf{s})$ is a 2×1 vector of referenced covariates containing an intercept and the width of the continental shelf at each location, and $\boldsymbol{\beta}_\mu$ is a 2×1 vector of regression coefficients. The reason for using shelf width as a covariate is justified by the theory of storm surges, which states that wider continental shelves lead to larger storm surges (43).

Finally, the logarithm of the scale parameter, $\log\sigma(\mathbf{s})$, is modeled as a spatial process:

$$\log\sigma(\mathbf{s}) \sim \text{GP}(\mathbf{x}^T(\mathbf{s})\boldsymbol{\beta}_\sigma, c(\mathbf{s}, \mathbf{s}'; \gamma_\sigma, \rho_\sigma)). \quad \textbf{[12]}$$

The scale parameter $\sigma(\mathbf{s})$ is restricted to be positive, and taking its logarithm ensures that such bound is enforced. This transformation is standard in Bayesian statistics.

For the covariance function $c(\cdot, \cdot)$ of all Gaussian processes, we assume a Matérn kernel with smoothing parameter $\nu = 5/2$ (42).

Model Validation

Checking that the model performs as expected is a crucial step to ensure that we can confidently use its outputs to learn about extremes. In this regard, we note that, aside from through parameters and processes, uncertainty can enter the model primarily through model inadequacy and lack of observations. It is important to consider these two sources of uncertainty. To this aim, we perform two validation experiments. First, we test the model on synthetic data generated under the same spatiotemporal model as the one used to fit the data. This allows us to isolate the influence from the sparseness of the observational record, since model inadequacy is eliminated by design. The second experiment involves testing the model on real tide gauge data and aims to quantify the real-world skill of the model, but it also allows us to assess, by comparison with the first experiment, the adequacy of the model in reality. For completeness, we perform an additional third experiment with reanalysis data from a dynamical surge model (see *Materials and Methods* for a description of the reanalysis). The surge reanalysis allows us to validate the model at more locations than it is possible using the tide gauge data and provides a means for further assessing the adequacy of the model. The results from the first and third experiments are presented in *SI Appendix*, whereas those from the validation with real tide gauge data are exposed in the following.

A few clarifications first. A description of the tide gauge dataset is given in *Materials and Methods*, while site locations are shown in *SI Appendix*, Fig. S1. Note that whenever we talk about estimates from the hierarchical model and their SEs we are referring to posterior distribution means and SDs, respectively.

Model skill is evaluated in terms of a number of diagnostics such as fractional differences (FDs), SEs, and Spearman's rank correlation for annual maxima predictions (see *Materials and Methods* for an explanation of these diagnostics).

The experiment with synthetic data (*SI Appendix*) shows that, given a perfectly adequate model, the hierarchical model is capable of characterizing the extreme value field at ungauged locations with high accuracy (median FDs of 0.09 and 0.10 for μ and σ, respectively, and a mean Spearman's rank correlation of 0.70 for annual maxima), despite the sparseness of the tide gauge record. To account for model inadequacy and quantify the real-world skill of the model, we perform the following experiment with real tide gauge data. We exclude one tide gauge site, then predict the extreme value field at the location of the omitted site using the remaining data sites, and compare this prediction with the true values. This procedure is repeated for each one of the 79 tide gauge sites. Since the true value of the GEV parameters is unknown, the predictions of μ and σ at the omitted sites are compared with the Bayesian estimates based on the full dataset. The predicted annual maxima are compared with the actual observed values.

Both μ and σ are well captured by the hierarchical model at most excluded data sites (Fig. 1 *A* and *B*), with median FDs of 0.14 and 0.11, respectively. This implies that the estimates of μ and σ at ungauged locations are accurate, on average, to within ~14 and ~11% of the "true" value, respectively. Notably, the FDs are fairly uniform across stations, indicating that model skill is largely independent of location. The predictive skill for annual maxima is also good (Fig. 1 *C* and *D*), with a mean Spearman's rank correlation of 0.62 and a fraction of annual maxima contained by the 1-sigma credible interval of 0.77 (mean value over the 79 stations). The latter is consistent with the theoretical expectation that a 1-sigma credible interval should contain the true value with a probability of

~0.7, and demonstrates that the hierarchical model yields realistic uncertainty estimates. This performance is similar to that estimated using synthetic data in a perfect model setting (*SI Appendix*), corroborating the adequacy of the hierarchical model. It is also comparable to the results from the validation with reanalysis data (*SI Appendix*), giving us additional confidence in the skill of the model.

