	Comments on "Technical Note: Tail behaviour of the statistical distribution of extreme storm surges" by Tom Howard (OSD)	
	This is a short technical note which attempts to make 3 points: (a) the shape parameter of extreme sea level curves at most UK sites is not zero (and usually negative) and so any parameterisation of the extreme level curve should accommodate its curvature, (b) in spite of that, an assumption of zero shape for a Gumbel distribution is reasonable for Hunter's allowance calculation, and (c) the shape parameters derived from short records are imprecise. These things were known already (or suspected anyway) but it does no harm to restate them in the same place.	Thank you for your thorough review. I have responded to each point as tabulated below, in most cases modifying the manuscript.
	I have no objections to the note's publication if the small things below can be attended to. The text is clearly written although the document itself is a little rough (hence some of the trivial comments below).	
1	line 6 - mean sea level rise here and mean-sea-level rise at line 40 (I said these were trivial comments but they suggest some lack of attention)	I have attempted to improve the consistency, hyphenating “sea level” only when it is a compound adjective.
2	15 - you don't present evidence that the shape parameter varies around the UK coast. You have a scatter plot in Figure 1 that shows there are clearly different values at different places but, unless you know where the UK place names refer to, you have no insight on how the shape varies around the actual coastline. A map is needed or at least a couple of sentences to say how it varies.	Have included a text description of the large-scale pattern and a more clear indication of where the pattern (and a map) are shown in HW21.
3	49 - not incompatible ==> compatible!	Modified.
4	50-52 - these lines would be better following on at line 39	Yes - done.
5	65 - I know this is a short technical note and there are many details in	The surge extrema are not “from the same years” at all. The simulation is a “free-running” climate model

	HW21, but it does no harm to give some essential minimum information. For example, presumably the surge extrema used for Figure 1 are from exactly the same years as the tide gauge extrema, or comparisons are not exact. So say so. Also say what the minimum record length of tide gauge record is employed.	control run - all of the atmospheric data is simulated. That is why I find it so interesting/surprising that the shape parameters correlate well with those diagnosed from the tide gauges. I have modified the text to make this clearer. I'm not sure that the length of the tide gauge records used is relevant to this particular result? I think it is more relevant to Fig.5, where it is mentioned.
6	After Figure 1 there should be a sentence to tell the reader that most of the UK shape parameters are negative. And that this observation is not new. For example, see Figure 9 of Marcos and Woodworth (JGR, 2017) which shows consistent negative shape parameters for both North Atlantic coasts. And Wahl et al. (2017) claim that 85% of records worldwide have negative shape parameters. As for the UK, I am sure the negative shapes will have been pointed out in older papers by Blackman, Horsburgh, Tawn etc. (although I have not checked which)	Thank you for reminding me of these two very relevant papers, both of which are cited elsewhere in the manuscript. But, regarding the shape parameters at UK tide gauges as discussed here, are they known to be negative? The recent CFB2018 report (with project team including Jon Tawn and Kevin Horsburgh) diagnosed a mixture of positive and negative shape parameters. I am not aware of a more recent publication specific to the UK which overturns that result.
7	69 - give a reference. For example chapter 7 of Pugh and Woodworth (2014). As well as the physics of wind stress etc., there is a general point that there is only so much water in the ocean, so one would imagine any extreme level curve to turn down at some point.	Reference: done. Regarding the curve turning down: I agree. A similar argument would apply to any physical process, I think. I contacted Simon Brown (a Met Office climate extremes expert) about this. He said: "*Ultimately all meteorological processes have a physical limit and so should have bounded distributions. However, we don't have perfect samples and we don't fit perfect statistical models. Both, I would argue, can lead to unbounded models seeming to be the best fit to data. As you say there seems to be no consensus on what to do about this nor even much of a discussion about the problem.* *My go to example is with wind extremes. An observational record will consist of samples of the warm conveyor belt, the cold conveyor belt and if you are lucky a sting jet. Each of these processes have different extreme behaviour but we fit a simple EV distribution. If there is just one sting jet sample, which is way above the others, the resultant EV fit will have a positive shape parameter even if all the three sub-processes have negative shape parameters. The fitted EV model does not reflect the underlying*

