# ACCOUNTING FOR SEASONALITY IN EXTREME SEA LEVEL ESTIMATION

# BY ELEANOR D'ARCY<sup>1,a</sup>, JONATHAN A. TAWN<sup>1</sup>, AMÉLIE JOLY<sup>2</sup> AND DAFNI E. SIFNIOTI<sup>2</sup>

<sup>1</sup>STOR-i Centre for Doctoral Training, Department of Mathematics and Statistics, Lancaster University, LA1 4YR, UK, <sup>a</sup>e.darcy@lancaster.ac.uk

<sup>2</sup>EDF Energy R&D UK Centre, Croydon, CR0 2AJ, UK

Reliable estimates of sea level return levels are crucial for coastal flooding risk assessments and for coastal flood defence design. We describe a novel method for estimating extreme sea levels that is the first to capture seasonality, interannual variations and longer term changes. We use a joint probabilities method, with skew surge and peak tide as two sea level components. The tidal regime is predictable but skew surges are stochastic. We present a statistical model for skew surges, where the main body of the distribution is modelled empirically whilst a non-stationary generalised Pareto distribution (GPD) is used for the upper tail. We capture within-year seasonality by introducing a daily covariate to the GPD model and allowing the distribution of peak tides to change over months and years. Skew surge-peak tide dependence is accounted for via a tidal covariate in the GPD model and we adjust for skew surge temporal dependence through the subasymptotic extremal index. We incorporate spatial prior information in our GPD model to reduce the uncertainty associated with the highest return level estimates. Our results are an improvement on current return level estimates, with previous methods typically underestimating. We illustrate our method at four UK tide gauges.

**1.** Introduction. Extreme sea levels pose an increasing risk to coastline communities. In the absence of any mitigation, the impacts can be severe: fatality, infrastructure damage and habitat destruction. Estimates of sea level return levels are fundamental for various purposes, including coastal flood defence design and flood risk assessments. A return level is the value we expect the annual maximum sea level to exceed with probability p. For a stationary series, this corresponds to a value exceeded once every 1/p years, on average. We are particularly interested in rare events, where  $p \in [10^{-4}, 10^{-1}]$  to cover sea levels that are important to a range of industries affected by coastal flooding. For example, we obtain data from the Heysham tide gauge: Heysham is a coastal town in north-west England that is home to two nuclear power stations. Nuclear regulators require accurate return level estimates for p = $10^{-4}$ . We also consider a further three gauges at Lowestoft, Newlyn and Sheerness; Sheerness is particularly important since it is located on the river Thames estuary where extreme sea levels can propagate down the river towards London. The annual return level, or equivalently its exceedance probability, changes with year if the series is non-stationary and discussions of appropriate design levels involve the intended lifespan of the defence (Rootzén and Katz, 2013). We focus on a fixed p and recognise that the return level varies across years.

Since the UK is regularly subject to coastal flooding, estimating extreme sea levels is crucial. The 1953 North Sea flood is the worst on record for the 20<sup>th</sup> century in the UK. Coastal defences were breached in 1,200 places leading to the evacuation of 30,000 people, damage to 24,000 properties and a death toll of 307 in England alone. The damage was

Keywords and phrases: Extreme sea levels, Generalised Pareto distribution, Joint probabilities method, Nonstationarity, Skew surge.

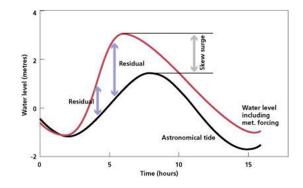


FIG 1. Sea levels during the passage of a surge for a single tidal cycle (NOC, 2021).

estimated at £1.2 billion in 2014 (Wadey et al., 2015). Following this event, coastal flood defences were upgraded around most of the UK. Recently, there has been growing concern regarding anthropogenic sea level rise due to climate change. Rises in the mean sea level, coupled with changes in storm frequency and size, can increase the likelihood of coastal flooding. Therefore, it is increasingly important to accurately estimate sea level return levels so that coastline communities are protected against events such as that in 1953. The latest best estimates for extreme sea levels can be found in Environment Agency (2018).

Pugh and Woodworth (2014) give a comprehensive overview of sea level processes. Sea levels are a combination of mean sea level, tide, surge and waves. We consider still water level (with waves filtered out), for simplicity we refer to this as sea level. The mean sea level trend is removed so that tide and surge are the only components. Tides are the regular and predictable changes in sea levels driven astronomically; these are well understood and perfectly forecast (Egbert and Ray, 2017). Surges define any departure from the predicted tidal regime, often resulting from meteorological forces such as storms and are hence sometimes called storm surges or non-tidal residual (as in Figure 1); these are stochastic. However, surges can also include gauge recording errors and tidal prediction errors, but these errors are typically negligible at well maintained gauges, such as those on the UK National Tide Gauge Network (see Section 3.1). Surges can also be influenced by tide-surge interaction, this is the change in the distribution of surge that is dependent on the tidal level, because the surge is essentially a wave which is influenced by the water depth. This interaction is more prominent in shallow water areas, with the largest surges typically occurring mid-tide on the rising tide and the smallest at high tide (Prandle and Wolf, 1978; Tawn, 1992).

Early methods to estimate extreme sea levels modelled the observed sea levels directly, ignoring the known tidal component. More recent approaches focus on the convolution of the surge and tide distributions. Due to the complex dependence between surge and tide, we instead consider skew surge and peak tide; skew surge is the difference between the maximum observed sea level and the maximum predicted tide (peak tide) within a tidal cycle, regardless of their timing (see Figure 1). There is a single skew surge value associated with each peak tide value (every 12 hours 26 minutes), as opposed to hourly or 15 minute interval observations, these components are preferred because they have much weaker dependence; Williams et al. (2016) demonstrate that it is reasonable to assume these are independent at most UK sites.

We build on the skew surge joint probabilities method (SSJPM) of Batstone et al. (2013), which assumes peak tide and skew surge are independent and that they are stationary processes within and across years, with extreme skew surges modelled by a generalised Pareto distribution (GPD). Our novel approach is able to reflect the realism of the sea level processes

by correcting the simplifying and false assumptions made in the SSJPM. We account for seasonality in the skew surges by adding a daily covariate to the rate of threshold exceedance and scale parameter of the GPD model. We also introduce a tidal covariate to capture skew surge-peak dependence. Skew surges also exhibit temporal dependence; we account for this using the extremal index (Tawn, 1992) but in a different way to previous analyses. Since tides are deterministic, we choose our tidal samples so that they account for monthly and interannual variations. We estimate return levels by deriving distributions for the annual maxima sea levels. We obtain the estimates using our proposed method and for the SSJPM, as well as some intermediate methods, at four sites on the UK coastline. We find that the SSJPM tends to underestimate return levels.

It is fundamental to recognise the uncertainty associated with return level estimates of the highest order to add value to the point estimates. We construct confidence intervals for our estimates using a stationary bootstrap procedure (Politis and Romano, 1994). This preserves the realism of the sea level processes. Our approach is the first to use a bootstrap procedure for uncertainty quantification on extreme sea level estimates. We incorporate prior information for the GPD shape parameter based on related spatial information (Environment Agency, 2018); we show that for the shape parameters and the sea-level return level estimates, this information substantially reduces the lengths of confidence intervals, without significantly changing point estimates.

We discuss the relevant extreme value theory and existing methodology for extreme sea level estimation in Section 2. In Section 3 we introduce the data and explore the seasonality of each component as well as the dependence between them. Section 4 describes the methodology for deriving the annual maxima distribution, starting with an idealised solution derived under simplified assumptions that are then relaxed. Our return level estimates are compared with previous methods in Section 5. Additional supplementary material is presented in D'Arcy et al. (2023); when cross referencing, the section, figure, table or equation number is preceded by 'S'.

## 2. Background to Method.

2.1. Relevant Extreme Value Methods. When deriving a model for extremes, it is natural to first consider the maximum  $M_n$  of a sequence of independent and identically distributed (iid) continuous random variables  $Z_1, \ldots, Z_n$ , i.e.,  $M_n = \max\{Z_1, \ldots, Z_n\}$ . This sequence has marginal distribution function F and upper end point  $z^F$ . If there exists sequences of constants  $\{a_n > 0\}$  and  $\{b_n\}$ , so that the rescaled block maximum  $(M_n - b_n)/a_n$  has a nondegenerate limiting distribution as  $n \to \infty$ , then the cumulative distribution function of this limit must have the form

(2.1) 
$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\},$$

where  $x_+ = \max\{x, 0\}$  for parameters  $(\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$  representing the location, scale and shape, respectively (Coles, 2001). This is the generalised extreme value distribution (GEV) that encompasses three families: for  $\xi > 0$  this corresponds to the Fréchet distribution,  $\xi < 0$  the Weibull and  $\xi = 0$  the Gumbel. Note  $\xi = 0$  should be interpreted as the limit as  $\xi \to 0$ . This result provides asymptotic motivation for using the GEV as a parametric model for observed block maxima and thus a basis to estimate and extrapolate to high return levels. However, this assumes an underlying iid process, which is unrealistic.

Now we relax the independence assumption, so that  $Z_1, \ldots, Z_n$  is a stationary sequence with the same marginal distribution function F. This corresponds to a series whose variables may be mutually dependent, but whose statistical properties are homogeneous through time.

The limiting distribution of the rescaled block maxima of a stationary process satisfying a long-range dependence condition, which ensures events long apart in time are near independent, is  $G^{\theta}(z)$  with G(z) as in equation (2.1) and  $\theta \in (0, 1]$  the extremal index (Leadbetter, Lindgren and Rootzén, 1983). For an independent series  $\theta = 1$ , but the converse is not true.

When a process exhibits extremal dependence, groups of extreme events form above high thresholds; these groups are called clusters. We define different clusters as those separated by some number of non-extreme values (i.e., below a high threshold z), this is called the run length r. Within a cluster, extreme events are considered as dependent whilst exceedances in different clusters are assumed to be independent. The extremal index tells us about clusters since it can be estimated empirically as the reciprocal of the mean cluster size of exceedances of z. This is known as the runs method (Smith and Weissman, 1994). Clusters are identified using an arbitrary choice of run length; this is the main pitfall with the approach. Ferro and Segers (2003) propose the intervals estimator, based on the limiting distribution of normalised times between exceedances of z. This distribution is exponential for independent random variables, otherwise it is a mixture distribution of an exponential with mean  $\theta^{-1}$  and a degenerate probability distribution at zero, with probabilities  $\theta$  and  $1 - \theta$ , respectively. Ferro and Segers (2003) also propose an automatic declustering scheme using the intervals estimate. Both methods only estimate  $\theta$  in the observed range of z. In practice the runs and intervals estimators of  $\theta$  are actually estimators of the subasymptotic extremal index, for threshold level z and run length r, defined by Ledford and Tawn (2003) as

(2.2) 
$$\theta(z,r) = \mathbb{P}(\max\{Z_2,\ldots,Z_r\} < z | Z_1 > z).$$

Then the extremal index is the limit of expression (2.2) as  $z \to z^F$  and  $r \to \infty$ , with z and r tending to their respective limits in a related fashion.

We can also define extremes as exceedances of a high threshold u. If expression (2.1) holds, then for an arbitrary term Z in the sequence  $Z_1, \ldots, Z_n$ , (2.3)

$$\mathbb{P}(Z > b_n + a_n z \mid Z > b_n + a_n u) \to H_u(z) \quad \text{where} \quad H_u(z) = \left[1 + \xi \left(\frac{z - u}{\sigma_u}\right)\right]_+^{-1/\xi}$$

for z > u as  $n \to \infty$ , with  $a_n$ ,  $b_n$  as previously and  $(\sigma_u, \xi) \in \mathbb{R}_+ \times \mathbb{R}$  the scale and shape parameters, respectively (Coles, 2001). The shape parameter is the same as that for the GEV. Whereas the scale is threshold dependent since  $\sigma_u = \sigma + \xi(u - \mu)$  for  $\mu$  and  $\sigma$  the GEV parameters. This is the generalised Pareto distribution (GPD). If  $Z_1, \ldots, Z_n$  are iid, then exceedances of a high threshold u are iid and have limiting GPD tail model

$$\mathbb{P}(Z > z) = \lambda_u \left[ 1 + \xi \left( \frac{z - u}{\sigma_u} \right) \right]_+^{-1/\xi},$$

for z > u where  $\lambda_u = \mathbb{P}(Z > u)$ . Again, if  $Z_1, \ldots, Z_n$  are stationary, a common approach is to identify clusters and decluster them to yield an approximately independent sequence of cluster maxima for which the GPD remains a valid model (Fawcett and Walshaw, 2007).

For non-stationary processes, it is common to allow the parameters of a stationary statistical model to vary with time or another covariate. In a block maxima framework, observations in a block are assumed to be iid so covariates in the GEV parameters cannot change within a block. In contrast, the GPD allows the covariates to vary uncontrolled over consecutive observations. A range of methods can be adopted to incorporate covariates in the model parameters  $\lambda_u, \sigma_u, \xi$ , including via harmonics (Coles, Tawn and Smith, 1994), splines (Jonathan et al., 2014) or generalised additive models (Chavez-Demoulin and Davison, 2005). 2.2. Existing Methodology. The earliest methods to estimate sea level return levels fit a GEV to the annual maxima (Graff, 1978; Coles and Tawn, 1990) or the annual *r*-largest observed sea levels (Tawn, 1988). These direct approaches ignore the known tidal component even though it induces non-stationarity into the sea level series, so that the GEV limit may not be a good approximation. Dixon and Tawn (1999) show that these direct methods underestimate extreme sea levels, especially at longer return levels for tidally dominant sites.

Pugh and Vassie (1978) were the first to exploit the decomposition of sea levels into tide and surge for extreme sea level estimation. This method is known as the joint probabilities method (JPM), where the probability distribution of extreme sea levels is derived by convolution of the distributions of these two components. The JPM gives consideration to all surge values in the data regardless of when they occurred relative to high tide. This approach forms the basis of the subsequent methods, but has some restrictive and unrealistic assumptions. Since the empirical surge distribution is used in the JPM, return level estimates are constrained by the sum of the highest predicted tide and the highest observed surge. Hourly surge observations are assumed to be independent, this is unrealistic as surge exhibits strong temporal dependence (Tawn and Vassie, 1989). Pugh and Vassie (1978) do account for dependence between tide and surge by dividing the tidal range into ordered bands of equal probability and estimating the surge distribution for each band, but this gives results which are sensitive to the choice of the number of bands and their boundary levels.

The revised joint probabilities method (RJPM) of Tawn (1992) attempts to address these limitations. For the upper tail of surges, an extreme value distribution is used to allow extrapolation beyond what has been observed, hence improving return level estimates. To account for temporal dependence, the extremal index is used. Tawn (1992) and Dixon, Tawn and Vassie (1998) model the surge-tide dependence by allowing parameters of the GEV to be functions of the tidal level. This is a difficult task because the relationship is complex (Prandle and Wolf, 1978). Dixon and Tawn (1994) and Haigh, Nicholls and Wells (2010) find that the RJPM is the best performing of these methods for observed and simulated data.

Batstone et al. (2013) propose the skew surge joint probabilities method (SSJPM) to overcome the requirement of modelling the dependence of surge on tide. Their approach assumes that skew surge and peak tide are independent. Williams et al. (2016) demonstrate this is a good approximation both physically and empirically, for many sites in Europe and the USA. The SSJPM fits a GPD to the upper tail of skew surges and the empirical distribution is used for the main body of the distribution. The extremal index is used to measure dependence for sea levels in adjacent tidal peaks. They find  $\theta \approx 1$  for all sites, suggesting no evidence of dependence in the upper tail. Baranes et al. (2020) adapt the SSJPM to account for interannual variations in the tidal regime, considering summer and winter separately and assuming peak tides and skew surge are stationary within each season.

All the methods discussed so far estimate extreme sea levels at a single site, but this gives uncertain return level estimates for sites with shorter observation periods. Spatial pooling can improve estimates at sites with limited or no data. Bernardara, Andreewsky and Benoit (2011) use regional frequency analysis to estimate extreme surges. This involves grouping statistically similar sites into homogeneous regions, then fitting an extreme value model with a constant shape parameter over all sites (Hosking and Wallis, 1997). More recent approaches account for uncertainty in the region selection (Asadi, Engelke and Davison, 2018; Rohrbeck and Tawn, 2021). Batstone et al. (2013) use hindcast sea level data to interpolate SSJPM estimates along the UK coast. Other methods allow parameters of the extreme value distribution to vary with spatial covariates. Coles and Tawn (1990) do this, allowing the location parameter of the GEV for sea level annual maxima to depend on the harmonic tidal constituents. Whereas, Dixon, Tawn and Vassie (1998) spatially smooth parameters of the extreme value model for surges used in the RJPM, using predictable tidal variations along the UK coastline.

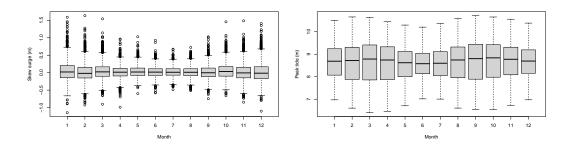


FIG 2. Monthly boxplots of skew surge (left) and peak tide (right) at Heysham.

## 3. Exploratory Analysis.

3.1. *Data.* We use data from the UK National Tide Gauge Network obtained from the British Oceanographic Data Centre; these undergo rigorous quality control before release. Sea level elevations are recorded at 44 sites along the UK coastline. We consider skew surge and peak tide observations, with one record every 12 hours and 26 minutes (i.e., one tidal cycle), at four sites: Heysham, Lowestoft, Newlyn and Sheerness. Heysham is located on the west coast of England and has records from 1964-2016, with 17% missing. Lowestoft is on the east coast of England, with data available from 1964-2020 (4% missing). Sheerness is at the Thames Estuary, also on the east coast of England and has records for England and has records from 1964-2020, with 17% missing. The data are in metres above chart datum.

