Response Letter

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

Journal: Geoscientific Model Development

We would like to thank the Editor and the referees for their detailed review of our manuscript and their positive feedback, constructive suggestions and criticisms. The responses to the Referees' 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).

Editor's comments:

In my role as Executive editor of GMD, I would like to bring to your attention our Editorial version 1.1:

http://www.geosci-model-dev.net/8/3487/2015/gmd-8-3487-2015.html

This highlights some requirements of papers published in GMD, which is also available on the GMD website in the 'Manuscript Types' section:

http://www.geoscientific-model-development.net/submission/manuscript_types.html

From your abstract and introduction I understand that you describe the newly developed Hydrological Emulator and evaluate its results. Therefore your paper is not an "Evaluation paper" but a "Development and Technical paper" and thus the criteria of this paper type are applied.

These are in particular:

Comment 1: "The main paper must give the model name and version number (or other unique identifier) in the title."

Response 1: We thank the Editor for all the comments and for allowing us to revise the manuscript. We will change the "Manuscript type" to "Development and Technical paper" during our submission of the revision, and we have added model name and version number in the title:

"*A Hydrological Emulator for Global Applications – HE v1.0.0*"

Comment 2: "All papers must include a section, at the end of the paper, entitled 'Code availability'.

Here, either instructions for obtaining the code, or the reasons why the code is not available should be clearly stated. It is preferred for the code to be uploaded as a supplement or to be made available at a data repository with an associated DOI (digital object identifier) for the exact model version described in the paper. Alternatively, for established models, there may be an existing means of accessing the code through a particular system. In this case, there must exist a means of permanently accessing the precise model version described in the paper. In some cases, authors may prefer to put models on their own website, or to act as a point of contact for obtaining the code. Given the impermanence of websites and email addresses, this is not encouraged, and authors should consider improving the availability with a more permanent arrangement. After the paper is accepted the model archive should be updated to include a link to the GMD paper."

Therefore please provide a version number and preferably the acronym used within the article (HE). Additionally, we strongly recomment to make the exact code version, your article refers to, available via a permanent archive providing a DOI (e.g. Zenodo).

Response 2: We have tried our best to meet the journal requirements in terms of code availability. First, we have created a repository in the open-source software site GitHub (https://github.com/JGCRI/hydro-emulator/) to make the hydrological emulator freely available. We have released the version of the specific HE v1.0.0 referenced in this paper on https://github.com/JGCRI/hydro-emulator/releases/tag/v1.0.0, where the source code (written in Matlab), all related data inputs and outputs, as well as the detailed Readme file are available. The repository is maintained by our organization, the Joint Global Change Research Institute (JGCRI), and long-term commitment for maintaining the repository is a standard practice. For example, both Le Page et al. (2016) and Hartin et al. (2015) published in Geoscientific Model Development (GMD) provided their codes on the GitHub site maintained by JGCRI (https://github.com/JGCRI/). Second, there is an ongoing effort to incorporate the hydrological emulator developed in this study to Xanthos (Li et al., 2017, https://github.com/JGCRI/xanthos), which is an open-source global hydrologic model, and the code for the HE referenced in this paper will also be freely available in the next version of Xanthos.

We have clarified it in the "Code and/or data availability" section as follows:

"*The hydrological emulator (HE) is freely available on the open-source software site GitHub (*https://github.com/JGCRI/hydro-emulator/). *We have released the version of the specific HE v1.0.0 referenced in this paper on https://github.com/JGCRI/hydro-emulator/releases/tag/v1.0.0, where the source code (written in Matlab), all related inputs, calibrated parameters and outputs for each of the global 235 basins, as well as the detailed Readme file are available.*"

References:

Hartin, C.A., Patel, P., Schwarber, A., Link, R.P. and Bond-Lamberty, B.P., 2015. A simple object-oriented and open-source model for scientific and policy analyses of the global climate system–Hector v1. 0. Geoscientific Model Development, 8(4), pp.939-955.

Le Page, Y., West, T.O., Link, R., Patel, P., 2016. Downscaling land use and land cover from the Global Change Assessment Model for coupling with Earth system models. Geoscientific Model Development, 9(9), p.3055.

Li, X., Vernon, C.R., Hejazi, M.I., Link, R.P., Feng, L., Liu, Y., Rauchenstein, L.T., 2017, Xanthos – A Global Hydrologic Model, Journal of Open Research Software, 5(1), p.21.

Reviewers' Comments to Author:
Reviewer 1

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 3: 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 3: 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."

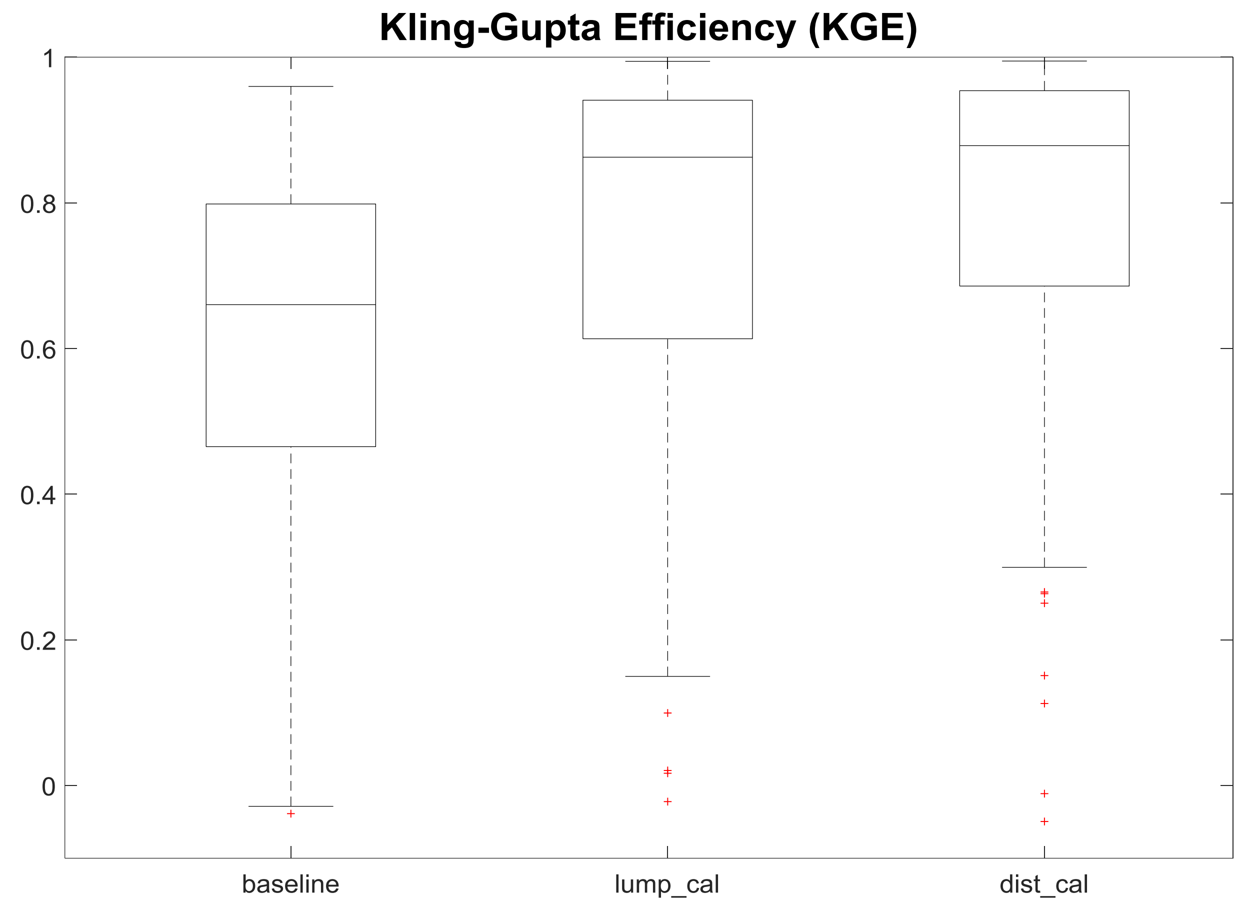


Figure 2. Kling-Gupta efficiency of the simulated basin-level total runoff across the global 235 basins (lump = lumped, dist = distributed, cal = calibration, the x-axis labels of "lump_cal" or "dist_cal" represent lumped/distributed scheme during calibration period).

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 4: 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 4: 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 5: 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 5: 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."

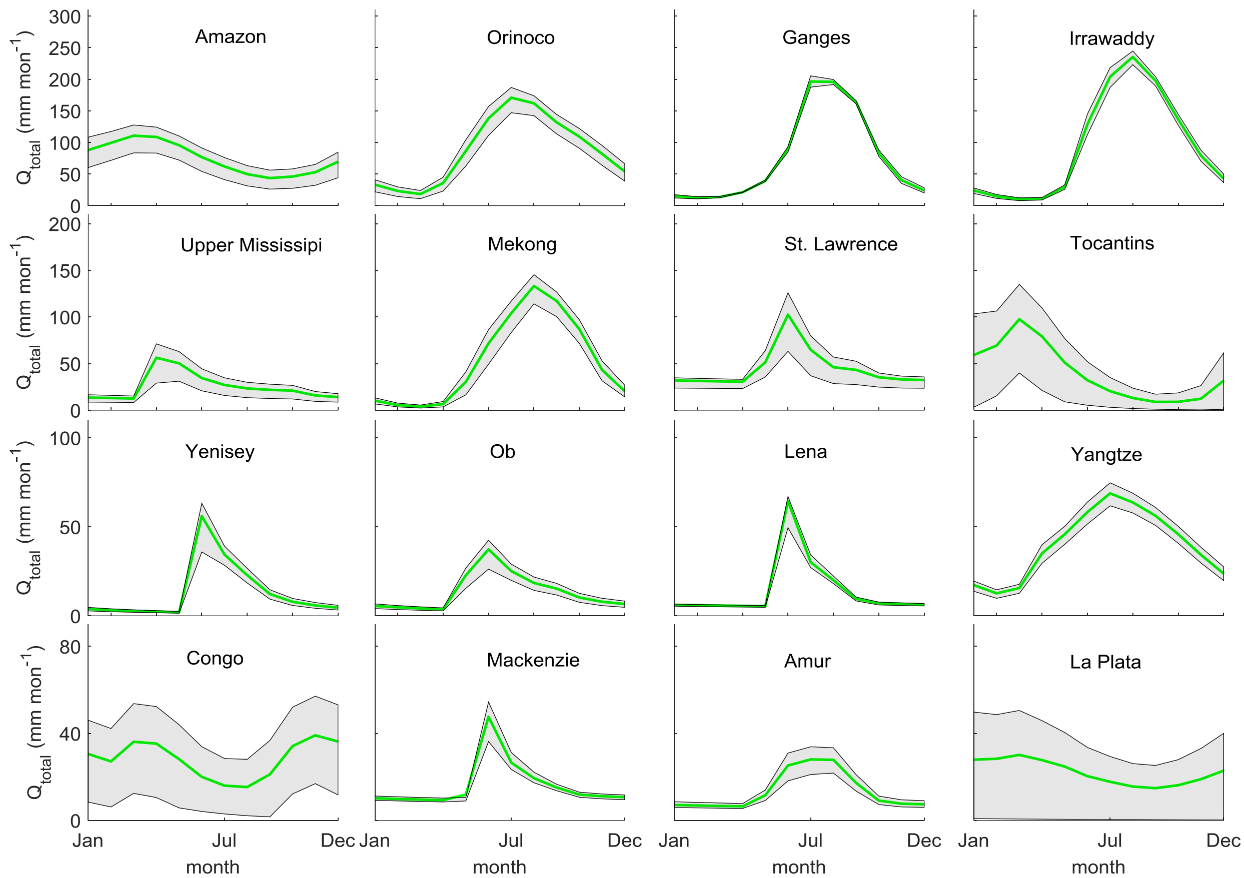


Figure 7 Parameter-induced uncertainty in total runoff for the world's sixteen river basins with top annual flow. The green line stands for simulated total runoff using the calibrated parameters, and the gray area represents the spread derived from variations in parameters.

