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

2   Yaling Liu[1,2], Mohamad Hejazi[1], Hongyi Li[3], Xuesong Zhang[1], Guoyong Leng[1]

3   [1]Joint Global Change Research Institute, Pacific Northwest National Laboratory, 5825
4   University Research Court, College Park, Maryland 20740, United States
5
6   [2] Department of Earth and Environmental Engineering, Columbia University, New York, NY
7   10027, United States
8
9   [3] Department of Land Resources and Environmental Sciences, Montana State University,
10   Bozeman, MT 59717, United States
11
12   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 construct both a lumped and a distributed scheme of the HE based on the monthly *"abcd"* model to explore the tradeoff between computational cost and model fidelity. 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 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

36   extraordinary advantage in computational efficiency. Our results suggest that the revised lumped

37   "*abcd*" model can serve as an efficient and acceptable HE for complex GHMs and is suitable for

38   broad practical use, and the distributed scheme is also an efficient alternative if spatial

39   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

63 quantification are often needed for decision making, but the users usually cannot afford to run a

64 large number of simulations with many GHMs like the VIC (also categorized as land surface

65 model (LSM)) due to their high computational expense (Oubeidillah et al., 2014). Another

66 situation is when the users seek reasonable estimates of water resources with minimal efforts

67 rather than acquiring highly accurate estimates through expensive inputs of time and efforts. For

68 example, when users seek to explore the hydroclimatology of a region and its long-term water

69 balance (Sankarasubramanian and Vogel, 2002), then GHMs with fine spatial (e.g., 1/8 degree)

70 and temporal resolution (e.g., hourly) are not necessarily needed. In this case, simple models

71 that possess reasonable predictability and are computationally efficient tend to be more suitable.

72 In addition, some studies have shown that GHMs/LSMs are sometimes outperformed by simple

73 empirical statistical models (Abramowitz, 2005; Abramowitz et al., 2008; Best et al., 2015),

74 suggesting that some GHMs/LSMs may underutilize the information in their climate inputs and

75 that model complexity may undermine accurate prediction. This also indicates the potential

76 advantages of simple model over complex GHMs/LSMs. Thus, constructing simple models that

77 can emulate the dynamics of more complex and computational expensive models (e.g.,

78 GHMs/LSMs) is warranted.

79   The motivation of this work arises from the need to construct a hydrological emulator

80 (HE) that can efficiently mimic the complex GHMs to address the abovementioned issues for

81 practical use, which provides the opportunity of speeding up simulations at the cost of

82 introducing some simplification. We develop a HE that is easy-to-use and efficient for any

83 interested groups or individuals to assess water cycle at basin/regional/global scales. This HE

84 possesses the following features: 1) minimum number of parameters; 2) minimal climate input

85 that is easy to acquire; 3) simple model structure; 4) reasonable model fidelity that captures both

86   the spatial and temporal variability; 5) high computational efficiency; 6) applicable in a range of

87   spatial scales; and 7) open-source and well-documented.

88   To achieve our goal of identifying a suitable HE, we have explored many hydrological

89   models to find one that may meet our needs. We start with a simple baseline model characterized

90   by mean seasonal cycle; i.e., the inter-annual mean value for every calendar day (Schaefli &

91   Gupta, 2007). Among others, we also explore the *"abcd"* model because: 1) it is widely-used

92   and proven to have reasonable predictability (Fernandez et al., 2000; Martinez and Gupta, 2010;

93   Sankarasubramanian and Vogel, 2002; Sankarasubramanian and Vogel, 2003; Thomas, 1981;

94   Vandewiele and Xu, 1992; Vogel and Sankarasubramanian, 2003); 2) it uses a monthly time step

95   and requires less computational cost than daily or hourly models; 3) it has solid physical basis

96   hence has potential to be extended to other temporal scales (Wang and Tang, 2014); 4) it requires

97   minimal and easily-available inputs; 5) it only involves 4-7 parameters; and 6) it can simulate

98   variables of interest such as recharge, direct runoff and baseflow that many other simple models

99   can't simulate (Vörösmarty et al., 1998) . This study marks the first time that the "*abcd*" based

100   model is applied globally, and also the first time the predictability and computational efficiency

101   for both the lumped and distributed schemes are evaluated. Below we describe the baseline and

102   the "*abcd*" models and data in Section 2; and we present the evaluation of the two models,

103   discuss their appropriateness of serving as a HE in Section 3; finally, in Section 4 we summarize

104   this work with concluding remarks.