To compare the relative importance of skew surge and peak tide across sites, we define a surge-tide index as the observed range of skew surges divided by the range of peak tides. Heysham and Newlyn are tidally dominant with indices 0.65 and 0.70, respectively. Lowest-oft and Sheerness are surge dominant with indices 3.36 and 1.66, respectively. We choose to study these sites because they are typically affected by different storms and all have a long observational duration. Heysham and Lowestoft are of particular interest because of their widely different surge-tide indices. Newlyn has the longest study period in the tide gauge network. Howard and Williams (2021) use physical model simulations to conclude that skew surge is dependent on peak tide at Sheerness, so we also study this site. The highest astronomical tide (HAT) observed is 10.72m, 2.92m, 6.10m and 6.26m for Heysham, Lowestoft, Newlyn and Sheerness, respectively. The data were preprocessed in an attempt to remove the linear non-stationary effect of sea level rise on skew surges caused by climate change and isostatic rebound, see Environment Agency (2018) for details. Therefore all results are presented relative to the mean sea level in 2017.

3.2. Seasonality and Temporal Dependence of Skew Surge and Peak Tide. Figures 2 and S9 show monthly boxplots of skew surge and peak tide observations at Heysham and the remaining sites, respectively. The median of the monthly skew surges remain relatively constant across months at Heysham, Newlyn and Sheerness but there is some variation at Lowestoft. The skew surge range varies across the year, exhibiting clear seasonality. Extreme skew surges and extreme peak tides typically occur in winter and at the equinoxes, respectively. Tides also exhibit interannual variability, including the 18.6 year lunar nodal cycle and the 8.85 year cycle of lunar perigee (Pugh and Woodworth, 2014).

To further assess seasonality, we estimate the probability that a randomly selected peak tide X is from month j where j = 1 - 12, given it is higher than some value  $x_q$ , the qth

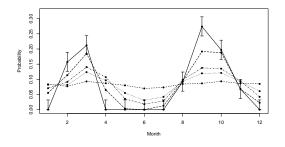


FIG 3. Estimates  $\tilde{P}_X(j; x_q)$  for j = 1 - 12 and q = 0.5 (dashed), 0.9 (dotted), 0.95 (dot-dashed), 0.99 (long dashed), 0.999 (solid) at Heysham, with 95% confidence intervals for q = 0.99.

quantile of the distribution of peak tides, with  $q \in [0, 1]$ . This is defined as

(3.1) 
$$\tilde{P}_X(j;x_q) = \hat{\mathbb{P}}(m(X) = j \mid X > x_q),$$

where m(X) denotes the month of occurrence of X and  $\hat{\mathbb{P}}(\cdot)$  is calculated empirically. If peak tides are identically distributed over the year and all months have equal duration, then  $\tilde{P}_X(j; x_q) = 1/12$  for all j and q. Of course, months vary in length, but a significant departure from  $\tilde{P}_X(j; x_q) = 1/12$  indicates that peak tides are not identically distributed over a year. Months with the largest peak tides will have higher values for  $\tilde{P}_X(j; x_q)$  for large q.

Figure 3 shows the estimates  $\tilde{P}_X(j; x_q)$  for a range of q at Heysham (see Figure S10 for the remaining sites). These show that  $\tilde{P}_X(j; x_{0.5}) \approx 1/12$  for all j = 1 - 12 at Heysham, Newlyn and Sheerness. At Lowestoft this varies, agreeing with the boxplot in Figure S9. For all  $q \ge 0.9$ , there is clear evidence that  $\tilde{P}_X(j; x_q)$  is largest in months nearest the equinoxes at all sites. At Heysham,  $\tilde{P}_X(j; x_{0.999}) > 0$  for j = 2, 3, 8, 9, 10; we find the same results at Newlyn but  $\tilde{P}_X(1; x_{0.999}) > 0$  also. However, at Lowestoft and Sheerness,  $\tilde{P}_X(j; x_{0.999}) = 0$ in months close to the spring equinox. Confidence intervals on these estimates are constructed by exploiting the property that the number of times that an event  $X > x_q$  occurs in month jfollows a multinomial distribution, with probabilities  $\tilde{P}_X(j; x_q)$ . Figures 3 and S10 also show 95% confidence intervals for  $\tilde{P}_X(j; x_{0.999})$  at each site. These indicate that the differences in  $\tilde{P}_X(j; x_{0.999})$  discussed above for each month are statistically significant.

Skew surge and peak tide both exhibit temporal dependence; we are only interested in modelling the former since peak tides are deterministic. Figure S13 shows the autocorrelation function (acf) plots for skew surge at each site. The correlation is stronger at lower lags, as expected, and tends to zero at higher lags; although this doesn't always reach zero due to seasonality. At Heysham and Newlyn, the correlation is significantly higher for all lags above 1, than at Lowestoft and Sheerness. In Section S.4 of the Supplementary Material we explore temporal dependence in the skew surges further by looking at the two mean measures of extremal dependence  $\chi$  and  $\overline{\chi}$  (Coles, Heffernan and Tawn, 1999) at different high thresholds and lags. We find similar patterns over lags, but weaker dependence, for the extreme values.

3.3. *Skew Surge and Peak Tide Dependence*. Prandle and Wolf (1978) examine the tidesurge interaction, both empirically and physically, for sites in the southern North Sea. They find that the most extreme surges occur on rising tides. This complex dependence structure motivates the use of skew surge and peak tide in the SSJPM (Batstone et al., 2013). Williams et al. (2016) conclude that the assumption of skew surge-peak tide independence is broadly well supported empirically and based on valid physical reasoning. However, Environment Agency (2018) and Williams et al. (2016) identify a weak correlation at Sheerness. Howard and Williams (2021) use oceanographic numerical model simulations to show that the highest skew surges tend to occur on lower peak tides at Sheerness. These findings raise questions about the validity of the independence assumption in the SSJPM. Ignoring the observed dependence would result in overestimates of return levels. We use a range of exploratory data analyses to investigate these claims in Section S.2 of the Supplementary Material. We find that it is reasonable to assume skew surge-peak tide independence at Heysham, Lowestoft and Newlyn, but not at Sheerness. Our results suggest that this dependence structure is changing throughout the year, with the strongest dependence found in May.

## 4. Novel Methodology.

4.1. Introduction. We begin by deriving the distribution of both the sea level monthly maxima  $M_j$  for j = 1 - 12 and annual maxima M under several simplifying assumptions that we subsequently relax. We use the fact that the peak sea level  $Z_i$  in tidal cycle i can be written as the sum of skew surge  $Y_i$  and peak tide  $X_i$  in that tidal cycle, for all i = 1, ..., T where T is the total number of tidal cycles. In Section 4.2 we develop a simplified model, assuming skew surges are iid and independent of peak tide. We relax these assumptions to develop a novel method that reflects the realism of the sea level process; accounting for skew surge seasonality (Section 4.3), skew surge-peak tide dependence (Section 4.4) and temporal skew surge dependence (Section 4.5). Unless stated otherwise, inference is conducted in a likelihood framework. We provide 95% confidence intervals for model parameter estimates based on the Hessian; this assumes extreme skew surges are temporally independent and we have shown otherwise. Stationary block bootstrap confidence intervals are preferable, but we do not use these for model selection as they are computationally demanding. However, we use this bootstrap approach for the confidence intervals of return level estimates (Section 5.4) to ensure that estimates that are used in practice have the optimal confidence intervals.

Extreme sea levels up to the  $\sim 20$  year return period can occur with various combinations of skew surge and peak tides, e.g., a typical skew surge value combined with the largest peak tide, through to an extreme skew surge with a typical peak tide. Since peak tides are bounded above by HAT, return periods over 20 years can only be achieved with extreme skew surges. The very largest return periods require skew surges bigger than already observed. We are interested in return levels corresponding to return periods of 1 year and above, so we require an estimate for the distribution of all possible skew surges, though we are particularly interested in modelling the upper tail for inference at high return periods. Thus in each variant of our method, we develop a model for the whole skew surge distribution under the associated assumptions about the skew surge process.

4.2. *Idealised Case.* Under the assumption that the skew surges are identically distributed, we estimate the distribution of skew surges below a threshold u using the empirical distribution  $\tilde{F}_Y$ . This is adequate because tide gauges on the UK National Tide Gauge Network typically have long observation lengths (>20 years) so the empirical distribution describes the main body of the data well. To enable extrapolation in the tail we use the GPD model (2.3), with constant parameters. Then, our model is

(4.1) 
$$F_Y(y) = \begin{cases} \tilde{F}_Y(y) & \text{if } y \le u \\ \\ 1 - \lambda_u \left[ 1 + \xi \left( \frac{y-u}{\sigma_u} \right) \right]_+^{-1/\xi} & \text{if } y > u, \end{cases}$$

where  $\lambda_u = 1 - \tilde{F}_Y(u)$  and  $\sigma_u, \xi$  are the parameters of the GPD. We take  $\lambda_u = 0.05$ . To simplify notation, we subsequently drop the *u* subscript on the scale  $\sigma$  and rate  $\lambda$  parameters.

Let  $T_j$  denote the number of tidal cycles in month j. We use sequential monthly peak tide samples  $\{X_{j_i}; j = 1 - 12, i = 1, ..., T_j\}$  where  $j_i$  denotes the *i*th peak tide in month j. Then, if skew surges are independent and peak tides repeat exactly on an annual cycle, the distribution of month j maximum sea level,  $M_j$ , is

(4.2) 
$$\mathbb{P}(M_j \le z) = \prod_{i=1}^{T_j} \mathbb{P}(Y_i \le z - X_{j_i}) = \prod_{i=1}^{T_j} F_Y(z - X_{j_i}).$$

Then the annual maxima skew surge distribution is given by,

(4.3) 
$$\mathbb{P}(M \le z) = \prod_{j=1}^{12} \prod_{i=1}^{T_j} F_Y(z - X_{j_i}).$$

Since peak tides form a deterministic sequence, exhibiting strong pairwise sample correlations, the peak tidal samples  $X_{j_i}$  that are used in expressions (4.2) and (4.3) must be from contiguous peak tides across the year.

Although cycles of periods up to a year dominate, peak tides have longer term periodicities. Previous methods, such as Tawn (1992), assumed peak tides are stationary within and across years so that the annual maxima distribution is given by

(4.4) 
$$P(M \le z) = \left(\prod_{i=1}^{T} F_Y(z - X_i)\right)^{1/K}$$

for T the total number of observations, K the number of years of observation and  $F_Y$  the stationary skew surge distribution (4.1). We incorporate interannual peak tide variations by taking the average of the yearly patterns over K years, where  $K \ge 19$  to address all nodal cycle variations. We denote peak tide on the *i*th tidal cycle in month j of year k by  $X_{j_i}^{(k)}$  for j = 1 - 12,  $i = 1, \ldots, T_j^{(k)}$  and  $k = 1, \ldots, K$ . Here  $T_j^{(k)}$  is the number of tidal cycles in month j and year k, which varies over k because the timing of the first cycle in the month can vary annually. Then the monthly and annual maxima sea level distributions are (4.5)

$$\mathbb{P}(M_j \le z) = \frac{1}{K} \sum_{k=1}^K \prod_{i=1}^{T_j^{(k)}} F_Y(z - X_{j_i}^{(k)}), \text{ and } \mathbb{P}(M \le z) = \frac{1}{K} \sum_{k=1}^K \prod_{j=1}^{12} \prod_{i=1}^{T_j^{(k)}} F_Y(z - X_{j_i}^{(k)}),$$

respectively. This no longer has the property that the distribution of the annual maxima is the product of the monthly maxima distribution, as the monthly maxima are now dependent due to associations in peak tides in different months across years.

4.3. Skew Surge Seasonality. As discussed in Section 3.2, skew surges are not identically distributed across a year. Here we describe how we capture seasonality in three ways: for values above the threshold, below the threshold and in the exceedance probability. We define extreme skew surges using a month-specific threshold  $u_j$ , defined as a quantile of each monthly distribution. We use the 0.95 monthly empirical quantiles; in Section 6 we discuss the merits of this approach relative to taking a threshold that varies on a daily scale. We consider a range of models for how the parameters of the GPD and exceedance probability separately vary over time. These models are fit independently of each other and to different data. To distinguish between the two types of model, we precede the model number by S (for the scale and shape parameter) and R (for the rate parameter).

Firstly we look at exceedances of the monthly thresholds and model these using the GPD. There are enough data within each month, over the different years, to justify the adoption of

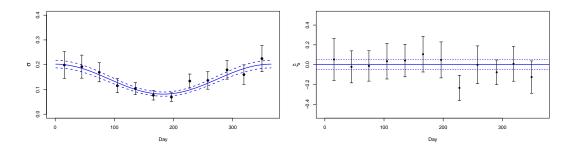


FIG 4. Scale (left) and shape (right) parameter estimates for Model S0 (black) and Model S2 (blue) at Heysham. 95% confidence intervals are added to Model S0 (error bars) and to Model S2 parameter estimates (dashed).

a GPD model for exceedances of a high threshold and to assess its fit. To account for seasonality we allow the parameters to change over time in a periodic fashion, as discussed in Section 2.1. We consider four models to describe how these parameters change with time, expressed in terms of a daily or monthly covariate. As a first approach to account for seasonality in extremes, Carter and Challenor (1981) suggested allowing each month to have a separate distribution. To do so we fit a GPD with a monthly covariate on the scale  $\sigma_i$  and shape  $\xi_i$  parameters; we refer to this as Model S0. We use this basic approach for comparison only. To obtain a more parsimonious model, we fix the shape parameter to be the same value over all months but keep a monthly covariate on the scale parameter; we call this Model S1. These models give discontinuities at transitions between months, they assume skew surges are identically distributed within a month and have a large number of parameters. Therefore we introduce a covariate d = 1 - 365 that denotes the day in year to capture within-year variations smoothly. For notational simplicity, we assume d is entirely defined using the tidal cycle index i and month j, so that there exists a function h where d = h(j, i). We consider two harmonic parameterisations of the scale parameter with a daily covariate. Model  $S^2$  uses a single harmonic defined by

(4.6) 
$$\sigma_d = \alpha_\sigma + \beta_\sigma \sin\left(\frac{2\pi}{f}(d-\phi_\sigma)\right),$$

for parameters  $\alpha_{\sigma} > \beta_{\sigma} > 0$ ,  $\phi_{\sigma} \in [0, 365)$  and periodicity f = 365. Model S3 uses two harmonics with periodicities of f and f/2. The shape parameter is equal across months for both of these models, so that Models S2 and S3 have 4 and 6 parameters, respectively.

We now fit the models outlined above to skew surge threshold exceedances at each site. Figures 4 and S12 show Model S0 parameter estimates for Heysham and the remaining sites, respectively. Since the 95% confidence intervals for monthly shape parameter estimates have considerable overlap at all sites, it is reasonable to have a common value across months as in Models S1 - 3. Table 1 reports the AIC and BIC scores for each model at all sites. These scores suggest that restricting the shape parameter to be an unknown constant, over months, improves model fit compared to Model S0 in all cases. Model S2 is selected as our final model as the evidence across all sites shows this is reasonable. Table 2 gives the parameter estimates for Model S2 at each site. Notice that the shape parameter estimates have 95% confidence intervals that encompass zero at each site. Figures 4 and S12 show the scale and shape parameter estimates graphically;  $\hat{\sigma}_d$  is higher in the winter and lower in the summer, as expected. The estimated scale parameters of Model S0 but do capture a smooth transition within each month. We use asymptotic normality of maximum likelihood estimators and the Delta method to add 95% confidence intervals to the scale parameter estimates.

is measured relative to Model S2.									
		Heysham		Lowestoft		Newlyn		Sheerness	
Model	No. Parameters	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
S1	13	-66.66	-119.52	-57.40	-118.72	-77.36	-143.83	-67.51	-123.52
S2	4	-69.5	-176.61	-60.47	-171.93	-88.89	-209.74	-76.18	-178.03
S3	6	-65.65	-161.98	-56.53	-156.85	-86.06	-194.82	-75.06	-166.73
S4	5	2.99	8.42	1.82	7.39	5.77	11.82	-181.57	3.53

 TABLE 1

 AIC and BIC scores at each site for each skew surge model relative to Model S0 scores, except Model S4 which is measured relative to Model S2.

We also considered using a log link function for the scale parameter instead of the identity link function for Models S2 and S3 but found that the identity link yields a better model fit. Furthermore the identity link function has the advantage of preserving the threshold stability property when covariates are included (Eastoe and Tawn, 2009).

Another fundamental part of capturing within year seasonality is through the rate parameter  $\lambda$ . So far, we have assumed this is constant as the threshold has been set at the same quantile across months. Here, we add a daily covariate d to capture smooth changes in  $\lambda$ throughout each month, whilst recognising that the average exceedance rate across a month is equal for all months. Let  $V_d$  be a binary random variable representing whether a skew surge value exceeds its month-specific threshold  $u_j$  or not, so that  $V_d \sim \text{Bernoulli}(\lambda_d)$ . Therefore, we use a logit link function  $g(\cdot)$  to model  $\lambda_d$  and capture daily changes using a generalised linear model (GLM). To account for within month variations, we relate  $g(\lambda_d)$  to the day in month  $d_j \in [1, 31]$ , standardised by the monthly mean day  $\bar{d}_j$ , so that  $(d_j - \bar{d}_j) \in [-15, 15]$ approximately. We parametrise the gradient using the day in year d to account for different gradients across the year using a harmonic with periodicity f = 365. We refer to this model as Model R0 and this is given by

(4.7) 
$$g(\lambda_d) = g(\lambda) + (d_j - \bar{d}_j)\beta_\lambda \sin\left(\frac{2\pi}{f}(d - \phi_\lambda)\right),$$

for  $\beta_{\lambda} > 0$ ,  $\phi_{\lambda} \in [0, 365)$  which are parameters to be estimated and  $\lambda$  the exceedance rate in a month (here  $\lambda = 0.05$ ). Fitting this model to our data demonstrates there is more variation in exceedance probabilities during spring and autumn compared to summer and winter. At all sites, the greatest range in  $\lambda_d$  was  $\sim 0.03$  in April and October. This agrees with our scale parameter model where the steepest gradient is found in spring and autumn (see Figures 4 and S12). At all sites, the fitted model for  $\lambda_d$  has negative gradient in months at the beginning of the year so that the exceedance probability is higher earlier in the month (closer to winter). The slope is positive later in the year.

