**Response Letter to Reviewer #1**

Title: A Hydrological Emulator for Global Applications – HE v1.0.0

Journal: Geoscientific Model Development

*We would like to thank the referee for their detailed review of our manuscript and their positive feedback, constructive suggestions and criticisms. The responses to the Referee's comments are shown in* blue *font below. All the line numbers indicated refer to the main text of the revised manuscript (clean version without tracking changes).*

---

**Reviewers' Comments to Author:**

The manuscript by Liu et al. addresses the interesting issue of model complexity needed for global hydrological simulations. They present a new simulation tool based on the existing abcd model, and show that their simulations show a fair performance when compared with simulations from the VIC model. While I am generally supportive of work aimed at finding optimum model complexities, I feel the current study will need additional work to further show and quantify the benefits of the current code. At the moment, the main message seems to be that a low-dimensional model can produce positive correlations at the monthly timescale with another model, and that the runtime of the simple model is shorter. Both findings are not particularly new, and, in my view, they are not enough to merit publication. The suggested benefits of a simpler model (the possibility of focussing on uncertainty and spatial heterogeneity) might be true, but none of this is actually shown in the paper and no model or code is presented that takes full advantage of these suggested benefits. I believe the authors should present more work in this direction before the manuscript can be accepted for publication in GMD. My main concerns are the following:

**Comment 1**: The motivation for choosing the abcd model is poor. Many simple models exist, and no objective criteria were used to select this particular model. The authors could have started with a simpler version, and adding components/complexity until a pre-defined threshold performance was reached. This would have made the selection less arbitrary. How does the modelled runoff for instance compare to a baseline "model" which is simply the monthly P−PET? The choice for the abcd model should be motivated better, but preferably a more systematic approach should be taken.

Response 1: We have clarified the motivation for choosing the "*abcd*" model in lines 85-96:

"*To achieve our goal of identifying a suitable HE, we have explored many hydrological models to find one that may meet our needs. We start with a simple baseline model characterized by mean seasonal cycle; i.e., the inter-annual mean value for every calendar day (Schaefli & Gupta, 2007). Among others, we also explore the "abcd" model because: 1) it is widely-used and proven to have reasonable predictability (Fernandez et al., 2000; Martinez and Gupta, 2010; Sankarasubramanian and Vogel, 2002; Sankarasubramanian and Vogel, 2003; Thomas, 1981; Vandewiele and Xu, 1992; Vogel and Sankarasubramanian, 2003); 2) it uses a monthly time step and requires less computational cost than daily or hourly models; 3) it has solid physical basis hence has potential to be extended to other temporal scales (Wang and Tang, 2014); 4) it requires minimal and easily-available inputs; 5) it only involves 4-7 parameters; and 6) it can simulate variables of interest such as recharge, direct runoff and baseflow that many other simple models can't simulate (Vörösmarty et al., 1998)."*

Further, in lines 137-146, we have described the modifications we have made to the "abcd" model:

"*In this study, we then adopt the "abcd" framework from Martinez and Gupta (2010) (Fig. 1); meanwhile, we make three modifications to suit the needs of a HE for global applications. First, in order to enhance the model efficiency with as least necessary parameters as possible, instead of involving three tunable snow-related parameters in the calibration process, we set the values for two of the parameters (i.e., temperature threshold above or below which all precipitation falls as rainfall or snow) from literature (Wen et al., 2013) and only keep one tunable parameter m – snow melt coefficient (0 < m < 1). Second, we introduce the baseflow index (BFI) into the calibration process to improve the partition of total runoff between the direct runoff and baseflow (see Section 2.4). Third, other than the lumped scheme as previous studies used, we first explore the values of model application in distributed scheme with a grid resolution of 0.5 degree*."

We have also enhanced our analysis according to the referee's suggestions in terms of a simpler model in Comment 3 and 6. Specifically, we have added a baseline model to better justify the appropriateness of constructing the hydrological emulator (HE) based on the "*abcd*" model. We have added the descriptions for the baseline model in lines 107-114:

"*2.1.1 Baseline model*

*Following the work of Schaefli & Gupta (2007), we explore a baseline model characterized by the inter-annual mean value for every calendar day, i.e., climatology. In this study, we adapt the baseline model to monthly scale by first calculating inter-annual mean value for every calendar day from daily runoff of the benchmark product during 1971-2010 (see Section 2.3.2), and then aggregating daily runoff to monthly runoff. The model uses climatology for prediction, for example, if the inter-annual mean runoff for July in the Amazon basin is 100 mm $mon^{-1}$, then the prediction of total runoff for July of every year is 100 mm $mon^{-1}$*."

