Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-273-RC2, 2017

Hydrology and Earth System Sciences
Discussions

Interactive comment on "A bootstrap method to estimate the influence of rainfall spatial uncertainty in hydrological simulations" *by* Ang Zhang et al.

D. Klotz (Referee)

daniel.klotz@boku.ac.at

Received and published: 11 September 2017

General Comments

The present study proposes a bootstrap based quantification of the spatial rainfall uncertainty for rainfall-runoff modelling. The main body of the work uses aforementioned bootstrap method is used to select between different numbers of stations and different groups of stations. The influence of the number and grouping is then quantified via the Nash Sutcliffe Efficiency and exponential fit is used to derive a heuristic for the estimation of the optimal densities of rainfall measurement stations. The whole analysis is based on three manually selected rainfall events from 2001, 2002 and 2006,

Printer-friendly version

Discussion paper

with data from of the Quingjian basin. Additionally, rainfall data from a second basin (Longxi) is used to provide a basic comparison about the scale and uniqueness of the conclusions.

The topic as such is certainly relevant for the study of the rainfall-runoff relationship as such and would fit a HESS. I do also conceive that the endeavor of the authors poses major challenges. Rainfall process can be extremely heterogeneous and as the authors state a quantification of the uncertainties depends multiple interplaying phenomena (such as the position of the meteorological stations, the unique futures of the studied events or the elevation profiles of a catchment). So, the authors did probably a good job, if we consider the possible difficulties that are inherent in such a study. I am nevertheless left with many concerns. The most important wants are listed in the following points, minor comments and questions are given later.

1. First and foremost, I believe that considerable attention should be given to improve the introduction in special and the readability of the manuscript in general. The whole manuscript needs a clearer structure, the aims must be defined in a clearer fashion (maybe less focus on different aspects and more focus on the main topics) and the manuscript as a whole should meander less between topics. Arguments about the computational expense, for example, can be found throughout the script, but no chapter is devoted to it.

2. In reference to my last point, I have to say that I truly believe that is worth to discuss the computational expense of the method as such. Currently hints about the computational efficacy are spread throughout the manuscript, but I miss a dedicated chapter and quantitative arguments. The method seems quite computationally expensive as the hydrological model needs to be executed approximately N*K times. Now what I would like to know is what it does mean quantitatively. How long does a run of the used r-r model take, and how long does an execution of the bootstrap need. Especially the last point provides a plethora of possible topics/experiment: Influence of shorter and longer time series, usage of the bootstrap without and with parallelization, usage of the

Printer-friendly version

Discussion paper

model with and without parallelization, on a cluster or a normal pc, etc.

3. The actual time-resolution of the modelling needs some further discussion and explanations. I think the main problem here is that the descriptions are not specific enough. The authors say that “most of the data” was at a resolution of 2hours and of 6 minutes respectively. It is difficult for readers to infer what that means exactly. A table that indicates which stations have fine/coarse resolutions and sparse data would be very helpful. In current manuscript also forces readers infer the magnitude of the distortion that is induced by the different resolutions. How exactly is the coarse scaled data disaggregated? How does the comparison work between aggregated (hence smoothed) fine scale data and disaggregated data? For many readers (including myself) it will seem strange that a 6 minutes interval for simulating the runoff data is used. The authors claim that the reason for that is to emphasize the main hydrological process. They do however not explain how works if the input data has a resolution of two hours only.

4. The used method need further underpinning. Why are Thiessen Polygons used and not a more sophisticated method (e.g. Kriging)? Especially when the authors conceive that the method can cause problems (Page 9, Lines 14-16). I have similar concerns regarding the use of the genetic algorithm. Why are not more modern and conventional techniques used. In the domain of rainfall-runoff modelling Shuffled Complex Evolution and Dynamically Dimensioned Search come to mind. In general, there also exist more modern versions of genetic/evolutionary algorithms.

5. Finally, I would be ready to be convinced otherwise, but I am not sure if the NSE is a well-chosen objective criterion for this sort of study. The low NSE values (that results, according to the authors, from the high resolution) make it difficult to use conventional intuition about the objective measure. Additionally, the chosen events differ largely in from, magnitude, runoff-coefficient and error structure. The amount of evaluated data with low runoffs (before and after the events) also seems to vary. Nonetheless, I believe that in this particular use of the NSE is suboptimal, since information about magnitude

of the events and errors are hidden by its bounded. For this particular case I would therefore propose to use the MSE instead. This would directly express the goodness of fit and the improvements due to the respective calibrations can be seen as relative improvements. Also no additional explanations would be needed to report the jumps of improvements reported at page 9, line 25. Alternatively, it would be useful to provide a sensitivity measure with regard to the individual results.

6. Figure 4 needs more explanations and better quality. For some reasons the observation-dots seem not to be equidistant! It is not explained why the model reacts faster than the real system (at least for event 2001 and 2006). At least the strange fit for the event of 2001 seems to be related to this issue, since the "independently calibrated" hydrograph might only be so low because of the large errors at the raising limp. The current resolution of the figure is not good and the depiction is difficult to read. For me it was, for example, not possible to discern what the cluster of observations means at the start of the raising limp of the 2002 event.

—

Specific Comments:

S1) This might be nitpicky, but the authors state that the other methods for spatial rainfall uncertainty quantification are not applicable to other basins. Why is that and why does the proposed method not exhibit this problem? The test is only conducted for one basis.

S2) Bootstrapping is a very, very robust method, so the following concern might be of less importance: As far as I know bootstrapping does (still) assume independence of samples. Is this given for rainfall stations? The description of the cluster analysis seems to suggest elsewise. What influence does this have?

