Precipitation extremes on multiple time scales - Bartlett-Lewis Rectangular Pulse Model and Intensity-Duration-Frequency curves

–

Reply to Comments from Reviewer #3

Christoph Ritschel, Henning Rust, Uwe Ulbrich

July 26, 2017

We are very grateful to three anonymous reviewers for carefully reading and commenting thoroughly on our manuscript. We received highly valuable and constructive comments which very much helped to improve our work and led to new insights. We additionally got plenty ideas for further investigations.

In the following, we go point by point through all the comments and reply to them. Reviewers' comments are all repeated in this document, typeset in black. They are individually addressed, typeset in blue. Changes to the original manuscript as resulting from the reviewers comments are repeated here to ease the comparison with the original version; they are typeset in *blue italic*.

Due to some comments from the reviewers, we decided to exchange the abbreviation *BLRPM* to *OBL model* in order to distinguish the original Bartlett-Lewis model (*OBL*) from a modified version (*MBL*).

Reviewer #3

General Comments

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 OBL model 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 OBL model 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.

As the reviewer wrote, we use a parametric approach to obtain a consistent IDF curve based on a block-maxima approach and a duration-dependent GEV. This idea is based on work by Koutsoyiannis et al. [1998] and later taken up by Soltyk et al. [2014]. The main advantage is to exploit the smoothness in the IDF curve for a more robust estimation. Parameter estimation is carried out by numerically optimising an objective function based on an approximation to the likelihood; the problem of local minima is taken care of by using a latin-hypercube resampling of initial guesses for the parameter optimisation.

From continuous cell simulation, rainfall series have been obtained by aggregating cell rainfall to 1h (minimum duration) and further on to match the duration used for the observed series. We thus include the following sentences at the end of Sect. 2 and augment a sentence at the beginning of Sect. 5.2, respectively

Simulations with the OBL model are in continuous-time on the level of storms and cells. We aggregate the resulting cell rainfall series to hourly time series.

Monthly block-maxima for every month in the year are drawn for various durations (1h, 3h, 6h, 12h, 24h, 48h, 72h, 96h) from the observational time series and 1000 OBL model simulations of same length.

The authors note in Section 5.2 (lines 220-2) and in the conclusion (lines 292-3) that the OBL model 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?

Motivated already by the first comment of Reviewer #2, we related our findings to a broader spectrum of literature, please see our answer in the corresponding document, lines 59-83.

For users, an IDF curve gives a broad and immediate overview about how much (intensity) rain over a period of time (duration) is likely (frequency) to fall. Previous studies mainly focus on Gumbel plots in reference to extreme value analysis. Therefore we believe the presented framework using consistent IDF curves based on a duration-dependent GEV together with stochastic precipitation models can contribute to the community.

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'.

Thanks for pointing this potential problem out! We did not intend to motivate other researchers to take out a "suspicious" date as the winter storm Kyrill in January 2007. Instead we wanted to demonstrate the OBL model's inability to capture characteristics of an event which is much larger in magnitude than the majority of the other events. On the other hand, we showed that the model is generally able to reproduce extreme precipitation events if they are well represented in the underlying data. We augment the first paragraph of Sect. 5.3: *The convective cold front passage of Kyrill accounted for a maximum intensity of 24.8mm rainfall per hour, whereas the next highest value of the remaining Januaries would be 4.9mm rainfall per hour in 2002 and thus being more than 5 times lower than for Kyrill. We construct another data set without the extreme event due to Kyrill, i.e. without the year 2007. The intention of this experiment is* not *to motivate removal of an "unsuitable" value. We rather want to show that the OBL model is in generally able to reproduce extremes; it is, however, not flexible enough to account for a single event with magnitude far larger than the rest of the time series. …*

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.

We admit that an extreme value analysis would benefit from a longer time series, which is unfortunately not available for this case study. With respect to the POT approach, please see our answer to Reviewer #1, Mayor comment 2, lines 40-47 in the corresponding document. It is not our aim to use the OBL model as a relief of the short data series problem. As mentioned in the last comment/answer, we also need a long series to estimate OBL parameters in a way that extremes with a long return period are sufficiently well reproduced.

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.

Thanks for the hint! Indeed, we do not take the OBL as a representative for the observed rainfall but as a tool to obtain long artificial series to be used in a model-world study. We change the first sentence in Sect. 5.4 to: *In the frame of a model-world study, long time series simulated with the OBL model can be used to investigate adequacy of the dd-GEV model conditional on the simulated series.*

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.

