Precipitation extremes on multiple time scales – review

1. The focus on IDF curves as a characteristic of mechanistic models appears to be novel and of wide relevance to hydrological modelling, climate impact assessment and risk estimation. The focus on short duration (5 minute) extremes is also of particular relevance. I therefore think this research is suitable for this publication and would be of general interest to its readership.
2. The paper addresses three research questions which are clearly set out in the introduction. Each question is then addressed in turn in the discussion and conclusions. The questions are as follows:

   I. *"Is the BLRPM able to reproduce the intensity-duration relationship found in observations?"*

   The authors use a depth-dependent GEV distribution (dd-GEV) to estimate extremes across different durations – it is assumed that "across different durations" means "across different temporal scales". Optimisation of the dd-GEV parameters is performed using random sampling from a Latin-Hypercube which appears to be a new method for calibrating these models and is referred to as the depth-dependent GEV approach. This approach is used to construct IDF curves from the observations, and 1000 BLRPM realisations of the same length.

   Typically when we want to estimate extremes from a rainfall model we would sample annual maxima directly from long duration simulations without then using a second extreme value model such as GEV or GP. However, in this case it seems appropriate to apply the dd-GEV for two reasons: 1. to enable direct comparison with the IDF curves from observations, and 2. because the dd-GEV method uses extremes across different scales in fitting. That said, it is not clear from the methodology set out in 5.2 at what scales rainfall has been simulated; is it the same as those used in fitting (i.e., 1, 3, 12, and 24 hrs)? This could be made clearer by the authors.

   The authors note in Section 5.2 (lines 220-2) and in the conclusion (lines 292-3) that the BLRPM tends to under-estimate the extremes. The under-estimation of extremes by mechanistic rainfall models (both Bartlett-Lewis and Neyman-Scott variants), especially at fine temporal scales, is a known issue and the authors' findings are entirely consistent with this. The discussion would be greatly improved by drawing a broader interpretation of the results with comparison with other studies that show under-estimation of extremes by mechanistic models. In particular, is there something to be gained by estimating fine-scale extremes in this way?

   II. *"How are IDF curves affected by a singular extreme event which might not be reproducible with the BLRPM?"*

   BL model parameters are estimated using central moments of the rainfall data therefore it is very likely that this one single extreme will not have as much influence on the estimation of BL model parameters as it does on dd-GEV parameters from observations. And indeed, the authors show that the problem with January disappears when this event is taken out.

   The reader is however left with the impression that the implication is that this event is treated as suspicious information, i.e. that it is fine to take out this largest observation because it is so abnormally larger than any other observed hourly rainfall depth. I don't think that the authors meant this to be the case, but it should be clarified in the text that the section in which this largest value is taken out does not carry the implication that it is OK to take out the largest value because the event is in some sense 'abnormal'.

This issue brings us to an important problem with the authors' analyses: the data set of 13 years (then reduced to 12 years) is rather short to be doing extreme-value analysis (typically, a peak-over-threshold approach would normally be preferred for such a short dataset. Perhaps the authors' aim is to bring out the greater usefulness of making use of a rainfall model when the data set is not long enough, in which case this should be stated.

III. "Is the parametric extension of the GEV a valid approach to obtain IDF curves?"

Here the authors test the validity of the dd-GEV approach to estimating IDF curves by comparing IDF curves obtained from 50 realizations of 1000 years duration from the BL models with GEV estimates from the same simulations. There is an important underlying hypothesis here, namely that the BL model has now been adopted as an accurate representation of the distribution of rainfall (in particular extremes), but we know that this is not true from the problems identified in the analysis of BL's IDF curves. So it is important to qualify the scope of this third research question to make it clear that it is an analysis conditional upon a hypothesis that is only approximately true.

This issue also has a bearing upon the interpretation of the results. For instance, when they identify an under-estimation of 10 and 100 year hourly extremes in January and July, the authors conclude that this is due to poor representation of the dd-GEV IDF curves at these scales which is described as flattening. However, this result is also consistent with the known issue of mechanistic models under-estimating fine-scale (hourly and sub-hourly) extremes yet there is no discussion to this effect.

It is potentially encouraging that the estimation of fine-scale extremes with dd-GEV IDF curves from BL model simulations does not show the underestimation ordinarily obtained from mechanistic models, therefore the authors could explore this in their discussion.

A further issue potentially lies in the estimation of confidence intervals. There may be over-confidence in the extreme value estimates and IDF curves presented in Figure 8. Confidence intervals are estimated from 50 realisations from the BL models. However, GEV extreme value estimates from each realisation would have an associated credible interval which is not shown. It is possible that if this were, then there would be greater overlap in estimation by the two methods and the marginal differences would not be statistically significant.