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 6: 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 6: 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 3.

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).

Reviewers' Comments to Author:
Reviewer 2

In this study, the authors use a simple hydrological model "abcd" to emulate the behavior of more complex models (e.g. VIC). They modify the abcd model by including the baseflow index to better represent the partition of total runoff into direct runoff and baseflow. They present a lumped and a distributed version of the model, which are calibrated using the GA technique. They apply the model on global scale and compare the results against VIC simulations. Based on the results, they provide recommendations on the use of different versions of the model. Although the model used is not new and the concept of simplified emulator is an established one, however, the global-scale application of the model and its assessment over multiple basins across globe make it an interesting study. A simple and computationally efficient emulator that can work well on global scale is useful for several applications.

I think the manuscript at its current stage needs some more work. Some additional analyses need to be added. Therefore, I suggest moderate revisions for the manuscript before it is accepted in GMD. Following are my comments:

Comment 7: How reliable are the VIC simulations? Calibration of VIC can significantly change its streamflow outputs. So what type of simulations are used in this case for the comparison purpose? How were the soil and vegetation parameters calibrated/selected? All these points need to be discussed in greater details.

Response 7: The VIC runoff product (Hattermann et al., 2017; Leng et al. 2015) is used as a benchmark in this study, and its use is merely to demonstrate the capability of the hydrological emulator (HE) developed in this work to mimic complex global hydrological models (GHMs). Despite the potential bias in the VIC product, it does not affect the key findings of this work about the capability of the HE. We have added detailed descriptions about the VIC simulations in lines 186-206:

"*The VIC runoff product here is a global simulation with a daily time step and spatial resolution of 0.5 degree for the period of 1971-2010, and the VIC daily runoff is aggregated to monthly data to be consistent with the temporal scale of the "abcd" model. The VIC model settings used in this study are based on the University of Washington VIC Global applications (http://www.hydro.washington.edu/Lettenmaier/Models/VIC/Datasets/Datasets.shtml). The sub-grid variability of soil, vegetation and terrain characteristics are represented in sub-grid area-specific parameter classifications. Soil texture and bulk densities are derived by combining the World Inventory of Soil Emission Potentials database (Batjes, 1995) and the 5-min digital soil map of the world from the Food and Agricultural Organization (FAO, 1998). Based on the work of Cosby et al. (1984), the remaining soil properties (e.g. porosity, saturated hydraulic conductivity and unsaturated hydraulic conductivity) are derived. Vegetation type data are obtained from the global land classification of Hansen et al. (2000). Parameters including the infiltration parameter, soil layer depths and those governing the baseflow function were calibrated for major global river basins and transferred to the global domain as documented in Nijssen et al. (2001b), based on which Zhang et al. (2014) and Leng et al. (2015) conducted additional calibrations in the China domain. In this study, the VIC model was forced by WATCH climate forcing at the daily time step Weedon et al. (2011), based on the calibrated parameters from Nijssen et al. (2001b), Zhang et al. (2014) and Leng et al. (2015). The simulated runoff used in this study has recently been validated globally within the framework of the Inter-Sectoral Impact Model Intercomparison Project and shows reasonable performance compared to other hydrological models (Hattermann et al., 2017; Krysanova and Hattermann, 2017).*"

Further, we compared the VIC product to other products to corroborate its appropriateness. The comparison is presented in lines 219-229:

"The VIC runoff product also compares well to other products (see Fig. S1, S2), including the UNH/GRDC runoff product (Fekete and Vorosmarty, 2011; Fekete et al., 2002) and the global streamflow product (Dai et al., 2009). The scatterplot pattern of the VIC long-term annual runoff product vs. the streamflow product matches well with that of the UNH/GRDC runoff vs. the streamflow product (streamflow is transferred to the same unit as runoff via dividing by the basin area). Further, the

correlation coefficient of the VIC and the UNH/GRDC long-term annual runoff is as high as 0.83 across the global 235 basins. This suggests the reasonability of VIC runoff product, because the UNH/GRDC runoff is calibrated with the GRDC observations. At the same time, the discrepancies between the VIC runoff products and the streamflow products (Fig. S2) may be attributed to human activities, such as reservoir regulations and upstream water withdrawals, which are not embedded in the runoff but reflected in the streamflow."

References:

Batjes, N., 1995. A homogenized soil data file for global environmental research: A subset of FAO, ISRIC and NRCS profiles (Version 1.0), ISRIC.

Food and Agriculture Organization (FAO), 1998. Digital soil map of the world and derived soil properties. Land and 410 Water Digital Media Series 1, CD-ROM.

Hansen, M., DeFries, R., Townshend, J.R., Sohlberg, R., 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. Int. J. Remote Sens., 21(6-7): 1331-1364.

Leng, G., Tang, Q., Rayburg, S., 2015. Climate change impacts on meteorological, agricultural and hydrological droughts in China. Global Planet. Change, 126: 23-34.

Cosby, B.J., Hornberger, G.M., Clapp, R.B. and Ginn, T., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water resources research, 20(6), pp.682-690.

Krysanova, V., & Hattermann, F. F. (2017). Intercomparison of climate change impacts in 12 large river basins: overview of methods and summary of results. Climatic Change, 141(3), 363-379.

Hattermann, F. F. et al. (2017). Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. Climatic Change, 141(3), 561-576.

Nijssen, B.N., G.M. O'Donnell, D.P. Lettenmaier and E.F. Wood, 2001: Predicting the discharge of global rivers, Journal of Climate, 14(15), 3307-3323

Weedon, G. et al., 2011. Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. J. Hydrometeorol., 12(5): 823-848.

Zhang, X. (2014). A long-term land surface hydrologic fluxes and states dataset for China. Journal of Hydrometeorology, 15(5), 2067-2084.

Comment 8: How did VIC perform in the extreme climate regions, for example, in snow dominated catchments? This issue needs to be addressed properly. Maybe you can explore the following cases: Case-1: If both emulator (E) and model (M) are matching the observations (O) well then that's great. There could be some sub-cases for this case: (i) Both emulator and model match the observations well but from different directions (M – O – E). For example, they might have opposite (positive/negative) bias errors but the absolute values of the errors could be close. (ii) The model is matching the observations well and the emulator is matching the model well, all in one direction (O – M – E). (iii) The emulator is matching well both the model and the observations, but in different directions (M – E – O). Case-2: If none of them are matching the observations well but their own outputs match each other, then too, I think an emulator is serving its purpose in a way (although not quite useful). Case-3: If the emulator is matching the observations well but the model isn't then that's an interesting finding. Case-4: If the model is matching the observations well but the emulator isn't then there is a problem. Therefore, this needs to be explored in greater depth.

Response 8: We thank the referee for the detailed comments regarding the comparison between the VIC model and the hydrological emulator (HE). The essential point of this work is to deliver an open-source and easy-to-use hydrological emulator that can be used for emulating global hydrological models (GHMs) of interest. VIC is used as an example GHM in this study to demonstrate the capability of the HE to emulate complex and computationally expensive GHMs (see also Response 6). Exploring the sources of differences between the performance of the VIC and the HE is outside the focus of this work, and it would be incorporated in our future work.

Comment 9: At seasonal time scales, the model performance is expected to be better. It would be crucial to also check the results on daily time scale. Maybe you can produce a set of time series plots, scatter plots, and spatial contour plots for daily level, as done for the seasonal case.

Response 9: We agree with the reviewer on the better performance of monthly time scale than that of daily, however, due to the needs of high computational efficiency for the hydrological emulator (HE), it is simulated at monthly time step and a daily time series comparison is not even feasible in this case.

Comment 10: Figure 3: Any idea why there are those biases in the lower streamflow values? Is there any location-specific pattern of these biases?

Response 10: From the uncertainty analysis we added in Section 3.5 (see also Figure 7), it shows basins like Congo and La Plata are not as robust as other basins to changes in parameters – 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. We have added discussions in lines 415-421:

"*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)*."

Comment 11: Line 113: Which one's the other parameter you adopt the value of?

Response 11: We adapt two snow-related parameters from literature, and make the other one – snow melt coefficient – tunable during the calibration process, it has been clarified in lines 139-143:

"*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)*."

References:

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 12: Line 200: Did you try different weights on the two objectives?

Response 12: No, we use the same weights for the two objectives as we believe the two objectives are equally important.

Comment 13: Line 290: In order to do a fair comparison, VIC and the two versions of the models should be run on the same computer, preferably with good configuration.

Response 13: Yes, we have clarified this in lines 365-367:

"*Note that all of the simulations here are conducted on the Pacific Northwest National Laboratory (PNNL)'s Institutional Computing (PIC) Constance cluster using 1 core (Intel Xeon 2.3 GHz CPU) with the same configuration*."

Comment 14: Figure S1: I am not sure if you can say that all of them are comparing well. The discrepancies/mismatches should be clearly discussed in the manuscript. You are only showing the correlations here. What about the bias error?

Response 14: We thank the referee for the concern. Figure S1 is to illustrate the relationship of VIC and UNH/GRDC runoff product with streamflow measurements at gauge stations, and the similar scatter patterns between the upper and lower panel indicates the similarity of the two runoff products. This analysis is to reveal the appropriateness of the VIC runoff product as a benchmark product in this work. The discrepancies between runoff products and streamflow measurements are induced from the ignorance of river routing, reservoir regulations and upstream water withdrawals in the simulated runoff products. This has been recognized in the main text (lines 226-229):

"*At the same time, the discrepancies between the VIC runoff products and the streamflow products (Fig. S2) may be attributed to human activities, such as reservoir regulations and upstream water withdrawals, which are not embedded in the runoff but reflected in the streamflow.*"

However, exploring the sources and magnitudes of the discrepancies among them is outside the focus of this study.

Comment 15: X-axis marks are missing for the first two subplots of Figure S1. Use same axis for the scatter plots in Figure 2.

Response 15: The X-axis marks for Figure S1 have been fixed. For the previous Figure 2 (currently Figure 3 in the revised manuscript), we use different axis for total runoff, direct runoff and baseflow is because they have different magnitude, and this may make the figure and scatter points more discernable.

Comment 16: My comments about the manuscript: Writing: The manuscript is very well written. I don't have any suggestions on this part. Figures: Figures look good. Increase the legend in Figure 3. Tables: Table 2 can go to the supplementary materials.

Response 16: We have increased the legend in the previous Figure 3 (currently Figure 4 in the revised manuscript) and moved Table 2 to the supplementary materials as suggested.

