Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-26-AC1, 2020

Hydrology and Earth System Sciences
Discussions

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.

Daniel E. Ibarra et al.

dibarra@berkeley.edu

Received and published: 5 July 2020

Response to review by Jaivime Evaristo [Note our responses are listed as "RESPONSE: ...."]

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.

RESPONSE: We thank the reviewer for thoughtful comments and points to address. All of them will be incorporated into our revised manuscript.

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

RESPONSE: This was partially due to sample size constraints and because climate type does not translate into regions (see colored areas in Figure 1). Sample size is an issue particularly for Type 2 catchments, we are limited to 6 catchments, and catchment areas are not evenly distributed among the catchments shown in figure 2B (i.e., Type 1 has few small catchments and type 4 has few large catchments). Additionally, in balancing this reviewers suggestions with those of Anonymous Reviewer #2 who felt that we were overly optimistic in our statement of utility with respect to using GRUN for ungauged small tropical catchments we feel that only providing a national level bias correction is most appropriate for use at a regional/national scale.

(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

RESPONSE: This is an excellent suggestion, we have performed a bootstrap analysis on our regression and report the 68% (1 sigma) confidence intervals for bias correction, the text will now read: "Thus, we first add the mean log10(runoff) difference between the observations and the predicted values (0.117 $\pm$ 0.022). Following this, using the

lm function in R, we fit a linear regression between the observations and the GRUN predicted values (log10(runoff, observed) = m $\times$ 270 log10(runoff, predicted) + b) and correct the predicted values using the slope (m=0.774 $\pm$ 0.025) and intercept (b=0.099 $\pm$ 0.006) derived from this regression. Uncertainties reported here are 68% confidence intervals and were was assessed by bootstrap resampling 1,000 observation and prediction pairs without replacement 10,000 times."

(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).

RESPONSE: We agree with this comment entirely and thank the reviewer for pointing out these references. This is an important point, following the calculation of the bias correction values in our discussion we will add several sentences to discuss this potential bias and note that a correction factor (following Ferguson, 1986) may need to be applied. We will provide the value of the unbiased estimator of the variance (s in Ferguson, 1986) in the text. See also our response to comment 4 of Anonymous Reviewer 2.

(4) Other comments: L152-153: Perhaps, a more appropriate way of describing NSE $\leq$ 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”

RESPONSE: Agreed, as written it was inappropriate and convoluted. We will rewrite this sentence to read: “NSE values are useful (compared to VE) in that values less than zero indicate that the model is no better than using the mean value of the observed data as a predictor”

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

RESPONSE: Excellent point, that is likely due to training on GRUN’s large catchments elsewhere. Thus, we will rewrite this sentence to say both: “This suggests two possibilities: first, that particularly for small catchments which may have steeper average slope, GRUN underpredicts monthly runoff values associated with the wet season; and second model-data agreement improves with catchment size due to the fact that GRUN was trained on large (> 10,000 km2) catchments.”

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?

RESPONSE: This is a question of baseflow not being properly represented well. For large basins where the time of concentration (Tc) of any given flood event will be much longer, Flood peaks may be wider and subdued because of abstractions and infiltration into the shallow aquifer. This phenomenon is significantly less apparent in smaller basins where peak flows are expected to be higher because of less infiltration and

narrower peaks. We will add this explanation to the revised manuscript.

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.

RESPONSE: Thank you for pointing this out. We will instead say: “Since NSE puts more weight on large flow (Criss and Winston, 2008), it is not surprising that our NSE-log10 scores are in most cases significantly more skillful among our catchments because extremely wet months are weighted less than using the raw runoff values compared to raw NSE scores.”

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.

RESPONSE: Agreed. We will change to two sentences and change the parenthetical citation to: “As such, we suggest that GRUN, is a useful new tool for studying trends, seasonality and average runoff from tropical catchments (such as in previous work: e.g., Merz et al., 2011; Wanders and Wada, 2015) in the Philippines. However, we qualify this finding by noting that GRUN is not suitable for extreme value analyses associated with major tropical storms during the wet seasons unless suitable bias corrections (see next section) can be effectively carried out.”

Technical corrections L61: “(ref)”. Reference placeholder

RESPONSE: Thank you, this was a mistake on our part, we will cite: Hagemann et al. (2011, J. of Hydrometeorology), Davie et al. (2013, Earth System Dynamics) and Winsemius et al. (2016, Nature Climate Change) as examples in this parenthetical. Other examples welcome.

L99: URL is not working. Please check

RESPONSE: Thank you for checking, unfortunately this URL has gone defunct since

submission, we are contacting the DPWH who provided a new URL; however as of July 2020 this link is not yet working. All data used in this analysis is provided in our supplemental files.

L126: “that”.

RESPONSE: Will be removed.

Typo L206: “were”. Typo

RESPONSE: Will be changed to “where”

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-26, 2020.