Response to review by Andreas Sterl of *Towards using state-of-the-art climate models to help constrain estimates of unprecedented UK storm surges.*
*by T. Howard and S. Williams, submitted 2021*
*(henceforth HW21)*

This response: Tom Howard Aug 2021

Thank you for your review and for raising a very interesting question. I have amended the draft to address the minor points, and included a discussion of your main point further below. I will attempt to upload the amended draft to accompany this response.

Regarding the minor points first (quotes from the review are shown in red).

The approach has been introduced nearly 20 years ago by Van Den Brink et al. (2004): Improving 10^4 year surge level estimates using data of the ECMWF seasonal prediction system. Geophys. Res. Lett., 31, L17210, doi: 10.1029/2004GL020610. This fact should be acknowledged in the present paper.

I apologise for failing to cite this very relevant paper in the first draft. I have now acknowledged this contribution.

Discussing Fig. 2c you refer to the "spread of the shape parameters diagnosed [...] from the simulation", but I cannot see any spread in the simulation-derived shape parameters. Spread is only depicted for the CFB estimates. Please clarify.

Thank you for pointing this out. I was referring to the spread associated with the spatial variations, rather than the uncertainty. I have clarified this in the paper as follows:
"The spread (i.e. the size of the spatial variations) of the shape parameters…"

You introduce a kernel. How did you obtain it? A short explanation of the procedure would be helpful. I am also not sure about the purpose of the kernel. Is it to extend the length of an episode, or its magnitude?

I have added the following sentences explaining the choice and purpose of the kernel:
"The kernel was designed to represent the important features of the RACMO-driven surge-only simulation, i.e., the approximate duration and shape of the time series plot. The purpose of convolution with the kernel is to identify those events which correlate well (in terms of their time series plot) with the RACMO-driven simulation, in other words, events which not only produce a large surge, but are also of comparable duration to the RACMO-driven simulation. The kernel was not used to modify events, but simply to identify significant ones."

---

Turning now to your main point regarding whether to fix the shape parameter at zero. This is prompted by the two papers by van den Brink and Können:

van den Brink and Können (2011): Estimating 10000-year return values from short time series. Int. J. Climatol., 31:115-126, doi: 10.1002/joc.2047 , and their related 2008 paper (both referred to here as vdB&K).

Thank you for reminding me of vdB&K's very interesting approach. I spent a long time (back in 2017) studying their 2008 paper, but it is only now, in testing the method on our own data in order to complete this response, that I am beginning to understand it.

HW21 was rooted in the methodology of CFB2018, which forms the current UK guidance on sea level extremes:
https://assets.publishing.service.gov.uk/media/603652cce90e0740b7caac9d/Coastal_flood_boundary_conditions_for_the_UK_2018_update_-_technical_report.pdf

(full reference in HW21). CFB2018 was developed in consultation with Professor Jonathan Tawn:

https://www.maths.lancs.ac.uk/~tawn/

Jonathan advised me not to fix the shape parameter as this gives false confidence intervals. The position is laid out in the textbook by Stuart Coles (full reference in HW21). Coles provides an illustration of a case similar to the case in section 4.5 of vdB&K (2011), and Coles discusses it as follows (Coles page 64)<Quote>:

> **Reduction of uncertainty is desirable, so that if the Gumbel model could be trusted its inferences would be preferred. But can the model be trusted? The extremal types theorem provides support for modelling block maxima with the GEV family of which the Gumbel family is a subset. The data** [in Coles's example 3.4.1, which, like vdB&K(2011) example 4.5, has an estimated shape parameter close to zero] **suggest that a Gumbel model is plausible, but this does not imply that other models are not. Indeed, the maximum likelihood estimate within the GEV family is not in the Gumbel family (although, in the sense that the estimated shape parameter is close to zero, it is "close"). There is no common agreement about this issue, but the safest option is to accept there is uncertainty about the value of the shape parameter --- and hence whether the Gumbel model is correct or not --- and to prefer the inference based on the GEV model. The larger measures of uncertainty generated by the GEV model then provide a more realistic quantification of genuine uncertainties involved in model extrapolation.**

<End Quote> (The red highlighting is mine)

On the other hand, we know that unconstrained GEV fits to short record lengths can give implausible shape parameters. One way to fix this is to put a prior on the shape parameter (see for example Martins and Stedinger 2000, full reference in HW21). This was the approach used in CFB2018 (again under advice from Jonathan Tawn).