## Probabilistic Reanalysis of Storm Surge Extremes

We now present the probability distribution of surge extremes as estimated by the Bayesian hierarchical model from the real tide gauge data. The posterior distribution for the model parameters (α, τ, ξ, ρ_σ, ρ_{μ_0}, γ_σ, γ_{μ_0}, β_σ, and β_μ) is shown in *SI Appendix*, Fig. S2. Next we discuss the estimates of the GEV parameters. The values of the time-mean location parameter μ at gauged locations range from 0.4 m along the coastlines of Portugal and Spain to 2 m in the German Bight (*SI Appendix*, Fig. S3A). The scale parameter σ shows a similar spatial structure, with values <0.1 m along southwestern Europe and values as large as 0.5 m in the German Bight (*SI Appendix*, Fig. S3B). The SEs associated with μ and σ also vary with location (*SI Appendix*, Fig. S3 *C and D*) but, in all cases, are at least an order of magnitude smaller than the parameter values (mean values of 2.9 and 1.9 cm, respectively), indicating that the model is able to estimate the parameters at data sites with high precision. The spatial dependence structure underlying the location and scale parameters emerges even more clearly when looking at the gridded estimates (*SI Appendix*, Fig. S4). Both parameters show larger values in the North Sea, particularly in the German Bight, and a general tendency for smaller values as we move southward. SEs at interpolation sites are larger than at gauged locations (mean values of 18.8 cm for μ and 2.5 cm for σ), but they are still only a small fraction of the parameter values at most locations. In particular, SEs are, on average, only 20% of the parameter value, and smaller than 66% (μ) and 46% (σ) of the value at all locations.

The estimates of the GEV parameters can be used to compute extreme event probabilities. In engineering design, such probabilities are typically specified in terms of the N-year return level, which is the level that is exceeded on average once every N years. The N-year return level can be calculated simply as $Q(1 - 1/N; \mu, \sigma, \xi)$, where Q is the quantile function for the GEV distribution. As an illustration, we show gridded estimates of the time-mean 50-y return level for the European coastlines (Fig. 2). As expected, the smallest return levels are found along the coastlines of Portugal and Spain with values of about 0.7 m, whereas the largest values are found in the North Sea with values as large as 4.4 m in the German Bight (Fig. 2*A*). The uncertainty associated with the estimated return levels is small relative to the value of the return levels, with SEs that range from about 0.2 m almost anywhere outside the North Sea to about 0.3 to 0.4 m in the North Sea (Fig. 2*B*). Note, however, that these SEs refer to interpolation sites; estimates at gauged locations have even smaller uncertainties with a median SE of 0.1 m. To put these uncertainties into context, we note that a single-site GEV model based on maximum likelihood estimation yields estimates of the 50-y return levels with a median SE of 0.2 m, which is two times larger than the Bayesian errors at gauged locations and comparable to the errors at ungauged sites. Furthermore, the single-site model is unable to yield meaningful estimates at three locations (Ferrol in Spain, Le Crouesty in France, and Moray Firth in Scotland) due to the failure of the maximum likelihood estimator to converge.

