

Interactive comment on “Bias correction of Simulated Historical Daily Streamflow at Ungauged Locations by Using Independently Estimated Flow-Duration Curves” by William H. Farmer et al.

Anonymous Referee #2

Received and published: 30 April 2018

General comments

The work presents a method that aims to produce unbiased time series of daily discharges for ungauged basins. It consists in correcting the distributional bias in the time series of simulated discharges obtained for this ungauged basin from some simulation model (in the present case, the simulation model is a regression model where the daily discharge of each day is estimated via kriging from gauged stations in the neighborhood of the target ungauged station). The correction is a quantile –mapping correction where the reference statistical distribution is obtained from the Flow-Duration-Curve of

C1

the target location, estimated via regionalization again from the observed FDC estimated for the neighboring stations.

The issue is a very relevant one and the approach a promising one. The novelty of the approach with respect to other similar studies (mentioned in the manuscript) has to be clarified.

More specifically, the aim of the paper is not well explained and it has taken me more than 2 hours to understand what was done in the paper. One or possibly several graphical schemes are definitively required to figure out the methodology / objective of the paper and data used for the application of the method. A reformulation of the introduction and of the abstract is for me also necessary. Showing figures with hydrographs or FDC is also required (boxplots are not informative enough).

The formulation in the abstract for instance is somehow clumsy : “Based on an existing approach that separates the simulated streamflow into components of timing and magnitude, the timing component is converted into simulated nonexceedance probabilities and rescaled to new volumes using an independently estimated flow-duration curve (FDC).” I think that the authors do not rescale the timing component. They just scale each daily discharge of the time series with a multiplicative correction factor that depends on the non exceedance probability of that discharge. . . The correction function is estimated from the FDC of the target location, estimated from the observed FDC of gauged stations in the region. I would have understood instantaneously the objective / approach of the work that way (I am perhaps wrong, but I guess this is roughly what you do). This has definitively to be clarified (here and everywhere else in the manuscript).

It is also not clear that the authors want to estimate unbiased time series of discharge at ungauged sites. This has to be clearly stated. One thus needs to have a regional simulation model and a regional model for the FDC estimation. The interest of the configuration where an observed FDC is used to bias correct the simulated time series has to be made clear in this context. This approach is presented as an alternative

C2

estimation approach. It is actually not and this is really confusing.

The results of the study were rather expected. The authors find that the combined “time series simulation”+“FDC bias correction approach” is unable to correct the bias in the simulated hydrograph. The authors explain this is due to the bad quality of the regionalized FDC which present a rather significant distributional bias for all discharge percentiles. By construction, the “poor” result of the FDC bias correction was thus expected. The interest of the method would have been for a configuration where the distributional bias obtained from the “time series simulation” approach is larger than that of the regionalized FDC. Obviously, this is unfortunately here not the case. At least in average for the 1168 basins considered in the work. More convincing results could have perhaps been presented considering different configurations of basins with different relative performances of the two approaches considered here for the combination. We would have seen that the method works in configurations where regionalised FDC are good.

Focused Remarks :

P2. “As attested to by many researchers focused on the reproduction of historical streamflow, this bias commonly appears as a general overestimation of low streamflow and underestimation of high streamflow”. This statement is not true. Simulations with any hydrological model will lead to the overestimation of some high stream flow and the underestimation of the others. The same for low flows : some are over-, some are underestimated. In rainfall-runoff models, it clearly depends on the amount of rainfall estimated from the small number of raingauge available, and frequently not numerous enough to have the right estimate of rainfall input. It also depends on the quality of the model which is limited for the simulation of specific hydro processes. The fact that a given model systematically overestimate lowflows and underestimate highflows surely means that the model is not good enough and could likely be improved. As a matter of fact, some models present those limitations and this justify the present work. But a general statement can not be given there on such a case. (Note that this statement is

C3

invalidated in your work by the results and comments you mention (p6 – line 20/25). Please reformulate.

P2. Ln 22. The difference / novelty of the present methodology with that of Fennessey (1994) and Hughes and Smakhtin (1996) is not clear. Please clarify.

P3. The description of the content of the paper is missing at the end of the introduction.

P3. Simulated discharges. “Initial predictions of daily streamflow values for each streamgage were obtained by applying the pooled ordinary kriging approach (Farmer, 2016) to each 2-digit Hydrologic Unit”. The approach has to be better explained. It is not possible to understand how time series are obtained for any given gauge from what is said in the paragraph (“It builds a single, time-invariant semivariogram model of cross-correlation that is then used to estimate ungauged streamflow as a weighted summation of all contemporary observations”). A summary methodological scheme could be helpful. Is the variogram model estimated independently for each day ? for each target location from stations in its neighborhood (or do you use all stations of the United States to estimate the daily discharge of a given station) ? What is a 2-digit Hydrologic Unit ? Does the leave-one out procedure applies for the target location ?