A Hydrological Emulator for Global Applications – HE v1.0.0

Yaling Liu[1], Mohamad Hejazi[1], Hongyi Li[2], Xuesong Zhang[1], Guoyong Leng[1]

[1]Joint Global Change Research Institute, Pacific Northwest National Laboratory, 5825 University Research Court, College Park, Maryland 20740, United States

[2] Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717, United States

Correspondence to: Yaling Liu (cauliuyaling@gmail.com)

Abstract

While global hydrological models (GHMs) are very useful in exploring water resources and interactions between the Earth and human systems, their use often requires numerous model inputs, complex model calibration, and high computation costs. To overcome these challenges, we construct an efficient open-source and ready-to-use hydrologic emulator (HE) that can mimic complex GHMs at a range of spatial scales (e.g., basin, region, globe). More specifically, ~~We~~ we ~~then~~ construct~~ed~~ both a lumped and a distributed scheme of the HE based on the monthly *"abcd"* model to explore the tradeoff ~~, for users' choice – minimal~~between computational cost and ~~with reasonable~~ model fidelity~~, or heavier computational load with better predictability~~. Model predictability and computational efficiency were evaluated in simulating global runoff from 1971-2010 with both the lumped and distributed schemes. The results are compared against the runoff product from the widely-used Variable Infiltration Capacity (VIC) model. Our evaluation indicates that the lumped and distributed schemes present comparable results regarding annual total quantity, spatial pattern and temporal variation of the major water fluxes (e.g., total runoff, evapotranspiration) across the global 235 basins (e.g., correlation coefficient r between the annual total runoff from either of these two schemes and the VIC is >0.96), except for several cold (e.g., Arctic, Interior Tibet), dry (e.g., North Africa) and mountainous (e.g., Argentina) regions. Compared against the monthly total runoff product from the VIC (aggregated from daily runoff), the global mean Kling-Gupta efficiencies are 0.75 and 0.79 for the lumped and distributed schemes, respectively, with the distributed ~~one~~ scheme better capturing spatial heterogeneity. Notably, the computation efficiency of the lumped scheme is two orders of magnitude higher than the distributed one, and seven orders more efficient than the VIC model. A case study of uncertainty analysis for the world's sixteen basins with top annual streamflow is

conducted using 100,000 model simulations, and it demonstrates the lumped scheme's extraordinary advantage in computational efficiency. Our results suggest that the revised lumped "*abcd*" model can serve as an efficient and acceptable HE for complex GHMs and is suitable for broad practical use, and the distributed scheme is also an efficient alternative if spatial heterogeneity is of more interest.

1 Introduction

A global hydrological model (GHM) is an effective tool to understand how water moves between soil, plants and the atmosphere. In terms of spatial discretization, hydrological models can be classified into: 1) lumped models treating one basin as a homogeneous whole and disregarding spatial variations, such as the Sacramento Soil Moisture Accounting Model (Burnash et al., 1973); and 2) distributed models where the entire basin is divided into small spatial units (e.g., square cells or triangulated irregular network) to capture spatial variability, such as the PCRaster Global Water Balance (Van Beek and Bierkens, 2009) and the WASMOD-M (Widén-Nilsson et al., 2007). For simplicity, models with division of one basin into separate areas or sub-basins are also categorized as distributed ones here. The corresponding predictability and computational efficiency of GHMs may vary from model to model, due to difference in complexity and structure. Recent years have seen rapid progress in GHMs. They are widely used in assessing the impacts of climate change and land surface changes on the water cycle (Alcamo and Henrichs, 2002; Arnell and Gosling, 2013; Liu et al., 2013; Liu et al., 2014; Nijssen et al., 2001a), exploring spatial and temporal distribution of water resources (Abdulla et al., 1996; Alkama et al., 2010; Bierkens and Van Beek, 2009; Gerten et al., 2005; Tang et al., 2010), examining how human activities alter water demand and water resources (De Graaf et al., 2014; Döll et al., 2009; Hanasaki et al., 2008; Liu et al., 2015; Rost et al., 2008; Vörösmarty et al., 2000), and investigating the interactions between human activities and water availability by incorporating GHM with integrated assessment models (Kim et al., 2016).

Applying GHMs usually requires miscellaneous inputs, high computational costs, and a complex calibration process. These challenges stand out in practical situations, especially when the computational resources are limited. For instance, sensitivity analysis and uncertainty quantification are often needed for decision making, but the users usually cannot afford to run a large number of simulations with many GHMs like the VIC (also categorized as land surface model (LSM)) due to their high computational expense (Oubeidillah et al., 2014). Another situation is when the users seek reasonable estimates of water resources with minimal efforts rather than acquiring highly accurate estimates through expensive inputs of time and efforts. For example, when users seek to explore the hydroclimatology of a region and its long-term water balance (Sankarasubramanian and Vogel, 2002), then GHMs with fine spatial (e.g., 1/8 degree) and temporal resolution (e.g., hourly) are not necessarily needed. In this case, ~~GHMs~~ simple models that possess reasonable predictability and are computationally efficient tend to be more suitable. 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.

The motivation of this work arises from the need to construct a hydrological emulator (HE) that can efficiently mimic the complex GHMs to address the abovementioned issues for practical use, which provides the opportunity of speeding up simulations at the cost of introducing some simplification. We develop a HE that is ready-to-use and efficient for any interested groups or individuals to assess water cycle at basin/regional/global scales. This HE possesses the following features: 1) minimum number of parameters; 2) minimal climate input that is easy to acquire; 3) simple model structure; 4) reasonable model fidelity that captures both the spatial and temporal variability; 5) high computational efficiency; 6) applicable in a range of spatial scales; and 7) open-source and well-documented.

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); ~~3~~4) it requires minimal and easily-available inputs; ~~4~~5) it only involves 4-7 parameters; and ~~5~~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) ~~it can estimate several variables of interest to a wide range of users (e.g., total runoff, baseflow, direct runoff, groundwater recharge, evapotranspiration)~~. ~~For the first time we apply~~ This study marks the first time that the "*abcd*" based model is applied globally, and also the first time the predictability and computational efficiency for both the lumped and distributed schemes are evaluated. Below we describe the baseline and the "*abcd*" models and data in Section 2; and we present the evaluation of the two models, discuss their appropriateness of serving as a HE in Section 3; finally, in Section 4 we summarize this work with concluding remarks.

2 Methods and data

2.1 Model description

We examine two simple models – baseline and the "*abcd*" model (both lumped and distributed scheme) in order to identify a suitable one for serving as a HE.

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}.

2.1.2 The "*abcd*" model

The monthly *"abcd"* model was first introduced by Thomas (1981) to improve the national water assessment for the U.S., with a simple analytical framework using only a few descriptive parameters. It has been widely used across the world, especially for the U.S. (Martinez and Gupta, 2010; Sankarasubramanian and Vogel, 2002; Sankarasubramanian and Vogel, 2003). The model uses potential evapotranspiration (PET) and precipitation (P) as input. The model defines four parameters *a*, *b*, *c*, and *d* that reflect regime characteristics (Sankarasubramanian and Vogel, 2002; Thomas, 1981) to simulate water fluxes (e.g., evapotranspiration, runoff, groundwater recharge) and pools (e.g., soil moisture, groundwater). The parameters *a* and *b* pertain to runoff characteristics, and *c* and *d* relate to groundwater. Specifically, the parameter *a* reflects the propensity of runoff to occur before the soil is fully saturated. The parameter b is an upper limit on the sum of evapotranspiration (ET) and soil moisture storage. The parameter c indicates the degree of recharge to groundwater and is related to the fraction of mean runoff that arises from groundwater discharge. The parameter d is the release rate of groundwater to baseflow, and thus the reciprocal of d is the groundwater residence time. Snow is not part of the original "*abcd*" model, which may result in poor performance of the model in cold regions where snow significantly affects the hydrological cycle. 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)).

In this study, we 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.~~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. The detailed model descriptions and equations are presented in the Appendix A, and the descriptions and ranges of model parameters are listed in Table 1.

2.2 Model structure

In terms of the *"abcd"* model, we evaluate both the lumped and distributed model schemes, although most previous applications of the model are conducted in a lumped scheme (Bai et al., 2015; Fernandez et al., 2000; Martinez and Gupta, 2010; Sankarasubramanian and Vogel, 2002; Sankarasubramanian and Vogel, 2003; Vandewiele and Xu, 1992; Vogel and Sankarasubramanian, 2003). In the lumped scheme, each of the 235 river basins is lumped as a single unit, and each of the climate input (see Section 2.3.1) is the lumped average across the entire basin, and thus all the model outputs are lumped as well. In terms of the distributed one, however, each 0.5-degree grid cell has its own climate inputs, and likewise, the model outputs are simulated at the grid-level. Although the two schemes differ in the spatial resolution of their inputs and outputs, their within-basin parameters are uniform. We use basin-uniform rather than grid-specific parameters for the distributed scheme for two reasons: 1) to enhance computational efficiency; and 2) to avoid drastically different parameters for neighboring grid cells that may be unrealistic. Note that lateral flows between grid cells and basins are not included at this stage for the *"abcd"* model. For the baseline model, as it is derived from the benchmark product (see Section 2.3.2), which presents runoff estimates in a spatial resolution of 0.5-degree, and thus every grid cell of each basin has its own inter-annual mean monthly runoff estimates.

2.3 Data

2.3.1 Climate data

The climate data needed for the *"abcd"* model only involve monthly total precipitation, monthly mean, maximum and minimum air temperature. The data we use is obtained from WATCH (Weedon et al., 2011), spanning the period of 1971-2010, and it is 0.5-degree gridded global monthly data. The climate data is used for model simulation over the global 235 major river basins (Kim et al., 2016). Additionally, we use the Hargreaves-Samani method (Hargreaves and Samani, 1982) to estimate potential evapotranspiration (PET), which is a required input for the *"abcd"* model, and it needs climate data of mean, maximum and minimum temperatures for the calculation.

2.3.2 Benchmark runoff product

In this study, the "*abcd*" model is tested for its ability to emulate the naturalized hydrological processes of a reference model since the "true" naturalized hydrological processes are unknown. The "perfect model" approach is well adopted in climate modeling studies where one model is treated as "observations" while the others are tested for their ability to reproduce "observations" (Murphy et al., 2004; Tebaldi and Knutti, 2007). Here, we use the process-based VIC model as the "perfect model", which was also driven by the WATCH climate forcing.

The VIC runoff product here is a global simulation with a daily time step and spatial resolution of 0.5 degree for the period of 1971-2010, and the VIC daily runoff is aggregated to monthly data to be consistent with the temporal scale of the "*abcd*" model. The VIC model settings used in this study are based on the University of Washington VIC Global applications (http://www.hydro.washington.edu/Lettenmaier/Models/VIC/Datasets/Datasets.shtml). The sub-grid variability of soil, vegetation and terrain characteristics are represented in sub-grid area-specific parameter classifications. Soil texture and bulk densities are derived by combining the World Inventory of Soil Emission Potentials database (Batjes, 1995) and the 5-min digital soil map of the world from the Food and Agricultural Organization (FAO, 1998). Based on the work of (Cosby et al., 1984), the remaining soil properties (e.g. porosity, saturated hydraulic conductivity and unsaturated hydraulic conductivity) are derived. Vegetation type data are obtained from the global land classification of (Hansen et al., 2000). Parameters including the infiltration parameter, soil layer depths and those governing the baseflow function were calibrated for major global river basins and transferred to the global domain as documented in (Nijssen et al., 2001b), based on which Zhang et al. (2014) and Leng et al. (2015) conducted additional calibrations in the China domain. In this study, the VIC model was forced by WATCH climate forcing at the daily time step (Weedon et al., 2011), based on the calibrated parameters from Nijssen et al. (2001b), (Zhang et al., 2014) and (Leng et al., 2015). The simulated runoff used in this study has recently been validated globally within the framework of the Inter-Sectoral Impact Model Intercomparison Project and shows reasonable performance compared to other hydrological models (Hattermann et al., 2017; Krysanova and Hattermann, 2017).

The VIC runoff product (Hattermann et al., 2017; Leng et al., 2015) is then used as a benchmark for calibrating and validating the "*abcd*" model due to two reasons. First, VIC runoff has been evaluated across many regions of the globe and is proved to be reasonably well (Abdulla et al., 1996; Hattermann et al., 2017; Maurer et al., 2001; Nijssen et al., 1997; Nijssen et al., 2001b). Second, since we have not involved river routing, reservoir regulations and upstream water withdrawals in the "*abcd*" model, the simulated monthly runoff is more representative of "natural conditions", and as such it tends to be more reasonable to compare the simulated runoff against the VIC runoff product rather than observed streamflow data from stream gauges (Dai et al., 2009; Wilkinson et al., 2014). 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.