Finally, we use a month-specific empirical distribution  $\tilde{F}_Y^{(j)}$  for skew surges below the threshold. Bringing this together with the parameterisations of  $\sigma_d$  and  $\lambda_d$  (expressions (4.6) and (4.7), respectively) for the upper tail model, the final full skew surge distribution  $F_Y$  is dependent on month j and day of the year d, and is given by

(4.8) 
$$F_Y^{(d,j)}(y) = \begin{cases} \tilde{F}_Y^{(j)}(y) & \text{if } y \le u_j \\ \\ 1 - \lambda_d \left[ 1 + \xi \left( \frac{y - u_j}{\sigma_d} \right) \right]_+^{-1/\xi} & \text{if } y > u_j. \end{cases}$$

As  $F_Y^{(d,j)}(y)$  changes with day d and d = h(j,i), it also changes for every tidal cycle i. Consequently the estimated monthly and annual maxima distributions of sea levels are

(4.9) 
$$\mathbb{P}(M_j \le z) = \frac{1}{K} \sum_{k=1}^K \prod_{i=1}^{T_j^{(k)}} F_Y^{(d,j)}(z - X_{j_i}^{(k)}),$$

 TABLE 2

 Parameter estimates for the scale parameter for Models S2 and S4, and the rate parameter for Model R1 with 95% confidence intervals, at each site.

		5					
	Heysham	Lowestoft	Newlyn	Sheerness			
Model	S2						
$\alpha_{\sigma}$	0.14 (0.13, 0.15)	0.15 (0.14, 0.16)	0.076 (0.073, 0.080)	0.11 (0.10, 0.12)			
$\beta_{\sigma}$	0.060 (0.050, 0.070)	0.080 (0.070, 0.090)	0.024 (0.020, 0.028)	0.052 (0.043, 0.061)			
$\phi_{\sigma}$	271.51 (262.77, 280.23)	266.32 (260.01, 272.63)	273.58 (265.06, 282.10)	272.11 (262.66, 281.56)			
ξ	0.002 (-0.042, 0.051)	0.024 (-0.023, 0.071)	-0.040 (-0.074, 0.006)	0.037 (-0.029, 0.10)			
Model S4							
$\gamma_{\sigma}^{(x)}$	0.002 (-0.005, 0.009)	-0.0051 (-0.040, 0.030)	0.0048 (-0.00048, 0.010)	-0.012 (-0.026, 0.0011)			
Model R1							
$\beta_{\lambda}$	0.0087 (0.0004, 0.017)	0.022 (0.015, 0.030)	0.024 (0.018, 0.030)	0.022 (0.014,0.032)			
$\phi_{\lambda}$	155.66 (100.74, 210.59)	175.16 (155.86, 194.46)	209.50 (195.48, 223.52)	184.31 (160.94,207.69)			
$\alpha_{\lambda}^{(x)}$	-0.13 (-0.18, -0.079)	-0.055 (-0.101, 0.009)	-0.063 (-0.099, 0.108)	-0.32 (-0.37,-0.26)			
$ \begin{array}{c} \alpha_{\lambda}^{(x)} \\ \beta_{\lambda}^{(x)} \\ \phi_{\lambda}^{(x)} \end{array} $	0.14 (0.068, 0.21)	-0.016 (-0.084, 0.051)	0.061 (0.014, 0.108)	0.23 (0.14, 0.31)			
$\phi_{\lambda}^{(x)}$	311.86 (281.78, 341.94)	359.95 (265.77, 454.15)	352.38 (299.32, 405.44)	278.54 (260.44, 293.63)			

(4.10) 
$$\mathbb{P}(M \le z) = \frac{1}{K} \sum_{k=1}^{K} \prod_{j=1}^{12} \prod_{i=1}^{T_j^{(k)}} F_Y^{(d,j)}(z - X_{j_i}^{(k)})$$

4.4. Skew Surge Dependence on Peak Tide. In Section 3.3 we conclude that skew surgepeak tide independence is a reasonable assumption at Heysham, Lowestoft and Newlyn, but not at Sheerness. Here, we describe how we account for this dependence in our skew surge model. We do this for the upper tail by adding a tidal covariate to the GPD scale and rate parameters, but not to the shape parameter  $\xi$  to avoid additional uncertainty. For values below the threshold, we use different empirical distributions of skew surges depending on their associated peak tide band. This is a similar approach to Pugh and Vassie (1978).

We first consider how the threshold exceedance probability varies with peak tide. Here, we extend the GLM parametrisation of expression (4.7) by adding a peak tidal covariate xto the rate parameter  $\lambda_d$ . We linearly standardise peak tide via  $(x - \bar{x})/s_x$ , where  $\bar{x}$  is the mean and  $s_x$  the standard deviation of all peak tide observations at each site. We parametrise the gradient in the same way as expression (4.7), using the day in year d and a harmonic with periodicity f = 365 to capture smooth daily changes within a year. This captures the time varying dependence structure between skew surge and peak tide (see Section S.2 of the Supplementary Material). We use the notation  $\lambda_{d,x}$  and the following model, denoted R1,

(4.11) 
$$g(\lambda_{d,x}) = g(\lambda_d) + \left(\frac{x-\bar{x}}{s_x}\right) \left[\alpha_{\lambda}^{(x)} + \beta_{\lambda}^{(x)} \sin\left(\frac{2\pi}{f}(d-\phi_{\lambda}^{(x)})\right)\right],$$

for  $g(\cdot)$  the logit link function,  $g(\lambda_d)$  defined by expression (4.7) and  $\alpha_{\lambda}^{(x)} \in \mathbb{R}$ ,  $\beta_{\lambda}^{(x)} > 0$  and  $\phi_{\lambda}^{(x)} \in [0, 365)$  being parameters to be estimated. The full model for  $\lambda_{d,x}$  has 5 parameters.

We fit this GLM to model the exceedance probabilities at each site and give parameter estimates in Table 2. Figure 5 shows these results graphically at Sheerness for March, June, September and December (see Figure S18 for the other sites). As expected the estimated exceedance probability is lower at higher tides, this result is more significant in months where we found skew surge-peak tide dependence to be stronger (April - September). As in Section 4.3,  $\lambda_{d,x}$  changes most with day in spring and autumn corresponding to the greatest range of skew surges within a month. Figure S17 shows the gradient of the tidal covariate for each month, when averaged over days. At Heysham, Newlyn and Sheerness, the gradients are greatest near the equinoxes; at Lowestoft there is little variation across months.

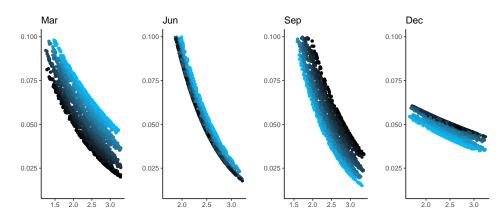


FIG 5. Estimated exceedance probability  $\lambda_{d,x}$  (y-axis) under Model R1 in March, June, September and December with respect to x being peak tide (x-axis in metres) and  $d_j$  being day in month at Sheerness. Darker (lighter) points represent days later (earlier) in the month.

We compare the Model R1 (4.11) with Model R0 (4.7) using AIC and BIC scores. Model R1 minimises the AIC score at Heysham, Newlyn and Sheerness; this is most notable at Sheerness with a reduction of 139. We obtain almost identical AIC scores at Lowestoft. BIC is also minimised by Model R1 at Heysham and Sheerness, but not at Lowestoft and Newlyn. We also compare Model R1 with the model in expression (4.11) but with  $\beta_{\lambda}^{(x)} = 0$ , so that the gradient term for tide does not vary with day. We found that AIC favoured Model R1 at all sites indicating that the dependence structure varies seasonally.

We also investigate adding a tidal covariate x to the scale parameter of the GPD for extreme skew surges, building on the existing parametrisation of Model S2 (4.6) to give Model S4,

(4.12) 
$$\sigma_{d,x} = \alpha_{\sigma} + \beta_{\sigma} \sin\left(\frac{2\pi}{f}(d-\phi_{\sigma})\right) + \gamma_{\sigma}^{(x)}x,$$

for  $\alpha_{\sigma} > \beta_{\sigma} > 0$ ,  $\phi_{\sigma} \in [0, 365)$ ,  $\gamma_{\sigma}^{(x)} \in \mathbb{R}$  parameters to be estimated, f = 365 the periodicity and d the day in year. Since we found evidence that the dependence of skew surge on peak tide changes within a year, we also considered parameterisations that allow the tidal effect to vary with day or month but found no improvement in fit relative to the loss of parsimony.

Table 2 gives estimates of  $\hat{\gamma}_{\sigma}^{(x)}$  at each site; this tells us how  $\sigma_{d,x}$  changes with tide (see Table S3 for all Model S4 parameter estimates). At Lowestoft and Sheerness  $\hat{\gamma}_{\sigma}^{(x)} < 0$  which agrees with our results of Section 3.3 that extreme skew surges tend to occur on lower tides, although their 95% confidence intervals contain 0. To formally compare the fit of Model S4 with S2 we use AIC and BIC scores (see Table 1). Model S4 is only favoured by the AIC at Sheerness, this suggests the simpler model without a tidal covariate is preferred elsewhere. Given that our findings of skew surge-peak tide dependence were different at Sheerness in comparison to the other sites, we examined this further to check it wasn't an artifact of the data measurement or tide extraction processes. We explored the fits of Models R1 and S4 on a 483 year data set from a hydrodynamical model driven by a regional climate model (Howard and Williams, 2021). We find that adding peak tide covariates to the scale and rate parameters improves fit, with similar estimates to the observations. See Section S.3 of the Supplementary Material for details. Therefore, we proceed with Model S4 for the scale parameter at all sites.

Below the threshold, we found it sufficient to capture the dependence by splitting the empirical distribution of skew surges into three associated peak tide bands, i.e., (4.13)

$$\tilde{F}_{j,x}(y) = \tilde{F}_{j}^{(1)}(y) \text{ if } x \le x_{0.33}^{(j)}, \text{ or } \tilde{F}_{j}^{(2)}(y) \text{ if } x_{0.33}^{(j)} < x \le x_{0.67}^{(j)}, \text{ or } \tilde{F}_{j}^{(3)}(y) \text{ if } x > x_{0.67}^{(j)},$$

where  $x_q^{(j)}$  is the q quantile of the peak tide distribution in month j and  $\tilde{F}_j^{(l)}$  for l = 1, 2, 3 is the empirical distribution of skew surges in month j which are associated with the lowest (l = 1), medium (l = 2) and highest (l = 3) band of peak tides. The choice of 3 tidal bands is somewhat arbitrary, but appears sufficient given the weak dependence on peak tide. We could have used a kernel smoother to ensure continuity across bands, as in Bashtannyk and Hyndman (2001). Since our interest lies with the extreme values, we did not explore this further as this would not have made a practical difference to our return level estimates.

So our skew surge model, that is dependent on peak tide, is given by

(4.14) 
$$F_Y^{(d,j,x)}(y) = \begin{cases} \tilde{F}_{j,x}(y) & \text{if } y \le u_j \\ \\ 1 - \lambda_{d,x} \left[ 1 + \xi \left( \frac{y - u_j}{\sigma_{d,x}} \right) \right]_+^{-1/\xi} & \text{if } y > u_j, \end{cases}$$

with  $\tilde{F}_{j,x}(\cdot)$ ,  $\sigma_{d,x}$  and  $\lambda_{d,x}$  defined in expressions (4.13), (4.12) and (4.11), respectively.

4.5. Skew Surge Temporal Dependence. So far, we have assumed that skew surges are independent. We now describe how we account for their temporal dependence across tidal cycles. As discussed in Section 2.1, temporal dependence causes clusters of events above high thresholds; if this is ignored, sea level return levels of annual maxima will be overestimated. The best measure of dependence for the extreme values of a stationary sequence is the extremal index  $\theta$ . Tawn (1992) and Batstone et al. (2013) use the extremal index of the highly non-stationary sea level series. We model the sub-asymptotic extremal index  $\theta(y, r)$  (defined by equation (2.2)) of skew surges using varying thresholds (levels) y with a fixed run length r. Then the distribution of the monthly and annual maxima sea levels are given by

$$\mathbb{P}(M_{j} \leq z) = \frac{1}{K} \sum_{k=1}^{K} \prod_{i=1}^{T_{j}^{(k)}} \left[ F_{Y}^{(d,j,x)}(z - X_{j_{i}}^{(k)}) \right]^{\hat{\theta}(z - X_{j_{i}}^{(k)},r)},$$

$$(4.15) \qquad \mathbb{P}(M \leq z) = \frac{1}{K} \sum_{k=1}^{K} \prod_{j=1}^{12} \prod_{i=1}^{T_{j}^{(k)}} \left[ F_{Y}^{(d,j,x)}(z - X_{j_{i}}^{(k)}) \right]^{\hat{\theta}(z - X_{j_{i}}^{(k)},r)}.$$

The empirical estimates  $\tilde{\theta}(y, r)$  of the sub-asymptotic extremal index are shown in Figures 6 and S16 for Heysham and the remaining sites, respectively. These demonstrate substantial variation with y. We are interested in  $\theta(y, r)$  across all ranges of skew surge, but particularly in values greater than those observed to allow extrapolation for return level estimation.

We estimate  $\theta(y,r)$  in two stages. For  $y \leq v$  where v is a high threshold, we use the empirical runs estimate  $\tilde{\theta}(y,r)$  as this is smooth over y in this range. For computational efficiency purposes, we evaluate  $\tilde{\theta}(y,r)$  on a regular grid of y values in the range  $[y_F,v]$  where  $y_F$  is the minimum observed skew surge. Then we linearly interpolate these for any y < v of interest. For y > v, the empirical estimate is increasingly variable so we adopt a parametric model for  $\theta(y,r)$ . We propose

(4.16) 
$$\hat{\theta}(y,r) = \begin{cases} \theta(y,r) & \text{if } y \le v \\ \theta - [\theta - \tilde{\theta}(v,r)] \exp\left[-\frac{(y-v)}{\psi}\right] & \text{if } y > v, \end{cases}$$

where  $\psi > 0$  and  $\tilde{\theta}(v, r) \le \theta \le 1$  are parameters to be estimated. This parametric form ensures the estimate asymptotes to the extremal index  $\theta$  and is continuous at v.

The parameters  $\psi$  and  $\theta$  are estimated using a weighted least squares approach with weight  $w(y) = \sqrt{c(y,r) - 1}$  where c(y,r) is the number of clusters above y separated by run length

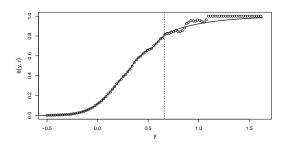


FIG 6. Estimates of the subasymptotic extremal index  $\theta(y,r)$  for different skew surge levels y and run length r = 2 at Heysham using the runs estimate (points) and our model estimate (line) of expression (4.16). The dashed black line is the 0.99 skew surge quantile.

r. This gives a greater weight when there are more clusters, and a weight of zero for a single cluster. We choose v to be the 0.99 quantile of all skew surges but this choice is subjective. The run length r represents the approximate duration of a storm across a single site. We take r = 2, 10, 1 and 10 for Heysham, Lowestoft, Newlyn and Sheerness, respectively. These choices are supported by estimating the run length using the intervals estimator for each season, where we expect the stationary assumption to be reasonably justified (Ferro and Segers, 2003). At Heysham,  $\hat{\theta} = 1$  and  $\hat{\psi} = 0.40$ ; the estimates for the other sites are given in Section S.4 of the Supplementary Material. Figures 6 and S16 show this model fit for Heysham and the remaining sites, respectively, compared with the entirely empirical estimates.

## 5. Results.

5.1. *Introduction*. Using the distribution of the monthly and annual maxima sea levels derived in Section 4, we estimate return levels by solving

$$\mathbb{P}(M_j \leq z_{j,p}) = 1 - p \quad \text{and} \quad \mathbb{P}(M \leq z_p) = 1 - p,$$

respectively for  $p \in [0, 1]$ . We are interested in return levels up to the 10,000 year level, corresponding to annual exceedance probability  $p = 10^{-4}$ . In the monthly case, this is the level we expect the monthly maxima  $M_j$  to exceed, on average, every 1/p of that particular month (for example, every 10,000 Januarys). Whilst in the annual case, this is the level we expect the annual maxima to exceed every 1/p years, on average.

To assess the importance of each of the novel modelling steps in Section 4, we derive return level estimates from each stage; accounting for non-stationarity in each component, the dependence between them and skew surge temporal dependence. Each stage is detailed below in a nested list numbered (i)-(vii), so that each model below builds on the previous method, except (iv) from (iii). We compare these to the current approach used in practice and a baseline approach, both derived under simplifying and false assumptions. We subsequently refer to each model by name, given below, using italics. The notation follows from Section 4.

- (i) **Current**: The methodology currently used in practice, where skew surges are assumed to be iid and peak tides are stationary, with annual maxima distribution (4.4).
- (ii) **Baseline**: As in (i) but we recognise interannual variations in the tide by averaging over yearly tidal samples. The annual maxima distribution is given by

(5.1) 
$$P(M \le z) = \frac{1}{K} \sum_{k=1}^{K} \prod_{i=1}^{T^{(k)}} F_Y(z - X_i^{(k)})$$

where  $X_i^{(k)}$ ,  $i = 1, ..., T^{(k)}$  represents an annual tidal sample for year k = 1, ..., K, where we choose K = 100 arbitrary but contiguous samples.

