

Interactive comment on “A Hybrid Stochastic Rainfall Model That Reproduces Rainfall Characteristics at Hourly through Yearly Time Scale” by Jeongha Park et al.

Anonymous Referee #2

Received and published: 27 July 2018

This manuscript reports on a new rainfall model that is an extension of the commonly employed Bartlett-Lewis approach. The model was applied to 29 rain gages, which at first glance seems like perhaps not enough. The authors argue that the approach is very laborious and they justify that this is a good start. Overall the work will make a positive contribution. Some suggestions for improvement are included below.

Major comments 1. The model performance should be assessed based not just on “effectiveness” (fitting data), but also on “efficiency” (how many parameter). The latter criteria is quantified using AIC or similar metrics. Such criteria are already embedded within ARIMA frameworks so it is clear that the authors should be familiar with the

C1

importance of AIC etc. However, the authors have chosen not to describe how many parameters the overall model uses? This is very important information that should be explicitly compared against the MBLRP model, which very clearly uses 6 parameters. The proposed model is acknowledged by the authors (p.26, line 20) as “not easy to implement”. Yet it is argued throughout the paper to perform better than the ‘standard’ MBLRP model, which is easy to implement. So the critical question is whether the improved performance is worth the cost in complexity. AIC or similar should be used to compare the gains in model performance vs the cost of increased complexity. It seems that the current model is perhaps far more complex than the MBLRP, yet this complexity is currently hidden from the reader. 2. Twenty-nine gages were selected. But the spatial distribution is not uniform. For example, most of the gages are in a similar climate region in the Midwestern USA with only two gages in the southeastern USA. What was the reason for this disproportionate representation? It needs to be specified if these are the only gages in the study region where sufficient data are available or whether this is just a sub-section of available data. If it is a sub-selection of available data, then the disproportionate representation needs some explanation. 3. I do not understand the necessity of the first model module (SARIMA model). From Figure 4 it seems that the input to the second module is “monthly rainfall mean”. If this is the only output from module 1 that is then used in module 2, then why is synthetic data created at all? It seems that the measured monthly means should be used here. 4. Also, From Figure 4 it seems that the second module is generating more outputs than are supported by the model inputs. If only monthly means are input, then it seems impossible to determine just from this all of the outputs that are listed. This wide array of outputs would have to be non-unique given only monthly means as an input. So something must be missing here. 5. Several figures use data from one selected gage as illustration. However, different gages are used between the figures. In such cases it should be specified whether similar trends are observed at all gages or whether this is only seen at this particular gage. The reader should not be led to suspect that the “best” gage is being selected for each figure.

C2

Other comments 1) P.1,Line 11, “fine tuned” here is vague and not descriptive. Suggest more specific description of the approach that is used. 2) P.1,Line 11, “fine scale” is similarly vague, so I suggest to be specific. 3) P.1, Line 14, change to “at 29 gages” 4) P.2,lines 1-2, Change to “provide rainfall input data to modeling studies for risk. . .” 5) P.2, lines 8-21, This direct quote from another paper was to me much too long to be appropriate. Either summarize briefly in your own words, or just cite the other paper. Do not extract large sections of another paper – even if some authors are shared between the papers. 6) P. 3, lines 9-12, This are far too many references cited here. Generally three references are enough to support a point. It is not necessary to cite in one sentence every study that has ever been done. 7) P. 4, line 31, Here it is not specified what model parameters were used nor how those parameters were determined. At a minimum it should say here that these were determined “as specified below/in section XX”. 8) P. 9, line 9, The justification for 24 is not clear here 9) P. 10, line 11, This seems to be equation 1 re-stated 10) P. 13, line 4, cite figure number, and describe which data these are based on 11) P.15, line 29, It is not clear here what data are meant by “generated fine time scale statistics” and “statistics of the generated synthetic hourly rainfall”. These sound too similar. 12) P. 16, lines 11-12, It is not appropriate to switch between R2 and R when the R2 value is not high enough to support your point. 13) P. 18, section 4.3, It is not clear why variance is described in section 4.2 and then standard deviation in section 4.3. These are not two independent parameters. Pick only one. 14) P.17, line 12, The equation should not be invoked here when it has not been presented yet 15) P. 20, line 2, It is not clear how the 100 or 200 year recurrence intervals are determined from measured data 16) P. 20, lines 7-9, These sentences should probably be removed. Readers of this journal do not need to be told that R2 approaching 1 means the model is good. 17) P. 22, lines 10-11, Revise this sentence. Currently “rainfall” is repeated four times. 18) P. 24, all of this seems like it should be in the methods rather than discussion. 19) Figure 2 – the meaning of the gray boxes should be described in the caption 20) Figure 3 – the caption should explain the importance of the BIC values 21) Figure 5 – the different shading of the boxes should

C3

be described in the caption 22) Figure 6 – the caption should specify which data sets these results correspond to 23) Figure 9 – it is not clear if these are the mean for all months or if there is one data point for each month. 24) Figure 13- make the symbol sizes bigger. I could not see triangles 25) Figure 14 – change the y-axis limits to better show the data range. Since the lower limit is already not equal to zero, just make it 0.7 and the upper limit 1.1 or 1.2

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2018-267>, 2018.

C4