We have also added the comparison of performances between the baseline and the "abcd" model in lines 287-301 to elaborate its superiority over the baseline model:

"*Generally, we find baseline model performs worse than the "abcd" model (Fig. 2). The baseline model exhibits a lower global mean KGE value (0.61) than the lumped and distributed schemes of the "abcd" model (0.75 and 0.79, respectively). In addition, our analysis indicates that the incorporation of BFI into the objective function leads to significant improvement in the partition of total runoff between direct runoff and baseflow (Fig. S4), without compromising predictability for total runoff, i.e., the global mean KGE values for modeled total runoff with or without the incorporation of BFI are almost the same (0.75 vs 0.76). Specifically, for the case of involving both the total runoff and BFI in the objective function, the correlation efficiencies (r) between the long-term annual benchmark and modeled direct runoff, and between benchmark and baseflow from the lumped scheme across global basins are 0.97 and 0.96, respectively, which are much higher than those of 0.86 and 0.72 in the case of only involving the total runoff in the objective function (Fig. S4).*

*Given the superiority of the "abcd" model over the baseline model, we focus in the following sections on evaluating the predictability and computational efficiency of the "abcd" model and its potential to serve as a HE.*"

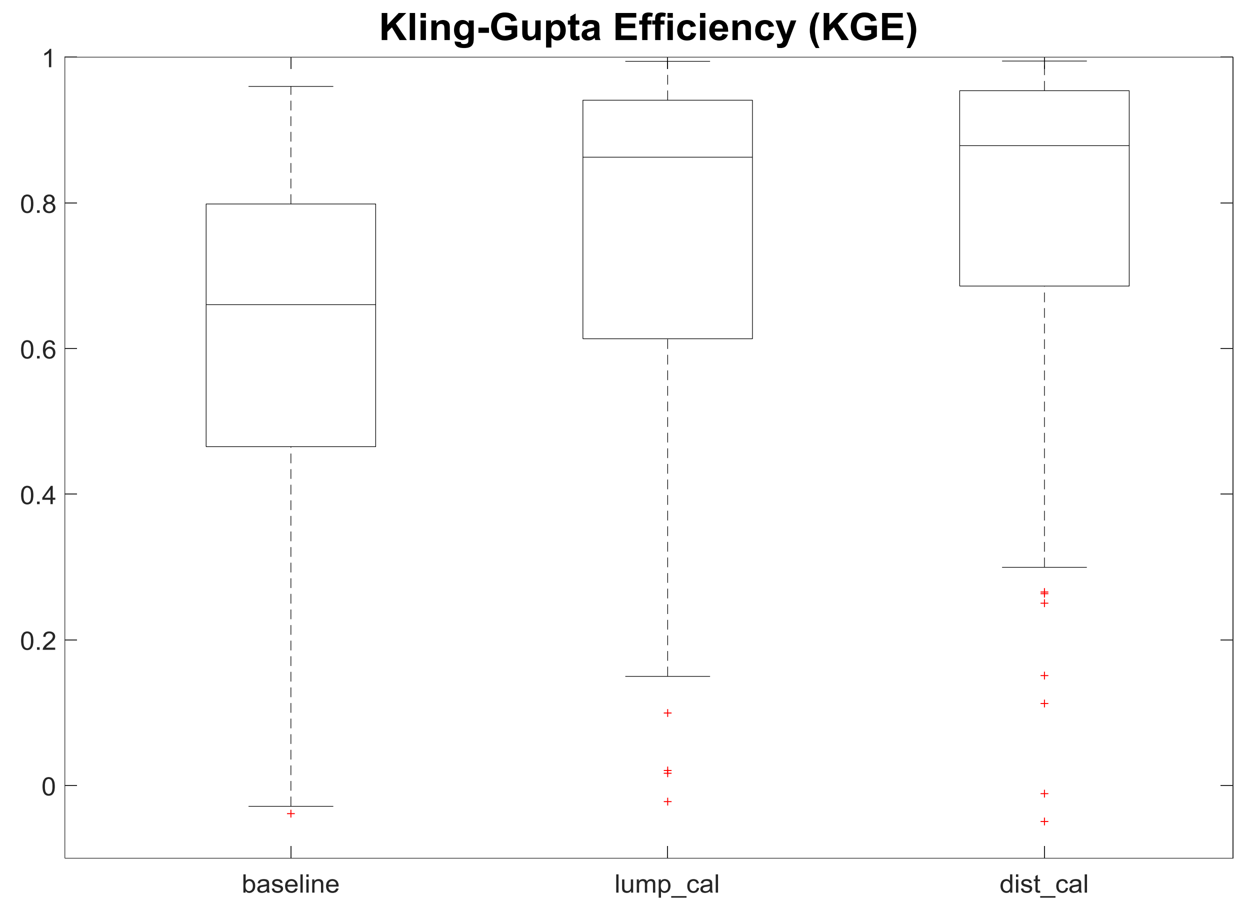


Figure 2. Kling-Gupta efficiency of the simulated basin-level total runoff across the global 235 basins (lump = lumped, dist = distributed, cal = calibration, the x-axis labels of "lump_cal" or "dist_cal" represent lumped/distributed scheme during calibration period).