Please note, that we now take a single fixed set of parameters to simulate 50 very long (1000yrs) series of rainfall surrogates. Based on these series, we compare two strategies for estimating return levels for different durations: the duration-dependent GEV (dd-GEV) and individual duration GEV approach. Problems of the OBL to represent observed extremes do not play a role here. However, we suggest that the observed effect for short durations indeed needs to be explored in a further analysis.

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.

Here, we assume that the reviewer uses the term "credible intervals" for the statistical uncertainty intervals, typically associated with any estimator, e.g., here for estimated GEV parameters (or the return levels derived from them). These intervals represent sampling uncertainty, i.e. the uncertainty due to having a particular sample and not the full population available. These estimates can and will vary if another – equally likely but different – sample had been observed. It is exactly this effect which we cover with presenting various samples – i.e. various pseudo-observations – to the GEV estimator. We can do so only in this model-world experiment where we have the model to generate these series. This way of presenting sampling uncertainty is equivalent (at least in interpretation) to the uncertainty intervals based on asymptotic properties of the maximum-likelihood estimator. The latter are typically associated with the GEV or other estimators. However, these asymptotic properties do not hold for the dd-GEV approach and we need a different approach to quantify sampling uncertainty: in this model-world study, we have the possibility to obtain more than one sample and can thus estimate the sampling uncertainty directly from different samples.

Specific Comments:

- 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.

References will be included in the revised manuscript as discussed in an answer to Reviewer #2, please see the corresponding document, lines 59-83.

- 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.

  We used the original BL model to gain an understanding of this type of stochastic precipitation models as we plan to use it in a non-stationary setting Kaczmarska et al. [2015], please see also our answer to Reviewer #2, first comment, lines 47-58.

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

  Thanks, changed.

- 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.

  Neyman and Scott developed a model to represent galaxies that tend to cluster. Later the very same model was found to be useful in other contexts, such as rainfall. We decide to leave this original reference in the text as it shows the origins of this model. References to Poisson-cluster models for rainfall are to be found in various places in our manuscript.

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

  We extended this part as follows in the revised manuscript: *Due to known drawbacks of the OBL model several improvements and extensions have been made in the past: Rodriguez-Iturbe et al. [1988] introduced the random parameter model, allowing for different type of cells, and additionally Onof and Wheater [1994] used a jitter and a gamma-distributed intensity parameter to account for a more realistic irregular shape of the cells. Cowpertwait et al. [2007] improved the representation of sub-hourly time scales by adding a third layer, pulses, to the model. Non-stationarity has been addressed by Salim and Pawitan [2003] and Kaczmarska et al. [2015]. Applications of these kind of models include the implementing of copulas to investigate wet and dry extremes [Vandenberghe et al., 2011, Pham et al., 2013], regionalisation [Cowpertwait et al., 1996a,b, Kim et al., 2013] and accounting for interannual variability [Kim et al., 2014].*

- 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?

  The top part of the figure represents a typical OBL simulation of cell clusters, drawn in red, and cells, drawn in blue. The red color corresponds to the life time of the cell cluster or usually referred to as storm. Hereby the vertical extensions of the storm has no physical meaning and only serves for better illustration. During its life time the storm generates rainfall cells (blue). Horizontally illustrated is the cell's life time and during its life time its constant intensity is illustrated by the vertical extension of the cell.

- 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 $w_i = 100$ 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)

  As the same point has been risen by Reviewer #2, we refer to our answer in the corresponding document, lines 122-133.

- 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.

  We did check the non-identifiability and came to the conclusion the symmetrised objective function is less likely to lead the optimization algorithm into local minima. In five out of six cases the numerical optimization lead to the same (and likely the global) minimum with same parameter values. To our understanding, profile objective functions would inform about *sampling uncertainty* for the given minimum of the OF.

- XVII. Line 151: The notation should read $\mathrm{IDF}_{T_2}(d) > \mathrm{IDF}_{T_2}(d)$.

  Thanks for the hint!

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

  After re-parametrising the GEV parameters to $\tilde{\mu} = \mu/\sigma_d$, $\tilde{\mu}$ and the shape parameter ξ are approximately independent of the duration d [Koutsoyiannis et al., 1998]. Please note, that these are only approximations but in the mentioned study it has been shown, that these approximations seem to be well justified.

- 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.

  Given equations (3) and (4), one can introduce the duration dependent scale parameter σ_d into equation (3). It results:

$$F(x; \tilde{\mu}, \sigma_d, \xi) = \exp\left\{ -\left[1 + \xi \left(\frac{x}{\sigma_d} - \tilde{\mu} \right) \right]^{\frac{-1}{\xi}} \right\}. \tag{1}$$

  Please note, that in the first version of this manuscript the tilde over μ was missing. This derivation has been made by Koutsoyiannis et al. [1998] and used, e.g. by Soltyk et al. [2014]. This is mentioned in the manuscript.