The VIC runoff product also compares well to other products (see Fig. S1, S2), including the UNH/GRDC runoff product (Fekete and Vorosmarty, 2011; Fekete et al., 2002) and the global streamflow product (Dai et al., 2009). The scatterplot pattern of the VIC long-term annual runoff product vs. the streamflow product matches well with that of the UNH/GRDC runoff vs. the streamflow product (streamflow is transferred to the same unit as runoff via dividing by the basin area). Further, the correlation coefficient of the VIC and the UNH/GRDC long-term annual runoff is as high as 0.83 across the global 235 basins. This suggests the reasonability of VIC runoff product, because the UNH/GRDC runoff is calibrated with the GRDC observations. At the same time, the discrepancies between the VIC runoff products and the streamflow products (Fig. S2) may be attributed to human activities, such as reservoir regulations and upstream water withdrawals, which are not embedded in the runoff but reflected in the streamflow.

2.~~3~~ 4 Model calibration

Typically, most applications of the “*abcd*” model utilize single-objective optimization for total runoff (or streamflow) during the calibration process to minimize the difference between measured and simulated streamflow (Bai et al., 2015; Martinez and Gupta, 2010; Sankarasubramanian and Vogel, 2002). While this may lead to a good fit for simulated total runoff, however, it may result in inappropriate partition of total runoff between direct runoff and baseflow. To improve the accuracy of the simulated total runoff and the partition between direct runoff and baseflow, we introduce the baseflow index (BFI) into the objective function.

Unlike the baseline model, the “*abcd*” model requires a calibration step for reasonable parameterization so as to enable good prediction. As mentioned above, we incorporate BFI into the objective function during the calibration process. On one side, we maximize Kling-Gupta efficiency (KGE) (Gupta et al., 2009), which is used as a metric to measure the accuracy of the simulated total runoff relative to the VIC benchmark runoff. The KGE is defined as the difference of unity and the Euclidian distance (ED) from the ideal point, thus we maximize KGE through minimizing the ED. The KGE and ED are calculated as follows (Gupta et al., 2009):

$$KGE = 1 - ED \quad (1)$$

$$ED = \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (2)$$

$$r = \frac{Cov_{so}}{\sigma_s \cdot \sigma_s} \quad (3)$$

$$\alpha = {\sigma_s}/{\sigma_o} \quad (4)$$

$$\beta = {\mu_s}/{\mu_o} \quad (5)$$

where r, α, β, and Cov_{so} are relative variability, bias, correlation coefficient, and covariance between the simulated and observed values (here we treat the VIC runoff as the observed), respectively; μ and σ represent the mean and standard deviation (subscript "s" and "o" stand for simulated and observed values). On the other side, we also nudge the simulated BFI towards the benchmark BFI (here we treat the benchmark BFI as the observed) – the mean BFI of the four products from (Beck et al., 2013). Then, the objective function is as follows:

$$\min(ED + abs(BFI_{obs} - BFI_{sim}))$$

where *min* stands for minimizing the value in the parenthesis, *abs* represents absolute value, ED is the Euclidian distance between the simulated and observed total runoff (Gupta et al., 2009), BFI_{obs} and BFI_{sim} are the observed and simulated BFI, respectively. Here we treat the benchmark runoff from the VIC and BFI from Beck et al. (2013) as observed values. We then minimize the objective function for parameter optimization by utilizing a Genetic Algorithm (GA) routine

(Deb et al., 2002). Note that for the distributed model scheme, we aggregate the grid-level total runoff estimates to basin-level and then nudge it toward basin-level benchmark total runoff during the calibration process.

2.~~4~~5 Model simulations

To evaluate the predictability and efficiency of the baseline and the "*abcd*" model so as to identify a suitable one to serve as a HE, we have conducted a series of simulations. Specifically, for the baseline model, no simulations are needed as it uses inter-annual mean value for each month – 12 monthly values – as prediction, so we just replicate the 12 monthly runoff for 1971-2010 and for each of the global 235 basins, and then compare against the benchmark runoff product. For the "*abcd*" model, two sets of model simulations across the global 235 basins are conducted, with one set for calibration and the other one for validation, for both the lumped and distributed model schemes. For the first set, we run the model for each basin for the period of 1971-1990 to get basin-specific parameters by using the GA approach (see Section 2.~~2~~4). For the second set, using the parameters identified in the first set of simulation, we run the model for the period of 1991-2010 to validate the model predictability and also evaluate the computational efficiency. Model inputs and outputs in the distributed scheme are at a spatial resolution of 0.5-degree, whereas those in the lumped scheme are all in lumped single unit for each basin. All model simulations are conducted in a monthly time step. Note that broad users can run the identified HE for global 235 basins, or for as many basins as they want for either scheme, as all the related basin-specific input data and calibrated parameters for both schemes are open-source.

3 Results and discussions

3.1 Comparison of performances between the baseline and the "*abcd*" 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.

3.~~1~~ 2 Evaluation of model predictability

In terms of total runoff, we find the lumped and distributed schemes are comparably capable in simulating long-term mean annual quantity, temporal variations and spatial patterns for the vast majority of river basins globally (Fig. ~~2~~3-~~4~~5). Estimates of long-term mean annual total runoff from both the lumped and distributed schemes match very well with that of VIC total runoff across the 235 basins, with a correlation coefficient (r) of higher than 0.96, for both the calibration and validation period (Fig. ~~2~~3). Similarly, the basin-level estimates of long-term mean annual direct runoff and baseflow also match well with those of the VIC across the globe, for both schemes and both periods (Fig. ~~2~~3). This suggests both schemes possess the capability in partitioning total runoff. ~~Also, we find introduction of BFI into the objective function has improved the partition of total runoff between direct runoff and baseflow (Fig. S4). Specifically, for the case of involving both the total runoff and BFI in the objective function (see Section 2.2), the correlation efficiencies (r) between the long-term annual benchmark and modeled direct runoff and baseflow from the lumped scheme across global basins are 0.97 and 0.96, respectively. However, for the case of only involving the total runoff in the objective function, the r values are 0.86 and 0.72, respectively (See Fig. S4).~~

Furthermore, both schemes display good capability in capturing the seasonal signals of the total runoff (Fig. ~~3~~4). Meanwhile, although the spatial patterns of annual total runoff from the lumped scheme present a general match with that of the VIC, it does not reflect the spatial variations inside a basin that is however captured by the distributed scheme (Fig. ~~4~~5). Therefore, the distributed scheme provides overall slightly higher KGE (Fig. ~~4 5~~6), with a global mean KGE value of 0.79 as compared to 0.75 for the lumped scheme (Fig. ~~S3~~2).

To ensure good model predictability for the major water fluxes, we also evaluate the modelled ET estimates. The modelled ET compares reasonably well with the VIC ET product as well as with the mean synthesis of the LandFlux-EVAL ET product (Mueller et al., 2013), displaying similar spatial variations (Fig. S5). Likewise, the distributed *"abcd"* scheme tends to have better capability in presenting spatial heterogeneity than the lumped one. Further, the good predictability of seasonality in runoff as illustrated in Fig. 4 also reflects similar performance for

ET, given the runoff and ET are the two major water fluxes in the water mass balance and the soil moisture changes are negligible over long-term.

The distributed scheme appears to outperform the lumped scheme in term of goodness-of-fit, especially in some cold (e.g., Arctic, Northern European, Interior Tibet) and in some dry (e.g., North Africa) regions (Fig. ~~5~~6). This is possibly because distributed inputs can reflect basin-level heterogeneity, and thus better capture the characteristic of the hydrological conditions in those regions. However, both schemes do not perform well in the southern end of the Andes Mountains (Fig. ~~5~~6). This may be attributed to the complex land surface characteristics in that mountainous area, which cannot be resolved due to the coarse spatial resolution. Moreover, the distributed scheme seems not performing very well in some cold regions (Fig. ~~5~~6), which is possibly due to lack of representation for permafrost in the model.

Previous studies investigating the credibility of lumped and distributed hydrological models indicate that, in many cases, lumped models perform comparably or just as well as distributed models (Asadi, 2013; Brirhet and Benaabidate, 2016; Ghavidelfar et al., 2011; Michaud and Sorooshian, 1994; Obled et al., 1994; Reed et al., 2004; Refsgaard and Knudsen, 1996; YAO et al., 1998). However, distributed models may have advantages for predicting runoff in ungauged watersheds (Reed et al., 2004; Refsgaard and Knudsen, 1996), for capturing spatial distribution of runoff due to heterogeneity in rainfall patterns or in land surface (Downer et al., 2002; Paudel et al., 2011; YAO et al., 1998), and for predicting flood peaks (Asadi, 2013; Brirhet and Benaabidate, 2016; Carpenter and Georgakakos, 2006; Krajewski et al., 1991). Our results on the predictability of lumped and distributed "*abcd*" model are in line with previous findings in the literature.

The good agreement between our modelled water fluxes, including total runoff, direct runoff, baseflow and ET, and the benchmark products provides confidence in the capability of both the lumped and distributed schemes in estimating temporal and spatial variations in major water fluxes across the globe. In addition, to identify a suitable HE, the required computation cost is another key factor as detailed below.

3.~~2~~ 3 Evaluation of computational efficiency

While the performance of model predictability is comparable for the lumped and distributed schemes as elucidated above, great disparity exists for runtime of the two schemes and the VIC model (Table ~~2~~S1). Take the Amazon basin that covers a total number of 1990 0.5-degree grid cells as an example, it takes 11.05 minutes for model calibration via the GA method in the distributed scheme but only 0.16 minute for the lumped one. Similar disparity is also found for model simulation with calibrated parameters, with runtime of 0.03 and 3.20 seconds for a 1000-year simulation of the Amazon basin for the lumped and distributed schemes, respectively. However, according to the authors' experience, it will take ~1 week for the VIC model to accomplish the same job, which is far more computationally expensive. In general, the computational efficiency of the lumped scheme is two orders of magnitudes higher than the distributed one, although that of the distributed one is still much higher than the VIC (~five orders of magnitude) and many other GHMs and LSMs. Note that all of the simulations here are conducted on the Pacific Northwest National Laboratory (PNNL)'s Institutional Computing (PIC) Constance cluster using 1 core (Intel Xeon 2.3 GHz CPU) with the same configuration.

3.~~3~~ 4 Potential application of the "*abcd*" model as a hydrological emulator

The good predictability and computational efficiency of both the distributed or lumped schemes as elucidated in Sections 3.~~1~~ 2 and 3.~~2~~ 3 suggest its suitability for serving as HEs that can efficiently emulate complex GHMs (e.g., the VIC or others). The source codes, input data, basin-specific parameters across the globe for both the lumped and distributed schemes are open-source and well-documented, which will make the HE ready to use and facilitate their wide and easy use with minimal efforts.

The choice of either the distributed or lumped scheme as HE depends on the user's specific needs. There is a tradeoff between the model predictability and computational efficiency. While the distributed scheme tends to better capture the spatial heterogeneity of water fluxes and can produce grid-level outputs that lumped scheme cannot, it incurs heavier computational cost than the lumped scheme. For applications that aim to strike a balance between predictability and computation cost, such as practical assessment of water resources, or estimation of water supply for integrated assessment models (IAMs), or quantification of uncertainty and sensitivity analyses, it would be reasonable to employ the lumped scheme as a HE. The lumped scheme is especially advantageous due to its minimal calibration and computational cost, parsimonious efforts for model implementation, and reasonable fidelity in estimating major water fluxes (e.g., runoff, ET). For users from the IAM community, the lumped scheme might be sufficiently suitable for their needs since 1) the lumped scheme can operate at the same spatial resolution at which IAMs typically balance water demands and supplies (Edmonds et al., 1997; Kim et al., 2006; Kim et al., 2016), and 2) the inherent uncertainty of the lumped scheme is likely comparable or even overshadowed by the intrinsic uncertainty of IAMs (Kraucunas et al., 2015; O'Neill et al., 2014). Similarly, for users who aim to conduct uncertainty and sensitivity analyses, the high computational efficiency of the lumped scheme allow the users to emulate the hydrological model of interest (e.g., GHMs, LSMs) and then run a large number of simulations to conduct their uncertainty and sensitivity analysis (Scott et al., 2016). Therefore, the high computational efficiency makes the lumped scheme more appealing as a HE in these cases. However, if the research questions hinge on the gridded estimates, or emphasize the spatial heterogeneity of the water fluxes or pools, it would be more desirable to deploy the distributed scheme as a HE instead.