- (iii) **Seasonal surge**: As in (ii) but we account for the within-year seasonality of skew surges using the skew surge distribution (4.8).
- (iv) **Seasonal tide**: As in (ii), but conversely to (iii) we account for within-year seasonality in peak tide and not in skew surge, so the skew surge model of expression (4.1) is used in the monthly and annual maxima distributions (4.5).
- (v) **Full seasonal**: As in (ii) but we account for within-year seasonality of both components, with monthly and annual maxima distributions (4.9) and (4.10), respectively.
- (vi) (Skew surge-peak tide) interaction: As in (v) but with skew surge-peak tide dependence captured, so that the skew surge distribution (4.14) is used.
- (vii) (Skew surge) temporal dependence: As in (vi) but accounting for temporal dependence in the skew surge series, with annual maxima distribution (4.15).

We compare our model estimates to empirical estimates. These are restricted to return periods within the range of data, but are useful for checking whether the model is capturing the distributional properties of observed sea levels. The empirical estimates act as a guide to the truth and we do not expect our model to fit these perfectly since each empirical estimate is specific to a particular annual tidal regime, whereas in our approaches (ii)-(vii) we account for tidal variations by averaging over different samples. The empirical estimates are also sensitive to missing data and so can be biased. In Sections 5.2 and 5.3, we make comparisons using point estimates to assess sensitivity to the model choice; in each case results in Section 4 identify statistically significant differences between these models. Once the model choice is finalised we present measures of uncertainty and assess model fit in Section 5.4. In Section 5.5 we use our model to find the probability that an extreme sea level occurs in a specific month, given it is higher than some level with a given return period of interest.

5.2. Return Levels: Accounting for Non-stationarity. We investigate how accounting for skew surge and peak tide seasonality influences sea level return level estimates by comparing monthly return level estimates from the baseline, seasonal surge, seasonal tide and full seasonal models. We are mainly interested in differences between the baseline and full seasonal estimates. We do not expect the full seasonal model to match the empirical estimates since we have not accounted for skew surge-peak tide dependence and temporal dependence yet. The intermediate solutions (seasonal surge and seasonal tide) allow us to understand which components' non-stationarity is influencing the return levels most in different months.

Figures 7 and S19 show monthly return level estimates at Lowestoft and the other sites, respectively. At all sites and in all months, the empirical estimates lie closer to *full seasonal* model than the *baseline*. The most noticeable feature is the difference between the *baseline* and *full seasonal* model in June, this reaches 1.54m at the 10,000 year level at Lowestoft. In June skew surges are particularly low relative to the rest of the year, so ignoring their seasonality leads to significant overestimation. This overestimation increases with return period. On the other hand, in December when skew surges are relatively high compared to the other months, ignoring seasonality results in underestimates of return levels. Since the tidal range at Lowestoft is narrow compared to the range of skew surges, differences between those from the *seasonal surge* and *full seasonal* models. For example, when the tidal range is largest at the autumn equinox in September, we observe a difference of 7cm at the 10,000 year level. Whereas at Heysham, where the tidal range is large relative to other sites, we observe a difference of 26cm at the 10,000 year level in September.

Figure 8 shows annual return level estimates from the *current*, *baseline* and *full sesaonal* models, compared to empirical estimates at all sites, up to the 100 year level (Figure S20

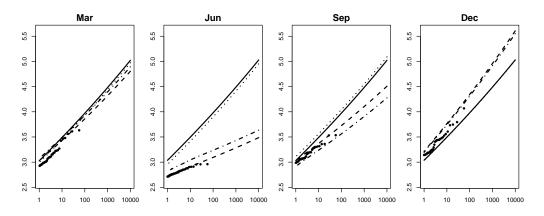


FIG 7. Monthly maxima sea level return level estimates (y-axis in metres) for different return periods (x-axis in years) at Lowestoft in March, June, September and December (from left to right), estimated using the baseline (black solid), seasonal surge (dot-dashed), seasonal tide (dotted) and full seasonal (dashed) models. Empirical estimates are shown by black points.

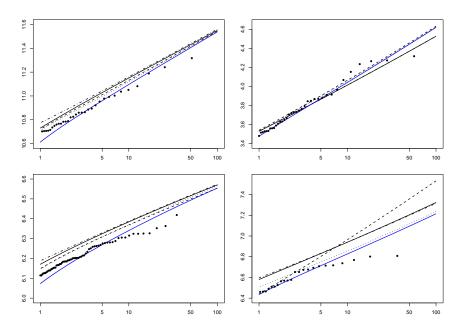


FIG 8. Annual maxima sea level return level estimates (y-axis in metres) for different return periods (x-axis in years) at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right), up to the 100 year level, estimated using the current (dot-dashed), baseline (black solid), full seasonal (dashed), interaction (dotted) and temporal dependence (blue solid) models. Empirical estimates are shown by the black points.

shows these up to the 10,000 year level). The *current* method gives similar results to the *baseline* approach for all return periods at all sites. This suggests assuming each year is identically distributed is not unreasonable. Even at Heysham, where year-to-year variations are large, the difference between the *baseline* and *current* method is small, with the *current* method giving a 1 year return level 4cm higher than the *baseline*. Since the *full seasonal* model lies closer to the empirical estimates than models (i)-(iv) at all sites, this highlights the importance of accounting for both forms of seasonality. At the 10,000 year level, the *baseline* gives a return level 6cm, 23cm and 58cm lower than the *full seasonal* method at Heysham, Lowestoft, and Sheerness, respectively, whereas it is 1cm higher at Newlyn.

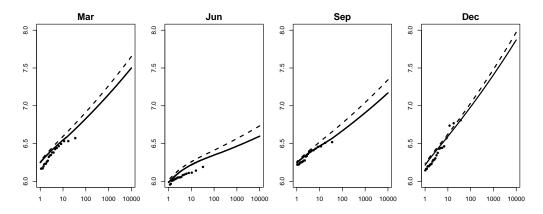


FIG 9. Monthly return level estimates (y-axis in metres) for different return periods (x-axis in years) at Sheerness in March, June, September and December, estimated using the full seasonal (dashed) and interaction (solid) models when we account for and ignore skew surge-peak tide dependence, respectively. Empirical estimates are shown by the black points.

5.3. *Return Levels: Accounting for Dependence.* Here, we build on the *full seasonal* model to capture skew surge-peak tide dependence (*interaction* model) and skew surge temporal dependence (*temporal dependence* model). We compare these with the empirical and *full seasonal* estimates. We present monthly maxima return levels of the *interaction* model to demonstrate the changing skew surge-peak tide dependence structure within a year, discussed in Section 3.3. Then we compare annual maxima return level estimates from both models.

Figure 9 shows monthly return level estimates in March, June, September and December for the *full seasonal* and *interaction* models, as well as the empirical estimates at Sheerness. We illustrate the results for Sheerness because this dependence is stronger compared to the other sites, as discussed in Section 4.4. Ignoring skew surge-peak tide dependence overestimates return levels compared to the *full seasonal* model for return periods greater than 10 years. At the 10,000 year level, these are overestimated by 15cm, 14cm, 17cm and 10cm in March, June, September and December, respectively, with similar values down to the 1 year level. This is slightly surprising as in Section 3.2 we found evidence that it is reasonable to assume skew surge and peak tide are independent in March, September and December. For Sheerness the results from the *interaction* model (see Figure 9); this suggests that accounting for this dependence is important and better reflects the process properties. At the other sites, the return level estimates from the *full seasonal* and *interaction* model are almost indistinguishable; this echoes that adding peak tide covariates into the skew surge model when independence is a reasonable assumption does not alter results significantly.

As skew surge temporal dependence is not yet accounted for in the *interaction* model, its return level estimates are anticipated to slightly overestimate the empiricals. Figure 8 compares annual return level curves for the *temporal dependence* and *interaction* models. These are very close for high return periods (>100 years). For lower return periods (<10 years) the *temporal dependence* estimates lie closer to the empiricals at all sites. This is a natural consequence of the extremal index model given by expression (4.16) for temporal dependence, since the estimated extremal index is closer to zero for lower skew surges, corresponding to shorter return periods. But for the highest skew surges, and for those greater than observed,  $\hat{\theta}(y,r) \rightarrow \hat{\theta} \approx 1$  and  $y \rightarrow y^F$ , so the largest return level estimates remain unchanged. Accounting for temporal skew surge dependence has the greatest influence at Heysham, where the acf estimates are higher compared to the other sites (see Figure S13). Here, the *temporal dependence* model gives a return values 10cm, 4cm and 2cm lower than the *interaction* model at the 1, 5 and 10 year return periods to better match the empirical estimates.

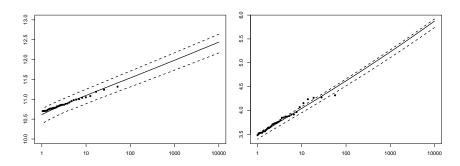


FIG 10. Return level estimates (y-axis in metres) for different return periods (x-axis in years) from the final model (temporal dependence) (solid line), with the maximum and minimum year-specific return level estimates (dashed line) and empirical estimates (points) at Heysham (left) and Lowestoft (right).

5.4. Assessment of Fit and Uncertainty. Figure 10 shows annual maxima return level estimates from the temporal dependence model, judged our best model, at Heysham and Lowestoft (see Figure S21 for the other sites). We also add the estimated maximum and minimum year-specific return levels, showing the effect of interannual peak tide variations, as given by (4.15) without averaging over K years. Figure S22 shows yearly estimates at each site. The range of yearly estimates for Lowestoft is very narrow since the tidal range is small, compared to Heysham which has the largest tidal range of the four sites. Empirical estimates may deviate away from the return level point estimates if that specific value occurs in a year with particularly low or high tides compared to the average. Figures 10 and S21 show that the empirical estimates lie within the bounds for each site, at most return periods. Rodríguez Enríquez et al. (2021) demonstrate that the 4.4 year perigean cycle has a greater effect on modulations in return levels than the nodal cycle in the UK. Our results are less clear (see Figure S23), with Sheerness year-specific return levels having an  $\sim 20$  year cycle.

We also assess goodness-of-fit by transforming the observed annual maxima to a uniform distribution using the probability integral transform with the distribution function of the annual maxima for their respective year. We do this for three cases; the *baseline* approach (5.1), our final model (4.15) and the year-specific final model. If the model fits well we expect the transformed annual maxima to be Uniform(0,1) distributed. Figure 11 shows PP plots for Sheerness, and the remaining sites are shown in Figure S27; a good fit is indicated by the empirical and model probabilities being equal so that the line of equality lies between the tolerance bounds. Figure 11 demonstrates the improvement in fit across the three models. We formally test this using the Kolmogorov-Smirnov test at each site for the final model and obtain p values 0.0066, 0.51, 0.0044 and 0.10 at Heysham, Lowestoft, Newlyn and Sheerness, respectively. At Heysham and Newlyn the p value is significant at the 5% level so we cannot conclude a good fit, but it is much better than for the baseline fits.

We also assess the model fit uncertainty by estimating 95% confidence intervals on the 1, 10, 100, 1000 and 10,000 year return levels using a stationary bootstrap procedure. We first transform the skew surges using our final model with their corresponding seasonal and tidal covariates, call this series  $\{U_i^Y\}$  for i = 1..., T for T the total number of observations, where  $U_i^Y = F_Y^{(d,j,x)}(Y_i)$ . This gives the basis for another approach to assess model fit. If the model is 'correct', these transformed observations will be Uniform(0,1) and we check this formally using the Kolmogorov-Smirnov test over all years, and for each year. We discuss this further in Section S.8 of the Supplementary Material. To account for temporal dependence, we sample  $\{U_i^Y\}$  using a stationary bootstrap with block length  $L \sim \text{Geometric}(1/10)$ , so the mean block length corresponds to 5 days (the maximum duration of a storm), and the total length of each sample is equal to that of the original data (Politis and Romano, 1994).

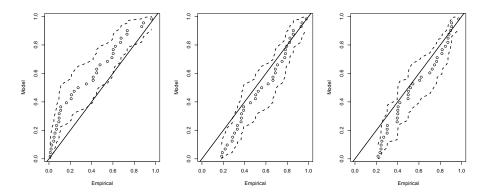


FIG 11. PP plots for the transformed annual maxima sea levels at Sheerness through the baseline (left), final (temporal dependence) (centre) and year specific final (right) distribution functions. The solid line shows the line of equality, y = x. 95% tolerance bounds are shown by dashed lines.

Call this bootstrapped series of uniform variables  $\{U_i^B\}$ . Then we transform the bootstrap sample back to the original scale using our final model, but preserving the original covariate information of  $\{Y_i\}$ , i.e.,  $\{X_{j_i}, d, j\}$ , call this series  $\{Y_i^B\}$ . Then we fit our tail model to the  $\{Y^B\}$  series to re-estimate all of the parameters  $\alpha_{\sigma}, \beta_{\sigma}, \phi_{\sigma}, \gamma_{\sigma}^{(x)}, \xi, \beta_{\lambda}, \phi_{\lambda}, \alpha_{\lambda}^{(x)}, \beta_{\lambda}^{(x)}, \phi_{\lambda}^{(x)}, \theta$ and  $\psi$ , as well as the monthly empirical distributions and thresholds. These estimates are then used to derive the annual maxima distribution of sea levels and estimate return levels.

We take 200 bootstrap samples and present 95% confidence intervals in Figures 12 and S26 at Sheerness and the other sites, respectively. The uncertainty associated with each return level increases with return period, as expected, since it becomes uncertain as we extrapolate. Uncertainty at the largest return periods can be attributed to the difficulty in accurately and precisely estimating the shape parameter of the GPD using data. This arises from using data at only a single site. It is a well-known limitation so a number of approaches have been developed to reduce the variation in estimates without introducing bias, with the most prominent being spatial pooling methods (see Section 6). We adopt a variant of this approach using the same strategy as Environment Agency (2018). They use a Normal $(0.0119, 0.0343^2)$  prior, derived from the sample of site specific shape parameter estimates for skew surges from over the entire UK network of coastal sites. We incorporate this prior information into our model, using a penalised likelihood framework for parameter estimation. Our shape parameter estimates are now -0.019 (-0.021, 0.059), 0.014 (-0.025, 0.052), -0.027 (-0.058, 0.004) and 0.008 (-0.039, 0.054) for Heysham, Lowestoft, Newlyn and Sheerness, respectively; 95% confidence intervals are given in parentheses. First note that this additional information is consistent with the data at each site since there is a close agreement in the point estimates of the shape parameters before and after including this information, with the point estimates all slightly closer to zero, whilst the scale and rate parameter estimates remain unchanged. More critically, we now have narrower confidence intervals for the shape parameters. By incorporating this extra spatial information about skew surges, we have reduced the uncertainty associated with shape parameter estimation and hence in the return level estimates.

Figure 12 demonstrates how adding the penalty (prior) information reduces uncertainty at Sheerness, since the distribution of  $\hat{\xi}$  over 200 bootstrap samples is much narrower. Figure 12 also shows the corresponding return level estimates and 95% bootstrap confidence intervals for this updated model; the return level curve is only affected at large return levels, as expected, where the estimates are less bounded. There is a dramatic reduction in the uncertainty associated with high return levels; at the 10,000 year level, adding prior information

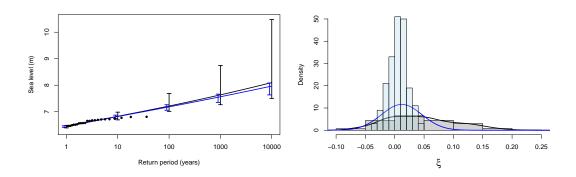


FIG 12. 95% bootstrap confidence intervals on the final (temporal dependence) return level estimates at Sheerness (left) before (black) and after (blue) adding a prior distribution to the shape parameter. Empirical estimates are shown by black points. Histograms of the shape parameter estimates and their densities (right) for these two models in their corresponding colours.

has reduced the range of confidence interval by 2.5m. Clearly, adding prior information on the shape parameter is important. However, we didn't do this sooner because it is important to allow the data to speak for itself when trying to reflect the realism of the sea level process in other aspects of our methodology. See Figure S26 for the updated return level estimates and confidence intervals at Heysham, Lowestoft and Newlyn.

In Section S.5 of the Supplementary Material we compare return level estimates for the sea level process from our final model across different quantiles used for the skew surge model to define extreme values. We find that the uncertainty associated with threshold selection is small compared to the other sources of uncertainty captured in the 95% confidence intervals, both with and without the prior information on the shape parameter.

So far, assessing sea level return level estimates derived across different statistical models has been restricted when comparing their agreement with empirical estimates, which are noisy, subject to bias from missing data and the tidal window over which they are observed, and the comparisons are limited to return periods up to the length of observation series. To aid with comparison to more reliable empirical estimates for short return periods and to enable comparison at longer return periods, we have also used the 483 year numerical model data set for Sheerness developed by Howard and Williams (2021). Specifically, inference is based on repeated samples which are identical in length to the observed data at Sheerness and are assessed relative to the inference using the full 483 years of data. We find that our approach gives smaller RMSE values than the SSJPM for all return periods.