105

106   **2 Methods and data**

107   **2.1 Model description**

108   We examine two simple models – baseline and the "*abcd*" model (both lumped and

109 distributed scheme) in order to identify a suitable one for serving as a HE.

110 2.1.1 Baseline model

111   Following the work of Schaefli & Gupta (2007), we explore a baseline model

112 characterized by the inter-annual mean value for every calendar day, i.e., climatology. In this

113 study, the baseline model is based on monthly climatology runoff, which comes from a model

114 simulation product – i.e., the runoff product from the Variable Infiltration Capacity (VIC) model

115 (Leng et al. 2015). Specifically, we first calculate grid-level inter-annual mean value for each of

116 the 365 calendar days from daily runoff of the benchmark product during 1971-2010 (see Section

117 2.3.2), and then aggregate daily climatology runoff to monthly climatology runoff at grid-level.

118 The baseline model here uses monthly climatology runoff for prediction. For example, if the

119 climatology runoff for July in one grid cell is 100 mm mon^{-1}, then the prediction of total runoff

120 for July of every year in that specific grid cell is 100 mm mon^{-1}.

121

122 2.1.2 The "*abcd*" model

123 The monthly *"abcd"* model was first introduced by Thomas (1981) to improve the national

124 water assessment for the U.S., with a simple analytical framework using only a few descriptive

125 parameters. It has been widely used across the world, especially for the U.S. (Martinez and

126 Gupta, 2010; Sankarasubramanian and Vogel, 2002; Sankarasubramanian and Vogel, 2003). The

127 model uses potential evapotranspiration (PET) and precipitation (P) as input. The model defines

128 four parameters *a*, *b*, *c*, and *d* that reflect regime characteristics (Sankarasubramanian and Vogel,

129 2002; Thomas, 1981) to simulate water fluxes (e.g., evapotranspiration, runoff, groundwater

130 recharge) and pools (e.g., soil moisture, groundwater). The parameters *a* and *b* pertain to runoff

131   characteristics, and *c* and *d* relate to groundwater. Specifically, the parameter *a* reflects the

132   propensity of runoff to occur before the soil is fully saturated. The parameter b is an upper limit

133   on the sum of evapotranspiration (ET) and soil moisture storage. The parameter c indicates the

134   degree of recharge to groundwater and is related to the fraction of mean runoff that arises from

135   groundwater discharge. The parameter d is the release rate of groundwater to baseflow, and thus

136   the reciprocal of d is the groundwater residence time. Snow is not part of the original "*abcd*"

137   model, which may result in poor performance of the model in cold regions where snow

138   significantly affects the hydrological cycle. The work of Martinez and Gupta (2010) has added

139   snow processes into the original "*abcd*" model, where the snowpack accumulation and snow

140   melt are estimated based on air temperature. Their work indicated that incorporation of the snow

141   processes in the monthly "*abcd*" model has significantly improved model performance in snow-

142   covered area in the conterminous United States (see Figure 4 in Martinez and Gupta (2010)).

143   In this study, we adopt the "*abcd*" framework from Martinez and Gupta (2010) (Fig. 1);

144   meanwhile, we make three modifications to suit the needs of a HE for global applications. First,

145   in order to enhance the model efficiency with as least necessary parameters as possible, instead

146   of involving three tunable snow-related parameters in the calibration process, we set the values

147   for two of the parameters (i.e., temperature threshold above or below which all precipitation falls

148   as rainfall or snow) from literature (Wen et al., 2013) and only keep one tunable parameter m –

149   snow melt coefficient ($0 < m < 1$). Second, we introduce the baseflow index (BFI) into the

150   calibration process to improve the partition of total runoff between the direct runoff and baseflow

151   (see Section 2.4). Third, other than the lumped scheme as previous studies used, we first explore

152   the values of model application in distributed scheme with a grid resolution of 0.5 degree. The

153   detailed model descriptions and equations are presented in the Appendix A, and the descriptions

154   and ranges of model parameters are listed in Table 1.

