Discussion of
Towards using state-of-the-art climate models to help constrain estimates of unprecendented UK storm surges

by Tom Howard and Simon David Paul Williams

Discussants: Eleanor D'Arcy and Jonathan Tawn

August 27, 2021

1 Overview

This paper proposes using climate model simulations to aid with still water level return level estimation and address problems that arise with the statistical approach when tide gauges have short record lengths. The model they present is the HadGEM3-GC3-MM to generate a dataset of 483-year present-day storm surges at sites on the UK tide gauge network. They compare the skew surge simulations to using only observations when fitting extreme value models by estimating parameters. The spatial distribution of the parameters for each dataset are generally well correlated. However, there is a negative bias in the simulation approach in the shape parameter estimate of the generalised Pareto distribution (GPD); the authors discuss this in detail and suspect it is due to a pitfall in the shelf sea model (CS3).

The paper also investigates the interaction between skew surge and peak tide at Sheerness. They study the effect that changes in timing between atmospheric forcings and tide has on skew surge. Additionally, they review the independence assumption of skew surge and peak tide used in the JPM. Using their model simulations, they show that extreme skew surges are more likely to occur on neap tide - this agrees with the results of Williams et al. (2016) (supplementary material).

2 General Comments

The paper is well written, in the proceeding sections we have discussed parts of the paper that pose interesting areas for future research and noted some technical corrections. The results are well presented and the figures well explained, giving a clear justification for the proposed model. The appendix provides strong support for ideas mentioned in the main paper. The authors have recognised potential downfalls with different aspects of the model and presented some initial investigation into these (for example, the discussion of why the shape parameter is more negative for simulations on pg. 15/16).

3 Specific Comments

3.1 Comparison of Model and Observed Data

Figure B1 shows some major departures between the model and observed data across the distribution of skew surges but particularly in the tails. In no sites is the model giving as high quantiles as the observed data. However, these departures have a systematic feature which is consistent over spatial regions, e.g., south-west and north-west UK. This suggests that it should be possible to account for these departures through a smooth spatial function which maps the differences in quantiles between the observations and model data. With this adjustment it is possible that the currently identified under-estimation may be corrected before making the tail based GEV/GPD comparisons you draw.

3.2 Shape Parameters with Similar Spatial Patterns

One of your exciting findings is that the spatial pattern of the shape parameter estimates is similar for the CFB2018 estimates from observed data and your estimates from the model, but with a systematic bias between them. This suggests using an alternative to the CFB2018 approach by explicitly exploiting this finding. Let $\xi_{obs}(x)$ and $\xi_{model}(x)$ be the shape parameters for the observed and model data respectively, for all gauged sites x. What you are saying is that you believe in the spatial variation of $\xi_{model}(x)$ but not its mean value. So you believe that $\xi_{obs}(x) = \xi_{model}(x) + \xi_{bias}$, where ξ_{bias} is a fixed constant that does not vary over x. Your estimates indicate that $\xi_{bias} > 0$. Fixing estimates of $\xi_{model}(x)$ for all x and fitting the function of $\xi_{obs}(x)$ over x now means only one parameter, ξ_{bias}, needs estimating. This could lead to substantial reductions in the uncertainty of $\xi_{obs}(x)$ estimates over x.

3.3 Penalised Likelihood

The shape parameter estimates in Figure 2 (d) based on the 483 years of model data show some site-to-site variations which are more pronounced than the broader smooth variations across coastlines. This suggests that they would also benefit from the penalty-based approach used in CFB2018 work.

At a few points you state that the prior/penalty is subjective and that this is a disadvantage. Yet you also point that the smooth pattern of the shape parameter estimates this gives for the observed data agrees well with the similar unpenalised estimates using model data. You say this is a really positive feature of the model data, we would also take this as supporting the value of the process of creating penalised estimates from the observed data. The penalty is giving something meaningful, so the effect of the claimed "subjectivity" is positive for the observed data analysis. The prior that was selected in the CFB2018 work was not subjective in the traditional sense of a subjective prior in Bayesian methods. It was actually a data-based prior which corresponds to an empirical Bayesian prior, using all the information that separately estimated shape parameters for UK skew surge provide. The effect of this was simply to move shape parameter estimates more towards the UK average, with the larger changes coming for sites with shorter record lengths.