The hierarchical model also produces estimates of the annual maxima in any given year and at any arbitrary location. These estimates enable us, for example, to analyze the magnitude and spatial extent of individual extreme events as they were actually observed. As an illustration, we show the surge levels during the
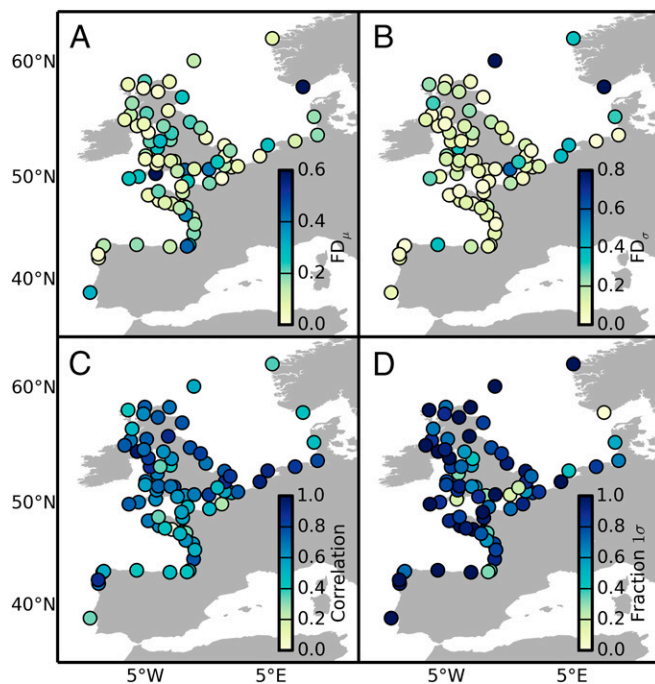
Fig. 1. Validation with real tide gauge data. FDs between the Bayesian estimates based on the full tide gauge dataset and the predicted values of the GEV time-mean location (*A*) and scale (*B*) parameters at omitted sites. The Spearman's rank correlation between the true and predicted annual maxima (*C*), along with the fraction of 1-sigma credible intervals that contains the true extreme value (*D*), are also shown.

Calafat and Marcos
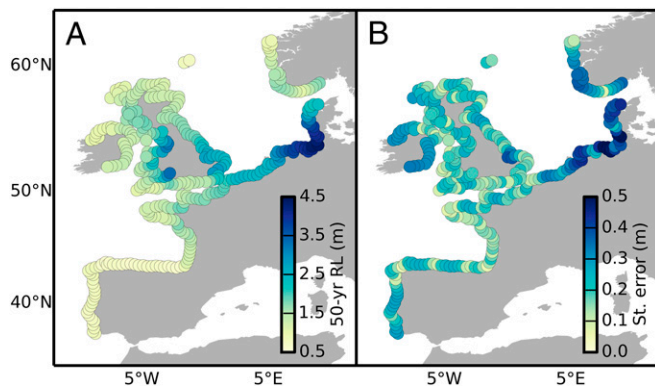
Fig. 2. Bayesian estimates of 50-y return levels. Gridded estimates of the 50-y return levels from the hierarchical model (*A*), along with their SEs (*B*). The time-mean value of the location parameter has been used.

passage of cyclones Xaver in December 2013 and Klaus in January 2009 (Fig. 3), which struck, respectively, the North Sea and Bay of Biscay coastlines and caused record-setting water levels at several coastal locations (e.g., ref. 44). The highest surge levels induced by Xaver occurred near the Germany–Netherlands border (Fig. 3*A*), with values as large as 3.5 m. Surge levels higher than 3 m affected a large section of coastline (>150 km), stretching from northwestern Netherlands to well within the central portion of the German Bight. To put these values into historical context, we also show the predicted time series of annual maxima for the period 1960 through 2013 in the German Bight along with the observed annual maxima at two nearby tide gauges (Fig. 3*C*). We see that, although Xaver is one of the largest events in the period 1960 through 2013, other events of similar magnitude are also evident in both the predicted and observed annual maxima. The series of annual maxima from the two tide gauge records are very coherent between them and also with the predicted annual maxima, giving us confidence in the predictive skill of the model in this region.