		mixture of processes and so when it is fitted we get a distribution that looks unphysical. *What happens is that generally there is some pragmatic choice that seems to fit the main objective on the analysis. I suppose being a good scientist one would look at the results with a free shape fit and compare with a constrained shape fit and discuss the merits of each. This can be quite subtle – with my wind example it is not clear that forcing the shape parameter to be <=0 will give you a more physical fit if the error of not including the mixture aspect of the sample is not fixed. It could easily be worse.*"
8	73 .. in [shape parameter] (reference needed. HW21 again?)	I have added a "pers. comm." type citation and a footnote. A colleague (Simon Brown again) pointed me to the scale-shape compensation issue last year. He can no longer remember the details of the paper where he read it. I had thought to maybe show analytically that the partial cross-derivative of the loglikelihood is usually negative, but it proved to be beyond my mathematical scope. So, I have settled for a footnote explaining that it can readily be confirmed numerically, and how to do so.
9	I don't understand why in practice you know there is spatial correlation in the location parameters. That can only be in model runs where the datum at every point is MSL. But if you are using real tide gauge data the location parameters will depend on the datums used at each site. (I hope you see what I mean.)	Yes, I do see what you mean. I have modified the caption of Fig 1 to clarify that we are dealing with the shape parameter of skew surge here: hence not dependent on the datum.
10	79 - why 'vector'? It seems an odd word to use here.	Now "data"
11	82 - say 'For the model data at each tide gauge site'. To make it clear you are using just the short model data sets here and not the 484 year set mentioned later.	I am using the 484 year data, There is no short model data set. Have modified the m/s: "The simulation takes atmospheric data from a free-running 484-year climate model control run…"
12	90 - from any other site. (?)	Actually from any site. Could in theory be the same one very occasionally, but since they are shuffled at random many times this does not materially affect the result.
13	104 - the long run of 484 years. And this is for the 44 (?) tide gauge sites	Yes. Have adjusted caption.
14	112 - .. not Gumbel-distributed as was known previously.	Changed to "re-emphasising…"
15	Figure 2 (a) and (b) should have (m) on each axis	Thank you.

16	line 4 of caption - .. the site of the 44 (?) tide gauges on ..	Yes. Have adjusted caption.
17	section 2.4 - I got the idea of this section although you have to read it a few times. It would help to fully explain things. For example, what does 'standard-uniform' (line 121) mean?	Have added a note explaining “standard uniform”, and reworded parts of the section. Have also tried to further emphasise that this is just a short informal description, for ease of reference, of the procedure that VdBK08 used. Have directed the reader to VdBK08 for full details.
18	126 - ... from a given site conforms to a precise GEV distribution.	Reworded.
19	130 - .. depends on the three GEV parameters.	Reworded.
20	133-134 - standard-uniform (as above)	Thank you.
21	151 - an average (?) optimum .. They preferred	They use the word “optimal”. I believe they performed their test with a fixed shape parameter across all sites, varying the shape parameter to see which value gave the “best” plot (closest, in some sense, to x=y on their plot).
22	155 - simulation as represented in Figure 1 (presumably)	Yes. Reworded.
23	Figure 3 - I don't understand why there are 4 plots here. Shouldn't there be 8? You have tide gauge data (shown here) and line 155 says you use model data also, so you need another 4 for the model data?	Figure 3 shows simulation-based results in top two panels and observation-based results (i.e. tide-gauge-based) in bottom two panels.
24	title caption should be VdBK and not VdB&K to be consistent with the text But I would remove that anyway and just have QQ plot to be consistent with PP plot on the right. Preumably the dots are ordered so as to be monotonic. Define in the caption the delta symbol on y-axis for QQ (differences at the outliers). Finally I don't understand why you call them 'theory'.	Have changed to VdBK08 to be consistent with the text. Yes, the dots are ordered so as to be monotonic (i.e. they are ranked). This is the usual approach in this kind of plot, as I understand it. The X-values of the PP plot are the expected values for a ranked set (of size m) of standard-uniform data: i/(m+1) (i: 1 to m). The Y-values are what we have for our sample set, given the fit. The QQ plot in this case is the same data, with both X- and Y-values transformed by -log(-log(.)) as described in the text. I have used the word “theory” for consistency with VdBK08: it shows the relationship you would see if every element of your sample took its expected value.
25	caption line 1 - this should be reworded as you say above for both that QQ and PP derive from VdBK	They are not both derived from VdBK08. See the line in the text which says “Instead of the plot described above, they...” VdBK08 show only what I call the QQ plot. They do not show the PP plot.
26	caption line 2 - at the 44 (?) sites of UK ...	Yes. Done.
27	172 - in general zero, as known already (refs).	Added.
28	Figure 4 - I thought you were using 44 sites (see caption figure 5). This	Amended: now 44.