In HW21 we find that the spatial variations in the shape parameter diagnosed from the tide gauges are also seen in the shape parameter as diagnosed from the simulation. We have argued that this finding supports the credibility of the spatial variations. Thus, it would be inconsistent to adopt an approach of fixing the shape parameter at zero within HW21, but for completeness I have tested a Gumbel fit to the annual maxima following the method of vdB&K and I show the results here.

I have used our data to make plots analogous to vdB&K 2008 Figure 3, i.e., their $\Delta\hat{X}_n$ vs. Gumbel variate plot, comparing Gumbel fits to the annual maxima with the fits used in our paper and in CFB2018. Fig. 1 is the plot for the observations:

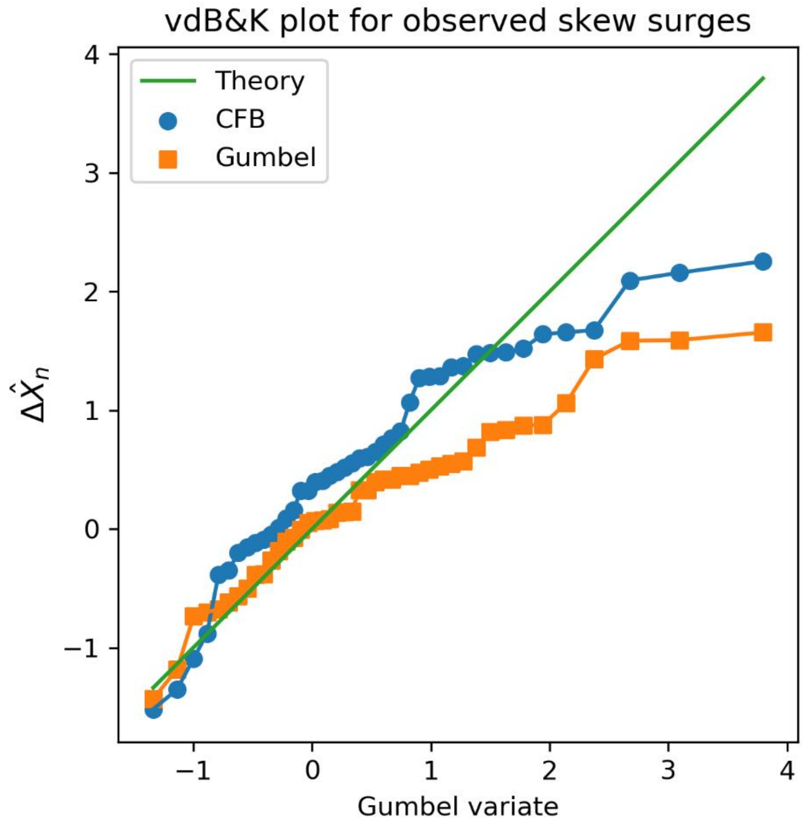


Fig. 1. Each data point represents a UK tide gauge. The points labelled "CFB" show $\Delta\hat{X}_n$ where $\hat{F}$ is determined by the CFB2018 method (constrained GPD fit to peaks over a threshold, as described in HW21), and the points labelled "Gumbel" show $\Delta\hat{X}_n$ where $\hat{F}$ is determined by a Gumbel fit to the annual maxima.

I can't see strong evidence here that the Gumbel fit is better overall. I was a bit concerned that I had not been rigorous about checking for independence of the events shown. This would not be straightforward to do with my current software setup, but a simple first fix is to miss out closely-neighbouring tide gauges. The following plots show results from every second tide gauge (Separation=2), every third tide gauge (Separation=3), etc. (Separation=1 is just a duplicate of the above plot).

