Authors note:

The authors would like to extend their thanks to the reviewers for taking the time to work through our manuscript and provide us with detailed and insightful reviews. The major criticism levelled by both was our choice not to explore uncertainty within the vulnerability component. This was a valid criticism and thanks to some further collaboration with Willis, we have now been able to incorporate their latest vulnerability uncertainty method into our study. This method therefore represents current industry practice and we believe it adds valuable context to the paper. The results may surprise some readers as, despite the vulnerability model imparting considerable uncertainty onto the exceedance probability curves, the choice of driving precipitation data remains the dominant source of uncertainty in this study.

Response to comments from anonymous referee #1:

1) We have now incorporated vulnerability uncertainty into the paper. The relevant literature review has been extended slightly in section 1.3, the method is described in 2.2.3, the results are presented in 3.4 and the discussion and conclusions have been updated. A new figure (figure 9) has been added.

2) Descriptions added as requested.

3) Unfortunately, due to the proprietary nature of the data to Willis, we are unable to print the exact values used in the stage damage functions. We believe it to be in the best interests of the paper and readership to present results generated using functions derived from real data and used by industry, rather to use fabricated values that can be printed but may not bear resemblance to those used in practice. We have provided a full description of the method in section 2.2.3 to enable others to test the method their own data.

4) Catchment average time series were used for several reasons, primarily because they are commensurate with the hydrological model input and because generating 500,000 year hourly spatial rainfall fields was beyond the scope of this paper. Orographic effects are important, particularly in the Dodder catchment, and the catchment average series does account for this by applying corrections to each cell according to the precipitation-altitude gradient during the inverse distance gridding process as described in section 2.1.1. We have further justified our choice in section 2.0

5) The station names have been added as requested.

6) We do briefly investigate the limited impact of moving from a 10 m hydraulic model to a 50 m model in the second half of section 2.2.2. Changing to a simpler form of hydraulic model (e.g. 1D or volume spreading) would undoubtedly alter the results again, although it is not possible to quantify the effect of this without further study. However, as one of the fundamental controls on the behaviour of a hydraulic model is the volume of water added to the domain, we would expect any credible hydraulic model to be sensitive to the fairly extreme differences in upstream boundary conditions produced by altering the driving data.

7) The reviewer notes that there are many influencing factors at play here, and it is difficult to isolate why it is that this particular ten year training series generates a curve that fits so well to then 1986

event, especially as the 1986 event does not fall within the training series. Ultimately the aim of this section is to demonstrate the sensitivity of the EP curve to the training record length, and we argue that this point is adequately made by the figure and discussion as it stands.

8) We now mention the alternative of evaluating flood frequency analyses in the context of a stochastic model in the discussion. However, we have also chosen to leave the original content in the conclusion as we feel it provides a useful starting point for future studies at the end of the paper.

9) Plot is now in colour.

Technical corrections:

1) Changed.

2) Corrected.

3) Industry term; definition added (loss before application of deductibles and/or reinsurance)

4) Corrected.

5} Corrected.

6) Figure labels changed as suggested.

Response to comments from anonymous referee #2:

1) Vulnerability model uncertainty now included. See response to question 1 from referee #1.

2) The reviewer is correct to identify the fact that we are almost certainly underestimating the uncertainty associated with the stochastic model by using the maximum likelihood GPD fits and not running multiple realisations with fits sampled from the 95% confidence intervals of the GPD parameters. Ultimately the reason this was not done was computational cost: calculating the uncertainty in losses derived from sampling of the GPD parameter spaces for both intensities and durations would have required many runs of the entire cascade and this was not something we were able to undertake. It is, however, something that should have been noted more clearly in the text and we have amended the discussion in recognition of this.

3) The following paragraph has been added to the discussion:

Spatial scales are an important consideration in the context of this study. The catchments modelled in this study are relatively small, and it is reasonable to suggest that the relatively coarse reanalysis and satellite products might perform better for major rivers where fluvial floods are driven by rainfall accumulations over longer time periods and large spatial areas. Some of their inherent traits, such as tendency for the reanalysis product to persistently 'drizzle' while underestimating storm rainfall accumulations, will negatively impact their applicability across most catchment scales although the severity of the effects may reduce as catchment sizes increase. However, it is wrong to assume that the dominant driver of flood risk is large events on major rivers. The majority of insurance losses resulting from the 2013 Central European Floods were termed 'off-floodplain' – that is to say they occurred either as a result of surface water (pluvial) flooding or as a result of fluvial flooding in small catchments (Willis, personal communication). This suggests that even when considering large events, the ability to produce realistic hazard footprints in small catchments remains critical and thus for practitioners concerned about such events, the findings of this paper remain relevant.

Technical corrections from annotated supplement:

p31) Although affiliations are the same, we thought it best to separate the details (including contact) of the corresponding author from others at the same institution

p33) Added reference to 2010 Indus flood; also added 2013 Central European Floods

p35) The inclusion of the vulnerability module analysis now covers these points; the literature review and discussion have been expanded and include new references.

P44) Threshold range added (threshold was allowed to vary depending on number of observations in each class)

P45) Generation method identical to Cameron et al (1999) – added to text.

P45) Observational data description added to text (Dodder catchment average from rain gauges).

P46) Clarified to state we are correcting between generation site and catchment mean records.

p55) Corrected

p56) End date added – it is different to the end date used in the preceding section as this analysis used only one type of data and therefore did not need to overlap with the other types.

P58) Vulnerability now considered. Flood defences comment added to discussion.

P58) Point added to discussion

P59) Reference added

P60) Reference to flood defence uncertainty added to discussion

P72) Table amended

P78) Figures 5, 7 and 9 now gridded with consistent y-axis

P79) Graph now in colour