General Comments

The motivation of the paper to examine the spatial and temporal coherence of flood risks is well justified and the authors' approach to examining this topic is efficient and effective. However, the justification for using raw climate data for examining floods resulting from extreme climates is not yet satisfactorily justified, but I believe an additional validation step would allow the authors to retain the approach they have already adopted. I find the manuscript to be overall well written and figures are mostly well presented (some improvements are needed for Figs 3, 4, 7, and 8 and Figure 5 needs to be revised). A number of corrections and clarifications are necessary to ensure that: consistent terminology is used; the language considers the global readership of this journal; descriptions are precise and objective; figures are easily interpreted; and referencing is accurate.

**Major comment**

I concur with the authors that the use of bias correcting in an attempt to capture rainfall extremes relevant to floods is difficult and introduces considerable uncertainty and the justification here is to avoid the introduced uncertainty and instead examine the differences in floods resulting from modelled historical and future climates. However, the justifications currently provided are inadequate and at times illogical.

The authors state that "due to the focus on [sic] the present work on extremes rather than the whole regime in general, bias correction is not applied here."
I'm afraid this justification is unsatisfactory. I'm certain the authors are knowledgeable of the fact that regional climate models are aimed at resolving large scale spatial and temporal variability, namely, what is referred to here as "the whole regime in general". A study that was aimed at large scale changes would therefore be well justified in using the raw RCM outputs. In contrast, extreme precipitation events are typically driven by synoptic scale events, which would justify the use of bias-corrected projections. Furthermore, flooding is sensitive to the spatial and temporal distribution of storms at even finer scales and have the added complexity of local factors that influence the flood response other than climate. The focus on extreme events is therefore a justification for introducing uncertainty to address the increased complexity of flood responses. Lines 65-66 therefore needs to be removed.

In their response letter, the authors make the case that they have kept the biases constant by comparing modelled historical and modelled projected floods and by capitalising on the dynamic nature of the GCMs. (however, it then follows that the same justification would not rule out the comparison of bias-corrected data – here too, the biases would be consistent between modelled historical and projected climates. Furthermore, bias corrected rainfalls are not constrained to the observation based datasets, as it appears to be suggested in the authors' response– both statistical and dynamical downscaling approaches allow for estimates outside the range of the observed records.)

This justification is sufficient and acceptable conditional upon the authors adopting the following suggestions, which should be easy to implement:

1. Repeat the analysis using observational data over the same period as the historical modelled data to quantify the degree to which the spatial and temporal coherence

of flooding is approximated and represented in the baseline case. This would then provide evidence of the fidelity of the baseline to which the *relative* changes in the number of widespread events could be estimated using the approach that has already been presented. Currently, the presentation of results dependent only on modelled climate data makes it easy for a reader to suppose that the results reflect similarities in catchment characteristics combined with the lack of spatial specificity in the modelled climate date. Having an observation-based analysis would provide context. Presenting these observed results in the SI would be adequate

2. Present the findings as relative changes in the number and spatial scale of the widespread events (more details of this in reference to figure 3 below)

**Specific comments (line numbers reference manuscript 2)**

Acronyms are often used before they are defined – the ones I noticed were: NRFA, RCM (this shouldn't be assumed knowledge since GCM is previously defined), PoE. There may be others.

It would be helpful to be clarify the terminology used to describe seasons - this is primarily for your readers in the southern hemisphere, please don't neglect us. Please consider including "boreal" when seasons are referenced. Alternatively, include the months in brackets after each season. It may seem superfluous for a northern hemisphere native, but it makes interpreting the results so much easier for those in the south.

The time slices are introduced as a "baseline" and "future", but the former is later referred to as "present". Please keep the terminology consistent, particularly since "present" is inaccurate when referring to a time period covering 1980-2010.

L 90: The term "percentiles" in reference to floods is used to describe distributions. The term that should be used here is *frequencies*, not *percentiles*.

L95: similarly, the notation $Q_x$ is widely used in hydrology to denote the x percentile of flow rather than a flood frequency. Please keep the terminology you have chosen consistent by referring to the 1 in 5 and 1 in 10 year probability event as POT 0.2 and POT 0.1 respectively. Alternatively, use the term Average Recurrence Interval of 5/10 years (ARI5 and ARI10).

L99: This is ambiguous as absolute values of flow are in fact being used as thresholds, and these thresholds are dependent on the distribution of the data: the definition of the thresholds are just location dependent. I suggest phrasing this as:
Note that the thresholds are not based on universally applied fixed values of flow magnitude but are instead dependent on thresholds defined by empirical flood frequencies.

