

Interactive comment on “Regional regression models of percentile flows for the contiguous US: Expert versus data-driven independent variable selection” by Geoffrey Fouad et al.

Anonymous Referee #2

Received and published: 10 March 2017

Overall, the problem addressed by this paper represents a fundamental challenge which remains unsolved in the sense that there is no generalized solution to the problem of estimation of FDC's at ungaged sites. However, there are many innovations in this area, with the current state of the art nicely summarized in the PUB book chapter by Castellarin et al. (2013) which was not referenced in this manuscript. What guidance did they give on this subject? Unfortunately, the approach taken in the manuscript, does not appear to include any fundamentally new approaches to the problem, because it simply combines commonly used approaches such as clustering and random forests with multivariate regression combined with very a large national dataset. I do not see any way to convert the approach taken in this paper into the type of scientific

C1

contribution required for publication in HESS, however, I do think that this paper would make an outstanding contribution to the Journal of Hydrology - Regional Studies, which is a subset of the Journal of Hydrology ideally suited to exactly this type of study.

This paper does provide a nice set of national regression models for estimation of FDC's for any location within the United States. However, even before this paper were to be published in the Journal of Hydrology - Regional Studies, I would ask the authors to address the following concerns:

1. What guidance do Castellarin et al. (2013) give for addressing this problem? Having reviewed ALL the literature on this problem, they should give some good guidance.
2. An existing national model for estimation of an FDC for ungaged sites in the continental United States exists, within the USGS STREAMSTATS program. This system is now operational for most regions of the US. How does your approach differ from the approach taken in STREAMSTATS? I believe STREAMSTATS takes a very similar approach to you, thus it is absolutely essential that you answer this question. I was very surprised that you never even mentioned the USGS STREAMSTATS system!
3. How does your approach compare with the results of STREAMSTATS for your validation stations, or at least a subset of your validation stations. There are also some recent USGS reports who have done some intercomparison studies which are not cited in your study.
4. All of your statistical analyses are based on very classical regression methods and goodness of fit procedures. This is both good and bad. It is good because your results will be understood by a wider audience. It is bad because you do not consider the new generation of 'influence statistics' which enable one to use ordinary regression procedures while simultaneously understanding the influence of outliers and more importantly the observations which have a large influence on the model coefficients. Please read the chapters in Helsel and Hirsch (2002) which explain how to use 'influence statistics' such as DFITS and Cooks D, and 'Prediction Rsquared' in addition to

C2

some of the statistics you used. I would never trust a national model that was not subject to this sort of analysis, because it is very likely that just a few anomalous stations are driving the entire model in each region.

5. At the very least, you should do the following intercomparison. To be able to determine if your model is an improvement over others, I suggest you do a very simple comparison. For each of your validation sites, simply use the drainage area discharge relationship to transfer the flows from the nearest gaged site to the validation site. Then construct the FDC for that site and compare it to your model. I wonder if your model is better than this very simple model! It is this type of comparison which makes your work credible and useful.

References Cited in Above Comments:

Castellarin, A., G. Botter, D.A. Hughes, S. Liu, T.B.M.J. Ouarda, J. Parajka, D.A. Post, M. Sivapalan, C. Spence, A. Viglione and R. M. Vogel, Prediction of flow duration curves in ungauged basins, Chapter 7 in Prediction in Ungauged Basins, Cambridge University Press, 496 pages. 2013.

Helsel, D. R., and Hirsch, R. M. (2002). Statistical methods in water resources, US Geological survey Reston, VA

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-639, 2016.