3.4 Investigation of Interaction between Tide and Skew Surge

We are really pleased to see the use of numerical models to explore systematically the widely claimed property that skew surges are independent of their associated peak tidal values. The work in Section 5 where you explore the timing effects of skew surge events relative to tides is very illuminating. It adds greatly to the existing empirical evidence for this feature at Sheerness. It would be interesting to have your opinions about what interactions are expected elsewhere in the UK given that the CFB2018 estimates over UK sites all assume independence. Since this analysis is based on simulations from HadGEM3-GC3-MM, this presents a physically-based justification for dependence between skew surge and peak tide at Sheerness. It suggests future research is required to investigate this. It may be that interaction occurs to some extent everywhere, but at a practical level it is not important apart from some locations.

In current work we have been investigating this feature empirically at a limited number of sites. Here we report some of the methods and findings using data from the tide gauges at Sheerness and Heysham (located close to Workington which you analyse). We study data from Heysham in 1964-2016, of which 17.5% is missing, and 1980-2016 at Sheerness, where 9.1% is missing.

We define extreme skew surges as exceedances of the 0.95 quantile at each site. Figure 1 shows scatter plots of extreme skew surges against their associated ranked peak tide; these are ranked so that 1 corresponds to the smallest observation and n is the largest, where n is the total number of observations. If the two components were independent, we expect extreme skew surges to be uniformly distributed over ranks. We test this using a Kolmogorov-Smirnov test for uniformity; the p value at Heysham is 0.0224 whilst at Sheerness it is 1.40×10^{-7}. Clearly the p value at Sheerness is much smaller, and provides statistical evidence that extreme skew surges are not independent of peak tide. However, at Heysham, there is not sufficient evidence to reject the claim of independence at the 0.01 significance level. This is clear in Figure 1, where more extreme skew surges occur on lower tides at Sheerness - this agrees with your findings and those of Williams et al. (2016).

We investigate this further by looking at skew surge and peak tide dependence on a month-by-month basis, using ideas from Williams et al. (2016). Figure 2 compares the distribution of peak tides associated

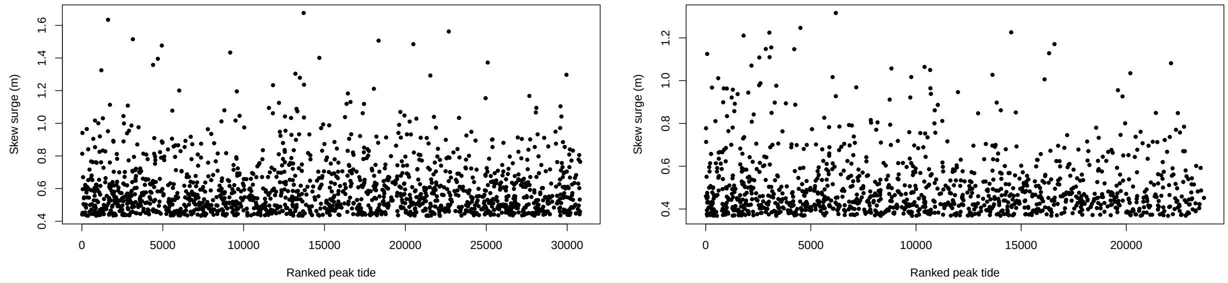

Figure 1: Extreme skew surge observations against ranked peak tide at Heysham (left) and Sheerness (right).