155

156   **2.2 Model structure**

157   In terms of the *"abcd"* model, we evaluate both the lumped and distributed model

158   schemes, although most previous applications of the model are conducted in a lumped scheme

159   (Bai et al., 2015; Fernandez et al., 2000; Martinez and Gupta, 2010; Sankarasubramanian and

160   Vogel, 2002; Sankarasubramanian and Vogel, 2003; Vandewiele and Xu, 1992; Vogel and

161   Sankarasubramanian, 2003). In the lumped scheme, each of the 235 river basins is lumped as a

162   single unit, and each of the data input (see Section 2.3.1) is the lumped average across the entire

163   basin, and thus all the model outputs are lumped as well. In terms of the distributed one, however,

164   each 0.5-degree grid cell has its own data inputs, and likewise, the model outputs are simulated

165   at the grid-level. Although the two schemes differ in the spatial resolution of their inputs and

166   outputs, their within-basin parameters are uniform. We use basin-uniform rather than grid-

167   specific parameters for the distributed scheme for two reasons: 1) to enhance computational

168   efficiency; and 2) to avoid drastically different parameters for neighboring grid cells that may be

169   unrealistic. Note that lateral flows between grid cells and basins are not included at this stage for

170   the *"abcd"* model. For the baseline model, as documented in Section 2.1.1, every 0.5-degree grid

171   cell of each basin has its own monthly climatology runoff estimates for each of the 12 calendar

172   months.

173

174   **2.3 Data**

175   2.3.1 Climate data

176   The climate data needed for the *"abcd"* model only involve monthly total precipitation,

177   monthly mean, maximum and minimum air temperature. The data we use is obtained from

178   WATCH (Weedon et al., 2011), spanning the period of 1971-2010, and it is 0.5-degree gridded

179   global monthly data. The climate data is used for model simulation over the global 235 major

180   river basins (Kim et al., 2016). Additionally, we use the Hargreaves-Samani method (Hargreaves

181   and Samani, 1982) to estimate potential evapotranspiration (PET), which is a required input for

182   the *"abcd"* model, and it needs climate data of mean, maximum and minimum temperatures for

183   the calculation.

184

185   2.3.2 Benchmark runoff product

186   In this study, the "*abcd*" model is tested for its ability to emulate the naturalized

187   hydrological processes of a reference model since the "true" naturalized hydrological processes

188   are unknown. The "perfect model" approach is well adopted in climate modeling studies where

189   one model is treated as "observations" while the others are tested for their ability to reproduce

190   "observations" (Murphy et al., 2004; Tebaldi and Knutti, 2007). Here, we use the process-based

191   VIC model as the "perfect model", which was also driven by the WATCH climate forcing.

192   The VIC runoff product here is a global simulation with a daily time step and spatial

193   resolution of 0.5 degree for the period of 1971-2010, and the VIC daily runoff is aggregated to

194   monthly data to be consistent with the temporal scale of the "*abcd*" model. The VIC model

195   settings used in this study are based on the University of Washington VIC Global applications

196   (http://www.hydro.washington.edu/Lettenmaier/Models/VIC/Datasets/Datasets.shtml). The sub-

197   grid variability of soil, vegetation and terrain characteristics are represented in sub-grid area-

198   specific parameter classifications. Soil texture and bulk densities are derived by combining the

199 World Inventory of Soil Emission Potentials database (Batjes, 1995) and the 5-min digital soil

200 map of the world from the Food and Agricultural Organization (FAO, 1998). Based on the work