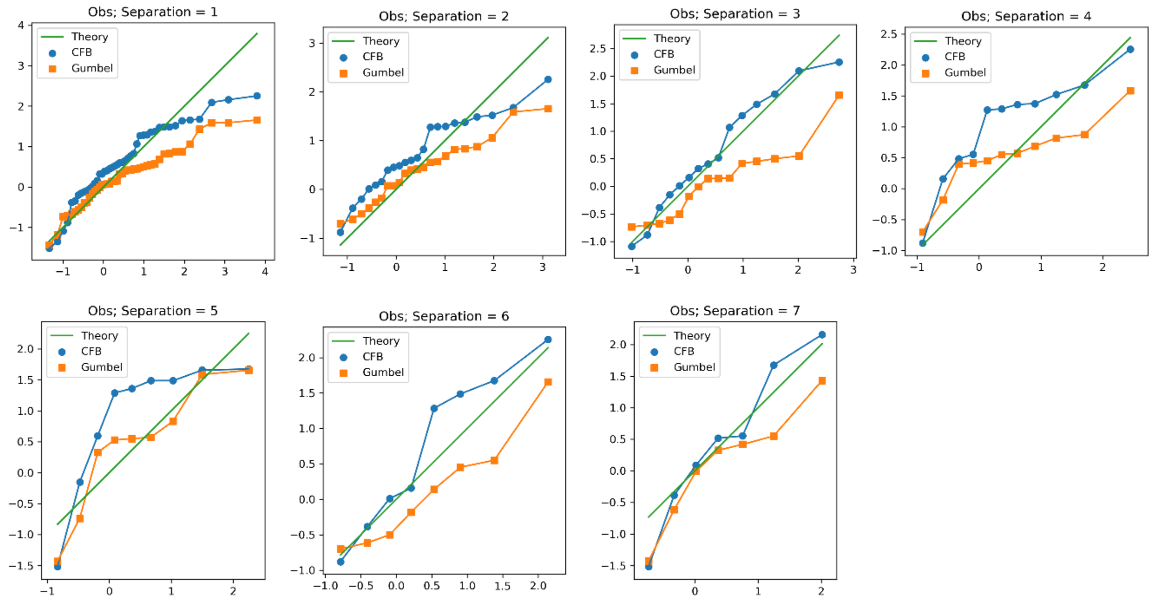


Fig. 2. See main text.

Again, there does not appear to be much support for preferring the Gumbel fit.

(continued…)

I followed the same procedure for the simulated skew surges to make Figs 3 and 4.