Regarding Cyclone Klaus, the maximum surge is found in southwestern France, with values of over 1 m affecting a 300-km section of coastline (Fig. 3*B*). Klaus appears to be the highest surge event over 1960 through 2013 in this region (Fig. 3*D*), both in the predicted and observed sequences of annual maxima. Note, however, that this is a region where the two nearest tide gauges (Arachon-Eyrac and Bocau) show less coherence than those in the North Sea (note, for example, differences in years 2009 and 2010), indicating that smaller length scales might be at work here. They also provide fewer data. The hierarchical model responds to this by widening the credible intervals and pulling the estimate toward the middle of the observed values. This means that events that are localized and show inconsistent values between neighboring stations might be underestimated by the hierarchical model.

Conclusions and Discussion

Understanding the risk of coastal flooding from sea-level extremes requires accurate estimates of their occurrence probabilities. Obtaining such estimates is, however, challenging because of the difficulties posed by the sparseness of the tide gauge record. There is, therefore, a need for new approaches that can address those difficulties and yield more precise estimates of event probabilities. Here, we have demonstrated that this can be achieved by modeling the surge extreme field as a continuous spatiotemporal process, as opposed to modeling data from each tide gauge station independently. Our approach has enabled us to generate observation-based estimates of event probabilities on a grid

covering the entire Atlantic and North Sea coasts of Europe for the period 1960 through 2013. When compared to the traditional single-site modeling approach, our model cuts the uncertainty in estimates of event probabilities (e.g., return levels) by half at gauged locations and yields estimates with comparable uncertainty at ungauged sites. These estimates will help coastal planners and stakeholders to make more confident decisions, particularly in regions with few or no observed data. An additional benefit is that, since our estimates are solely based on observations, they can be used to validate extreme data from dynamical surge models.

Our results are indeed encouraging, but a few remarks need to be made. The ability of the model to share information across space depends directly on the length scales of the extreme field; the smoother the field the more precise the estimates and the further away from a tide gauge we can interpolate with confidence. The GEV parameters tend to vary smoothly in space, following variations in the width of the continental shelf and changes in climate conditions by latitude, and so in general they can be estimated with confidence up to a few hundred kilometers away from any station. We expect this to be the case even in regions affected by tropical cyclones, which have relatively small length scales, since the spatial structure of the GEV parameters is defined by climatological rather than residual dependence. Conversely, coherence among annual maxima is directly related to the spatial extent of individual events and so it tends to show smaller length scales. This means that values of annual maxima at ungauged sites are, in general, more difficult to predict than those of the GEV parameters. The hierarchical model could be extended to capture small-scale effects, for example, by incorporating spatial information via prior distributions, but this would come at the cost of increased model complexity. Other factors that can affect model performance are tide–surge interactions and wave setup effects. The former might introduce errors during the extraction of the surge extremes by causing inaccuracies in the tidal predictions, although here we
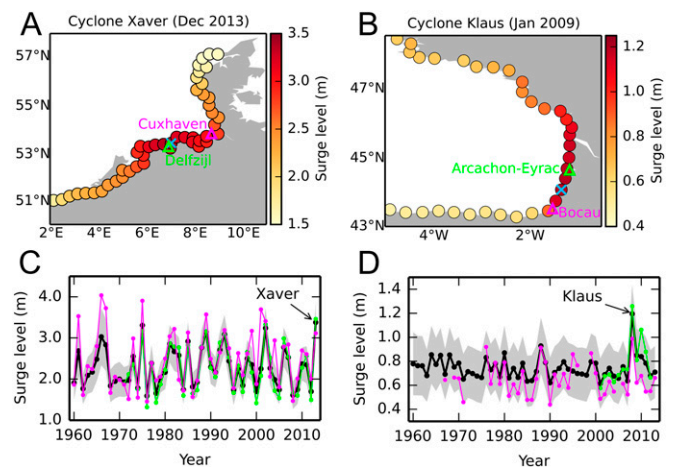
Fig. 3. Bayesian predictions of surge annual maxima. Surge levels induced by cyclones (*A*) Xaver in December 2013 and (*B*) Klaus in January 2009 as estimated by the Bayesian hierarchical model. The blue cross denotes the site with the maximum surge whereas the green and magenta triangles denote the location of the two closest tide gauges on either sides of the blue cross. The predicted sequence of annual maxima (thick black line) at the location showing the maximum surge (the blue cross) during (*C*) Xaver and (*D*) Klaus is also shown, along with the observed annual maxima (green and magenta lines) at the two tide gauges shown in *A* and *B*. The gray shading in *C* and *D* denotes the 1-sigma credible interval associated with the predicted annual maxima. Note that years in our analysis start in April, so Cyclone Klaus, which occurred in January 2009, falls into year 2008.