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.

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.

4 Conclusions

Toward addressing the issue that many global hydrological models (GHMs) are computationally expensive and thus users cannot afford to conduct a large number of simulations for various tasks, we firstly construct a hydrological emulator (HE) that possesses both reasonable predictability and computation efficiency for global applications in this work. Built upon the widely-used *"abcd"* model, we have adopted two snow-related parameters from literature rather than tuning them for parameter parsimony, and also have improved the partition of total runoff between the direct runoff and baseflow by introducing baseflow index into the objective function of the parameter optimization. We then evaluate the appropriateness of the model serving as an emulator for a complex GHM – the VIC, for both the lumped and distributed model schemes, by examining their predictability and computational efficiency.

In general, both distributed and lumped schemes have comparably good capability in simulating spatial and temporal variations of the water balance components (i.e., total runoff, direct runoff, baseflow, evapotranspiration). Meanwhile, the distributed scheme has slightly better performance than the lumped one (e.g., capturing spatial heterogeneity), with mean Kling-

Gupta efficiency of 0.79 vs. 0.75 across global 235 basins, and also it provides grid-level estimates that the lumped one incapable of. Additionally, the distributed scheme performs better in extreme climate regimes (e.g., Arctic, North Africa) and Europe. However, the distributed one incurs two more orders of magnitudes of computation cost than the lumped one. A case study of uncertainty analysis with 100, 000 simulations for each of the world's sixteen basins with top annual streamflow further demonstrates the lumped scheme's extraordinary advantage in terms of computational efficiency. Therefore, the lumped scheme could be an appropriate HE – reasonable predictability and high computational efficiency. At the same time, the distributed scheme could be a suitable alternative for research questions that hinge on grid-level spatial heterogeneity. Finally, upon open-sourcing and well-documentation, the HE is ready to use and it provides researchers an easy way to investigate the variations in water budgets at a variety of spatial scales of interest (e.g., basin, region or globe), with minimum requirements of efforts, reasonable model predictability and appealing computational efficiency.

Code and/or data availability

The hydrological emulator (HE) is freely available on the open-source software site GitHub (https://github.com/JGCRI/hydro-emulator/). We have released the version of the specific HE v1.0.0 referenced in this paper on https://github.com/JGCRI/hydro-emulator/releases/tag/v1.0.0, where the source code (written in Matlab), all related inputs, calibrated parameters and outputs for each of the global 235 basins, as well as the detailed Readme file are available. ~~The code and data are available on the GitHub open-source software site (). The repository includes the source code (written in Matlab), all related data inputs and outputs for global 235 basins, and a Readme file.~~

Appendix A: Descriptions and equations of the "*abcd*" model

The *abcd* model was first introduced by (Thomas, 1981), and Martinez and Gupta (Martinez and Gupta, 2010) added snow processes into the model. In this work, we adopted the snow scheme in Martinez and Gupta (2010):

$$Snow_i = \begin{cases} 0 & T^{rain} < T_i^{\min} \\ P_i \times \dfrac{T^{rain} - T_i^{\min}}{T^{rain} - T^{snow}} & T^{snow} < T_i^{\min} < T^{rain} \\ P_i & T_i^{\min} < T^{snow} \end{cases} \tag{1}$$

$$SP_i = SP_{i-1} - SNM_i + Snow_i \tag{2}$$

$$SNM_i = \begin{cases} 0 & T_i^{\min} < T^{snow} \\ (SP_{i-1} + Snow_i) \times m \times \dfrac{T^{rain} - T_i^{\min}}{T^{rain} - T^{snow}} & T^{snow} < T_i^{\min} < T^{rain} \\ (SP_{i-1} + Snow_i) \times m & T^{rain} < T_i^{\min} \end{cases} \tag{3}$$

where P_i, SP_i, SNM_i and $Snow_i$ are total precipitation, snowpack storage, snowmelt and the precipitation as snowfall at time step *i*, respectively, T^{rain} (or T^{snow}) stands for the temperature threshold above (or below) which all precipitation falls as rainfall (or snow), and $T_i^{\min}$ is the minimum temperature at time step *i*, and the parameter *m* is the snowmelt coefficient. Rather than keeping the three parameters T^{rain}, T^{snow} and *m*, we adopt the T^{rain} value of 2.5 °C and T^{snow} value of 0.6 °C (Wen et al., 2013) and thus only keep one snowmelt-related parameter *m* in the model, in order to alleviate the computation load during the parameter optimization process.

The model defines two state variables "available water" and "evapotranspiration opportunity", denoted as W_i and Y_i, respectively. The W_i is defined as:

$$W_i = SM_{i-1} + Rain_i + SNM_i \tag{4}$$

where SM_{i-1} is soil moisture at the beginning of time step i, $Rain_i$ and SNM_i are rainfall and snowmelt during period i.

Y_i stands for the maximum water that can leave the soil as evapotranspiration (*ET*) at period i, and it is defined as below:

$$Y_i = ET_i + SM_i \tag{5}$$

where ET_i is the actual ET at time period i and SM_i is soil moisture at the end of time step i. Further, Y_i has a non-linear relationship with W_i as:

$$Y_i = \frac{W_i - b}{2a} - \sqrt{(\frac{W_i - b}{2a})^2 - W_i \times b / a} \tag{6}$$

where a and b are parameters detailed in Section 2.1.2.

Allocation of W_i between ET_i and SM_i is estimated by assuming that the loss of soil moisture by *ET* will be proportional to *PET* as:

$$\frac{dS}{dt} = -PET \times \frac{SM}{b} \tag{7}$$

After integrating the above differential equation and assuming $S_{i-1} = Y_i$, SM_i can be derived as:

$$SM_i = Y_i \times \exp({}^{-PET_i}\!/_{b}) \tag{8}$$

Then, ET_i can be calculated through equation (2).

In the model framework, $W_i - Y_i$ is the sum of the groundwater recharge (*RE*) and direct runoff (Q_d), and the allocation is determined by the parameter c:

$$RE_i = c \times (W_i - Y_i) \tag{9}$$

$$Q_d = (1-c)\times(W_i - Y_i) \tag{10}$$

The baseflow from the groundwater (*GW*) pool is modeled as:

$$Q_b = d \times GW_i \tag{11}$$

where d is a parameter reflecting the release rate of groundwater to baseflow. Then the total runoff (Q_t) is the sum of the direct runoff and baseflow:

$$Q_t = Q_d + Q_b \tag{12}$$

The GW_i is the sum of groundwater storage at the end of last time step and the groundwater recharge minus the baseflow, and GW_i is derived as:

$$GW_i = \frac{GW_{i-1} + RE_i}{1+d} \tag{13}$$

Then, all the water fluxes and pools are solved.

Author contribution

Yaling Liu and Mohamad Hejazi designed this work, and all co-authors offered help through discussions. Yaling Liu developed the hydrological emulator and conducted the simulations and evaluations. Yaling Liu wrote the manuscript, and all co-authors contributed to the revision.

Competing interests

The authors declare that they have no conflict of interests.

Acknowledgement: This research was supported by the Office of Science of the U.S. Department of Energy through the Integrated Assessment Research Program. PNNL is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.

References

Abdulla, F.A., Lettenmaier, D.P., Wood, E.F., Smith, J.A., 1996. Application of a macroscale hydrologic model to estimate the water balance of the Arkansas ‐ Red River Basin. Journal of Geophysical Research: Atmospheres, 101(D3): 7449-7459.

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.

Alcamo, J., Henrichs, T., 2002. Critical regions: A model-based estimation of world water resources sensitive to global changes. Aquat. Sci., 64(4): 352-362.

Alkama, R. et al., 2010. Global evaluation of the ISBA-TRIP continental hydrological system. Part I: Comparison to GRACE terrestrial water storage estimates and in situ river discharges. J. Hydrometeorol., 11(3): 583-600.

Arnell, N.W., Gosling, S.N., 2013. The impacts of climate change on river flow regimes at the global scale. J. Hydrol., 486: 351-364.

Asadi, A., 2013. The Comparison of Lumped and Distributed Models for Estimating Flood Hydrograph (Study Area: Kabkian Basin). J. Electron. Commun. Eng. Res, 1(2): 7-13.

Bai, P., Liu, X., Liang, K., Liu, C., 2015. Comparison of performance of twelve monthly water balance models in different climatic catchments of China. J. Hydrol., 529: 1030-1040.

Batjes, N., 1995. A homogenized soil data file for global environmental research: A subset of FAO, ISRIC and NRCS profiles (Version 1.0), ISRIC.

Beck, H.E. et al., 2013. Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. Water Resour. Res., 49(12): 7843-7863.

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

Bierkens, M., Van Beek, L., 2009. Seasonal predictability of European discharge: NAO and hydrological response time. J. Hydrometeorol., 10(4): 953-968.

Brirhet, H., Benaabidate, L., 2016. Comparison Of Two Hydrological Models (Lumped And Distributed) Over A Pilot Area Of The Issen Watershed In The Souss Basin, Morocco. European Scientific Journal, 12(18).

Burnash, R.J., Ferral, R.L., McGuire, R.A., 1973. A generalized streamflow simulation system, conceptual modeling for digital computers, U.S. Department of Commerce, National Weather Service, and State of California, Department of Water Resources, Sacramento, CA.

Carpenter, T.M., Georgakakos, K.P., 2006. Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. J. Hydrol., 329(1): 174-185.

Cosby, B., Hornberger, G., Clapp, R., Ginn, T., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour. Res., 20(6): 682-690.

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.

De Graaf, I., Van Beek, L., Wada, Y., Bierkens, M., 2014. Dynamic attribution of global water demand to surface water and groundwater resources: Effects of abstractions and return flows on river discharges. Advances in Water Resources, 64: 21-33.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE transactions on evolutionary computation, 6(2): 182-197.

Döll, P., Fiedler, K., Zhang, J., 2009. Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. Hydrology and Earth System Sciences, 13(12): 2413-2432.

Downer, C.W., Ogden, F.L., Martin, W.D., Harmon, R.S., 2002. Theory, development, and applicability of the surface water hydrologic model CASC2D. Hydrol. Processes, 16(2): 255-275.

Edmonds, J. et al., 1997. An integrated assessment of climate change and the accelerated introduction of advanced energy technologies-an application of MiniCAM 1.0. Mitigation and adaptation strategies for global change, 1(4): 311-339.

FAO, 1998. Digital soil map of the world and derived soil properties. In: Organization, F.a.A. (Ed.). Land and 410 Water Digital Media Series 1.

Fekete, B., Vorosmarty, C., 2011. ISLSCP II UNH/GRDC Composite Monthly Runoff. ISLSCP Initiative II Collection, edited by: Hall, FG, Collatz, G., Meeson, B., Los, S., Brown de Colstoun, E., and Landis, D., Data set, available at: http://daac. ornl. go v/, from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, doi, 10.

Fekete, B.M., Vörösmarty, C.J., Grabs, W., 2002. High‐resolution fields of global runoff combining observed river discharge and simulated water balances. Global Biogeochem. Cycles, 16(3).

Fernandez, W., Vogel, R., Sankarasubramanian, A., 2000. Regional calibration of a watershed model. Hydrol. Sci. J., 45(5): 689-707.

Gerten, D. et al., 2005. Contemporary “green” water flows: Simulations with a dynamic global vegetation and water balance model. Physics and Chemistry of the Earth, Parts A/B/C, 30(6): 334-338.

Ghavidelfar, S., Alvankar, S.R., Razmkhah, A., 2011. Comparison of the lumped and quasi-distributed Clark runoff models in simulating flood hydrographs on a semi-arid watershed. Water Resour. Manage., 25(6): 1775-1790.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol., 377(1): 80-91.