5.5. Sea Level Seasonality. It is helpful to understand when in the year the most extreme sea levels might occur, such as for coastal defence maintenance planning. We use our final model (temporal dependence) to investigate the seasonality of extreme sea levels. We are interested in the probability that a randomly selected sea level annual maxima M is from month j given it is equal to some level z. We consider  $z_p$ , a sea level with an associated annual exceedance probability  $p \in [0, 1]$ , so that  $z_p$  is derived from expression (4.15). Then the probability of interest is defined by,

(5.2) 
$$P_M(j;z) = \mathbb{P}(m(M) = j|M = z),$$

where m(M) denotes the month of occurrence of M and  $\hat{\mathbb{P}}(\cdot)$  is under our final model. This is a similar metric to that used for assessing peak tide seasonality (Section 3.2), however, since we have a model for the sea levels we condition on M = z because we can obtain

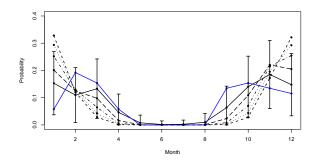


FIG 13. Estimates of  $\hat{P}_M(j; z)$  for months j = 1 - 12 at Sheerness, for p = 1 (black solid), 0.1 (long-dashed), 0.01 (dot-dashed), 0.001 (dotted), and 0.0001 (dashed). The blue line is the empirical estimate  $\tilde{P}_M(j; z_1)$ , 95% confidence intervals are for  $\hat{P}_M(j; z_1)$ .

the form of the probability density function. Increasing z to rare return levels uncovers the seasonal variations in extreme sea levels. For months where the most extreme sea levels occur,  $\hat{P}_M(j;z)$  takes its largest values, with  $\hat{P}_M(j;z) \to 0$  for months with the least extreme sea levels. We derive the form of  $\hat{P}_M(j;z)$  in Section S.7 of the Supplementary Material.

Figure 13 shows the estimates  $\hat{P}_M(j;z)$  to demonstrate sea level seasonality of extreme levels at Sheerness, as well as the empirical estimate for sea levels equal to the 1 year return level (see Figure S28 for the remaining sites). These empirical estimates tend to lie within the 95% confidence intervals for the corresponding model estimates, at the 1 year return level. This demonstrates that our model for extreme sea levels is capturing the seasonality well. Our model estimates allow us to extrapolate beyond the observed data. At Sheerness, the empirical and model estimates at the 1 year return level are influenced by the tidal seasonality since they peak near the equinoxes. However, once we extrapolate to higher return levels, the sea level seasonality is almost entirely influenced by skew surge seasonality. This is the case for all sites at high return levels, since tides are bounded above by HAT but this is less than the 10 year sea level return level for all four sites. Therefore, extremely large skew surges are required to exceed the highest return levels, which are most likely to occur in winter.

**6.** Discussion. We have developed a novel methodology for estimating extreme sea levels by accounting for seasonality in skew surge and peak tide, the dependence between them, and temporal dependence in skew surges. Our results show a significant improvement on current methods, which ignored these features of the sea level processes and instead made several simplifying assumptions. Our model also allows us to study the seasonality of sea levels exceeding levels previously unobserved; this can be useful for coastal defence maintenance planning. The return levels estimated from our model present a more accurate representation of future extreme levels, and should be used for future coastal defence upgrades. Our methodology can be applied to all UK National Tide Gauge Network sites with a sufficient data record length, and is not limited to UK locations. Our method could also be applied to estimate extreme sea currents distributions where a similar style framework for dealing with tide and surge components is used (Robinson and Tawn, 1997); this would provide substantial value for designing offshore structures to ensure required levels of safety. Seasonal variations are a common feature of many environmental variables and therefore should be accounted for when estimating return levels; the methodology discussed in Section 4.3 is applicable to a range of variables. We conclude by addressing some of the limitations with our model and suggest avenues for further exploration.

Our method assumes a steady state climate, since the existing mean sea level rise trend was removed before analysis, and we have not included any longer-term trends in our model. Note the mean sea level trend must be added back onto the return level estimates presented in this paper when used in practice. It is fundamental that the effects of anthropogenic induced climate change and sea level rise are accounted for. Whilst most of this is captured by mean sea level rise, it has been suggested there is a small, but different, trend in the extreme sea levels (Haigh, Nicholls and Wells, 2010; Menéndez and Woodworth, 2010). There are two approaches to do this; the preprocessing model of Eastoe and Tawn (2009) where trends are modelled and removed before analysis, then covariates can be added to the extreme value model to explain trends in the tail of the distribution that are different from that in the body. Alternatively, covariates can be added to model parameters straight away without an attempt to remove any trends from the data (Davison and Smith, 1990). D'Arcy, Tawn and Sifnioti (2022) adopt this approach to study changes in extreme sea level relative to the much easier estimated changes in mean sea levels. They extend the methodology developed here to incorporate a global mean temperature (GMT) anomaly covariate. They find changes with GMT increases are primarily experienced through an increased frequency of extreme skew surges.

We have shown that skew surges clearly exhibit within-year seasonality. We used a monthly threshold, defined as a quantile of the monthly skew surge distribution, to define extreme values and develop our non-stationary model. Using a quantile ensures there are a similar number of exceedances to model per month. This approach is similar to that of Carter and Challenor (1981), where we first assume stationarity within months and then build in the seasonal variation on a shorter temporal resolution through covariates in the GPD parameters. This meant that we were able to capture most of the non-stationarity, as well as skew surge-peak tide dependence, at the same stage of the modelling process. However, we could have considered a smoother threshold choice by using quantile regression (Northrop, Jonathan and Randell, 2016) but we did not try this as our monthly threshold appeared sufficient.

Spatial pooling provides a promising framework to capture longer-term trends due to climate change; single site trends are subtle but sharing information across sites could give more significant results. Spatial pooling also enables inference at locations with limited or no data, where our current singe site model would not perform well. In Section 5.4 we add a prior penalty for the shape parameter in our model, based on spatial information; this is a first approach to spatial pooling and drastically reduced the uncertainty associated with longer term return levels. Table S3 shows that the 95% confidence intervals of  $\xi$  for Model S4 (no penalty) overlap at all sites, therefore it may be reasonable to fix the shape parameter across sites, as an alternative approach for borrowing information. Fixing the shape parameter in homogeneous regions is a crucial step in regional frequency analysis, originally introduced by Hosking and Wallis (1997). This is also an assumption in the Bayesian hierarchical modelling framework presented by Sharkey and Winter (2019) and a common assumption for spatial extreme value theory (see Davison, Huser and Thibaud (2019)). We refer the reader to Batstone et al. (2013), Bernardara, Andreewsky and Benoit (2011) and Haigh, Nicholls and Wells (2010) for different approaches to spatial pooling for extreme sea level estimation. Whilst spatial pooling is clearly an important aspect of extreme sea level estimation, it is also fundamental that the marginal site estimates are accurate; therefore our approach for capturing the realism of the sea level process should be adopted within a spatial framework.

Acknowledgments. The authors would like to thank Jenny Sansom of the Environment Agency and Joanne Williams of National Oceanography Centre for providing the tide gauge data, and Tom Howard of the Met Office for supplying the physical model data.

**Funding.** This paper is based on work completed while Eleanor D'Arcy was part of the EPSRC funded STOR-i centre for doctoral training (EP/S022252/1).

### SUPPLEMENTARY MATERIAL

#### **Supplementary Material**

Supporting material is provided for the method discussed in the main text, as well as results from a simulation study. Additional figures, tables and derivations are also included.

## Code

Code to conduct our exploratory analysis and implement our methods can be found in the eleanordarcy/ESL\_estimation repository on GitHub and in the accompanying .zip file.

#### Data

The data used in this paper is a quality controlled version of the publicly available data: https://www.bodc.ac.uk/data/hosted\_data\_systems/sea\_level/uk\_tide\_gauge\_network/.

#### REFERENCES

- ENVIRONMENT AGENCY (2018). Coastal Flood Boundary Conditions for the UK: update 2018. Technical summary report. https://environment.data.gov.uk/dataset/6e856bda-0ca9-404f-93d7-566a2378a7a8. Accessed 01/10/21.
- ASADI, P., ENGELKE, S. and DAVISON, A. C. (2018). Optimal regionalization of extreme value distributions for flood estimation. *Journal of Hydrology* **556** 182–193.
- BARANES, H., WOODRUFF, J., TALKE, S., KOPP, R., RAY, R. and DECONTO, R. (2020). Tidally driven interannual variation in extreme sea level frequencies in the Gulf of Maine. *Journal of Geophysical Research: Oceans* 125 e2020JC016291.
- BASHTANNYK, D. M. and HYNDMAN, R. J. (2001). Bandwidth selection for kernel conditional density estimation. Computational Statistics & Data Analysis 36 279–298.
- BATSTONE, C., LAWLESS, M., TAWN, J. A., HORSBURGH, K., BLACKMAN, D., MCMILLAN, A., WORTH, D., LAEGER, S. and HUNT, T. (2013). A UK best-practice approach for extreme sea-level analysis along complex topographic coastlines. *Ocean Engineering* **71** 28–39.
- BERNARDARA, P., ANDREEWSKY, M. and BENOIT, M. (2011). Application of regional frequency analysis to the estimation of extreme storm surges. *Journal of Geophysical Research: Oceans* **116**.
- CARTER, D. and CHALLENOR, P. (1981). Estimating return values of environmental parameters. *Quarterly Journal of the Royal Meteorological Society* **107** 259–266.
- CHAVEZ-DEMOULIN, V. and DAVISON, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C* 54 207–222.
- COLES, S. G. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer, London.
- COLES, S. G., HEFFERNAN, J. and TAWN, J. A. (1999). Dependence Measures for Extreme Value Analyses. *Extremes* **2** 339–365.
- COLES, S. G. and TAWN, J. A. (1990). Statistics of coastal flood prevention. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* 332 457–476.
- COLES, S. G., TAWN, J. A. and SMITH, R. L. (1994). A seasonal Markov model for extremely low temperatures. *Environmetrics* **5** 221–239.
- D'ARCY, E., TAWN, J. A. and SIFNIOTI, D. E. (2022). Accounting for climate change in extreme sea level estimation. *Water* **14** 2956.
- D'ARCY, E., TAWN, J. A., JOLY-LAUGEL, A. and SIFNIOTI, D. E. (2023). Supplement to "Accounting for Seasonality in Extreme Sea Level Estimation".
- DAVISON, A. C., HUSER, R. and THIBAUD, E. (2019). Spatial Extremes. In *Handbook of Environmental and Ecological Statistics* 711–744. Chapman and Hall/CRC.
- DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds (with discussion). Journal of the Royal Statistical Society: Series B 52 393–425.
- DIXON, M. and TAWN, J. (1994). Extreme sea levels: modelling interaction between tide and surge. Statistics for the Environment 2: Water Related Issues 221–232.
- DIXON, M. J., TAWN, J. A. and VASSIE, J. M. (1998). Spatial modelling of extreme sea-levels. *Environmetrics* **9** 283–301.
- DIXON, M. J. and TAWN, J. A. (1999). The effect of non-stationarity on extreme sea-level estimation. *Journal of the Royal Statistical Society: Series C* 48 135–151.
- EASTOE, E. F. and TAWN, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. Journal of the Royal Statistical Society: Series C 58 25–45.
- EGBERT, G. D. and RAY, R. D. (2017). Tidal prediction. Journal of Marine Research 75 189-237.

- FAWCETT, L. and WALSHAW, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics* **18** 173–188.
- FERRO, C. A. T. and SEGERS, J. (2003). Inference for clusters of extreme values. Journal of the Royal Statistical Society: Series B 65 545–556.
- GRAFF, J. (1978). Concerning the recurrence of abnormal sea levels. *Coastal Engineering* 2 177–187.
- HAIGH, I. D., NICHOLLS, R. and WELLS, N. (2010). A comparison of the main methods for estimating probabilities of extreme still water levels. *Coastal Engineering* 57 838–849.
- HOSKING, J. R. M. and WALLIS, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press, Cambridge; New York.
- HOWARD, T. and WILLIAMS, S. D. P. (2021). Towards using state-of-the-art climate models to help constrain estimates of unprecedented UK storm surges. *Natural Hazards and Earth System Sciences* **21** 3693–3712.
- JONATHAN, P., RANDELL, D., WU, Y. and EWANS, K. (2014). Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects. *Ocean Engineering* **88** 520–532.
- LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, New York.
- LEDFORD, A. W. and TAWN, J. A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B* 65 521–543.
- MENÉNDEZ, M. and WOODWORTH, P. L. (2010). Changes in extreme high water levels based on a quasi-global tide-gauge data set. *Journal of Geophysical Research: Oceans* 115.
- NOC (2021). National Tidal and Sea Level Facility. https://www.ntslf.org/.
- NORTHROP, P. J., JONATHAN, P. and RANDELL, D. (2016). Threshold Modeling of Nonstationary Extremes. In *Extreme Value Modeling and Risk Analysis* 107–128. Chapman and Hall/CRC.
- POLITIS, D. N. and ROMANO, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association* **89** 1303–1313.
- PRANDLE, D. and WOLF, J. (1978). Surge-tide interaction in the southern North Sea. *Elsevier Oceanography Series* 23 161–185.
- PUGH, D. and VASSIE, J. (1978). Extreme sea levels from tide and surge probability. *Coastal Engineering* 16 911–930.
- PUGH, D. and WOODWORTH, P. (2014). Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes. Cambridge University Press.
- ROBINSON, M. E. and TAWN, J. A. (1997). Statistics for extreme sea currents. Journal of the Royal Statistical Society: Series C (Applied Statistics) 46 183–205.
- RODRÍGUEZ ENRÍQUEZ, A., WAHL, T., BARANES, H., TALKE, S. A., ORTON, P. M., BOOTH, J. F. and D HAIGH, I. (2021). Predictable changes in extreme sea levels and coastal flood risk due to nodal and perigean astronomical tidal cycles. *Earth and Space Science Open Archive*. https://doi.org/10.1002/essoar.10508511.1
- ROHRBECK, C. and TAWN, J. A. (2021). Bayesian Spatial Clustering of Extremal Behavior for Hydrological Variables. *Journal of Computational and Graphical Statistics* **30** 91–105.
- ROOTZÉN, H. and KATZ, R. W. (2013). Design life level: quantifying risk in a changing climate. *Water Resources Research* **49** 5964–5972.
- SHARKEY, P. and WINTER, H. C. (2019). A Bayesian spatial hierarchical model for extreme precipitation in Great Britain. *Environmetrics* 30 e2529.
- SMITH, R. L. and WEISSMAN, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society:* Series B 56 515–528.
- TAWN, J. A. (1988). An extreme-value theory model for dependent observations. *Journal of Hydrology* **101** 227–250.
- TAWN, J. A. (1992). Estimating probabilities of extreme sea-levels. *Journal of the Royal Statistical Society:* Series C 41 77–93.
- TAWN, J. and VASSIE, J. (1989). Extreme sea levels: the joint probabilities method revisited and revised. *Proceedings of the Institution of Civil Engineers* **87** 429–442.
- WADEY, M. P., HAIGH, I., NICHOLLS, R. J., BROWN, J. M., HORSBURGH, K., CARROLL, B., GALLOP, S. L., MASON, T., BRADSHAW, E. et al. (2015). A comparison of the 31 January–1 February 1953 and 5–6 December 2013 coastal flood events around the UK. *Frontiers in Marine Science* 2 84.
- WILLIAMS, J., HORSBURGH, K. J., WILLIAMS, J. A. and PROCTOR, R. N. (2016). Tide and skew surge independence: new insights for flood risk. *Geophysical Research Letters* 43 6410–6417.

# SUPPLEMENT TO "ACCOUNTING FOR SEASONALITY IN EXTREME SEA LEVEL ESTIMATION"

BY ELEANOR D'ARCY<sup>1,a</sup>, JONATHAN A. TAWN<sup>1</sup>, AMÉLIE JOLY<sup>2</sup> AND DAFNI E. SIFNIOTI<sup>2</sup>

<sup>1</sup>STOR-i Centre for Doctoral Training, Department of Mathematics and Statistics, Lancaster University, LA1 4YR, UK, <sup>a</sup>e.darcy@lancaster.ac.uk

<sup>2</sup>EDF Energy R&D UK Centre, Croydon, CR0 2AJ, UK

**S.1. Introduction.** This document outlines the supplementary material for "Accounting for Seasonality in Extreme Sea Level Estimation", subsequently referred to as the main paper. Firstly in Section S.2, we present the exploratory analysis to assess skew surge-peak tide dependence, discussed in Section 3.3 of the main paper; we demonstrate a time-varying relationship at Sheerness. We fit the tide dependent skew surge model presented in Section 4.4 of the main paper to a 483 year physical model dataset in Section S.3, to illustrate the physical justification of modelling skew surge-peak tide dependence. Then in Section S.4 we investigate temporal dependence in the skew surge series, this was discussed in Section 3.2 of the main paper and then accounted for in our methodology in Section 4.5. We investigate the sensitivity of our sea level return level estimates to the choice of the threshold used in the skew surge GPD model in Section S.5. We present the results of a simulation study in Section S.6, where we estimate sea level return levels using our proposed method and that used in practice (the skew surge joint probabilities method of Batstone et al. (2013), abbreviated to SSJPM) to samples of 483 years of physical model data. In Section S.7 we derive an analytical expression for the probability that a randomly selected sea level annual maxima is from a particular month, given it is equal to a return level (see equation (5.2) of the main paper); this is evaluated in Section 5.5 of the main paper to understand the seasonality of extreme sea levels. In Section S.8 we detail the process of transforming skew surges to uniform margins using the final model presented in the main paper, and we use this as a means of assessing skew surge model fit; this was used in the bootstrap procedure for uncertainty quantification on return level estimations in Section 5.4 of the main paper. Lastly, we present supplementary figures in Section S.9.

**S.2. Skew Surge-Peak Tide Dependence.** In this section, we present our exploratory analysis to demonstrate that it is reasonable to assume skew surge and peak tide are independent at Heysham, Lowestoft and Newlyn, but not at Sheerness. We perform various statistical tests at all sites. Firstly, we formally test if there is a relationship between extreme skew surges and their associated ranked peak tide, where extreme skew surges are defined as exceedances of different thresholds. Then we investigate if all peak tides come from the same distribution as those associated with extreme skew surges. We also do this on a monthly scale at Sheerness to understand how the dependence structure changes within a year. Lastly, we use a simple quantile regression technique to test if the quantile of skew surges associated with different ordered tidal bands varies.