S3) **For all Figures**: Please always mention the used basin explicitly in the figure captions. I understand that it is somewhat redundant because the evaluation takes

only place with regard to Quingjian river basins. But, it is very useful for readers who want to get an overview.

S4) **Page 2, Line 9f:** I think rules should be in plural here. Villarini et al. (2008) use a catchment of the size of 125km2, but seem to compare the gauging stations with satellite images with a resolution of 200km2. I therefore believe that the "rules of thumb" should approximate the size of the satellite pixel. I am not sure however, the argument can get finicky here. Lastly, I think one should mention that the study takes place in England.

S.5) **Page 6, Line 28**: What does "adopted" mean in this context? What was changed from the original one. Why is a genetic algorithm used and not a more prevalent method such as the shuffled complex evolution? Why is not a newer evolutionary/genetic algorithm?

S.6) **Page 6, Line 8**: What does it mean that the "influence of topography on rainfall is negligible"?

S.7) **Page 7, Line 1-2**: Please specify the "stop criterion" explicitly (number of generations, NSE, ?) How many generations did it take on average?

S8) **Page 9, Line 9**: What kind of independent calibration are we talking about here? Why are parameter interactions not seen as a problem in doing this? Why is only the K_{ZUS} optimized and not the most sensitive parameter group (e.g. all soil hydraulic conductivity parameters)?

S9) **Page 9, Lines 25-26**: I would disagree with the statement that the individually calibrated model runs prove that DYRIM is able to represent the rainfall-runoff events in a sufficient way, as long as no evidence is provided that the results are not just due to overfitting (Maybe evaluate the individually derived parameters for the other events to?). I would propose to see the NSE values as hints (if anything). One might also be able to argue that the generated hydrographs can be seen as some sort are upper boundaries or best case scenarios for the DYRIM simulations. Additionally, would it

Printer-friendly version

Discussion paper

not be possible to use this information to determine the possible upper bound for the relationship between NSE and measurement side density (Equation 2 and Figure 9)?

S10) **Page 11, Lines 1-11**: In my opinion this should be part of the method section and not of the results.

S11) **Page 13, Lines 29-31**. Here it is argued that the large errors of the hydrological simulations of the 2006 rainfall event are most likely due the structural and parameter deficiency. Is it possible to plot the cumulated rainfall alongside the cumulate measured runoff for this event (or for all of them)? On basis of the low runoff coefficient and the hydrographs I would (perhaps naively) assume that it there is a bias in the input or the runoff measurements.

—

Minor Remarks

Page 1, Line 25: Can you provide additional sources here? Beven, 2001 is a large tome, while Cibin 2014 seems to be focused on ungauged basins (which are not even mentioned as application example).

Page 1, Line 25-29: For me the first sentence of this passage is to intertwined. Could you divide the sentence into two? One point is that hydrologist try to improve the accuracy of simulations and predictions and another point is how this can be achieved (i.e. improving the model structure, better calibration method, better measurements of input data). From there on I think it would be worth to add an additional sentence, arguing why the precipitation is seen important factor for improving simulations and forecasts.

Page 2, Line 1: Please rethink this sentence. The comparison between radar and rainfall station is strangely formulated.

Page 2, Lines 11-13: Please provide additional sources. Four studies are not numerous.

Page2, Line 14: Please recalculate the units from square-miles to square kilometers and use that unit consistently throughout the manuscript.

Page 2, Line 18f. You described what Moulin et al. (2009), but not their results/conclusions. Readers will wonder why not, as they are provided for the previously mentioned studies. Could you expand on that?

Page 3, Lines 19-20: I do not understand what the citations are referring to. Are all these authors proving the advantages of simplicity and high-accuracy?

Page 3, Lines 7: Can somehow you remove the double and in the sentence (maybe use “as well as”? It would improve the readability of the sentence

Page 4, Lines 8-9: I find that argument a bit difficult, in my eyes the bootstrap is a generalization of the jackknife. Maybe compare it to another resampling technique?

Page 4, Line 13: Is it wise to put this argument forward like this. As far as I know, bootstrapping assumes independence of samples. That is not necessarily a property of the population per se, but the sentence could lead to misunderstandings.

Page 4, Line 19-22. You might want to split the sentence to improve readability.

Page 5, Line 13: Change to “are obtained”.

Page 5, Line 16: Can you make the following statement more explicit: “. . ., the bootstrap method is used to traverse most of the combinations of rainfall stations, . . .” (emphasis is my own). What does most mean?

Page 5, Line 30f: What exactly does “unit” mean in this context?

Page 6, Line 5: Remove “Then”.

Page 6, Line 8: Remove “End”.

Page 6, Line 31-31 I do not understand the sentence “This technique promises the parameters independent of the GA and easy to be optimized”

Page 8, Line 1: Write “One rainfall event, which occurred on . . .”.

Page 9, Lines 11-12: I do not understand this sentence.

Page 9, Lines 32-24: This statements needs a citation.

Page 10, Lines 3-4: Shouldn’t this be part of the results and discussion?

Page 10, Lines 27-31.: Please reformulate.

Page 11, Lines 27-28: Sentence is unclear.

Page 12, Line 10: Please reformulate.

Page 13, Lines 3-4: Write “that even if”

Figure 2: Please rework the plot. The legends are hard to read. It is difficult to grasp the extends of the basins.

Figure 9: Why is the “prediction curve with upper limit” in plot c, higher than the prediction curve without upper limits. The former appears to be fitted for less data. Also, is it possible to provide uncertainty bounds?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-273, 2017.

Printer-friendly version

Discussion paper