P4. Ln 1-8 : Please clarify. “Daily FDCs were developed independently of the stream-flow simulation procedure by following a regionalization procedure similar to that of Farmer et al. (2014). “You mention “These same percentiles were then estimated using a leave-one-out cross-validation of regional regression”. The objective/process of this estimation was not clear. Please make clear that the FDC used to bias correct your target station is obtained from a regional FDC model, obtained from all (or a part of) the stations close to this target station.

P4. Explanatory variables. Please comment the possible correlations between those. Have you looked / used for uncorrelated sets of explanatory variables ?

P4 – Ln 10 : what are “best-subsets regression”. Best sub-sets of what ? What is the

C4

difference with “three contiguous streamflow regimes” you mention in line 17.

P4; Please clarify. “the percentiles were grouped into a maximum of three contiguous streamflow regimes based on the behavior of the unit FDCs in the region”. What is a unit FDC ? How are defined the regimes ? Are they relative to different parts of the FDCs or to different sets of different basins within the region ? Do you group 95th and 90th percentiles for instance ?

P4. “The percentiles in each regime were estimated by the same explanatory variables, allowing only the fitted coefficients to change..” Do you mean that a regression is estimated for each of the Twenty-seven percentiles you considered ? If yes, a very significant risk on a non monotonic behavior of the FDC is likely. Why not working with a analytical model fitted to the FDC (and then propose a regionalization of the parameters of this model ?). This issue should be discussed somewhere.

P4. “Further details on the approach of percentile grouping this methodology can be explored in the associated data and model archive Farmer et al. (2018).” The given reference is a technical report. The soundness of the “percentile grouping approach” is thus uncertain. It has thus to be fully explained in the present manuscript or it has to be described in another a peer reviewed journal.

P4. The section on the bias correction approach has to be made clear. Especially how the bias correction works and with which data in the case of ungauged catchments. Again a synthetic graphical scheme of the BC approach would be useful. You say “The nonexceedance probabilities were then converted to standard normal quantiles and linearly interpolated along two types of independently 30 estimated FDCs: the regionally regressed FDCs and the observed FDCs determined by applying the Weibull plotting position.” It is something to do with quantile mapping ? I can imagine that a correction function can be established for any given station where you have both observations and simulations. Is the idea here to transfer the correction function obtained for a gauged catchment to neighboring ungauged ones ? If yes, how do you

C5

define the neighborhood ? how do you consider the different corrections functions you can obtained for the different gauges stations you may have in the neighborhood of your target location ?

P4. Ln – 29/30 : “The nonexceedance probabilities were then converted to standard normal quantiles and linearly interpolated along two types of independently estimated FDCs: the regionally regressed FDCs and the observed FDCs determined by applying the Weibull plotting position.” » This is confusing. In the ungauged basin configuration : only one FDC is expected to be used : the regionalized one. The observed FDC is not expected to be available in a ungauged catchment. You use it here only to estimate the added value of the quality of the regionalized FDC on the resulting bias corrected discharge time series. The “independently estimated FDC” you mention in line 31 (and basically everywhere else in your manuscript) should first refer, I guess, to the regionalized FDC (and not to the FDC from observations).

P5. Ln 3: reformulate : “by correcting the simulated volumes to an independently estimated FDC.”

P5. Evaluation :

Ln 10- the two different evaluations approaches were not clear for me at first. A reformulation would be worth Ln 20 & follow. Evaluation for the whole FDC, or for a given tail : what is the evaluation criterion : the mean of the bias estimated between each pair of percentiles ? the bias between the mean of the percentiles for the raw and corrected data respectively ?? Evaluation on observation-independent tails and observation dependant tails. A graphical scheme please to explain what is done in the second case, at least in a supplementary material ! Results : all results are given in the form of boxplots. This is likely not enough to understand how the methodology work and how good it is. To give for a selection of stations the different curves (observed / regressed / regressed+bias corrected) would be helpful (with a good performance station and a bad one for instance) P5 Ln33 and p7 . Ln 25. I find the term “timing”

C6

and “error in timing” not appropriate. You could perhaps say “ an error in timing of the percentiles”. This is however more an error in the temporal structure of your simulated time series. This results from an error in volume which is one day an over-estimation of the true volume and the day after an under-estimation.

P7 ln34. You say “These timing errors also almost result in errors in a particular direction: low for high flow and high for low flows”. You have perhaps such a mean behavior but as mentioned above, you have a number of low flows that are overestimated but also a number of low flows (less frequent) that are underestimated. . . Please reformulate to put this statement in perspective.