3. The following are a list of general and specific comments on the analysis presented:
   I. On the whole the English is very good although there are a number of spelling and grammatical errors which should be addressed with a thorough proof-read by the authors prior to final submission. Specific examples can be found on lines 12, 13, 38, and 74 although there may be more. Some errors are flagged in this list.
   II. Line 19: 'affects'
   III. I suggest 'the object of much research' rather than 'the subject of research'
   IV. Line 28: 'in' instead of 'into'
   V. The authors state on lines 44-5 that "Due to the high degree of simplification of the precipitation process, the model is known to have difficulties in the extremes." It is not clear that this is why mechanistic models have a tendency to under-estimate short duration extremes, and many hypotheses have been put forward to address this exact

problem in the literature since their inception in the late 1980s. The authors make a valid point, but it could be enhanced with some references and broader discussion.

VI. Line 54: add ‘of’ before ‘the BLRPM’

VII. Line 57: ‘estimates’; also, ‘ability’ rather than ‘capability’

VIII. Line 68: the author’s name is ‘Le Cam’, not ‘Cam’. This also needs to be changed in the list of references

IX. On line 73 the authors highlight that they have chosen to use the original 5 parameter BL model. It would be good to give some justification for using this model variant over the randomised versions of the models, especially given that Kaczmarska, Isham & Onof, (2014) present a new randomised model with enhanced estimation of fine-scale (sub-hourly) extremes.

X. Add ‘a’ before ‘set’

XI. On line 87 the authors refer to a “time continuous step function”. Should this be “continuous-time”?

XII. On line 94 the authors comment that the Neyman-Scott model is “…motivated from observations of the distribution of galaxies in space”. This sounds fascinating although its relevance to rainfall simulation is perhaps somewhat removed. This statement should be reformulated with an appropriate reference.

XIII. The sentence on lines 97-9 requires further elaboration.

XIV. Figure 2:What is the meaning of the red? Is it the duration of the cell generating time (the time during which the storm is active)? And how does it contrast with the blue?

XV. In Section 2 the authors introduce the BL models and their chosen calibration strategy. On lines 108-10 they highlight their choice of weights with ω_i = 100 being applied to the first moment T_i (mean). In my experience the mean is usually very well represented by the BL model therefore it is unclear why the authors should want to up-weight this moment so much compared with the others. Given that the authors appear to be using a Generalised Method of Moments, it might be better to weight the summary statistics by the inverse of their observed variance (see http://rsfs.royalsocietypublishing.org/content/1/6/871.figures-only)

XVI. In lines 123-6 the authors discuss non-identifiability of model parameters although they don’t mention if they’ve checked this for their own calibrations. This could be done by estimating parameter uncertainty or producing profile objective functions on model parameters.

XVII. Line 151: The equation should read $IDF_{T_2}(d) > IDF_{T_1}(d)$

XVIII. Line 160: What is meant by ‘such a shape parameter ξ? Is the claim that ξ is also independent of the scale (duration d)? Is that true?

XIX. It’s not clear from the information provided exactly how equation 5 is derived. If this is derived in a previous publication this should be clearly stated and referenced.

XX. Line 164: It is not clear why there are two extra parameters. It would seem that you are placing several GEV fits (one for each scale) with 3 parameters each, by one fit with 4 parameters (?)

XXI. In Section 4 it would be useful to identify the gauge resolution. It would also be useful to provide a sentence justifying the choice of gauge location.

XXII. Line 177: ‘records’ instead of ‘registers’

XXIII. Line 178: explain why a data set with 13 years only was chosen

XXIV. Line 200: ‘for instance’ instead of ‘examplary’

XXV. In Section 5.2, line 210 the authors point the reader to a dotted line in Fig. 5 for IDF curves from observations. In the figure legend, the dotted line is for the IDF curves from BLRPM simulations. This needs to be corrected.

XXVI. In Section 5.2, line 227 the authors point the reader to February in their discussion of IDF curves in Fig. 5. I think the authors mean January as curves are only presented for January, April, July and October. The authors do the same on line 293 in the conclusions.

XXVII. Line 233: ‘to impact’ instead of ‘for’

XXVIII. Line 259: add ‘to’ before ‘resulting’

XXIX. Line 260: ‘simulations’

XXX. Line 282: ‘of the year’

XXXI. Line 289: ‘extent’

XXXII. In the conclusions on lines 314-7 the authors state that they do not find the BLRPM producing unrealistically high precipitation amounts as discussed for the random-η model by Verhoest et al., (2010). The generation of unrealistically high extremes by the modified (random-η) model is specific to that model and is therefore not relevant here as the authors have used the original 5 parameter model.

References

Kaczmarska, J., Isham, V. & Onof, C. (2014) Point process models for fine- resolution rainfall. *Hydrological Sciences Journal.* 59 (11), 1972-1991.

Verhoest, N. E., Vandenberghe, S., Cabus, P., Onof, C., Meca-Figueras, T. & Jameleddine, S. (2010) Are stochastic point rainfall models able to preserve extreme flood statistics? *Hydrological Processes.* 24 (23), 3439-3445.