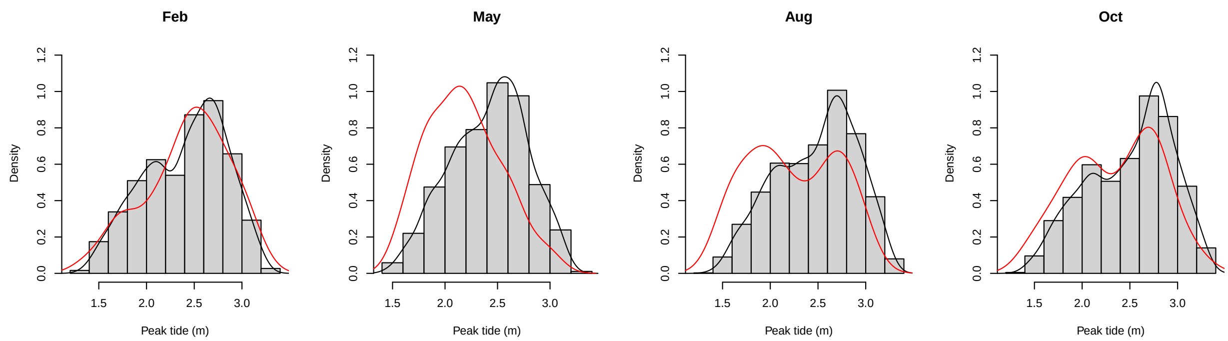

Figure 2: Monthly distributions of peak tides at Sheerness in February, May, August and October. The probability density function of all peak tides (black) and peak tides associated with extreme skew surge (red) are interpolated onto each distribution.

with all skew surges and the distribution of peak tides associated with extreme skew surges for February, May, August and October. If peak tide and skew surge are independent, these two distributions should be the same, up to sampling variation. We estimate the probability density function (pdf) using a Gaussian kernel density estimate, and use an Anderson-Darling test to check if peak tides come from the same distribution as peak tides associated with extreme skew surges. Figure 2 highlights how the dependence between skew surge and peak tide is changing with the time of year; the distribution of peak tides and the peak tides associated with extreme skew surge are most different in May and least different in February. In May, the mode of the distribution of peaks 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 for every month tell us there is insufficient evidence to reject the null hypothesis that peak tides and the peak tides associated with extreme skew surge come from the same distribution in February, March, September, November and December. In these months, we conclude it is reasonable to assume skew surge and peak tide are independent. In the remaining months, we find sufficient evidence to suggest the two components are dependent. We believe this poses a really interesting area for further research.

In D'Arcy et al. (2021) we fit a GPD (see Coles (2001) for details) to extreme skew surges that accounts for seasonal variations, since they are more extreme in the winter. Our results show this is an improvement on a standard GPD fit, where skew surges are assumed to be independent of peak tide. Here, we investigate adding a covariate of peak tide into our skew surge model to account for the dependence found at Sheerness. D'Arcy et al. (2021) defines extreme skew surges as exceedances of the monthly 0.95 quantile. Non-stationarity is accounted for in the scale parameter of the GPD through a daily covariate $d = 1, \ldots, 365$:

$$\sigma_d = a + b \sin\left(\frac{2\pi}{365}(d - \phi)\right) \tag{3.1}$$

for $a, b, \phi \in \mathbb{R}$ parameters to be estimated. Note that the shape parameter of the GPD is fixed across months. To account for the dependence between skew surge and tide, we now also consider the following parameterisation on the scale parameter:

$$\sigma_{d,t} = a + b \sin\left(\frac{2\pi}{365}(d - \phi)\right) + ct \tag{3.2}$$