P6. Ln 2 : “Figure 4 and Table 1 summarize the tail bias in all approaches to streamflow simulation considered here.”. What are the 3 different approaches ? This has to be clearly explained previously. The BC-Obs is not an “approach” similar to the 2 others as you do not know in principle the observed discharge for the target ungauged basin. It just allows you to identify the influence of the quality of the regionalised FDC. A reformulation is required when relevant in all the manuscript. The presentation of the method has probably also to be restructured to make it clearer.

P7. 1. The analysis of the second paragraph in this section 3.1 is clumsy. The Observed FDC is in principle perfect and thus the bias in simulation for the observation independent evaluation case should fully vanish after correction. You should have nothing or roughly nothing to comment here. The results should be perfect. Why is there some remaining bias with the BC-obs approach ?? Please comment. (could it be a difference in the time period used for the simulation and the time period used to identify the observed FDC? Is it something related to the epsilon value you add to discharge data for the logarithmic transformation issue ? to the reduced number of percentile used to describe the FDC ? something else ?) The comment on this in the conclusion section has also to be modified accordingly (p12 – ln 1/10). P7. 24: “To understand the effect of errors in timing further, consider Figure 6, which shows the mean error in the nonexceedance probabilities of the observation-dependent upper and lower tails.”

C7

I can not understand (just guess) what is referred to here ? Please clarify.

Fig 2 and following : how many data are used for each boxplot (#stations x #percentiles ???). Is there one point for each station/percentile ?

Discussion : what about the likely seasonal dependence of the correction function ? Is there some potential for improvement here ? The estimated FDCs are composed of 27 quantiles, of which the upper and lower tails contain only the eight values with nonexceedance probabilities 95% and larger and 5% and smaller, respectively.). A comment on the number of quantiles used to describe the FDC would be worth (a sensitivity analysis of the results to this number could be also included in a discussion section)

What is the influence of the duration of the observation time period used to estimate the observed FDC > on the quality of the FDC estimation and then on the quality of the bias correction ?

Is there any dependence of the results to the hydroclimatic context of the basin ? How is it structured in space in US ?

Minor remark : P2. Ln10. The interest of the “long term forecast term” here is not clear. It seems to be out of the scope of the work. To be better explained. What is long term ?

P2. 15/20 : » this paragraph is not clear »> to be clarified / double-formulate. The “interpolation of non-exceedance probabilities along the FDC” is a rather clumsy formulation. What does it mean ?

p2. Ln 33 : please clarify what is meant there P2. Ln 35 and following. This does not belong to the introduction but to a discussion section. The discussion should probably give the evaluation of the other method suggested here (Additional research to explore if estimating nonexceedance probabilities directly, as opposed to the conversion of simulated streamflow to nonexceedance probabilities used here, might further improve

C8

nonlinear spatial interpolation using FDCs or simulation more generally.

p3. Ln 3 : give the structure of the paper

p5. “The root-mean-squared error of the common logarithms of streamflow and the differences therein were used to quantify accuracy.” Do you mix streamflow and differences between streamflow in the computation of a single RMSE criterion ? If yes, I fear it is not relevant or please clarify / justify.

P6. Ln 9 – add a subsection title “simulated hydrographs without correction”

P6 – 25. “These results show upward bias in lower tails and downward bias in upper tails.” No, this is not the case in general. See your paragraph above.

P9 “For the observation-independent case, the errors are removed almost completely, and the remaining errors in the observation-dependent case mimic the timing (nonexceedance probability) errors.” This is not true (only if the observed FDC is used)

P10. 2 : “The change in the absolute bias of the observation-independent lower tail has a 0.72 Pearson correlation with the absolute bias of the lowest eight percentiles of the FDC estimated with regional “regression.” I do not understand what is meant here.

Figures P10- 6 “as regional regression is not the only tool for estimating FDCs, improved methods for DC estimation would further increase the impact of this bias-correction procedure.” Mention other such methods.

P10.15 “It may not always be possible to determine the accuracy with which a given FDC estimation technique might perform, making it difficult to determine if these results can be generalized.” There is no reason why the accuracy of any FDC regionalization approach could not be assessed (this is done in all FDC regionalization study). At least with a leave-one out procedure. Before applied for the bias correction of any time series simulation model, this quality should be checked and estimated to be better than that of the simulation model.

C9

Fig. 1 – What is meant here : “The outlines of 2-digit Hydrologic Units are provided for further context. Fig. 2 : BC-RR and BCObs have to be defined in the main text. Figures and tables : simplify the captions : a number of repetitions could be removed (and a reference to the caption details of one reference figure added in the caption of all other figures) Are the 3 tables usefull ?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2018-30>, 2018.

C10