Hanasaki, N. et al., 2008. An integrated model for the assessment of global water resources–Part 2: Applications and assessments. Hydrology and Earth System Sciences, 12(4): 1027-1037.

Hansen, M., DeFries, R., Townshend, J.R., Sohlberg, R., 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. Int. J. Remote Sens., 21(6-7): 1331-1364.

Hargreaves, G.H., Samani, Z.A., 1982. Estimating potential evapotranspiration. Journal of the Irrigation and Drainage Division, 108(3): 225-230.

Hattermann, F. et al., 2017. Cross‐scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. Clim. Change: 1-16.

Kim, S.H., Edmonds, J., Lurz, J., Smith, S.J., Wise, M., 2006. The O bj ECTS framework for integrated assessment: Hybrid modeling of transportation. The Energy Journal: 63-91.

Kim, S.H. et al., 2016. Balancing global water availability and use at basin scale in an integrated assessment model. Clim. Change, 136(2): 217-231.

Krajewski, W.F., Lakshmi, V., Georgakakos, K.P., Jain, S.C., 1991. A Monte Carlo study of rainfall sampling effect on a distributed catchment model. Water Resour. Res., 27(1): 119-128.

Kraucunas, I. et al., 2015. Investigating the nexus of climate, energy, water, and land at decision-relevant scales: the Platform for Regional Integrated Modeling and Analysis (PRIMA). Clim. Change, 129(3-4): 573-588.

Krysanova, V., Hattermann, F.F., 2017. Intercomparison of climate change impacts in 12 large river basins: overview of methods and summary of results. Clim. Change, 141(3): 363-379.

Leng, G., Tang, Q., Rayburg, S., 2015. Climate change impacts on meteorological, agricultural and hydrological droughts in China. Global Planet. Change, 126: 23-34.

Liu, Y. et al., 2015. Agriculture intensifies soil moisture decline in Northern China. Scientific reports, 5: 11261.

Liu, Y. et al., 2013. Response of evapotranspiration and water availability to changing climate and land cover on the Mongolian Plateau during the 21st century. Global Planet. Change, 108: 85-99.

Liu, Y. et al., 2014. Response of evapotranspiration and water availability to the changing climate in Northern Eurasia. Clim. Change, 126(3-4): 413-427.

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).

Maurer, E.P., O'Donnell, G.M., Lettenmaier, D.P., Roads, J.O., 2001. Evaluation of the land surface water budget in NCEP/NCAR and NCEP/DOE reanalyses using an off‐line hydrologic model. Journal of Geophysical Research: Atmospheres, 106(D16): 17841-17862.

Michaud, J., Sorooshian, S., 1994. Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. Water Resour. Res., 30(3): 593-605.

Mueller, B. et al., 2013. Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. Hydrology and Earth System Sciences.

Murphy, J.M. et al., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature, 430(7001): 768-772.

Nijssen, B., Lettenmaier, D.P., Liang, X., Wetzel, S.W., Wood, E.F., 1997. Streamflow simulation for continental‐scale river basins. Water Resour. Res., 33(4): 711-724.

Nijssen, B., O'donnell, G.M., Hamlet, A.F., Lettenmaier, D.P., 2001a. Hydrologic sensitivity of global rivers to climate change. Clim. Change, 50(1-2): 143-175.

Nijssen, B., O'Donnell, G.M., Lettenmaier, D.P., Lohmann, D., Wood, E.F., 2001b. Predicting the discharge of global rivers. J. Clim., 14(15): 3307-3323.

O’Neill, B.C. et al., 2014. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. Clim. Change, 122(3): 387-400.

Obled, C., Wendling, J., Beven, K., 1994. The sensitivity of hydrological models to spatial rainfall patterns: an evaluation using observed data. J. Hydrol., 159(1-4): 305-333.

Oubeidillah, A.A., Kao, S.-C., Ashfaq, M., Naz, B.S., Tootle, G., 2014. A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US. Hydrology and Earth System Sciences, 18(1): 67-84.

Paudel, M., Nelson, E.J., Downer, C.W., Hotchkiss, R., 2011. Comparing the capability of distributed and lumped hydrologic models for analyzing the effects of land use change. Journal of Hydroinformatics, 13(3): 461-473.

Reed, S. et al., 2004. Overall distributed model intercomparison project results. J. Hydrol., 298(1): 27-60.

Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. Water Resour. Res., 32(7): 2189-2202.

Rost, S., Gerten, D., Heyder, U., 2008. Human alterations of the terrestrial water cycle through land management. Advances in Geosciences, 18: 43-50.

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).

Scott, M.J. et al., 2016. Sensitivity of future US Water shortages to socioeconomic and climate drivers: a case study in Georgia using an integrated human-earth system modeling framework. Clim. Change, 136(2): 233-246.

Tang, Q. et al., 2010. Dynamics of terrestrial water storage change from satellite and surface observations and modeling. J. Hydrometeorol., 11(1): 156-170.

Tebaldi, C., Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 365(1857): 2053-2075.

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

Van Beek, L., Bierkens, M.F., 2009. The global hydrological model PCR-GLOBWB: conceptualization, parameterization and verification. Utrecht University, Utrecht, The Netherlands.

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.

Vörösmarty, C.J., Green, P., Salisbury, J., Lammers, R.B., 2000. Global water resources: vulnerability from climate change and population growth. science, 289(5477): 284-288.

Wang, D., Tang, Y., 2014. A one ‐ parameter Budyko model for water balance captures emergent behavior in darwinian hydrologic models. Geophys. Res. Lett., 41(13): 4569-4577.

Weedon, G. et al., 2011. Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. J. Hydrometeorol., 12(5): 823-848.

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.

Widén-Nilsson, E., Halldin, S., Xu, C.-y., 2007. Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation. J. Hydrol., 340(1): 105-118.

Wilkinson, K., von Zabern, M., Scherzer, J., 2014. Global Freshwater Fluxes into the World Oceans, Tech. Report prepared for the GRDC. ‐ Koblenz, Federal Institute of Hydrology (BfG),(GRDC Report No. 44. doi: 10.5675/GRDC_Report_44, 23pp.[Available from h ttp://www. bafg. de/GRDC/EN/02_srvcs/24_rprtsrs/report_44. pdf].

YAO, H., HASHINO, M., TERAKAWA, A., SUZUKI, T., 1998. Comparison of distributed and lumped hydrological models. PROCEEDINGS OF HYDRAULIC ENGINEERING, 42: 163-168.

Zhang, X.-J., Tang, Q., Pan, M., Tang, Y., 2014. A long-term land surface hydrologic fluxes and states dataset for China. J. Hydrometeorol., 15(5): 2067-2084.

Figure Caption

Figure 1 Schematic diagram of the "*abcd*" model, with enhancements of snow and partition of total runoff between direct runoff and baseflow.

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).

Figure ~~2~~ 3 Comparison of basin-specific long-term annual total runoff, direct runoff and baseflow estimates from both the lumped and distributed "*abcd*" model schemes against VIC products, across global 235 basins and for the calibration period of 1971-1990 and validation period of 1991-2010. The labels are denoted as combination of model scheme and period, where lump and dist stand for lumped and distributed model scheme, cal and val represent the calibration and validation period, respectively. These denotations remain the same for all figures in this work. Note that the basin-level VIC baseflow is derived by multiplying the gridded VIC long-term annual total runoff and the mean of the four gridded baseflow index products from Beck et al. (2014), and then aggregating from grid-level to basin-level. The basin-level VIC direct runoff is then calculated by subtracting baseflow from the total runoff.

Figure ~~3~~ 4 Time series of basin-specific total runoff (Q_{total}) from the VIC product, the lumped and distributed "*abcd*" schemes for the world's sixteen river basins with top annual flow (Dai et al. 2009) during 1981-1990. KGE_l and KGE_d stand for KGE value for the lumped and distributed scheme, respectively.

Figure ~~4~~ 5 Spatial patterns of long-term annual total runoff (mm yr^{-1}) across global 235 basins: a) VIC runoff product; b) total runoff estimates from the lumped "*abcd*" scheme (lump = lumped); and c) total runoff estimates from the distributed "*abcd*" scheme (dist = distributed).

Figure ~~5~~ **6** The spatial pattern of Kling-Gupta efficiency (KGE) for the total runoff estimates of the global 235 basins for the calibration period of 1971-1990: a) the lumped "*abcd*" scheme; and b) the distributed "*abcd*" scheme.

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.

Figure 1

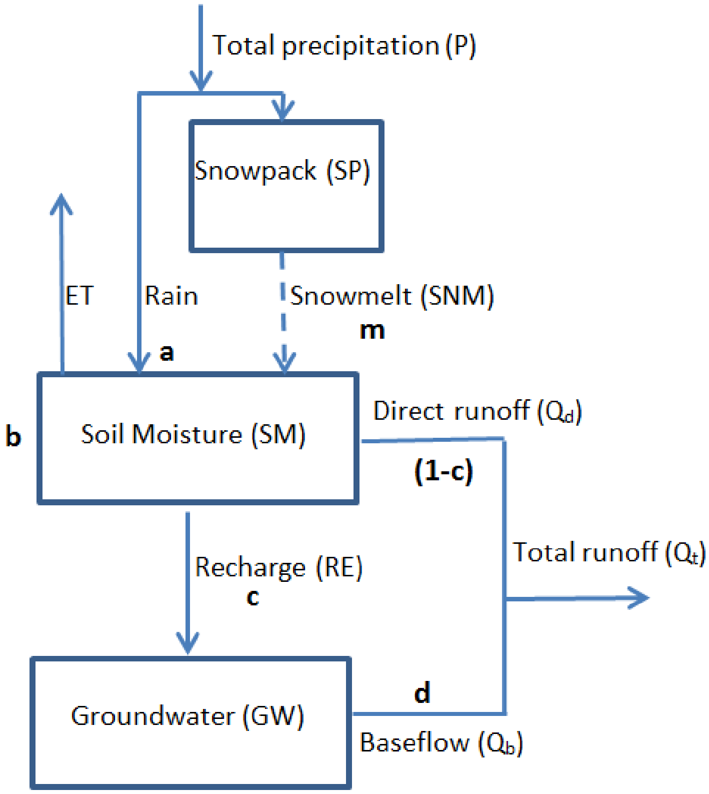

Total precipitation (P)
Snowpack (SP)
ET
Rain
Snowmelt (SNM)
m
a
b
Soil Moisture (SM)
Direct runoff (Q_d)
(1-c)
Total runoff (Q_t)
Recharge (RE)
c
Groundwater (GW)
d
Baseflow (Q_b)

