HESSD

Interactive
comment

Interactive comment on "Technical Note: Evaluation and bias correction of an observations-based global runoff dataset using historical streamflow observations from small tropical catchments in the Philippines" *by* Daniel E. Ibarra et al.

Jaivime Evaristo (Referee)

j.evaristo@uu.nl

Received and published: 13 April 2020

General comments: It has been an absolute pleasure reading through this contribution. By using monthly runoff observations from 55 catchments in the Philippines with more than 10 years of data between 1946 and 2014, Ibarra et al. evaluated the possible utility (and veracity) of a recently published global runoff product GRUN_v1 (Ghiggi et al., 2019). They showed significant albeit weak correlation between their data and

GRUN model predictions, and somewhat improved model-data correspondence using volumetric efficiency (VE) and log-transformed NSE criteria. Among others, Ibarra et al. demonstrated systematic over- and underprediction of baseflow during the dry months, and underprediction of peak flow in some wet months in most catchments. To go above and beyond a simple demonstration of model-data correspondence, the authors proposed a two-step bias correction procedure that particularly addresses GRUN underpredictions during the wettest months. The authors suggested that the utility of GRUN can be extended to other ungauged tropical basins if a similar bias correction methodology is applied.

While GRUN_v1 was trained and validated using GISM and GRDC, respectively, none of the corresponding GISM and GRDC data from the Philippines was used in GRUN_v1. Thus, it is worth noting that this contribution by Ibarra et al. is indeed an independent test of GRUN_V1 runoff reconstruction.

There is no doubt that the broader community will stand to benefit from Ibarra et al.'s analysis. The scientific and engineering literature on water resources continues to "suffer" from a mid- to high-latitude bias. Ibarra et al.'s work represents a substantial contribution to reducing this bias and increasing our understanding of tropical hydrology, particularly with respect to the implications of their work for the ungauged tropical basins. Moreover, I can only hope that the community will also commend Ibarra et al. for making these Philippine datasets publicly available, which may prove useful for similar and sundry purposes. These favorable comments notwithstanding, I raise some [relatively minor] points that when addressed may only serve to improve this contribution.

Specific comments: (1) On bias correction at the national scale: Is there any particular practical significance for the bias correction at the national level, as opposed to, say, at the basin level or per climate types? For example, the per-climate-type analysis seems to show some interesting patterns (Figs. 3&4), and so as at the basin level or catchment size (Fig. 2B). This comment of course assumes that a sufficiently wide

range of flows are similarly captured at these levels of abstraction as at the national level, thereby, making log-transformations meaningful. Such seems to be the case per climate type based on the scatterplots in Figure 3. In any case, it might be useful to know why the bias correction was performed at the aggregate national level and not at [or not in addition to] sub-national levels.

(2) Parameter uncertainty: I would encourage the authors to also perform uncertainty estimation on their slope (m=0.774) and intercept (b=0.099) parameters, possibly via bootstrapping. This would make their proposed bias correction method more robust and bounded. Suggested references follow:

Efron, B. (1981), "Nonparametric Standard Errors and Confidence Intervals," The Canadian Journal of Statistics 9:2, 139–158.

Rubin, D. (1981), "The Bayesian Bootstrap", The Annals of Statistics, 9:1, 130–134.

(3) On transformation bias in curve fitting: One utility of Ibarra et al.'s work is on possibly using GRUN for other ungauged basins in the tropics and applying a similar bias correction as proposed (L301). Because these corrections are in log-log space, the user may then need to back-transform (or antilog) to obtain the "corrected" runoff values. Such toggling has long been shown to carry some inherent statistical bias that [also] needs to be corrected, as succinctly discussed by Ferguson (1986). This bias can be non-trivial and results from the use of least squares regression in estimating the logarithms of, say, runoff in ungauged basins. Without repeating here the arguments that Ferguson most effectively articulated in 1986 (and Miller 1984), the authors may find it worthwhile to reflect on the implications of this [possible] statistical bias on their proposed method for bias correction.

Ferguson, R. I. River Loads Underestimated by Rating Curves. Water Resour. Res. 22, 74–76 (1986).

Miller, D. M. Reducing Transformation Bias in Curve Fitting. Am. Stat. 38, 124 (1984).

(4) Other comments: L152-153: Perhaps, a more appropriate way of describing NSE ≤ 0 would be that the "model is no better than using the mean value of the observed data as a predictor" (e.g. Gupta et al. 2009, Journal of Hydrology). This conveys a somewhat different meaning than how it is presently written, i.e. "values less than zero indicate that the mean value of the observed data is a better predictor than the hydrologic model"

L196-197: Or alternatively, that model-data agreement improves with catchment size

L226-227: "These catchments experience distinct wet and dry seasons in the northwest Philippines." Can the authors comment on the implication of this sentence for catchments (outside the Philippines) with distinct wet and dry seasons vis-à-vis the physical significance (e.g. of rainfall-runoff transfer functions) that the VE criterion represents?

L230-231: Please qualify/rewrite because while Criss and Winston (2008) underlined that NSE tends to put more weight on large flows, they did not particularly discuss or say anything regarding NSE-log10.

L240-243: Please consider rewriting this long sentence for clarity. Also, this sentence refers particularly to GRUN (published in 2019) yet it cites two papers that predates GRUN. Please qualify for congruence.

Technical corrections

L61: "(ref)". Reference placeholder

L99: URL is not working. Please check

L126: "that". Typo

L206: "were". Typo

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-

26, 2020.