We test if ranked peak tides associated with extreme skew surge (defined as exceedances of the 0.95 quantile) are uniformly distributed, using a Kolmogorov-Smirnov test (see Figure S11). If the two components are independent, these will be distributed Uniform(0,T)

Kolmogorov-Smirnov test p values for uniformity of ranked peak tides associated with extreme skew surges, defined exceedances of different quantiles of the data (0.9, 0.95, 0.975, 0.999). Average p values, after repeated bootstrapping, are shown in parentheses.

Site	Block size	0.9	0.95	0.975	0.99
HEY	19	0.0052 (0.012)	0.044 (0.083)	0.1813 (0.11)	0.32 (0.12)
LOW	5	0.63 (0.17)	0.72 (0.22)	0.78 (0.27)	0.33 (0.23)
NEW	20	0.070 (0.075)	0.091 (0.087)	0.15 (0.17)	0.12 (0.11)
SHE	6	$1.5 \times 10^{-8} (4.1 \times 10^{-8})$	$1.1 \times 10^{-4} (2.9 \times 10^{-4})$	$9.9 \times 10^{-4} \ (0.0011)$	0.0013 (7×10-4)

where T is the total number of tidal cycles. The standard version of the Kolmogorov-Smirnov test falsely assumes that peak tides are temporally independent; we use the bootstrap procedure of Politis and Romano (1994) to account for this. Average p values over 100 iterations are reported in Table S1, with expected block sizes inferred from the site specific autocorrelation function (acf) plots (see Figure S13). At Sheerness, we find strong evidence to reject the null hypothesis that ranked peak tides associated with extreme skew surges are uniformly distributed, with p value  $2.89 \times 10^{-3}$ . At the other three sites, we find sufficient evidence to reject this claim at the 5% level. We also explore the sensitivity of this test to the choice of threshold used to define extreme skew surges, the p values are reported in Table S1. Lower thresholds typically correspond to lower p values, suggesting we are more likely to reject the independence hypothesis when there are more exceedances.

We also test if the distribution of peak tides associated with extreme skew surges is the same as the distribution of all peak tides, using the Anderson-Darling test. Figure S1 shows these distributions at all sites with their associated probability density functions, estimated using a Gaussian kernel density estimator. If peak tide and skew surge are independent, these two distributions should be identical up to sampling variation. We find insufficient evidence to reject the null hypothesis that these are from the same distribution at Heysham, Lowestoft and Newlyn at the 0.01 significance level (p values are 0.014, 0.083 and 0.215, respectively). However, this is not the case at Sheerness, with p value 0.00025.

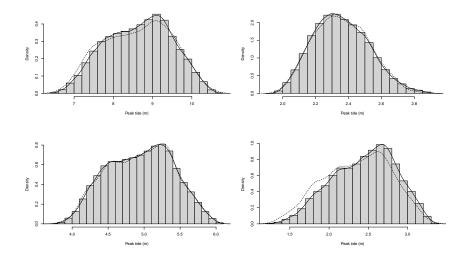


FIG S1. Histogram of all peak tides at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right). Probability density function of all peak tides (solid) and peak tides associated with extreme skew surges (dashed) are interpolated onto each distribution.

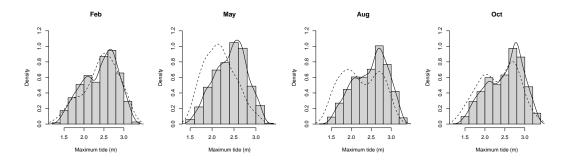


FIG S2. Monthly distributions of peak tides at Sheerness in February, May, August and October. The probability density function of all peak tides (solid) and peak tides associated with extreme skew surge (dashed) are interpolated onto each distribution.

We investigate this relationship at Sheerness further by studying the dependence on a monthly basis to understand how the skew surge-peak tide dependence changes throughout the year. Again, we compare the distributions of all peak tides and extreme skew surge-related peak tides. Figure S2 shows this for February, May, August and October. In May, the mode of the distribution of peak tides associated with extreme skew surges has shifted to a lower value than the distribution of all peak tides. Results from the Anderson-Darling test suggest there is significant evidence, at the 1% level, to reject the null hypothesis that peak tides and the peak tides associated with extreme skew surge come from different distributions in 7 months (including May, August and October from Figure S2). Therefore, we conclude the skew surge-peak tide independence assumption is not valid for these months, at least. When modelling the dependence of skew surge and peak tide in Section 4.4 of the main paper, we recognise that this relationship is changing across the year.

To further investigate skew surge-peak tide dependence, we partition the ordered peak tide series into blocks of 100 (so that block 1 corresponds to the 100 smallest peak tide observations). For each block we estimate the 0.95 quantile of the associated skew surges. The estimated quantiles are shown in Figure S3. There is no immediate relationship between the skew surge quantile and block number. However, when we fit a linear model, there is a significant trend at the  $< 10^{-4}\%$  level at Sheerness and the 0.1% level at Newlyn, but no significant trend at Lowestoft and Heysham. Since our other tests have not found dependence at Newlyn, this relationship is likely to be a physically small relationship but statistically significant due to the length of the data series. However, at Sheerness, it is likely this is due to skew surge-peak tide dependence based on our other findings.

**S.3.** Physical Model Data. In Section S.2 here (and Section 3.3 of the main paper) we identify weak dependence between skew surge and peak tide at Sheerness; we account for this in our skew surge model. Our results show that incorporating peak tide as a covariate on the rate and scale parameter of the GPD for extreme skew surges improved the model fit. In this section, we make the same comparisons on a 483 year data set from a hydrodynamical model driven by a regional climate model (HadGEM3-GC3-MM). Howard and Williams (2021) present this model to generate a dataset of 483 year present-day surges at sites on the UK National Tide Gauge Network. They use their simulations to review the skew surge-peak tide independence assumption at Sheerness, and demonstrate that extreme skew surges are more likely to occur on larger peak tides. D'Arcy and Tawn (2021) evaluate this assumption using a simplified version of the method presented in the main paper. For the scale parameter, we compare the models given in equations (4.6) and (4.12), and we refer to these as Models S2 (without a tide covariate) and S4 (with a tide covariate), respectively, as in the main paper.

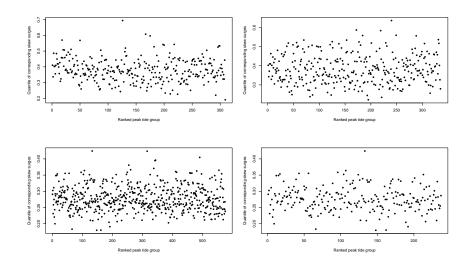


FIG S3. 0.95 quantile estimates of skew surges associated with ranked peak tide groups of size 100, at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right).

Models for the rate parameter are given by equations (4.7) and (4.11), we refer to these as Model R0 (without a tide covariate) and R1 (with a tide covariate), respectively.

For the physical model skew surge data, we begin by comparing models for the scale parameter only. Model S4 reduces both AIC and BIC by 28.75 and 21.03, respectively, relative to Model S2 suggesting that the extra parameter that captures variation with peak tide tide is necessary. The likelihood ratio test strongly agrees with this, giving a significant p value of the order  $10^{-8}$ . We estimate the tidal coefficient to be  $\hat{\gamma}_{\sigma} = -0.013$  with 95% confidence interval (-0.018, -0.0081). This suggests that more extreme skew surges occur on lower peak tides, as found in our exploratory analysis. Since the confidence interval doesn't contain 0 the tidal coefficient is significant. This parameter estimate is reassuringly close to the corresponding estimate for the observed data of  $\hat{\gamma}_{\sigma} = -0.012$  (-0.026, 0.0011) estimated in Section 4.4 of the main paper. Whilst this suggests threshold excesses occur on lower peak tides, the confidence interval contains 0 so this result for the observed data at Sheerness, which are based on physical reasoning with no data measurement issues and no issues with changes in the tide gauge and estuary over time.

Next, we compare models for the rate parameter on the physical model data. We find that Model R1 reduces AIC by 88.4 and BIC by 56.3 when compared with Model R0, this suggests that adding a peak tidal covariate to the rate parameter is important. Therefore, we make the same conclusions on 483 years of physical model data, as we do on 37 years of observed data. This shows the rate model parameterisation is supported empirically and physically. When we cannot reasonably assume skew surge and peak tide are independent, accounting for their dependence is important as it can lead to practical differences in the sea level return level estimates, as shown in Section 5.3 of the main paper.

**S.4. Temporal Skew Surge Dependence.** Here we provide further details on skew surge temporal dependence discussed in Section 3.2 of the main paper. We study the two key measures of extremal dependence, these are a measure of asymptotic dependence  $\chi$  and of asymptotic independence  $\bar{\chi}$  for each site. For random variables  $Y_i$  and  $Y_{i+\tau}$  separated by lag  $\tau$  from a stationary sequence, with  $y^F$  the upper endpoint of their distribution F, Coles, Heffernan

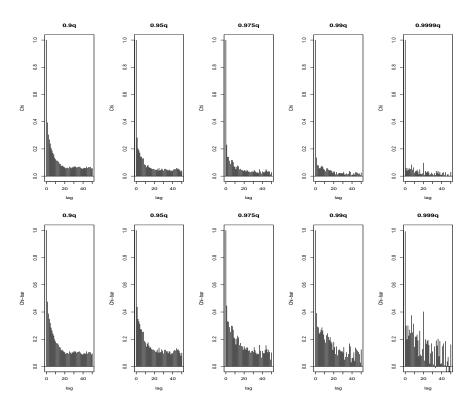


FIG S4. Estimates of  $\chi$  (top row) and  $\bar{\chi}$  (bottom row) for Heysham for exceedances of the 0.9, 0.95, 0.975, 0.99 and 0.999 quantiles (from left to right column) at various lags.

and Tawn (1999) define

(S.1) 
$$\chi_{\tau} = \lim_{y \to y^F} \mathbb{P}(Y_{i+\tau} > y | Y_i > y)$$

as a measure of asymptotic dependence where  $\chi_{\tau} \in [0, 1]$ . If  $\chi_{\tau} \in (0, 1]$ , we say that  $Y_i$  and  $Y_{i+\tau}$  are asymptotically dependent; this means there is non-zero probability of  $Y_{i+\tau}$  being large when  $Y_{i+\tau}$  is large at all extreme levels. Whereas  $\chi_{\tau} = 1$  and  $\chi_{\tau} = 0$  correspond to perfect dependence and asymptotic independence, respectively. Therefore,  $\chi_{\tau}$  fails to signify the level of asymptotic independence, so Coles, Heffernan and Tawn (1999) also define the measure  $\bar{\chi}_{\tau}$  as

(S.2) 
$$\bar{\chi}_{\tau} = \lim_{y \to y^F} \frac{2\log \mathbb{P}(Y_i > y)}{\log \mathbb{P}(Y_i > y, Y_{i+\tau} > y)} - 1,$$

where  $\bar{\chi}_{\tau} \in (-1, 1]$ . Asymptotic dependence and asymptotic independence correspond to  $\bar{\chi} = 1$  and  $\bar{\chi} < 1$ , respectively, whilst  $0 < \bar{\chi}_{\tau} < 1$  and  $-1 < \bar{\chi}_{\tau} < 0$  correspond to positive and negative association, respectively, and  $\bar{\chi}_{\tau} = 0$  corresponds to near independence. We evaluate both measures with y at different quantiles of the distribution in Figure S4 for Heysham, and Figures S14 and S15 for the remaining sites.

Firstly, we use estimates of  $\chi$  and  $\bar{\chi}$  to choose a high threshold y to empirically estimate the extremal index  $\theta$  using the runs method. We are interested in choosing a threshold where these measures tend to zero without significant noise; the 0.95 quantile is sufficient for each site. We explore the sensitivity of our estimate to this threshold choice in Figure S5; as we increase the threshold, the estimate of  $\theta$  increases. This is what we expect since exceedances of lower thresholds are likely to exhibit more dependence than exceedances of higher thresholds. We

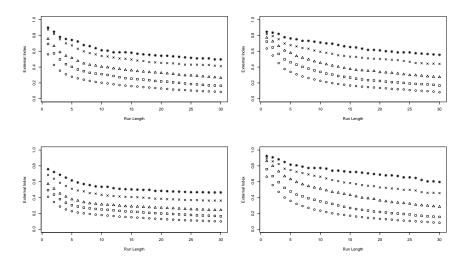


FIG S5. Empirical (runs) estimates of the extremal index  $\theta$  for skew surge, at various quantiles and run lengths at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right). The estimates use thresholds which are taken to be quantiles 0.9 (circle), 0.95 (square), 0.975 (triangle), 0.99 (cross) and 0.999 (star).

also explore the sensitivity to run length r in Figure S5. Estimates of the extremal index are generally higher at Newlyn and Sheerness, which agrees with the acf values in Figure S13. The choice of run length and threshold level are important because they have a significant influence on the  $\theta$  estimate, so we choose these carefully.

Since our empirical estimates are sensitive to the threshold level, we develop a parametric model for the subasymptotic extremal index  $\hat{\theta}(y, r)$ , that is dependent on the skew surge level y; this is given in equation (4.16) of the main paper. Figure 6 of the main paper shows a consistent model fit with the empiricals at Heysham, with parameter estimates  $\hat{\theta} = 1$  and  $\hat{\psi} = 0.33$  for r = 2. Figure S16 shows the fit at the remaining sites and these also match closely with the empiricals. The parameter estimates are  $\hat{\theta} = 1, 0.95, 0.89$  and  $\hat{\psi} = 0.42, 0.17, 0.21$  for r = 10, 1, 10 at Lowestoft, Newlyn and Sheerness, respectively. As expected,  $\hat{\theta}$  lies close to 1 in all cases, this represents the limiting extremal index and the case of independence. The estimate of  $\psi$  tells us about the speed of convergence to  $\hat{\theta}$ , so that Newlyn converges fastest whilst Lowestoft is the slowest.

**S.5. Threshold Sensitivity.** In this section, we explore the sensitivity of the sea level return level estimates derived from the proposed model (4.15) to the choice of threshold u for the skew surge model (4.14); above this threshold we fit a non-stationary GPD and the empirical distribution is used below. Recall that our modelling approach uses a time-varying threshold, so we have a different threshold  $u_j$  for each month j = 1, ..., 12. In the main paper, we use the 0.95 quantile of monthly skew surges, here we compare these results with using the 0.9 and 0.99 quantile to define extreme skew surges.

Figure S6 shows the return level estimates from the proposed model where the 0.9, 0.95 and 0.99 quantiles are used to define the extreme values at Sheerness. The 95% confidence intervals are the same as those in the main paper where the 0.95 quantile is used, both before and after we add a prior distribution (see Section 5.4 of the main paper for details). We can see that the estimates for the method using the 0.9 and 0.99 quantile of skew surges lie within the confidence intervals for the original estimates (0.95 quantile). The estimated values from these two new thresholds are much closer to the point estimate for the 0.95 quantile threshold

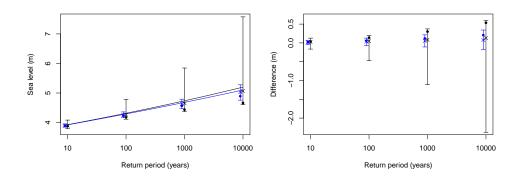


FIG S6. Return level estimates from the final (temporal dependence) model with 95% confidence intervals before (black) and after (blue) adding a prior distribution to the shape parameter at Sheerness, where the 0.95 quantile is used for the skew surge distribution (left); estimates are also shown at the 10, 100, 1000 and 10,000 year levels when the 0.9 (crosses) and 0.99 (dots) quantiles are used in the skew surge model, both with (blue) and without (black) the prior distribution on the shape parameter of the GPD. The corresponding difference between the estimates from the original approach (0.95 quantile) with the 0.9 (crosses) and 0.99 (dots) quantiles, compared with confidence intervals (right).

than the endpoints of the associated 95% confidence intervals, suggesting that the impact of threshold uncertainty is small compared to other sources of uncertainty in the modelling framework. This is the case for the model with and without the prior distribution on the shape parameter of the GPD, where the uncertainty for the former is much smaller but still larger than the uncertainty attributable to threshold selection. We find similar results at the remaining sites, see Figure S29.

**S.6. Simulation Study.** So far, we have compared models when fit to observed data at the tide gauges therefore comparison with the empirical estimates are restricted to the length of observed data. Here, we use samples from the 483 year physical model data at Sheerness (introduced in Section S.3) to fit our proposed model (given in expression (4.15) of the main paper, without a prior on the shape parameter) and the SSJPM of Batstone et al. (2013), and compare these with empirical estimates. Howard and Williams (2021) introduce the data, which are generated from a widely-used physical model using a long-run hydrodynamical model driven by a long-run climate model (HadGEM3-GC3-MM) without any drivers for climate change. This gives us an entirely realistic dataset as the associated hydrodynamical model has been calibrated to give realistic skew surges and peak tides, whilst the climate model ensures features such as seasonality and temporal dependence are captured.

From the 483 year dataset, we randomly select 30 samples of 37 continuous years so that they approximately match the observed length of data at Sheerness. Note we say approximately here because the physical model data benefit from no missing data, whilst the observed tide gauge data has 9% missing at Sheerness. Additionally, months are simplified to 12 months of 30 days each in the physical model data, so years are only 360 days long. For each sample, we fit the SSJPM and our proposed model. The latter averages over the annual maxima distribution for each specific year in the data to capture interannual tidal variations. When fitting this to a sample of physical model data, we average over all 483 years of tidal data (which can be predicted from the observed 37 years of sea level data) so that each years' annual tidal regime is captured, but the skew surge model parameters are estimated from the 37 year sample data. This ensures that longer-term tidal variations are accounted for and the results are more comparable with empirical estimates. Since the SSJPM assumes peak tides

 TABLE S2

 Sample standard deviation (SD), bias and RMSE (in cm) of the 1, 10 and 100 year sea level return level

 estimates from 30 samples of 37 years from the physical model data using our model and the SSJPM (Batstone et al., 2013). Here we take the truth as empirical estimates based on the 483 years of physical model data.