have aimed to minimize such errors as explained in *SI Appendix*. Wave setup, if present in the tide gauge records after removal of the tide and the mean sea level, might reduce spatial coherence in the extreme values leading to increased uncertainty, but we expect this effect to be small since wave setup is generally small inside harbors where most tide gauges are located. In any case, our model provides realistic uncertainty estimates and these should always be taken into account when using our estimates.

Materials and Methods

Tide Gauge Data. Hourly sea-level observations were obtained from the Global Extreme Sea Level Analysis (GESLA) tide gauge data set (45), which consists of 1,355 records of variable quality and length, obtained from the international databases at the University of Hawaii Sea Level Center and the Global Sea Level Observing System, complemented with additional observations from national and subnational data providers. The surge annual maxima are extracted from each tide gauge record as described in *SI Appendix*.

Inference in the Hierarchical Model. The posterior distribution in our model does not admit a closed-form expression and thus we must resort to MCMC sampling to perform inference. However, inference in our Bayesian hierarchical model is challenging and standard MCMC methods can fail to converge in a reasonable amount of time. Therefore, here we use the No-U-Turn Sampler (NUTS) (46) as implemented by the Stan probabilistic programming language (47). NUTS is a variant of Hamiltonian Monte Carlo with adaptive optimization, and is capable of fitting our model without difficulty, providing fast MCMC mixing and convergence. We run the sampler with four MCMC chains of 2,000 iterations each (warm-up = 1,000). A number of convergence and performance diagnostics for the MCMC sampler are presented in *SI Appendix*.

Storm Surge Reanalysis from a Dynamical Model. Daily storm surge maxima from the Global Tide and Surge Reanalysis (GTSR) numerical model (14) along the European coastlines have been used. GTSR provides near-coast time series of storm surges and tides globally at the spatial resolution of the Dynamic Interactive Vulnerability Assessment model (48), spanning the period 1979 through 2014. Storm surge data are simulated with a hydrodynamic model forced with 10-m wind speed and atmospheric pressure from the ERA-Interim global atmospheric reanalysis (49) and run over an unstructured grid with spatial resolution down to 5 km near the coasts.

Skill Diagnostics. The skill of the model in estimating the GEV parameters is evaluated in terms of FDs, which are defined as $FD = |(x_{true} - \hat{x})/x_{true}|$, where x_{true} and $\hat{x}$ are the true and estimated values of the parameter, respectively. Note that, since this metric compares true and estimated values, it provides a measure of the accuracy of the hierarchical model. The SEs yielded by the model, on the other hand, are a metric for model precision.

To estimate the skill of the model in interpolating the annual maxima at ungauged locations, we use two metrics: 1) the Spearman's rank correlation between the true and predicted extreme values; and 2) the fraction of true extreme values that fall within the 1-sigma credible interval. That is, at each location, we calculate how many of the 54 true extreme values over 1960 through 2013 are contained by the 1-sigma credible intervals, and then divide the result by 54.

Data and Code Availability. The probabilistic surge reanalysis data presented in this paper and the code to implement the hierarchical model are both available via Zenodo, DOI: 10.5281/zenodo.3471600; DOI: 10.5281/zenodo.3442167.