201 of (Cosby et al., 1984), the remaining soil properties (e.g. porosity, saturated hydraulic

202 conductivity and unsaturated hydraulic conductivity) are derived. Vegetation type data are

203 obtained from the global land classification of (Hansen et al., 2000). Parameters including the

204 infiltration parameter, soil layer depths and those governing the baseflow function were

205 calibrated for major global river basins and transferred to the global domain as documented in

206 (Nijssen et al., 2001b), based on which Zhang et al. (2014) and Leng et al. (2015) conducted

207 additional calibrations in the China domain. In this study, the VIC model was forced by WATCH

208 climate forcing at the daily time step (Weedon et al., 2011), based on the calibrated parameters

209 from Nijssen et al. (2001b), Zhang et al., (2014) and Leng et al., (2015). The simulated runoff

210 used in this study has recently been validated globally within the framework of the Inter-Sectoral

211 Impact Model Intercomparison Project and shows reasonable performance compared to other

212 hydrological models (Hattermann et al., 2017; Krysanova and Hattermann, 2017).

213   The VIC runoff product (Hattermann et al., 2017; Leng et al., 2015) is then used as a

214 benchmark for calibrating and validating the "*abcd*" model due to two reasons. First, VIC runoff

215 has been evaluated across many regions of the globe and is proved to be reasonably well

216 (Abdulla et al., 1996; Hattermann et al., 2017; Maurer et al., 2001; Nijssen et al., 1997; Nijssen

217 et al., 2001b). Second, the simulated monthly runoff by the "*abcd*" model is more representative

218 of "natural conditions" because human activities (e.g., reservoir regulations and upstream water

219 withdrawals) are currently not represented in the model. Thus it tends to be more reasonable to

220 compare the simulated runoff against the VIC natural runoff product rather than comparing

221 against observed streamflow data from stream gauges (Dai et al., 2009; Wilkinson et al., 2014).

222   Despite potential bias in the VIC runoff product, using it as a benchmark here is to demonstrate

223   the capability of the HE developed in this work to mimic complex GHMs. Furthermore, the

224   application of the HE is not tied to the VIC model and should be able to emulate other GHMs.

225   The VIC runoff product compares well to other products (see Fig. S1, S2), including the

226   University of New Hampshire/Global Runoff Data Centre (UNH/GRDC) runoff product (Fekete

227   and Vorosmarty, 2011; Fekete et al., 2002) and the global streamflow product (Dai et al., 2009).

228   The scatterplot pattern of the VIC long-term annual runoff product vs. the GRDC product

229   (GRDC, 2017) matches well with that of the UNH/GRDC runoff vs. the GRDC product

230   (streamflow is transferred to the same unit as runoff via dividing by the basin area), which means

231   the behavior of the VIC runoff product is similar to that of the UNH/GRDC product. Further, the

232   correlation coefficient of the VIC and the UNH/GRDC long-term annual runoff is as high as 0.83

233   across the global 235 basins (Fig. S2). This suggests the reasonableness of VIC runoff product,

234   because the UNH/GRDC runoff is calibrated with the GRDC observations. At the same time, the

235   discrepancies between the VIC runoff products and the streamflow products (Fig. S2) may be

236   attributed to human activities, such as reservoir regulations and upstream water withdrawals,

237   which are not embedded in the runoff but reflected in the streamflow. This is because the VIC

238   model simulates runoff at natural conditions, and then a stand-alone routing model can be used to

239   route these flows downstream (Nijssen et al., 2001b). The routing model may account for human

240   activities such as water extractions and reservoir operations (Haddeland et al., 2014). However,

241   here we use the VIC runoff under natural conditions as the benchmark product, which explains

242   the discrepancies between the VIC runoff and observed streamflow products.