where $c \in \mathbb{R}$ is another parameter to be estimated, and t the associated peak tide observation. We fit a GPD with both (3.1) and (3.2) formulations to extreme skew surges at Heysham and Sheerness. The Akaike information criteria (AIC), frequently used for model selection, suggests that formulation (3.1) gives a better model fit at Heysham, i.e., an independence conclusion is supported by this analysis. Whereas, AIC suggests that the parametrisation in equation (3.2) yields a better fit at Sheerness. By ordering peak tide observations from smallest to largest and calculating $\sigma_{d,t}$ at its winter peak ($d = 365$) for each value, we observe a 15% reduction in the scale parameter as tides increase at Sheerness, which corresponds to an equal percentage reduction in the skew surge quantiles for excesses of the December skew surge threshold. We also compare the model fits at each site using a Likelihood Ratio Test. At Heysham the p value is 0.657 which provides insufficient evidence to reject the null hypothesis, that is the simpler model (in equation (3.1)) is sufficient. Whereas, at Sheerness we get a much smaller p value of 0.059, which provides statistically significant evidence at the 0.1 significance level to reject the null hypothesis and conclude that the more complex model is required. This highlights the importance of accounting for skew surge and peak tide dependence at sites where the independence assumption is not justified.

3.5 Ungauged Sites

We feel you do not do full justice to your developments given the focus is on comparisons made at gauged sites with long and trustworthy records. The real value in modelled data is the ability to give estimates at other sites. This could be the focus of a natural follow up paper.

3.6 Other Comments

1. In the abstract, you say "results suggest an event of this magnitude has an expected frequency of about 1 in 500 years at [Sheerness]" when referring to the North Sea floods of 1953. In the paper, the only result to show this is presented in Section 6.2: "the fact that the 483-year surge-only simulation produces more than one event of comparable magnitude to the simulated 1953 event suggest the return period of the 1953 atmospheric forcing is less than 483 years." If this is the justification, it would be better to more formally quantify this using your statistical models.

2. In the Introduction (line 51) you list assumptions required to fit extreme value models as 'events are effectively random and statistically independent of each other.' But what about events being identically distributed (or stationary), it is clear that tide and storm surge (or skew surge) are not stationary as both process exhibit seasonality. Instead of "effectively random" it would be more mathematically correct to say "stochastic." Extreme value methods also handle dependence, so "independent of each other" is not formally required.

3. On line 84 climate change is discussed. Tide gauge observations will exhibit an approximately linear mean trend due to sea level rise, it is unclear whether this has been removed before comparison with the simulations in Section 4. It is likely that removing this change will not change the results significantly, but it is important to remove this non-stationary effect before fitting extreme value models.

4. Figure C1 shows the reduction in uncertainty on shape parameter when record lengths are increased in the tide gauge network, but it would be nice to see this for more record lengths that are equally spaced (say 10 to 500 years in increments of 10 years) to really highlight this - with measures of uncertainty as in Figure C1 (d).

4 Technical Corrections

1. Table 1 gives a list of useful acronyms and symbols, it would be help to include what 'GC3' and 'MM' stand for in 'HadGEM3-GC3-MM' as it is not mentioned in text.

2. Equation (2) has a surplus close bracket.

3. Equation (6), should this derivative be evaluated at $L = \log(u)$ rather than $y = u$ since y is the return level (as in equation (1))?

4. Figures 2 (c) and (d) would benefit from having 95% confidence intervals on shape parameters using the model data.

5. Line 283: We think the wrong figure is referenced here.

6. Line 319: "They used ....time-series." This has nothing to do with how CFB2018 estimate the shape parameters and should be cut. It only links to the derivation of SWL return levels.

7. Line 349: You hint that the reason for the disagreement between the methods could be due the short observed data series. But Figure B1 shows major departures between the observed data and the model data in the body of the distributions to a level that could in no way be due to short observed series and must be that the model data does not reproduce well the observed data. Linked to this, in discussing Figure 1 you say the agreement at both sites is "excellent". With the amount of data at Sheerness it is clear that there are major disagreements that cannot be accounted by sampling variations.

8. Line 421: You suddenly mention a "kernel" to spread the duration of events but fail to provide any information. A little detail would be helpful here.

9. Figure E2: A description of the different points would be useful.

References

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

D'Arcy, E., Tawn, J., Joly-Laugel, A., and Sifnioti, D. E. (2021). Accounting for seasonality in extreme sea level estimation *(in preparation)*.

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(12):6410–6417.