	should be mentioned at the places in the text I pointed out above. However here in Figure 4 there are 46 locations given.	
29	188 - why did CFB2018 take +0.0119 as its prior shape parameter when all the evidence from previous publications and your Figure 1 has it negative? And at line 192 why did you use a prior of +0.0119 ?	Good question! This is discussed by Jon Tawn and Eleanor D'Arcy in section 3.3 ("Penalised Likelihood ") of their comments on HW21, which are in the public domain at https://nhess.copernicus.org/preprints/nhess-2021-184/nhess-2021-184-RC3-supplement.pdf The relevant paragraph says: *"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."* Considering the scatter plot (Figure 1 of the manuscript under review here), as you say an overall negativity is seen in the shape parameters. But that is only in the shape parameters diagnosed **from the simulation** (Y-axis), **and not** from the tide gauges (X-axis). Even without the penalty function, the unconstrained shape parameter estimates based on the tide gauge data have a **positive** mean (+0.0119). Their estimates are shown in figure E.1 of the CFB 2018 report https://www.gov.uk/government/publications/coastal-flood-boundary-conditions-for-uk-mainland-and-islands-design-sea-levels Their prior is formed from the distribution of their unconstrained estimates. I used a prior with a mean of +0.0119 to be consistent with the CFB2018 approach. Incidentally, the negative bias in the simulation-based shape parameters compared to the observation-based remains unexplained. It is discussed at some length in HW21. I have checked that it is not simply a sign issue (unfortunately authors and statistics packages use differing conventions regarding what is meant by a positive shape parameter).

30	Could you explain Figure 5 a bit better? If the data really has a non-zero shape, and the choice of prior is reasonable, then wouldn't you expect the right-hand side to be tighter than the left for Gumbels?	I have added the following phrase: "*This figure shows that, even though we believe that the data represent distributions with non-zero shape parameters, the likely inaccuracies associated with unconstrained shape parameters are more serious than the likely inaccuracies associated with the over-constraint of insisting the shape parameters be zero (Gumbel fitting). In other words, we see the importance of choosing an appropriate prior constraint on the shape parameter, for typical real-world record lengths*." As the existing text already explains, we are choosing to take the constrained fit as "truth" for the purpose of this experiment.
31	198 - .. is negative in common with most UK sites (Figure 1) and worldwide (Marcos and Woodworth, 2017; Wahl et al., 2017).	Again, I'm not so confident about asserting that here. As mentioned above in response to your comment number 29, the negative shape parameters in my Figure 1 are only for the simulation. See also Fig E.1 of the CFB2018 publication: they diagnose a fairly balanced distribution of +ve and -ve
32	Figure 6 left - the Hinkley plot is described in some detail in Batstone et al. (2013)	Thanks. I have added a citation. Although, incidentally, I regard Batstone et al. as having been largely superseded by CFB2018.
33	Acknowledgements - define BEIS and Defra	Done.
34	221 - Climate, 231 - Research Letters, 235 - Communications	Amended. Thanks.