- 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 (?)

  Introducing the duration-dependent scale parameter σ_d into the GEV framework leads to two additional parameters (θ and η) and a total of five parameters. These additional parameters describe the dependence of the scale parameter σ_d on the duration d. As the reviewer mentions, it is indeed possible with this formulation to estimate the IDF relationships over all durations consistently with one single model. Benefit of this approach is a) consistency, in the sense that different quantiles cannot cross along the duration axis, and b) strength in parameter estimation is borrowed from neighbouring durations.

- 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.

  The gauge resolution is one minute, see Sect. 4. The location is chosen due to its vicinity to our institute and interest in local rainfall characteristics, as well as the easy data availability.

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

  As mentioned we were interested in local rainfall characteristics in the vicinity of our workplace. Therefore we chose a time series from our weather station in botanical garden Berlin. Also we were interested in the question if a short time series like this can be used for this kind of studies and if it would be sufficiently long enough to gain information about its extreme value distribution. It is known that long rainfall time series with such a high temporal resolution are sparse and many stations do not have long records and thus it is an interesting problem if extreme value distributions can already be obtained from short series. Thus, this study helps in investigating this issue.

- 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.

  Thanks for the hint, this mistake was 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.

  Yes, February was put wrongly here and January was meant. This is corrected in the revised manuscript.

- 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.

  To our knowledge the occurrence of unrealistically high extremes as mentioned by Verhoest et al. (2010) was never investigated for the OBL model and thus we gave it a check. This point was also raised by Reviewer #2, please see our answer in the corresponding document, lines 90-93 and lines 164-175.

References

P.S.P. Cowpertwait, P.E. O'Connell, A.V. Metcalfe, and J.A. Mawdsley. Stochastic point process modelling of rainfall. i. single-site fitting and validation. *J. Hydrol.*, 175:17–46, 1996a.

P.S.P. Cowpertwait, P.E. O'Connell, A.V. Metcalfe, and J.A. Mawdsley. Stochastic point process modelling of rainfall. ii. regionalisation and disaggregation. *J. Hydrol.*, 175:47–65, 1996b.

P.S.P. Cowpertwait, V. Isham, and C. Onof. Point process models of rainfall: Developments for fine-scale structure. Research Report 277, Departement of Statistical Science, University College London, June 2007.

J.M. Kaczmarska, V.S. Isham, and P. Northrop. Local generalised method of moments: an application to point process-based rainfall models. *Environmetrics*, 26:312–325, 2015.

D. Kim, F. Olivera, H. Cho, and S.A. Scolofsky. Regionalization of the modified bartlett-lewis rectangular pulse stochastic rainfall model. *Terr. Atmos. Ocean Sci.*, Vol.24, No. 3:421–436, 2013.

D. Kim, J. Kim, and Y.-S. Cho. A poisson cluster stochastic rainfall generator that accounts for the interannual variability of rainfall statistics: validation at various geographic locations across the united states. *Journal of Applied Mathematics*, 2014, 2014.

D. Koutsoyiannis, D. Kozonis, and A. Manetas. A mathematical framework for studying rainfall intensity-duration-frequency relationships. *J. Hydrol.*, 206(1):118–135, 1998.

C. Onof and H.S. Wheater. Improvements to the modelling of british rainfall using a modified random parameter bartlett-lewis rectangular pulse model. *J. Hydrol.*, 157:177–195, 1994.

M. T. Pham, W. Vanhaute, S. Vandenberghe, B. De Baets, and N. Verhoest. An assessment of the ability of bartlett–lewis type of rainfall models to reproduce drought statistics. *Hydrology and Earth System Sciences*, 17(12):5167–5183, 2013.

I. Rodriguez-Iturbe, D.R. Cox, and V. Isham. A point process model for rainfall: further developments. *Proc. R. Soc. Lond.*, A417:283–298, 1988.

A. Salim and Y. Pawitan. Extensions of the bartlett-lewis model for rainfall processes. *Statistical Modelling*, 3(2):79–98, 2003.

S. Soltyk, M. Leonard, A. Phatak, E. Lehmann, et al. Statistical modelling of rainfall intensity-frequency-duration curves using regional frequency analysis and bayesian hierarchical modelling. In *Hydrology and Water Resources Symposium 2014*, page 302. Engineers Australia, 2014.

S. Vandenberghe, N. Verhoest, C. Onof, and B. De Baets. A comparative copula-based bivariate frequency analysis of observed and simulated storm events: A case study on bartlett-lewis modeled rainfall. *Water Resources Research*, 47(7), 2011.