References:
Schaefli, B. and Gupta, H.V., 2007. Do Nash values have value? Hydrological Processes, 21(15), 2075-2080.
Fernandez, W., Vogel, R., Sankarasubramanian, A., 2000. Regional calibration of a watershed model. Hydrol. Sci. J., 45(5): 689-707.

Martinez, G.F., Gupta, H.V., 2010. Toward improved identification of hydrological models: A diagnostic evaluation of the "abcd" monthly water balance model for the conterminous United States. Water Resour. Res., 46(8).

Sankarasubramanian, A., Vogel, R.M., 2002. Annual hydroclimatology of the United States. Water Resour. Res., 38(6).

Sankarasubramanian, A., Vogel, R.M., 2003. Hydroclimatology of the continental United States. Geophys. Res. Lett., 30(7).

Thomas, H., 1981. Improved methods for national water assessment. Report WR15249270, US Water Resource Council, Washington, DC.

Vandewiele, G., Xu, C.-Y., 1992. Methodology and comparative study of monthly water balance models in Belgium, China and Burma. J. Hydrol., 134(1-4): 315-347.

Vogel, R.M., Sankarasubramanian, A., 2003. Validation of a watershed model without calibration. Water Resour. Res., 39(10).

Vörösmarty, C.J., Federer, C.A., Schloss, A.L., 1998. Potential evaporation functions compared on US watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling. J. Hydrol., 207(3-4): 147-169.

Wang, D. and Y. Tang (2014), A one-parameter Budyko model for water balance captures emergent behavior in Darwinian hydrologic models, Geophysical Research Letters , 41, doi:10.1002/2014GL060509.

Wen, L., Nagabhatla, N., Lü, S., Wang, S.-Y., 2013. Impact of rain snow threshold temperature on snow depth simulation in land surface and regional atmospheric models. Adv. Atmos. Sci., 30(5): 1449-1460.

**Comment 2:** The notion that simple models can do a good job in describing the output of more complex models is not new. In particular, Gab Abramovic has written numerous papers on this topic. This work should be considered and used in the interpretation/motivation.

Response 2: We thank the referee for pointing out the useful references. We have added relevant papers of Gab Abramovic in the Introduction section (lines 69-75) to better justify our work of exploring a simple model:

"*In addition, some studies have shown that GHMs/LSMs are sometimes outperformed by simple empirical statistical models (Abramowitz, 2005; Abramowitz et al., 2008; Best et al., 2015), suggesting that some GHMs/LSMs may underutilize the information in their climate inputs and that model complexity may undermine accurate prediction. This also indicates the potential advantages of simple model over complex GHMs/LSMs. Thus, constructing simple models that can emulate the dynamics of more complex and computational expensive models (e.g., GHMs/LSMs) is warranted.*"

References:

Abramowitz, G., 2005. Towards a benchmark for land surface models. Geophys. Res. Lett., 32(22).

Abramowitz, G., Leuning, R., Clark, M., Pitman, A., 2008. Evaluating the performance of land surface models. J. Clim., 21(21): 5468-5481.

Best, M.J. et al., 2015. The plumbing of land surface models: benchmarking model performance. J. Hydrometeorol., 16(3): 1425-1442.

**Comment 3:** The motivation for the study is weak. In the current work, the authors only show a single application of their model (at grid and basin scales) and argue this is a good alternative to more complex models. But why not use the output of these complex models directly if the main goal is a best assessment of monthly average predictions of water balance partitioning? Such (multi-model) output is readily available at the global scale and does not require the running of even a simple model. Of course a simple model can be used for sophisticated uncertainty assessment (important advantage), but the authors did not yet do any work in this direction. This should be part of a revised version.