	Our model			SSJPM			
Return period	SD	Bias	RMSE	SD	Bias	RMSE	
1	2.4	-2.5	3.4	2.1	4.8	5.3	
10	4.8	-0.4	4.7	5.3	3.2	6.1	
100	8.6	11.3	14.2	11.8	13.6	17.9	

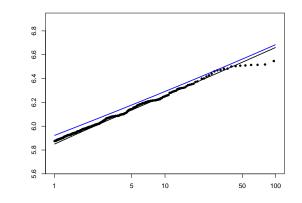


FIG S7. Return level estimates (y-axis in metres) for different return periods (x-axis in years) from the final (temporal dependence) model (black) and the SSJPM (blue) (Batstone et al., 2013), averaged over 30 samples of 37 years from the physical model data (Howard and Williams, 2021). Empirical estimates from all 483 years of physical model data are shown by black points.

are stationary and only uses tidal information from a single nodal cycle, we only use the tides in the 37 year sample, rather than all 483 years of tides, to better represent their approach and to help show that the major improvements come from our modelling of seasonality. We then estimate sea level return levels corresponding to return periods between 1 and 100 years, and compare them to the empirical estimates; these are shown in Figure S7, averaged over each sample for each return period. The average estimates from our model lie much closer to the empirical than the SSJPM, especially for lower return periods where they match almost perfectly. At the 100 year level, our estimate still lies closer but the empirical estimates are less reliable, even though they are taken from 483 years of data. At the 1, 10 and 100 year level, the RMSE over all samples for our model is 0.03m, 0.05m and 0.14m, respectively. For the SSJPM the RMSE values are greater by factors of 1.54, 1.29 and 1.26, respectively. Notice the difference between estimates from the two methods is greater for lower return periods, with our method always being superior. A summary of the 1, 10 and 100 year return level estimates is given in Table S2 for both methods; the SSJPM has a greater bias and variance over the 30 samples than our method. Note that our method slightly underestimates at the 1 and 10 year return levels, whilst the SSJPM always overestimates and by a greater magnitude.

S.7. Derivation of Expression for Probability (5.2). We are interested in deriving an analytical expression for the probability  $\hat{P}_M(j;z)$  that a randomly selected sea level (annual maxima) M is from month j given it equal to some level z, given by,

(S.1) 
$$\hat{P}_M(j;z) = \hat{\mathbb{P}}(m(M) = j|M = z),$$

where m(M) denotes the month of occurrence of the variable M and  $\hat{\mathbb{P}}(\cdot)$  is under our final model of Section 4.5 in the main paper. This is probability (5.2) in the main paper and we

use this in Section 5.5 to evaluate extreme sea level seasonality. We are specifically interested when  $z = z_p$ , a level with an associated annual exceedance probability  $p \in [0, 1]$ , so that z is a return level derived from expression (4.15).

As the distribution of M varies with year k due to the tidal variations, we begin by conditioning on a fixed year k, so that we only look at sea levels within a specific year and consider the probability  $\hat{P}_{M^{(k)}}(j;z)$ . We rewrite this in terms of the distribution and density of the month j maxima sea level,  $F_{M_j^{(k)}}$  and  $f_{M_j^{(k)}}$ , respectively, and the density of the annual maxima sea level  $f_{M^{(k)}}$ , each conditional on some year  $k = 1, \ldots, K$ , where K is the total number of years of observation,

$$\hat{P}_{M^{(k)}}(j;z) = \frac{f_{M_{j}^{(k)}}(z) \prod_{\substack{J=1,\ldots,12\\J \neq j}} \mathbb{P}(M_{J}^{(k)} < z)}{f_{M^{(k)}}(z)}$$

$$= \frac{1}{f_{M^{(k)}}(z)} \left[ \left( \frac{f_{M^{(k)}_j}(z)}{F_{M^{(k)}_j}(z)} \right) F_{M^{(k)}}(z) \right].$$

Therefore, it follows to find an expression for each of these terms to simplify the above expression. To find the form of the density of the monthly maxima sea levels in a given year  $f_{M^{(k)}}$ , we must differentiate the corresponding distribution function,

$$F_{M_j^{(k)}}(z) = \mathbb{P}(M_j^{(k)} \le z) = \prod_{i=1}^{T_j^{(k)}} \left[ F_Y^{(d,j,x)}(z - X_{j_i}^{(k)}) \right]^{\hat{\theta}(z - X_{j_i}^{(k)}, r)},$$

where  $F_Y^{(d,j,x)}$  is the skew surge distribution function given by equation (4.14) and  $\hat{\theta}(\cdot, r)$  is our extremal index model given by equation (4.16) for fixed run length r. Differentiating this gives

$$(\textbf{S.2}) \qquad \qquad f_{M_{j}^{(k)}}(z) \approx F_{M_{j}^{(k)}}(z) \sum_{i=1}^{T_{j}^{(k)}} \frac{f_{Y}^{(d,j,x)}(z - X_{j_{i}}^{(k)})\hat{\theta}(z - X_{j_{i}}^{(k)}, r)}{F_{Y}^{(d,j)}(z - X_{j_{i}}^{(k)})},$$

when we ignore smaller order terms arising from  $\hat{\theta}'(\cdot, r)$ . The density of skew surges  $f_Y^{(d,j,x)}$  is given by

(S.3) 
$$f_Y^{(d,j,x)}(y) = \begin{cases} \hat{f}_j(y) & \text{if } y \le u_j \\ \frac{\lambda_{d,x}}{\sigma_{d,x}} \Big[ 1 + \xi \Big( \frac{y - u_j}{\sigma_{d,x}} \Big) \Big]_+^{-\frac{1}{\xi} - 1} & \text{if } y > u_j \end{cases}$$

where  $\lambda_{d,x}$  and  $\sigma_{d,x}$  are given by expressions (4.11) and (4.12), respectively. Below the monthly threshold  $u_j$ , we estimate the derivative of the monthly empirical distribution using a Gaussian kernel density estimator and denote this  $\hat{f}_j$ . We find the density of the annual maximum sea levels, for a fixed year k, by differentiating in similar way and ignoring smaller order terms, so that,

(S.4) 
$$f_{M^{(k)}}(z) \approx F_{M^{(k)}}(z) \sum_{J=1}^{12} \sum_{i=1}^{T_J^{(k)}} \frac{f_Y^{(d,J,x)}(z - X_{J_i}^{(k)})\hat{\theta}(z - X_{J_i}^{(k)}, r)}{F_Y^{(d,J,x)}(z - X_{J_i}^{(k)})}.$$

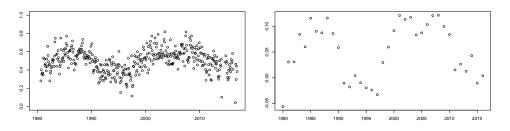


FIG S8. Monthly means (y-axis in meters) of transformed skew surge observations through the final skew surge distribution function (4.14) (left) and annual mean skew surges (right), both at Sheerness, against the year of observation (x-axis).

Using equations (S.2) and (S.4), we can simplify  $\hat{P}_{M^{(k)}}(j;z)$  to

$$\hat{P}_{M^{(k)}}(j;z) = \frac{A^{(i,j,k)}(z)}{\sum\limits_{J=1}^{12} A^{(i,J,k)}(z)} \quad \text{where} \quad A^{(i,j,k)}(z) = \sum\limits_{i=1}^{T_j^{(k)}} \frac{f_Y^{(d,j,x)}(z - X_{j_i}^{(k)})\hat{\theta}(z - X_{j_i}^{(k)}, r)}{F_Y^{(d,j,x)}(z - X_{j_i}^{(k)})}$$

Clearly  $\sum_{j=1}^{12} \hat{P}_{M^{(k)}}(j;z) = 1$ . Then the probability over all K years is given by  $\hat{P}_M(j;z) = \frac{1}{K} \sum_{k=1}^{K} \hat{P}_{M^{(k)}}(j;z)$ .

**S.8. Transforming Skew Surges to Uniform Margins.** In Section 5.4 of the main paper we assess the fit of our final model for sea level annual maxima (4.15) by looking at year-specific distributions and bootstrap confidence intervals on the return level estimates. Here, we test the goodness-of-fit for the final skew surge model (expression (4.14)), using the probability integral transform; if our model fits well, transforming the observations through the fitted distribution function will give a sample of identically distributed Uniform(0, 1) values. This was also a step in our stationary bootstrap procedure, but here we check if these transformed values  $\{U_i^Y\}$  are Uniform using a Kolmogorov-Smirnov test.

Figure S8 shows the transformed skew surges at Sheerness. We can immediately see these are not Uniformly distributed. This is supported by results of yearly Kolmogorov-Smirnov tests for uniformity, where the p values are almost all > 0.05. Instead there appears to be a cyclic sinusoidal trend, following the trend of the 18.6 year nodal cycle. We suspect this is a data issue, perhaps the tidal series was not correctly removed from the sea level observations when the skew surges are obtained. Figure S8 also shows the annual mean skew surges at Sheerness, we can see there is a similar cyclic trend here and that the means are not centred at zero (see Figure S24 for the remaining sites). To correct for this we re-centre the data at zero, by removing the corresponding annual mean from the observations. This is an ad hoc approach and the data should be investigated further. We did not correct for this trend at an earlier stage in the modelling process because it does not have a significant effect on the extreme values. However, once we correct for this trend, we find that the transformed data are uniformly distributed in 32 years (out of 37) at Sheerness, indicating a good model fit for skew surges. We find similar results at the remaining sites, where a cyclic trend is first observed in  $\{U_i^Y\}$  but once the annual means are removed, the transformed data can be reasonably assumed as Uniform(0,1). Figure S25 shows the p values for each year at each site.

## S.9. Supplementary Figures.

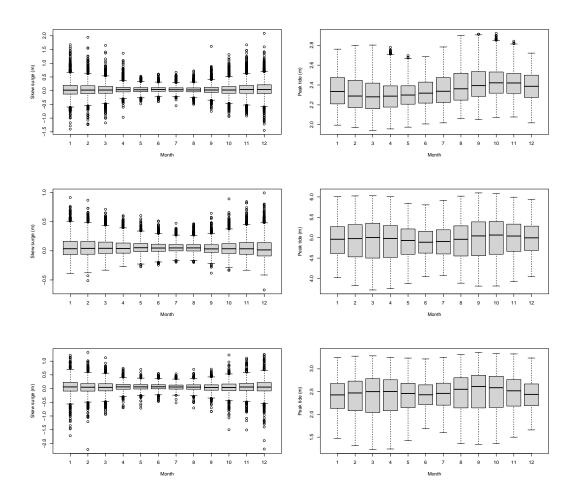


FIG S9. Monthly box plots of skew surge (left column) and peak tide (right column) at Lowestoft (top row), Newlyn (middle row) and Sheerness (bottom row).

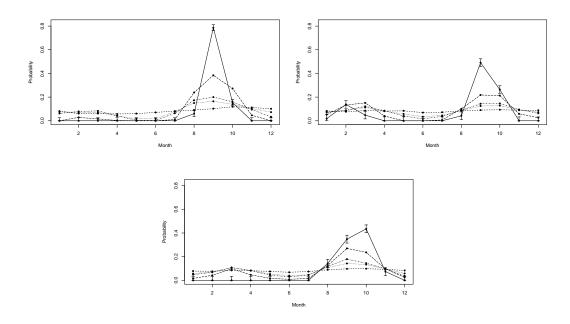


FIG S10. Estimates of  $\tilde{P}_X(j; x_q)$  (expression (3.1) of the main paper) for months j = 1 - 12 and q = 0.5 (dashed), 0.9 (dotted), 0.95 (dot-dashed), 0.99 (long dashed), 0.999 (solid) at Lowestoft (top left), Newlyn (top right) and Sheerness (bottom), with 95% confidence intervals when q = 0.99.

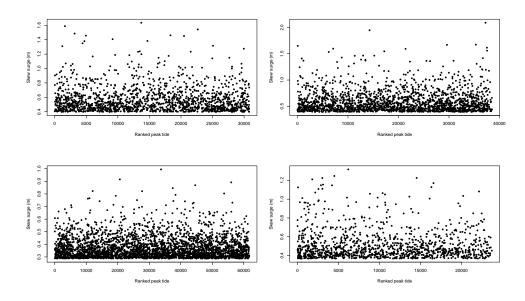


FIG S11. Scatter plot of extreme skew surge observations (exceedances of 0.95 quantile) against associated ranked peak tides at Heysham (top right), Lowestoft (top left), Newlyn (bottom left) and Sheerness (bottom right).

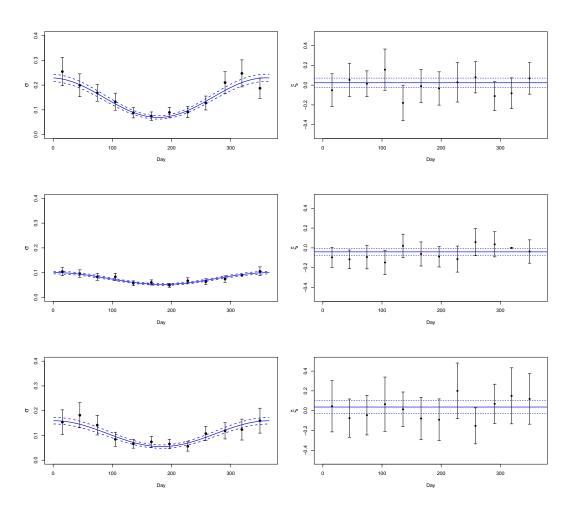


FIG S12. Scale (left column) and shape (right column) parameter estimates for Model S2 (blue) and Model S0 (black) at Lowestoft (top row), Newlyn (middle row) and Sheerness (bottom row). 95% confidence intervals are added to Model S0 (black error bars) and to Model S2 parameter estimates (blue dashed lines).

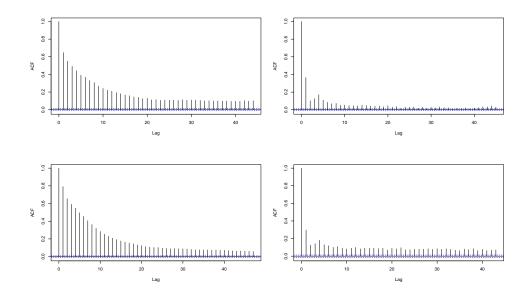


FIG S13. Autocorrelation function (acf) plots for Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right).

 TABLE S3

 Parameter estimates for the scale parameter Models S2 and S4, and the rate parameter Model R1 with 95% confidence intervals, at each site.

confluence intervals, a cach site.				
	Heysham	Lowestoft	Newlyn	Sheerness
Model S2				
$\alpha_{\sigma}$	0.14 (0.13, 0.15)	0.15 (0.14, 0.16)	0.076 (0.073, 0.080)	0.11 (0.10, 0.12)
$\beta_{\sigma}$	0.060 (0.050, 0.070)	0.080 (0.070, 0.090)	0.024 (0.020, 0.028)	0.052 (0.043, 0.061)
$\phi\sigma$	271.51 (262.77, 280.23)	266.32 (260.01, 272.63)	273.58 (265.06, 282.10)	272.11 (262.66, 281.56)
ξ	0.002 (-0.042, 0.051)	0.024 (-0.023, 0.071)	-0.040 (-0.074, 0.006)	0.037 (-0.029, 0.10)
Model S4				
$\alpha_{\sigma}$	0.13 (0.063, 0.19)	0.16 (0.078, 0.24)	0.053 (0.026, 0.79)	0.14 (0.074, 0.14)
$\beta_{\sigma}$	0.060 (0.049, 0.071)	0.080 (0.070, 0.090)	0.024 (0.020, 0.027)	0.053 (0.043, 0.061)
$\phi\sigma$	272.20 (263.31, 281.09)	266.10 (259.65, 272.56)	278.97 (270.32, 287.61)	271.37 (262.91, 281.32)
$\gamma_{\sigma}^{(x)}$	0.002 (-0.005, 0.009)	-0.0051 (-0.040, 0.030)	0.0048 (-0.00048, 0.010)	-0.012 (-0.026, 0.0011)
ξ	0.0049 (-0.044, 0.054)	0.024 (-0.023, 0.071)	-0.037 (-0.071, -0.003)	0.033 (-0.033, 0.099)
Model S4 with prior on shape				
ξ	0.019 (-0.021, 0.059)	0.014 (-0.024, 0.052)	-0.027 (-0.058, 0.004)	0.008 (-0.039, 0.054)
Model R1				
$\beta_{\lambda}$	0.0087 (0.0004, 0.017)	0.022 (0.015, 0.030)	0.024 (0.018, 0.030)	0.022 (0.014,0.032)
$\phi_{\lambda}$	155.66 (100.74, 210.59)	175.16 (155.86, 194.46)	209.50 (195.48, 223.52)	184.31 (160.94,207.69)
$\alpha_{\lambda}^{(x)}$	-0.13 (-0.18, -0.079)	-0.055 (-0.101, 0.009)	-0.063 (-0.099, 0.108)	-0.32 (-0.37,-0.26)
$\beta_{\lambda}^{(x)}$	0.14 (0.068, 0.21)	-0.016 (-0.084, 0.051)	0.061 (0.014, 0.108)	0.23 (0.14, 0.31)
$ \begin{array}{c} \alpha_{\lambda}^{(x)} \\ \beta_{\lambda}^{(x)} \\ \phi_{\lambda}^{(x)} \end{array} $	311.86 (281.78, 341.94)	359.95 (265.77, 454.15)	352.38 (299.32, 405.44)	278.54 (260.44, 293.63)

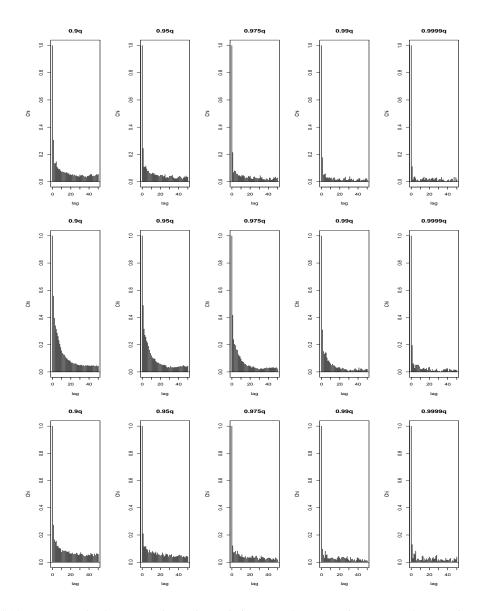


FIG S14. Estimates of  $\chi$  for Lowestoft, Newlyn and Sheerness (from top to bottom row) for exceedances of the 0.9, 0.95, 0.975, 0.99 and 0.999 quantiles (from left to right column) at various lags.