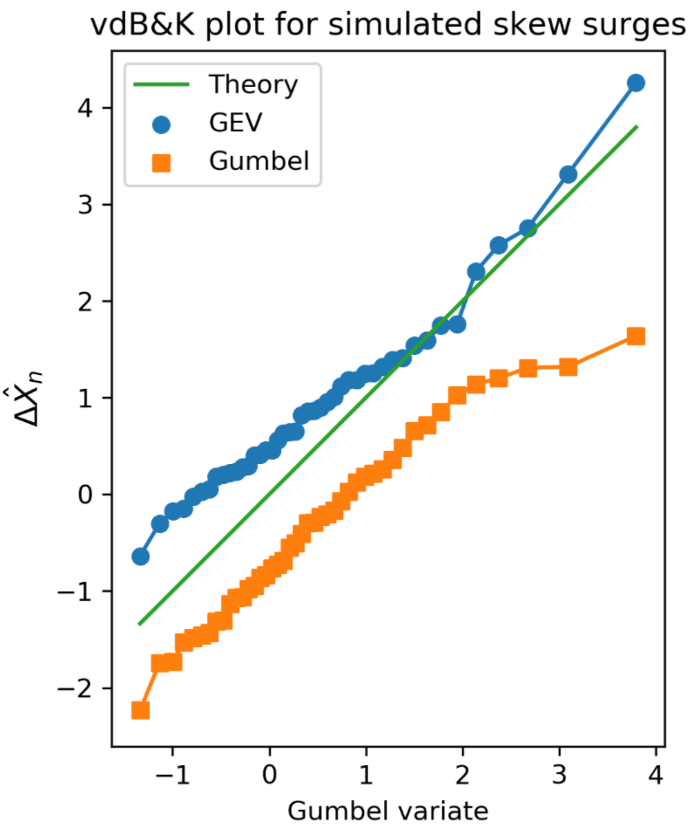


Fig. 3. As Fig. 1, but for the simulated skew surges. Each data point represents a UK tide gauge. The points labelled "GEV" show $\Delta\hat{X}_n$ where $\hat{F}$ is determined by a GEV fit to the annual maxima, and the points labelled "Gumbel" show $\Delta\hat{X}_n$ where $\hat{F}$ is determined by a Gumbel fit to the annual maxima.

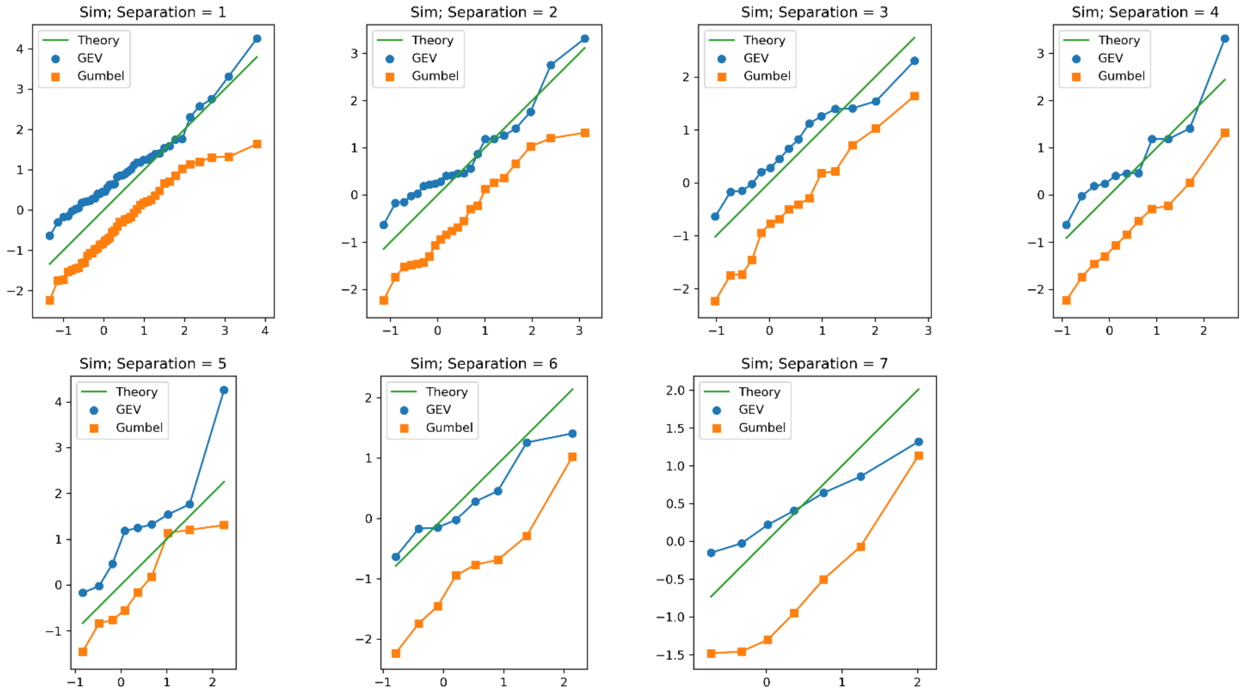


Fig. 4. As Fig. 2, but for the simulated skew surges.

Again, this does not seem to me to give support for preferring the Gumbel fit. Even if it did, I would not think it defensible to fix the shape parameter at a single value (e.g., zero). Looked at from the point of view of applying a prior to the shape parameter, fixing it at a single value seems to be equivalent to asserting that we are sure that no other value is plausible, and we are sure that the shape parameter does not vary by location. I cannot support either of those assertions.

Note to self: internal reference for figures: fig_KK