Response 3: The main merit of a hydrological emulator (HE) is its capability of emulating complex global hydrological models (GHMs). Multi-model projects, such as ISI-MIP, do provide outputs like global runoff, but the available products are still very limited. The HE developed in this study provide an easy-to-use and open-source tool for the community to emulate GHMs of interest and simulate any scenarios of interest with reasonable predictability and high computational efficiency, which is a capability that is computationally prohibitive for multi-model projects using GHMs. Some related explanation has been added in lines 215-218:
"*Despite potential bias in the VIC runoff product, using it as a benchmark here is to demonstrate the capability of the HE developed in this work to mimic complex GHMs. Furthermore, the application of the HE is not tied to the VIC model and should be able to emulate other GHMs*."

We also have clarified the usage of the HE in lines 399-404:
"*Based upon our open-source HE and the validated basin-specific parameters across the globe, researchers can easily investigate the variations in water budgets at the basin/ regional/global scale of interest, with minimum requirements of input data, efficient computation performance and reasonable model fidelity. Likewise, researchers can utilize the framework of the HE with any alternative input data, or recalibrate the HE to emulate other complex GHMs or LSMs of interest, to meet their own needs.*"

Further, we have followed the reviewer's suggestion and have conducted an uncertainty analysis (UA) to demonstrate the advantage of the HE. We have added the UA in Section 3.5 as follows:

"*3.5 Case study for uncertainty analysis*
*To demonstrate the capability of the examined "abcd" model serving as a HE, we use the lumped scheme to conduct parameter-induced uncertainty analysis for the runoff simulation at the world's sixteen river basins with top annual flow (Dai et al. 2009). Specifically, for each of the sixteen basins, we first apply ±10% change to each of the five calibrated parameters (a, b, c, d, m) to compose varying ranges; note that we just truncate the range to those valid in Table 1 if the ±10% change exceeds the valid range. Then we randomly sample the five parameters from corresponding ranges for 100,000 times (i.e., 100,000 combinations of parameters). After that, we run the lumped scheme 100,000 times for each basin with the 100,000 combinations of parameters to examine the parameter-induced uncertainty in total runoff. The uncertainty analysis indicates that most basins are robust to changes in parameters, other than the Tocantins, Congo and La Plata (Fig. 7). In other words, for basins Congo and La Plata, slight changes in parameters may lead to large changes in runoff estimates. Then the uncertainty in the calibrated parameters for the two basins may lead to large bias in the simulated runoff, which may more or less explain why modelled runoff for the two basins tend to have higher biases than other basins (Fig. 4). Notably, the 100,000 times of simulations only takes ~80 seconds on a Dell Workstation T5810 with one Intel Xeon 3.5 GHz CPU, which demonstrates the extraordinary computational efficiency of the lumped scheme and its advantage for serving as a HE.*"

