We are very grateful to Anonymous Referee #2 for their comments and suggestions on our manuscript. We have prepared the following responses to the points raised and indicate our proposed changes to the manuscript.

Anonymous Referee #2

Received and published: 9 October 2017

General comments

The research work presented in the manuscript develops a new methodology to estimate fine-scale rainfall extremes. Although there has been a substantial amount of work done, by many authors over the years, on stochastic point process models for rainfall, most of the models proposed tend to underestimate the rainfall extremes at fine-scales. Estimation or reproduction of extreme rainfall at hourly and sub-hourly scales is a well-known problem. In this context, this paper attempts to address this problem by using a censored approach to model rainfall extremes. This is in a way similar to the Excess Over Threshold (EOT) method commonly used in extreme value modelling, but here a stochastic mechanistic model is used along with this idea. Application of this novel idea of censured modelling approach is illustrated in the estimation of fine-scale rainfall extremes from two different regions to provide an improved representation of extremes.

The paper gives an excellent coverage of the history of work carried out in this area to convey the rationale for the need to study or explore alternative methods for fine-scale extremes.

The success of this new approach, of course depends heavily on the choice of the censor level and, hence, emphasis was placed on finding appropriate value of the censor for the application. If the estimation of extreme rainfall is the main objective of the study then using this censored approach is certainly a useful tool and worthwhile addition to the existing methods.

One drawback in the proposed approach might be the amount of fine-tuning required to get the best set of potential estimates for the extreme rainfall with respect to model, its parameterisation, censor, statistics used in fitting as well as aggregation levels. This level of tweaking or fine-tuning might prove to be a lot to generate sufficient interest amongst practitioners. The rationale behind the need to make these choices, however, has been explained in the manuscript though.

Specific comments.

Line 257: Would be useful to give a reason for the assumption of rain cells starting at the storm origin.

In the Bartlett-Lewis rectangular pulse (BL) models, it is assumed that rain cells start at the storm origin largely for mathematical convenience. Because the BL cluster mechanism is defined by the interval between successive cells, a starting point is required. Therefore, it is convenient to assume that rain cells start at the storm origin. In contrast, the Neyman-Scott rectangular pulse (NS) cluster process is defined by the temporal distance between storm and cell origins and typically assumes that rain cells do not start at the storm origin. Again, this is largely for mathematical convenience.

In the case of the BL models, the assumption of rain cells starting at the storm origin prevents the simulation of empty storms which can occur if the first rain cell starts after the end of the storm. This issue does not arise in NS models because the number of cells per storm is a model parameter to be fitted, therefore a minimum value of one can be specified thus preventing the simulation of empty storms.

Line 316-319: Can appreciate the reason given for the choice of fitting statistics used for model calibration, but the question now is that how do the parameter estimates compare when the same fitting stats are used for uncensored fitting? Has this been explored?

This is an interesting question but one that we haven't explored. We will look into this although we don't feel that it will change the analysis presented in this paper.

Line 358: Perhaps you need to explain what you meant by behavioural parameters for the readers.

This was also highlighted by Anonymous Referee #1 therefore the response below is the same as that provided to referee #1.

We have used the term "behavioural parameters" by analogy with Beven and Binley (1992). We have used the term to refer to well identified models. We have found that for well identified parameters with narrow 95% confidence intervals, simulation bands on the extreme value estimates are correspondingly narrow. As the parameters become less well identified, their 95% confidence intervals increase giving rise to extreme value estimates which deviate significantly from the observations, which in turn results in significant deviation of the simulation band upper limit. This effect is shown in Fig.11 resulting from the very large parameter uncertainty shown in Fig.12.

We will remove the reference to "behavioural parameterizations" in the context of this research and change all references to well identified parameters.

Line 358: 95% confidence intervals: unless you are using the standard errors of the estimates, I am not sure whether "confidence" interval is the appropriate terminology here. Simulation bands?

This was also highlighted by Anonymous Referee #1 therefore the response below is the same as that provided to referee #1. We agree with the suggestion and will change all occurrences in the manuscript.

There are in fact two issues here: if we were doing very long simulations with practically no random noise (so that another simulation would yield practically the same result), then we would have identified approximate confidence intervals. But with the shorter simulation length, both parameter uncertainty and the randomness of the model are combined in the spread we observe in the simulated statistics, so that 'simulation bands' is indeed a better descriptor.

Line 396: Not sure why your validation for Atherstone was based on 0.6mm censor which seem to contradict your statement on lines 375-380. Some insight/explanation would be useful to the readers.

We agree that it would have been more consistent to select 0.2 mm for validation of the Atherstone hourly censored model. We will change the selection in Table.2 to include 0.2 mm for the hourly resolution at Atherstone (as below). We will then revise the validation plots to suit.

Table 2 Censor selection for model validation.

	5 minutes	15 minutes	60 minutes
Bochum	0.5 mm	1.0 mm	1.0 mm
Atherstone	0.6 mm	0.6 mm	0.2 mm

Line 396: Table 2. Different censor for different sites is understandable. However, why do you need to use the same sensor at 3 different aggregation level for Atherstone while using different censors for the 3 levels of aggregation for Bochum?

The model is fitted separately for each temporal resolution which explains why we have different censors. Because of the effect of aggregation, we cannot use a model censored at one resolution to estimate rainfall extremes at a coarser one. Therefore, censored model parameters are scale dependent which is explained in section 3.

The censors given in Table.2 were chosen for validation. Our analyses show that there are a range of censors that could be applied giving improved estimation of extremes. In the case of the two sites investigated, the gauge resolution at Atherstone is much coarser than that at Bochum. We note on lines 504-7 (page 28) that that a censor of 0.5 mm for 15 minute rainfall at Atherstone gives very similar extreme value estimation to the selected 0.6 mm censor, implying that it may be sufficient at this site to limit the censor to the gauge resolution. At Bochum, the finer gauge resolution will capture rainfall amounts with greater accuracy than at Atherstone. Therefore, we expect that there is greater capacity for the Bochum models to give improved estimation of extremes with increasing censors hence the different censors selected in Table.2.

Fig 8: row 2. The nice seasonal pattern observed in the mean rainfall for Atherstone at 5 and 15 minutes has become less prominent or disappeared at 60 minutes. Can you comment on why? No observation or comment was made about this.

The plots in Fig.8 show the summary statistics for censored rainfall with different censors applied in each column. While the censors chosen for validation in Table.2 are the same (0.6 mm), their effect on model fitting is different because they are applied to each temporal scale. Hence, when we look at the mean monthly rainfall in validation, we are looking at the seasonal variation in the rainfall after censoring.

Without censoring, the seasonal variation in mean monthly rainfall will only change in magnitude between scales. For a constant censor between scales as shown in panels d, e and f, the seasonal variation in mean monthly rainfall will vary between scales because there is a higher proportion of low observations at short temporal scales removed by the censors. The greater prominence in seasonal variation shown in plots d and e indicates that the summer months (approx. Apr - Oct) are more prone to short intense bursts of rain, and the winter months longer periods of low rainfall intensity. This is consistent with there being more convective rainfall in the summer, and stratiform rainfall in the winter.

References

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process., 6, 279-298, 1992.