Figure 2

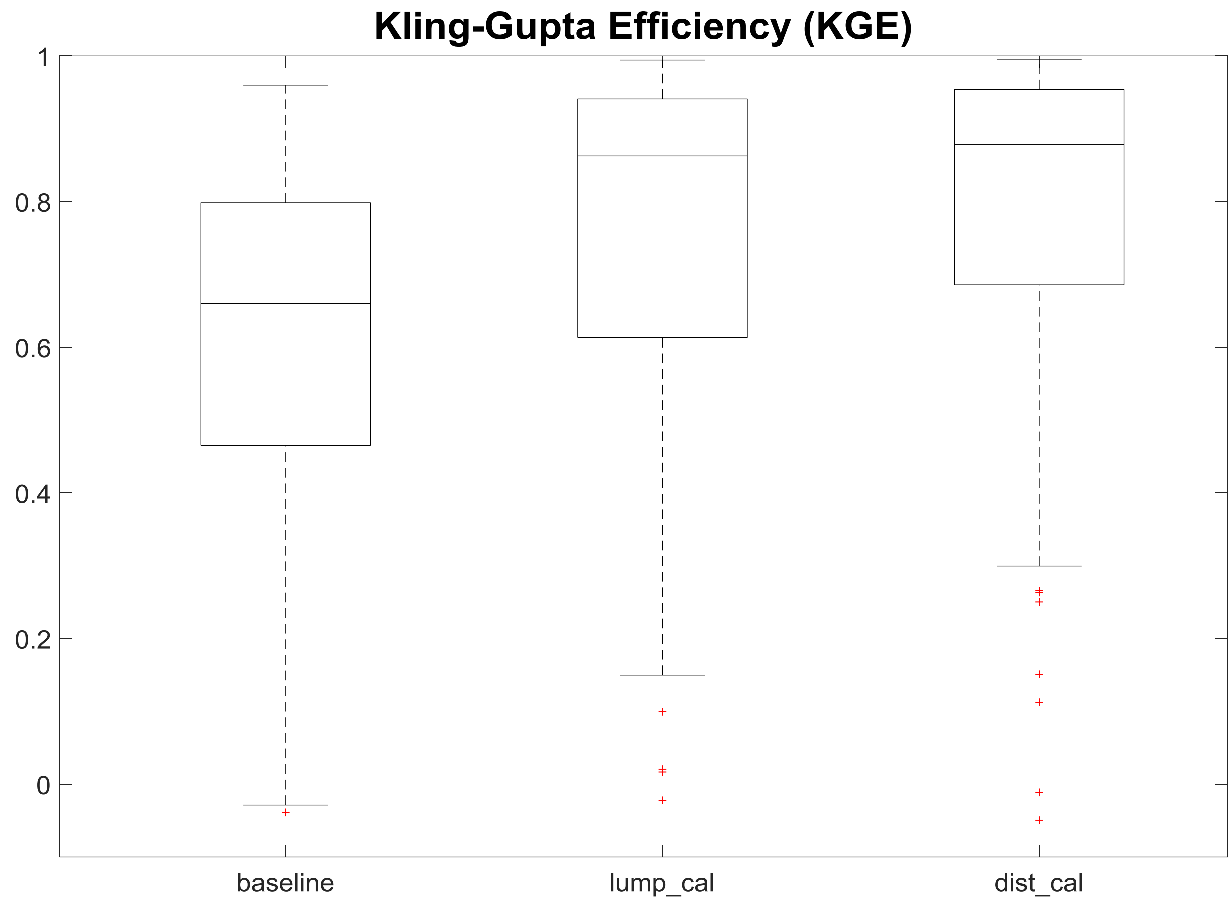

Kling-Gupta Efficiency (KGE)
1
0.8
0.6
0.4
0.2
0
baseline
lump_cal
dist_cal

Figure ~~2~~3

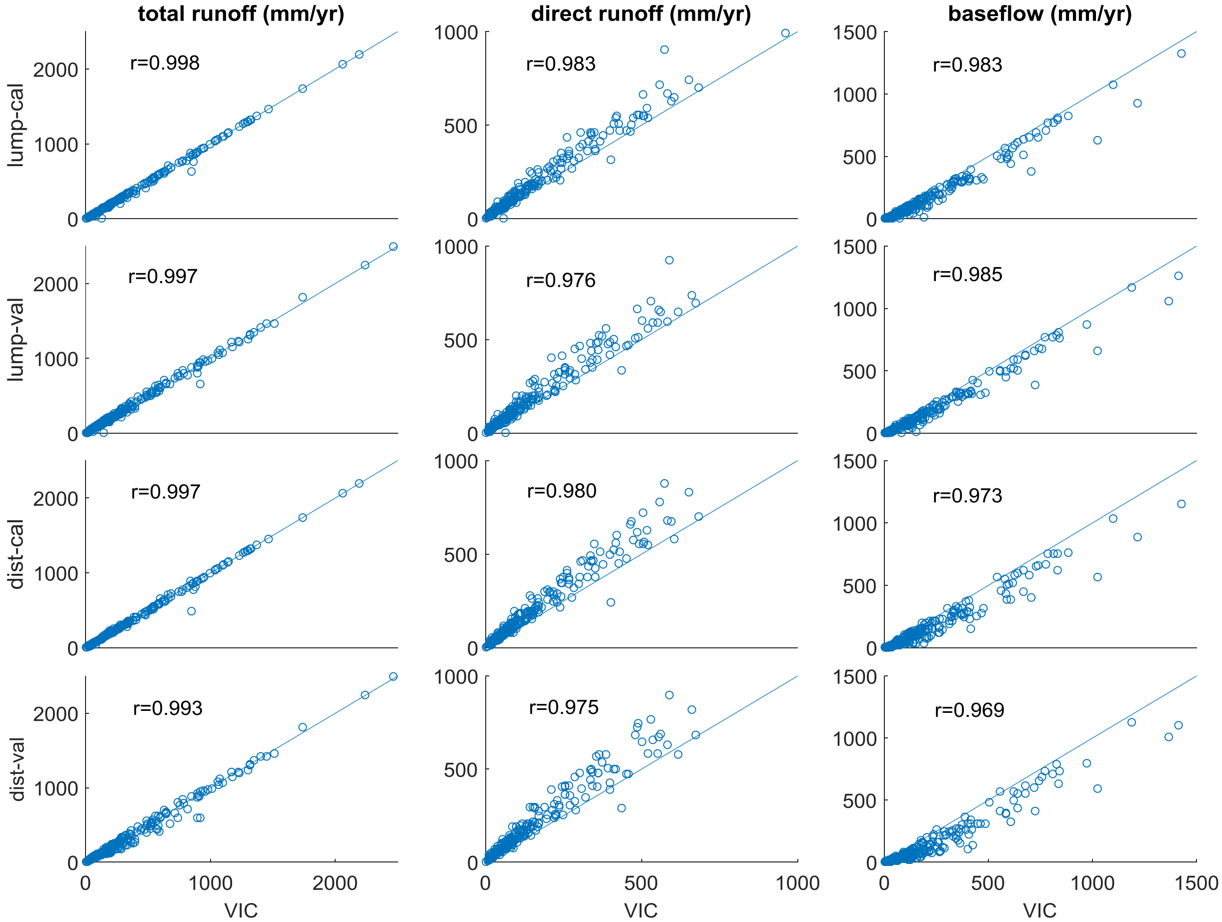

total runoff (mm/yr)
direct runoff (mm/yr)
baseflow (mm/yr)
lump-cal
lump-val
dist-cal
dist-val
r=0.998
r=0.983
r=0.983
r=0.997
r=0.976
r=0.985
r=0.997
r=0.980
r=0.973
r=0.993
r=0.975
r=0.969
0
1000
2000
500
1500
VIC
VIC
VIC

Figure ~~3~~4

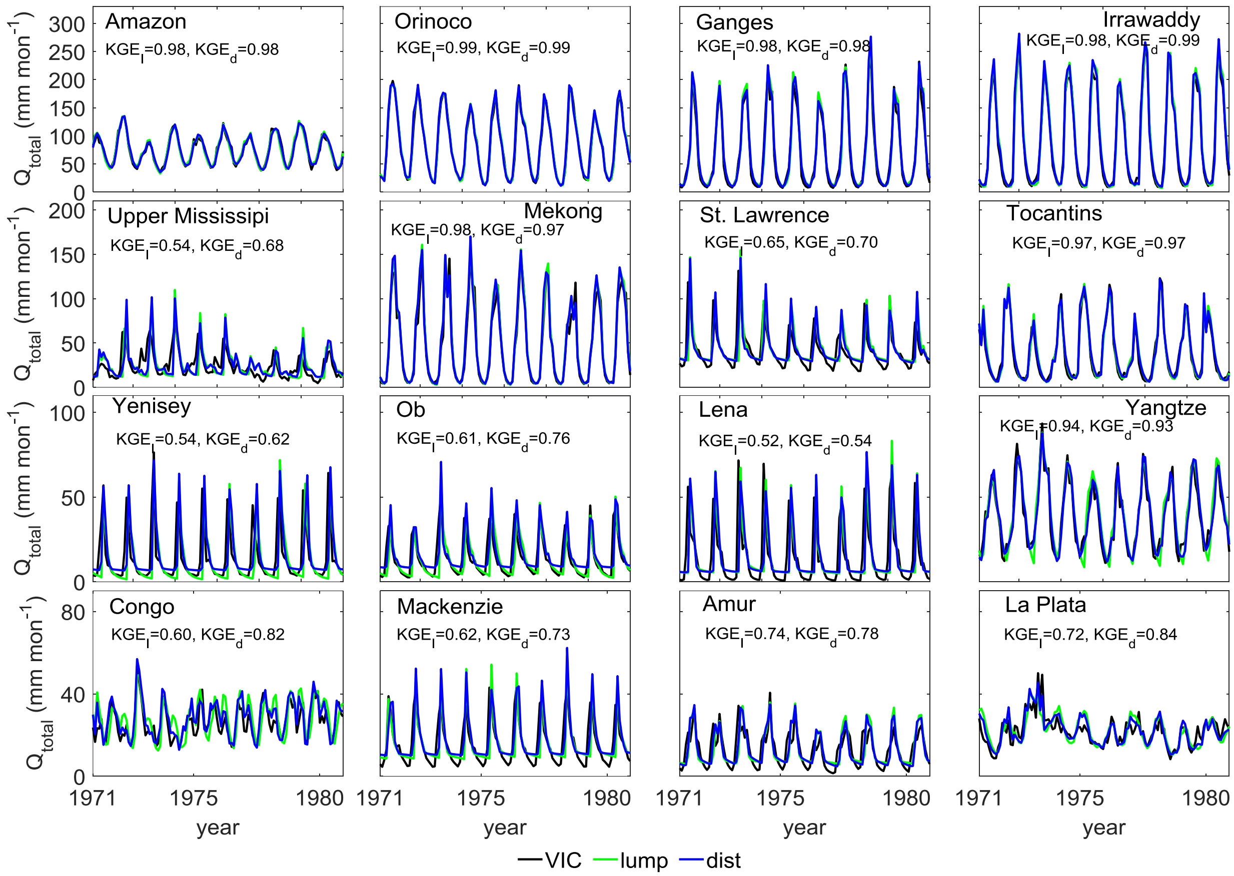

Amazon
KGE_l=0.98, KGE_d=0.98
Orinoco
KGE_l=0.99, KGE_d=0.99
Ganges
KGE_l=0.98, KGE_d=0.98
Irrawaddy
KGE_l=0.98, KGE_d=0.99
Upper Mississipi
KGE_l=0.54, KGE_d=0.68
Mekong
KGE_l=0.98, KGE_d=0.97
St. Lawrence
KGE_l=0.65, KGE_d=0.70
Tocantins
KGE_l=0.97, KGE_d=0.97
Yenisey
KGE_l=0.54, KGE_d=0.62
Ob
KGE_l=0.61, KGE_d=0.76
Lena
KGE_l=0.52, KGE_d=0.54
Yangtze
KGE_l=0.94, KGE_d=0.93
Congo
KGE_l=0.60, KGE_d=0.82
Mackenzie
KGE_l=0.62, KGE_d=0.73
Amur
KGE_l=0.74, KGE_d=0.78
La Plata
KGE_l=0.72, KGE_d=0.84
Q_total (mm mon^-1)
1971
1975
1980
year
VIC
lump
dist