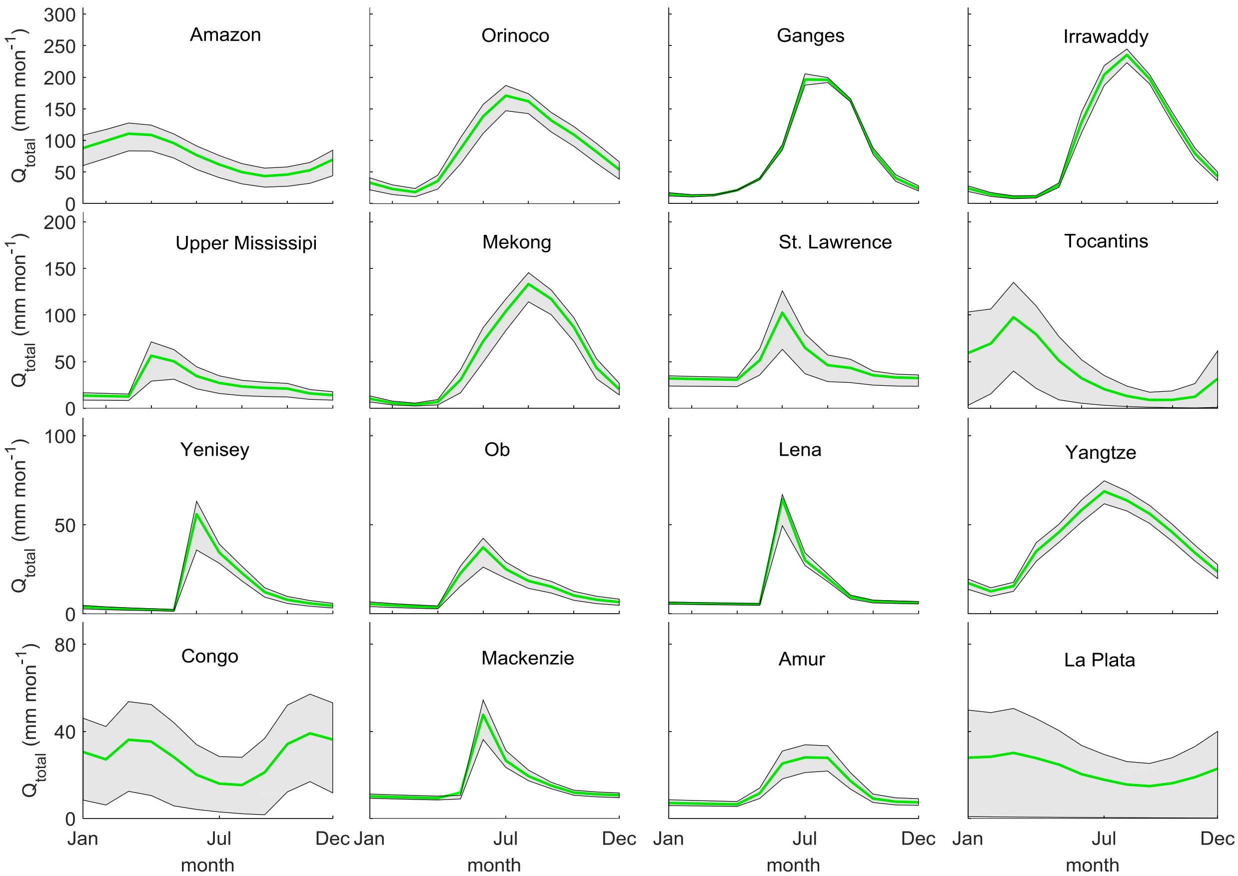


Figure 7 Parameter-induced uncertainty in total runoff for the world's sixteen river basins with top annual flow. The green line stands for simulated total runoff using the calibrated parameters, and the gray area represents the spread derived from variations in parameters.

References:

Dai, A., Qian, T., Trenberth, K.E., Milliman, J.D., 2009. Changes in continental freshwater discharge from 1948 to 2004. J. Clim., 22(10): 2773-2792.

**Comment 4:** The choice for the VIC model is poorly motivated. While I agree that some studies have shown that VIC produces positive NSE scores against observations, many of these studies evaluated their results at very course time resolutions at which nearly any model would show a good performance (in particular because at monthly timescales the seasonal cycle dominates, which is easy to reproduce). The VIC model will generally not work well when evaluated at hourly or daily timesteps, even when calibrated. Related to this point is the issue of temporal resolution. It can be questioned whether nonlinear processes such as snow accumulation and melt can be modelled at a monthly timestep and at course spatial scales (see Melsen et al., Hydrol. Earth Syst. Sci. doi:10.5194/hess-20-1069- 2016). In order to show that this is indeed possible, the authors should show that their model is able to outperform a baseline model consisting of, for instance, a mean seasonal cycle (as in Schaefli & Gupta, Hydrol. Process. 21, 2075–2080).

Response 4: We agree with the reviewer about the performance of the VIC at fine time-steps (e.g., hourly). The essential point of this work is not to emulate VIC, but using VIC as an example to demonstrate the HE developed in this study could be used to emulate any global hydrological models (GHMs) of interest. We have clarified this in lines 215-218:

"*Despite potential bias in the VIC runoff product, using it as a benchmark here is to demonstrate the capability of the HE developed in this work to mimic complex GHMs. Furthermore, the application of the HE is not tied to the VIC model and should be able to emulate other GHMs.*"

Due to the requirement of high computational efficiency in addition to reasonable predictability, daily or sub-daily time step is not suitable for the HE, so we use monthly time step. In terms of the processes such as snow accumulation and melt, Martinez and Gupta (2010) have shown that the incorporation of snow processes in the monthly "*abcd*" model significantly improves the model performance in regions with snow cover. This is why we adopt the "*abcd*" version with the snow module (Martinez and Gupta 2010) in this study. We have clarified this in lines 132-136:

"*The work of Martinez and Gupta (2010) has added snow processes into the original "abcd" model, where the snowpack accumulation and snow melt are estimated based on air temperature. Their work indicated that incorporation of the snow processes in the monthly "abcd" model has significantly improved model performance in snow-covered area in the conterminous United States (see Figure 4 in Martinez and Gupta (2010)).*"

Other than that, we have followed the reviewer's suggestions and have added a baseline model in this work to reveal the superiority of the adopted "*abcd*" model over the baseline model, for details please see the Response 1.

References:

Martinez, G.F., Gupta, H.V., 2010. Toward improved identification of hydrological models: A diagnostic evaluation of the "abcd" monthly water balance model for the conterminous United States. Water Resour. Res., 46(8).