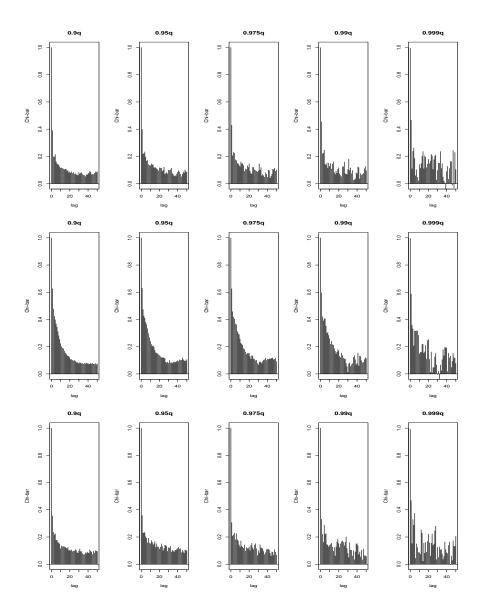


FIG S15. Estimates of  $\bar{\chi}$  for Lowestoft, Newlyn and Sheerness (from top to bottom row) for exceedances of the 0.9, 0.95, 0.975, 0.99 and 0.999 quantiles (from left to right column) at various lags.

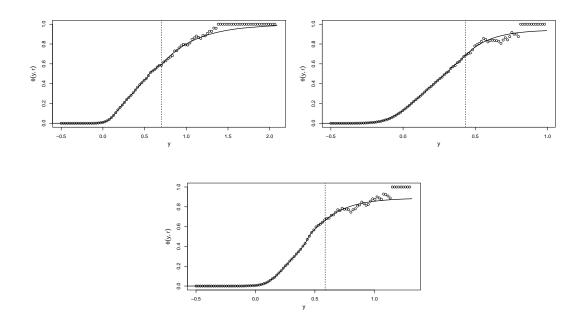


FIG S16. Estimates of the subasymptotic extremal index  $\theta(y, r)$  for different skew surge levels using the runs estimate (black) and our model estimate (red) (expression (4.16) of the main paper) at Lowestoft (top left), Newlyn (top right) and Sheerness (bottom). Run lengths are chosen as 10, 2 and 10, respectively. The threshold v is chosen as the 0.99 skew surge quantile for all sites (black dashed line).

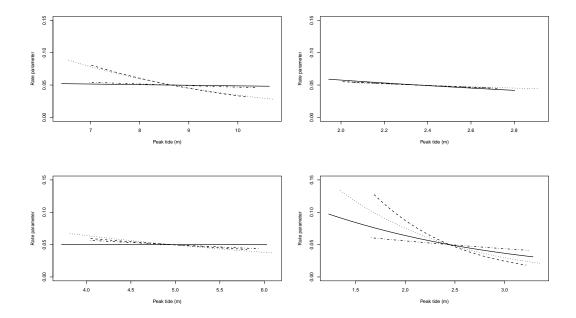


FIG S17. Estimated exceedance probability  $\lambda_{d,x}$  (expression (4.11)) per month with respect to peak tide x and day in month  $d_j$ , averaged over day at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right). Trends for March (solid), June (dashed), September (dotted) and December (dot-dashed) are shown here.

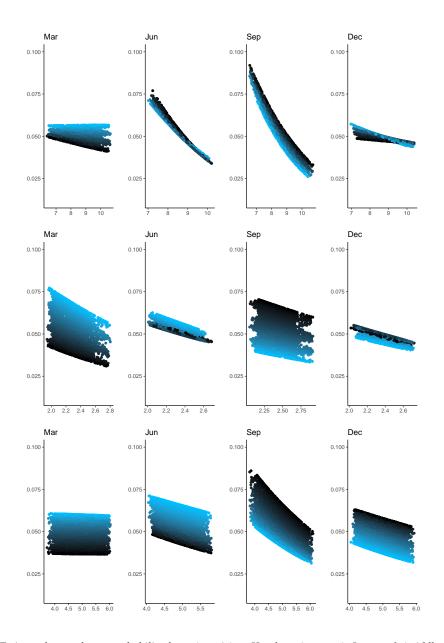


FIG S18. Estimated exceedance probability  $\lambda_{d,x}$  (y-axis) at Heysham (top row), Lowestoft (middle row) and Newlyn (bottom row), in March, June, September and December (from left to right by column) with respect to x being peak tide (x-axis in metres) and  $d_j$  being day in month at Sheerness. Darker (lighter) points represent days later (earlier) in the month.

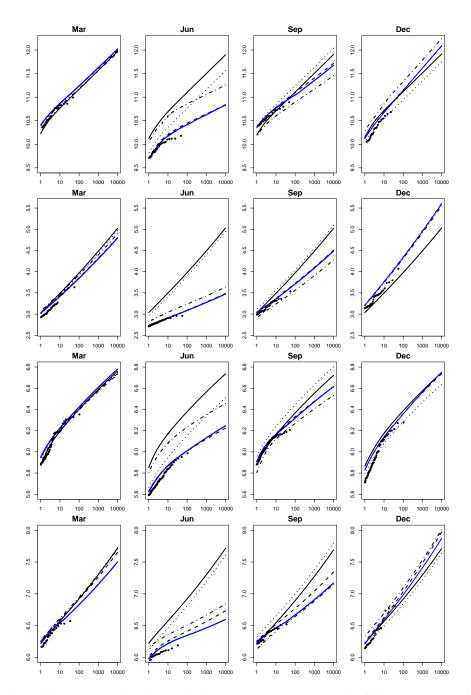


FIG S19. Monthly maxima sea level return level estimates (y-axis in metres) for different return periods (x-axis in years) at Heysham, Lowestoft, Newlyn and Sheerness (from top to bottom row) in March, June, September and December (from left to right by column) estimated using the baseline (black solid), seasonal surge (dot-dashed), seasonal tide (dotted), full seasonal (dashed) and interaction (blue solid) models. Empirical estimates are shown by black points.

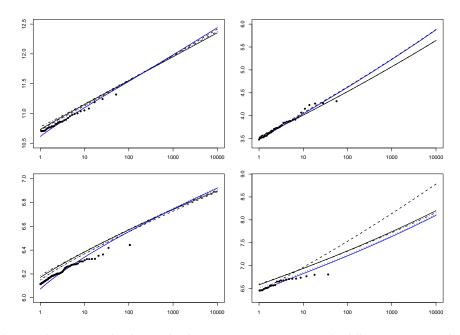


FIG S20. Annual maxima sea level return level estimates (y-axis in metres) for different return periods (x-axis in years) at Heysham (top left), Lowestoft (top right), Sheerness (bottom left) and Newlyn (bottom right), estimated using the

textitcurrent (dot-dashed), baseline (black solid), full seasonal (dashed), interaction (dotted) and temporal dependence (blue solid) methods. Empirical estimates are shown by black points.

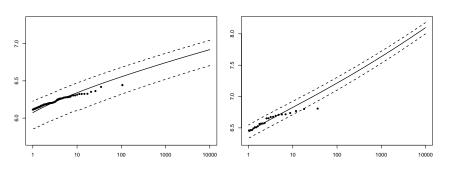


FIG S21. Return level estimates (y-axis in metres) for different return periods (x-axis in years) from the final model (temporal dependence) (solid), with the maximum and minimum year-specific return level estimates (dashed) and empirical estimates (black) at Newlyn (left) and Sheerness (right).

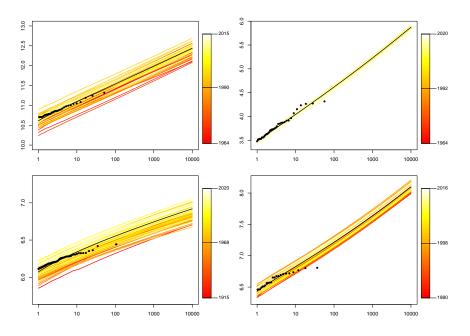


FIG S22. Annual maxima sea level return level estimates (y-axis in metres) for different return periods (x-axis in years) at Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right) estimated using the final model (temporal dependence) shown by the solid black line, with year-specific return levels shown by the red-yellow lines. Empirical estimates are shown by black points.

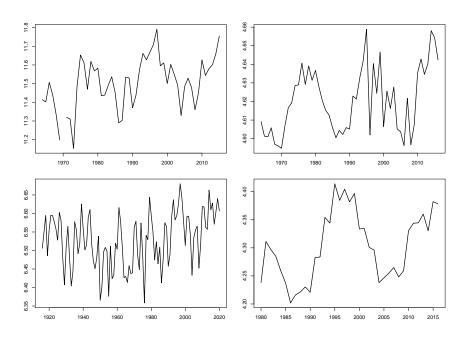


FIG S23. 100 year return level estimates (y-axis in meters) for Heysham (top left), Lowestoft (top right), Newlyn (bottom left) and Sheerness (bottom right) estimated using the year-specific final model over the years of observation (x-axis).

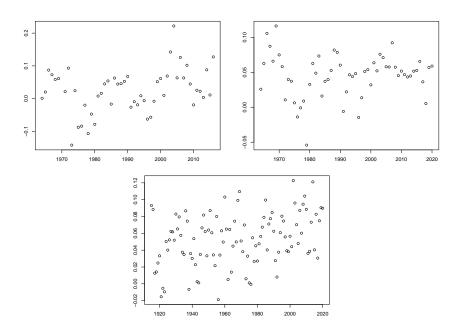


FIG S24. Annual mean skew surges (y-axis in meters) against year (x-axis) at Heysham (top left), Lowestoft (top right) and Newlyn (bottom).

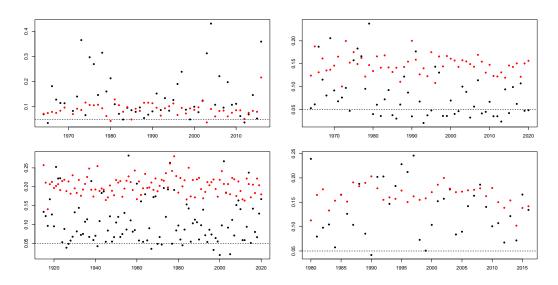


FIG S25. p values from the Kolmogorov-Smirnov test for uniformity (y-axis) of yearly (x-axis) samples of the transformed skew surge observations through the final skew surge distribution function (expression 4.14) at Heysham (top right), Lowestoft (top left), Newlyn (bottom left) and Sheerness (bottom right) before (black) and after (red) we remove the annual mean trends of the skew surge series. The 5% significance level is shown by the black dashed line.

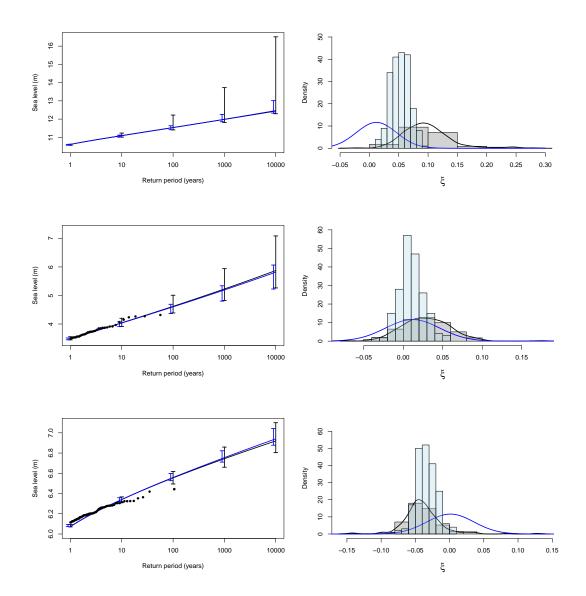


FIG S26. 95% bootstrap confidence intervals on the final (temporal dependence) return level estimates at Heysham (top row), Lowestoft (middle row) and Newlyn (bottom row) before (orange) and after (blue) adding a prior distribution to the shape parameter (left). Empirical estimates are shown by black points. Histograms of the shape parameter estimates and their densities (right) for these two models in their corresponding colours.

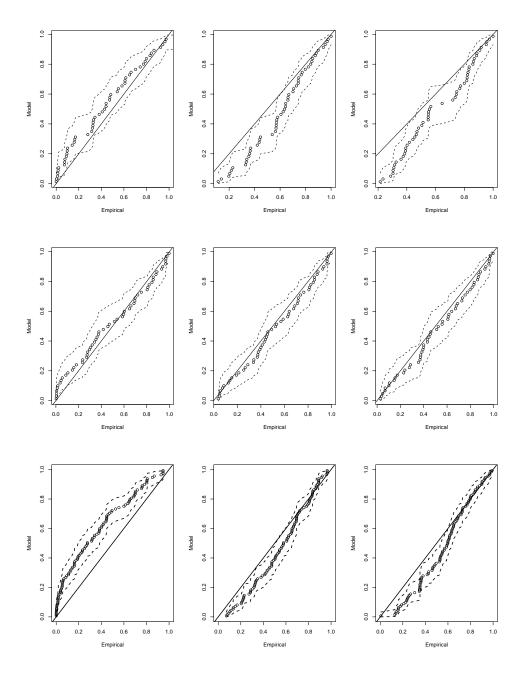


FIG S27. PP plots for the transformed annual maximum sea levels to a uniform scale at Heysham (top row), Lowestoft (middle row) and Newlyn (bottom row). These are transformed using the baseline (left column) and final (temporal dependence) (central column) distribution function for the annual maxima, as well as the year specific final model (right column). The black line shows the line of equality, y = x.

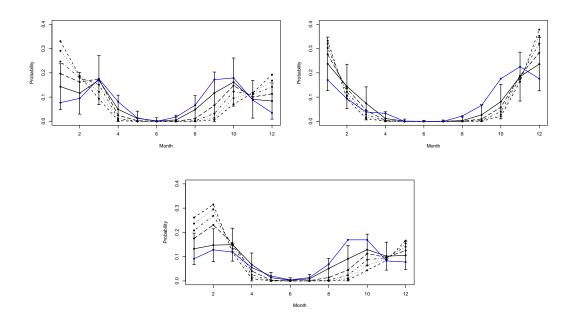


FIG S28. Estimates of  $\hat{P}_M(j;z)$  for months j = 1 - 12 at Heysham (top left), Lowestoft (top right) and Newlyn (bottom), for p = 1 (black solid), 0.1 (long-dashed), 0.01 (dot-dashed), 0.001 (dotted), and 0.0001 (dashed). The blue line is the empirical estimate  $\tilde{P}_M(j;z_1)$ , 95% confidence intervals are for  $\hat{P}_M(j;z_1)$ .

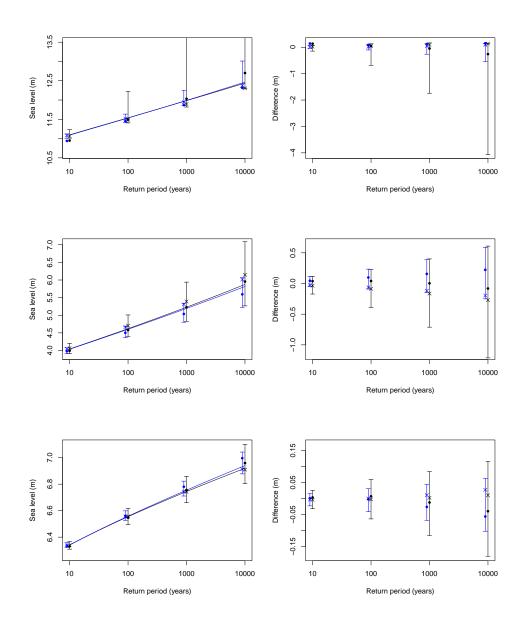


FIG S29. Return level estimates from the final (temporal dependence) model with 95% confidence intervals before (black) and after (blue) adding a prior distribution to the shape parameter at Heysham (top row), Lowestoft (middle row) and Newlyn (bottom row), where the 0.95 quantile is used for the skew surge distribution (left column); estimates are also shown at the 10, 100, 1000 and 10,000 year levels when the 0.9 (crosses) and 0.99 (dots) quantiles are used in the skew surge model, both with (blue) and without (black) the prior distribution on the shape parameter of the GPD. The corresponding difference between the estimates from the original approach (0.95 quantile) with the 0.9 (crosses) and 0.99 (dots) quantiles, compared with confidence intervals (right column).

## REFERENCES

- COLES, S. G., HEFFERNAN, J. and TAWN, J. A. (1999). Dependence Measures for Extreme Value Analyses. *Extremes* **2** 339–365.
- D'ARCY, E. and TAWN, J. A. (2021). Discussion of "Towards using state-of-the-art climate models to help constrain estimates of unprecedented UK storm surges" (by Howard, T. and Williams, S. D. P.). *Natural Hazards and Earth System Sciences Discussions* **21** 3693–3712.
- HOWARD, T. and WILLIAMS, S. D. P. (2021). Towards using state-of-the-art climate models to help constrain estimates of unprecedented UK storm surges. *Natural Hazards and Earth System Sciences* **21** 3693–3712.
- POLITIS, D. N. and ROMANO, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association* **89** 1303–1313.