L150: I had to read this several times and I'm still unclear. The text implies that $\bar{\chi}$ applies to large flows and conversely that $\chi$ applies to flows of all magnitudes (which I do not believe is the intended message). I suggest the following (but I'm unsure whether I've interpreted the text correctly. Please revise as necessary)
χ describes the level of asymptotic dependence; if $\chi \neq 0$ then the variables are asymptotically dependent. A value of $\chi = 0$ represents asymptotic independence. Asymptotically independence is also represented by a value of $\bar{\chi} \neq 1$. A value of $\bar{\chi} = 1$ represents flows that are dependent but nor asymptotic.

L185: please justify why these four events are selected. Why “four of” and why not “the four largest”? Are they selected to demonstrate that most events are spatially contiguous? (As an aside to the major comment above, is the spatial coherence an attribute derived from the modelled climate data, or is the pattern present in observed data?)

Figure 2: If I’m understanding this correctly, this figure shows the degree of spatial inundation with the colour scale showing the equivalent severity of the flood for events identified using the POT2 threshold, equivalent to a return period of 0.5. If this is the case, why are more frequent events with return periods of 0.2 to 0.5 shown? Perhaps they are not, but the light yellow scale is difficult to distinguish between greater or less than 0.5. I suspect it is simply a case of the legend needing to be updated to show grey between 0.2 and 0.5. Also, the text size of the labels on this and the next figure are rather disproportionately large. Please reduce the text size.

Figure 3: Comparing the changes between the baseline and future is not easy with the way this figure is presented – there are a lot of vertical bars in different ensemble members to compare and lining up the number of events for different return periods is challenging, while the colour selection implies that A and B are showing a different variable to C and D. I strongly suggest combining the information from A and C, and B and D by showing these results as the difference between time slices. If there is a very good reason to not do this, please at the very least change the colours to paired colours: e.g. light blue for A, blue for C, light green for B and green for D, so that light colours correspond to the baseline, darker colours to the future, blues for area, and greens for return period.

L 203-2007: The choice of using the word “may” makes the sentence sound speculative. These sentences could be revised to reflect the certainty of the results. Suggest the following or similar:
However, the increase in widespread events is confined to the boreal autumn (SON) and winter (DJF) with a decrease in events between March and August. The decrease in future boreal spring and summer events could be due to overall projections of drier summers (Murphy et al., 2019), or could result from a spatial contraction of summer floods in the future, which have historically resulted from short-duration, high intensity storms.

L 212: It’s unclear what the term “methods” is in reference to. Is it the method of climate modelling or is this meant to mean differences in flood responses to different storm types?

Figure 4: The left figure is superfluous. Please remove this.

L227-230: Describing the changes in reference to changes in return periods is unconventional and is probably due to the way the figure axes are configured (see comment on Figure 5). Please change this description to one that describes the shift in event duration with respect to frequency.

Figure 5: is there actually a heavier tail in the future in SON? (L223)
Also, the return period *must* be plotted on the x axis, as the duration of the event is dependent on the return period not the other way around. This isn’t just a formatting choice – the reader is forced to attempt to transpose figure 5 as the convention in referring

to tails is the distribution along the horizontal (with a few exceptions). In addition, there isn't a good reason to present Figures 5 and 6 in a way that is disjointed. The reason given in response to the previous reviewer of duration having a smaller range than return period (and likewise in Figure 6 the return period having a smaller range than area) is disputable: one could easily represent the return period in log10 years and have an effective scale of 0-4 (just making a point – I'm not suggesting that return periods be presented in this way).

Figure 7: labels a, b, c, d are missing from Figure 7 (b is referenced in the text). The caption needs to describe what is in each of the four figures (i.e. different time slices and different measures). This figure is also inconsistent with previous figures that have shown the baseline time slice on the top row and the future on the bottom (this can be fixed and the legends could simply be placed horizontally under the respective columns). Using a different color scheme for each metric would aid in interpreting the different scale of results.

Figure 8: as per Figure 7 regarding layout and caption. Only one legend is needed.

**Minor comments**
Grammar and style
L66: "of the present work"
L175: "shown" instead of "show"
L231-233: missing "there"; "matched" not "matches"; "dynamics" is not an appropriate adjective here. I think what is meant is: but there is variability amongst the ensemble members in the relationship between event duration and frequency.

L237: "is not surprising": please replace this with objective language e.g. statistically plausible

Referencing: Following on from a previous reviewer's comment, please check the accuracy of all references. One example is that of Tawn et al. 2018: it is referenced as 2019 in the text and the list of authors in the reference list is incorrect. The dates for Towe are also reference incorrectly in text. There may be others, so it may be prudent to check the referencing system.