1. J. Hinkel *et al.*, Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3292–3297 (2014).
2. R. E. Kopp *et al.*, Probabilistic 21st and 22nd century sea-level projections at a global network of tide-gauge sites. *Earth's Future* **2**, 383–406 (2014).
3. M. I. Vousdoukas *et al.*, Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard. *Nat. Commun.* **9**, 2360 (2018).
4. E. M. Gnedenko, Sur la distribution limite du terme maximum d'une serie aleatoire. *Ann. Math.* **44**, 423–453 (1943).
5. D. McFadden, Modeling the choice of residential location. *Transp. Res. Rec.* **672**, 72–77 (1978).
6. J. Pickands, Statistical inference using extreme order statistics. *Ann. Stat.* **3**, 119–131 (1975).
7. J. A. Church *et al.*, "Sea level 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*, T. F. Stocker *et al.*, Eds. (Cambridge University Press, Cambridge, 2013), pp. 1137–1216.
8. M. Menéndez, P. L. Woodworth, Changes in extreme high water levels based on a quasi-global tide-gauge data set. *J. Geophys. Res.* **115**, C10011 (2010).
9. M. Marcos, F. M. Calafat, A. Berihuete, S. Dangendorf, Long-term variations in global sea level extremes. *J. Geophys. Res. Oceans* **120**, 8115–8134 (2015).
10. T. Wahl, D. P. Chambers, Evidence for multidecadal variability in US extreme sea level records. *J. Geophys. Res. Oceans* **120**, 1527–1544 (2015).
11. M. Marcos, P. L. Woodworth, 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).
12. D. Conte, P. Lionello, Characteristics of large positive and negative surges in the Mediterranean Sea and their attenuation in future climate scenarios. *Global Planet. Change* **111**, 159–173 (2013).
13. F. M. Calafat *et al.*, The ability of a barotropic model to simulate sea level extremes of meteorological origin in the Mediterranean Sea, including those caused by explosive cyclones. *J. Geophys. Res. Oceans* **119**, 7840–7853 (2014).
14. S. Muis, M. Verlaan, H. C. Winsemius, J. C. J. H. Aerts, P. J. Ward, A global reanalysis of storm surges and extreme sea levels. *Nat. Commun.* **7**, 11969 (2016).
15. M. I. Vousdoukas, E. Voukouvalas, A. Annunziato, A. Giardino, L. Feyen, Projections of extreme storm surge levels along Europe. *Clim. Dyn.* **47**, 3171–3190 (2016).
16. M. A. Merrifield, A. S. Genz, C. P. Kontoes, J. J. Marra, Annual maximum water levels from tide gauges: Contributing factors and geographic patterns. *J. Geophys. Res. Oceans* **118**, 2535–2546 (2013).
17. J. Rohmer, G. Le Cozannet, 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).
18. M. N. Tsimplis, D. Blackman, Extreme sea-level distribution and return periods in the Aegean and the Ionian Seas. *Estuar. Coast. Shelf Sci.* **44**, 79–89 (1997).
19. P. L. Woodworth, D. L. Blackman, Changes in high waters at Liverpool since 1768. *Int. J. Climatol.* **22**, 697–714 (2002).
20. S. Coles, J. Tawn, Bayesian modelling of extreme surges on the UK east coast. *Philos. Trans. A Math. Phys. Eng. Sci.* **363**, 1387–1406 (2005).
21. M. Marcos, M. N. Tsimplis, A. G. P. Shaw, Sea level extremes in southern Europe. *J. Geophys. Res.* **114**, C01007 (2009).
22. S. Dangendorf *et al.*, North Sea storminess from a novel storm surge record since AD 1843. *J. Clim.* **27**, 3582–3595 (2014).
23. X. Feng, M. N. Tsimplis, Sea level extremes at the coasts of China. *J. Geophys. Res. Oceans* **119**, 1593–1608 (2014).
24. I. D. Haigh *et al.*, Estimating present day extreme water level exceedance probabilities around the coastline of Australia: Tides, extra-tropical storm surges and mean sea level. *Clim. Dyn.* **42**, 121–138 (2014).
25. S. A. Talke, P. Orton, D. A. Jay, Increasing storm tides in New York Harbor, 1844–2013. *Geophys. Res. Lett.* **41**, 3149–3155 (2014).
26. A. Cid, T. Wahl, D. P. Chambers, S. Muis, Storm surge reconstruction and return water level estimation in Southeast Asia for the 20th century. *J. Geophys. Res. Oceans* **123**, 437–451 (2018).
27. C. Tebaldi, B. H. Strauss, C. E. Zervas, Modelling sea level rise impacts on storm surges along US coasts. *Environ. Res. Lett.* **7**, 014032 (2012).
28. M. K. Buchanan, R. E. Kopp, M. Oppenheimer, C. Tebaldi, Allowances for evolving coastal flood risk under uncertain local sea-level rise. *Clim. Change* **137**, 347–362 (2016).
29. A. Gelman *et al.*, *Bayesian Data Analysis* (Chapman and Hall/CRC Press, Boca Raton, FL, ed. 3, 2014).
30. D. Cooley, D. Nychka, P. Naveau, Bayesian spatial modeling of extreme precipitation return levels. *J. Am. Stat. Assoc.* **102**, 824–840 (2007).
31. A. C. Davison, S. A. Padoan, M. Ribatet, Statistical modeling of spatial extremes. *Stat. Sci.* **27**, 161–186 (2012).
32. B. J. Reich, B. A. Shaby, A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat.* **6**, 1430–1451 (2012).
33. B. A. Shaby, B. J. Reich, Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics* **23**, 638–648 (2012).
34. A. G. Stephenson, E. A. Lehmann, A. Phatak, A max-stable process model for rainfall extremes at different accumulation durations. *Weather Clim. Extrem.* **13**, 44–53 (2016).
35. J. Tawn, R. Shooter, R. Towe, R. Lamb, Modelling spatial extreme events with environmental applications. *Spat. Stat.* **28**, 39–58 (2018).
36. L. Bardet, C.-M. Duluc, V. Rebour, J. L'Her, Regional frequency analysis of extreme storm surges along the French coast. *Nat. Hazards Earth Syst. Sci.* **11**, 1627–1639 (2011).