243   Uncertainties arising from the runoff process in the VIC model should be acknowledged.

244   Implementation of different runoff generation schemes (e.g. TOPMODEL) within the same

245   modeling framework is an alternative that can be adopted in the future to explore the uncertainty

246   range. A recent inter-model comparison study shows that the VIC model falls within the range of

247   large model ensembles (Hattermann et al. 2017). Notably, groundwater and its interaction with

248   river and land surface are not represented in the model. Thus, the model may not be able to fully

249   capture the hydrologic responses in areas where lateral flow and the three way streamflow-

250   aquifer-land interactions are important. Further, vegetation dynamics and water management that

251   may affect runoff are not considered in the model simulations. Nonetheless, the use of the HE

252   documented here is not tied to the VIC, and it could be used to emulate other GHMs of interest.

253

254   **2.4 Model calibration**

255   Typically, most applications of the "*abcd*" model utilize single-objective optimization for

256   total runoff (or streamflow) during the calibration process to minimize the difference between

257   measured and simulated streamflow (Bai et al., 2015; Martinez and Gupta, 2010;

258   Sankarasubramanian and Vogel, 2002). While this may lead to a good fit for simulated total

259   runoff, however, it may result in inappropriate partition of total runoff between direct runoff and

260   baseflow. To improve the accuracy of the simulated total runoff and the partition between direct

261   runoff and baseflow, we introduce the baseflow index (BFI) into the objective function.

262   Unlike the baseline model, the "*abcd*" model requires a calibration step for reasonable

263   parameterization so as to enable good prediction. As mentioned above, we incorporate BFI into

264   the objective function during the calibration process. On one side, we maximize Kling-Gupta

265   efficiency (KGE) (Gupta et al., 2009), which is used as a metric to measure the accuracy of the

266   simulated total runoff relative to the VIC benchmark runoff. The KGE is defined as the

267   difference of unity and the Euclidian distance (ED) from the ideal point, thus we maximize KGE

268   through minimizing the ED. The KGE and ED are calculated as follows (Gupta et al., 2009):

269      $KGE = 1 - ED$ (1)

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

271      $r = \dfrac{Cov_{so}}{\sigma_s \cdot \sigma_s}$ (3)

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

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

274   where $r, \alpha, \beta,$ and Cov_{so} are relative variability, bias, correlation coefficient, and covariance

275   between the simulated and observed values (here we treat the VIC runoff as the observed),

276   respectively; μ and σ represent the mean and standard deviation (subscript "s" and "o" stand for

277   simulated and observed values). On the other side, we also nudge the simulated BFI towards the

278   benchmark BFI (here we treat the benchmark BFI as the observed) – the mean BFI of the four

279   products from (Beck et al., 2013). Then, the objective function is as follows:

280      $\min(ED + abs(BFI_{obs} - BFI_{sim}))$ (6)

281   where min stands for minimizing the value in the parenthesis, abs represents absolute value, ED

282   is the Euclidian distance between the simulated and observed total runoff (Gupta et al., 2009),

283   BFI_{obs} and BFI_{sim} are the observed and simulated BFI, respectively. Here we treat the benchmark

284   runoff from the VIC and BFI from Beck et al. (2013) as observed values. We then minimize the

285   objective function for parameter optimization by utilizing a Genetic Algorithm (GA) routine

286   (Deb et al., 2002). Note that for the distributed model scheme, we aggregate the grid-level total

287   runoff estimates to basin-level and then nudge it toward basin-level benchmark total runoff

288   during the calibration process.

289

2.5 Model simulations

291   To evaluate the predictability and efficiency of the baseline and the "*abcd*" model so as

292   to identify a suitable one to serve as a HE, we have conducted a series of simulations.

293   Specifically, for the baseline model, no simulations are needed as it uses inter-annual mean value

294   for each month – 12 monthly values – as prediction, so we just replicate the 12 monthly runoff

295   for 1971-2010 and for each of the global 235 basins, and then compare against the benchmark

296   runoff product. For the "*abcd*" model, two sets of model simulations across the global 235 basins

297   are conducted, with one set for calibration and the other one for validation, for both the lumped

298   and distributed model schemes. For the first set, we run the model for each basin for the period

299   of 1971-1990 to get basin-specific parameters by using the GA approach (see Section 2.4). For

300   the second set, using the parameters identified in the first set of simulation, we run the model for

301   the period of 1991-2010 to validate the model predictability and also evaluate the computational

302   efficiency. Model inputs and outputs in the distributed scheme are at a spatial resolution of 0.5-

303   degree, whereas those in the lumped scheme are all in lumped single unit for each basin. All

304   model simulations are conducted in a monthly time step. Note that broad users can run the

305   identified HE for global 235 basins, or for as many basins as they want for either scheme, as all

306   the related basin-specific input data and calibrated parameters for both schemes are open-source.