Figure 45

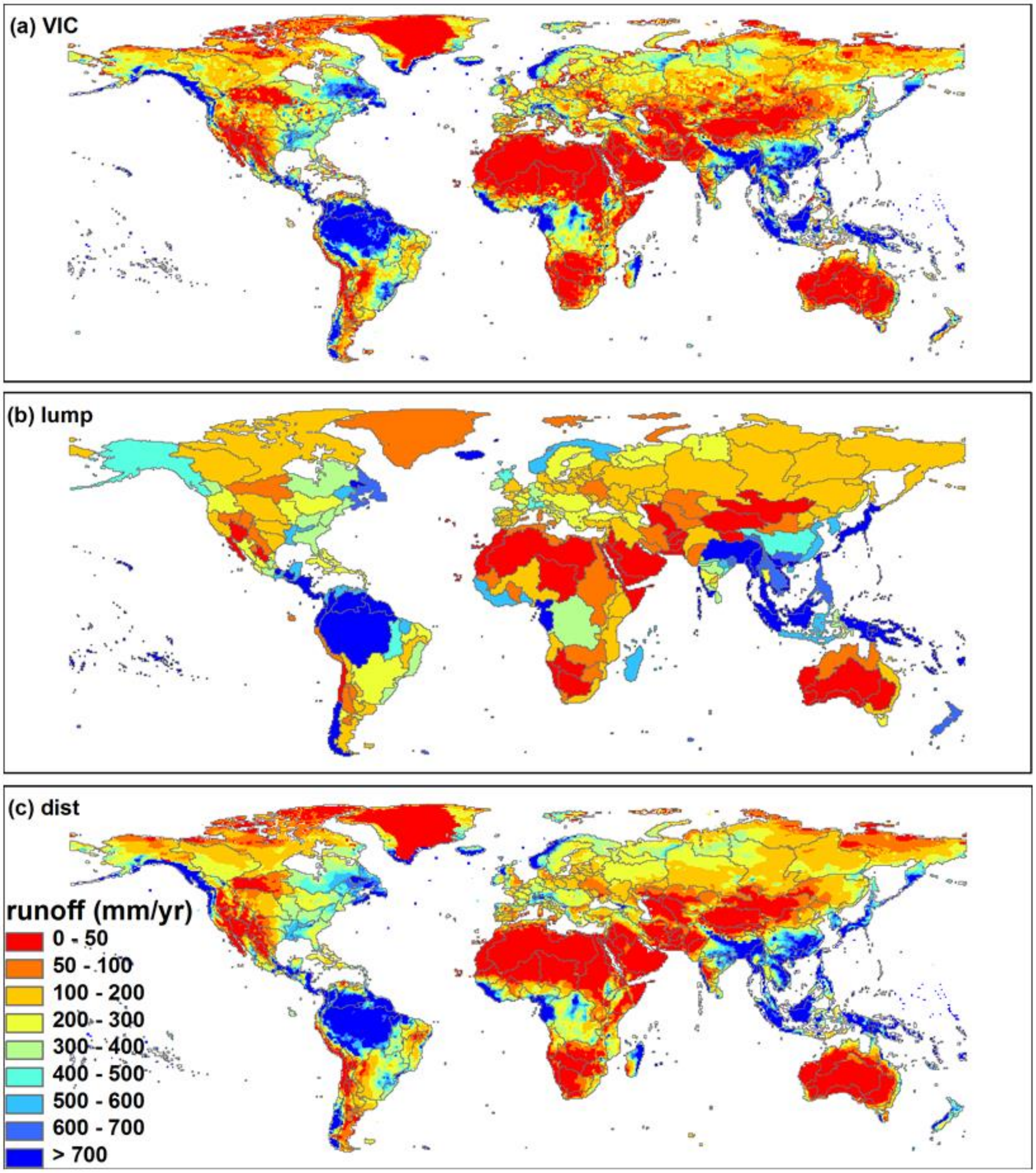
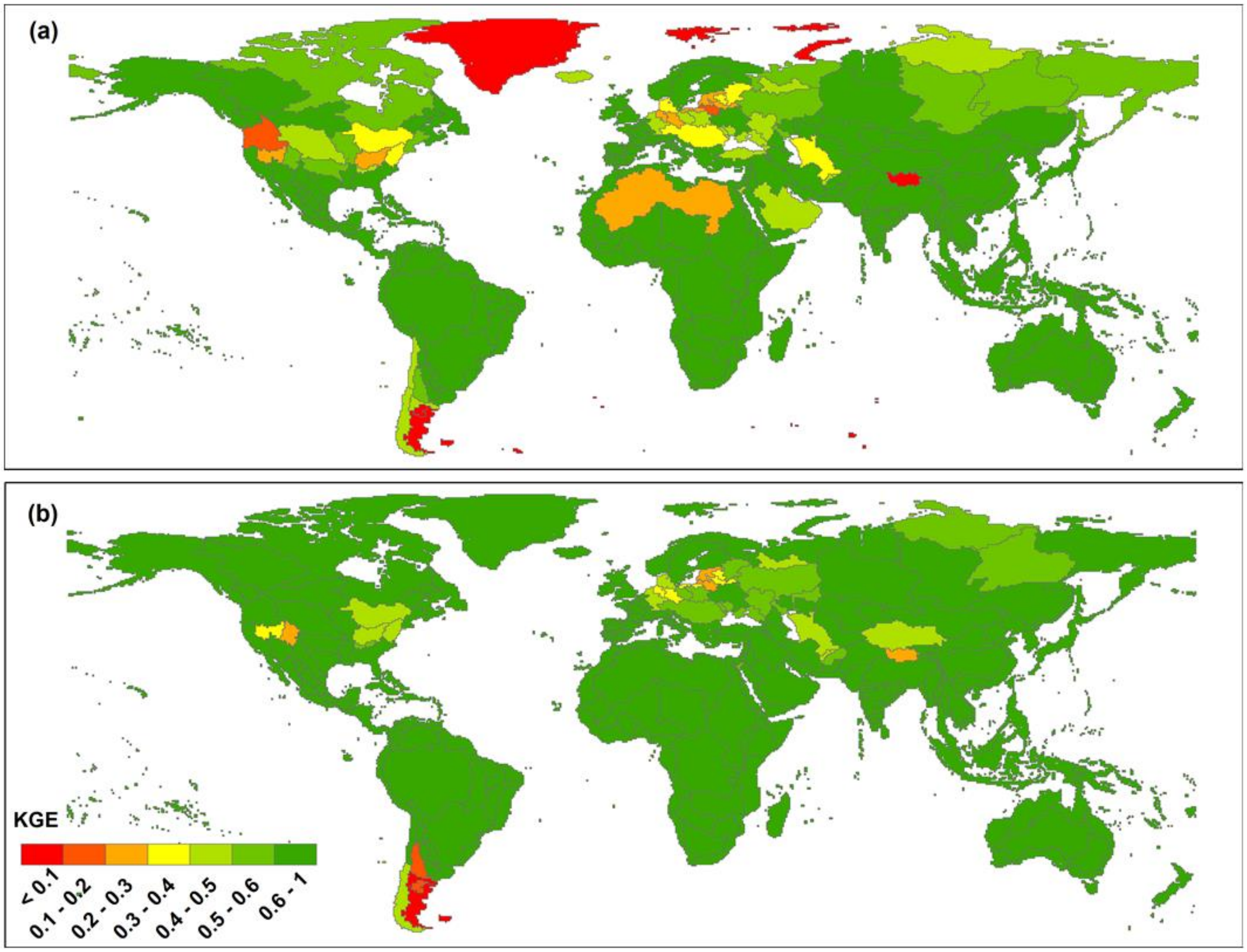
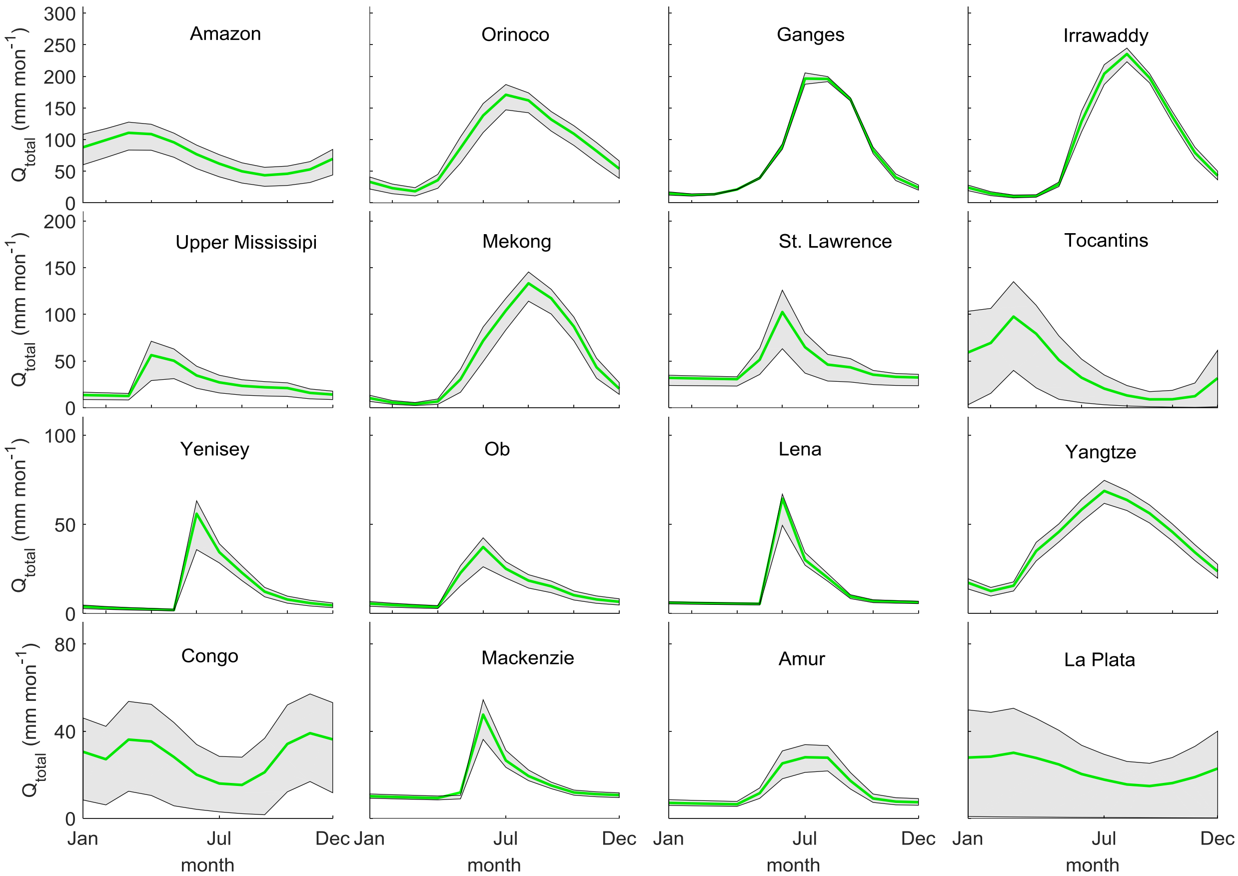

(a) VIC
(b) lump
(c) dist
runoff (mm/yr)
0 - 50
50 - 100
100 - 200
200 - 300
300 - 400
400 - 500
500 - 600
600 - 700
> 700

Figure ~~5~~6



751 Figure 7



752

Table 1 Parameters description and ranges for the "*abcd*" model (the parameters *a,c,d* and *m* are dimensionless, and the unit for parameter *b* is mm)

parameter	**description**	**range**	**references**
a	Propensity of runoff to occur before the soil is fully saturated	0-1	(Alley, 1984; Martinez and Gupta, 2010; Sankarasubramanian and Vogel, 2002; Vandewiele and Xu, 1992)
b	Upper limit on the sum of evapotranspiration and soil moisture storage	0-4000	
c	Degree of recharge to groundwater	0-1	
d	Release rate of groundwater to baseflow	0-1	
m	Snow melt coefficient	0-1	(Wen et al., 2013)

~~Table 2 Runtime for model calibration and simulation at Amazon basin for the lumped (lump) and distributed (dist) "*abcd*" model scheme, as well as for the VIC model.~~

	~~calibration~~	**~~1000 years' simulation~~**
~~lump~~	~~0.16 min~~	~~0.03 s~~
~~dist~~	~~11.05 min~~	~~3.20 s~~
~~VIC~~	~~N/A~~	~~~ 1 week~~