Calafat and Marcos

37. P. Bernardara, M. Andreewsky, M. Benoit, Application of regional frequency analysis to the estimation of extreme storm surges. *J. Geophys. Res.* **116**, C02008 (2011).
38. J. Weiss, P. Bernardara, M. Benoit, Modeling intersite dependence for regional frequency analysis of extreme marine events. *Water Resour. Res.* **50**, 5926–5940 (2014).
39. M. Schlather, Models for stationary max-stable random fields. *Extremes* **5**, 33–44 (2002).
40. F. M. Calafat, M. Marcos, Probabilistic reanalysis of storm surge extremes in Europe. Zenodo. https://doi.org/10.5281/zenodo.3471600. Deposited 3 October 2019.
41. N. Cressie, C. Wikle, *Statistics for Spatio-Temporal Data* (John Wiley & Sons, New York, 2011).
42. C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).
43. A. E. Gill, *Atmosphere-Ocean Dynamics* (Academic, San Diego, 1982).
44. S. Dangendorf, A. Arns, J. G. Pinto, P. Ludwig, J. Jensen, The exceptional influence of storm "Xaver" on design water levels in the German Bight. *Environ. Res. Lett.* **11**, 054001 (2016).
45. P. L. Woodworth *et al.*, Towards a global higher-frequency sea level data set. *Geosci. Data J.* **3**, 50–59 (2017).
46. M. D. Hoffman, A. Gelman, The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
47. B. Carpenter *et al.*, Stan: A probabilistic programming language. *J. Stat. Softw.* **76**, 10.18637/jss.v076.i01 (2017).
48. A. T. Vafeidis *et al.*, A new global coastal database for impact and vulnerability analysis to sea-level rise. *J. Coast. Res.* **244**, 917–924 (2008).
49. D. P. Dee *et al.*, The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).

EARTH, ATMOSPHERIC, AND PLANETARY SCIENCES