307

3 Results and discussions

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

310   Generally, we find baseline model performs worse than the "*abcd*" model (Fig. 2). The

311   baseline model exhibits a lower global mean KGE value (0.61) than the lumped and distributed

312   schemes of the "*abcd*" model (0.75 and 0.79, respectively). In addition, our analysis indicates

313   that the incorporation of BFI into the objective function leads to a significant improvement in the

314   partition of total runoff between direct runoff and baseflow (Fig. 3, Fig. S4), without

315   compromising predictability for total runoff, i.e., the global mean KGE values for modeled total

316   runoff with or without the incorporation of BFI are almost the same (0.75 vs 0.76). Specifically,

317   for the case of involving both the total runoff and BFI in the objective function, the correlation

318   efficiencies (r) between the long-term annual benchmark and modeled direct runoff, and between

319   benchmark and modeled baseflow from the lumped scheme across global basins are both 0.98

320   (Fig. 3), which are much higher than those of 0.86 and 0.72 in the case of only involving the total

321   runoff in the objective function (Fig. S4). Given the superiority of the "*abcd*" model over the

322   baseline model, we focus in the following sections on evaluating the predictability and

323   computational efficiency of the "*abcd*" model and its potential to serve as a HE.

324

325   **3.2 Evaluation of model predictability**

326   In terms of total runoff, we find the lumped and distributed schemes are comparably

327   capable in simulating long-term mean annual quantity, temporal variations and spatial patterns

328   for the vast majority of river basins globally (Fig. 3-5). Estimates of long-term mean annual total

329   runoff from both the lumped and distributed schemes match very well with that of VIC total

330   runoff across the 235 basins, with a correlation coefficient (r) of higher than 0.96, for both the

331   calibration and validation period (Fig. 3). Similarly, the basin-level estimates of long-term mean

332   annual direct runoff and baseflow also match well with those of the VIC across the globe, for

333 both schemes and both periods (Fig. 3). This suggests both schemes possess the capability in

334 partitioning total runoff.

335   Furthermore, both schemes display good capability in capturing the seasonal variations of

336 the total runoff for both the calibration and validation period (Fig. 4, Fig. S5). Meanwhile,

337 although the spatial patterns of annual total runoff from the lumped scheme present a general

338 match with that of the VIC, it does not reflect the spatial variations inside a basin that is however

339 captured by the distributed scheme (Fig. 5). Likewise, overall much lower percentage differences

340 between the modeled runoff from the distributed scheme and the VIC runoff product than those

341 between the VIC and the lumped one further corroborate the significantly better performance of

342 the distributed scheme (Fig. S6). Both schemes still show large percentage differences in some

343 dry (e.g., North Africa) or cold regions (e.g., Tibet Plateau). This is because the runoff there is at

344 a low magnitude and thus small changes in runoff will lead to large percentage differences.

345 Therefore, the distributed scheme provides overall slightly higher KGE (Fig. 6), with a global

346 mean KGE value of 0.79 as compared to 0.75 for the lumped scheme (Fig. 2).

347   To ensure good model predictability for the major water fluxes, we also evaluate the

348 modelled ET estimates. The modelled ET compares reasonably well with the VIC ET product as

349 well as with the mean synthesis of the LandFlux-EVAL ET product (Mueller et al., 2013),

350 displaying similar spatial variations (Fig. S7). Likewise, the distributed *"abcd"* scheme tends to

351 have better capability in presenting spatial heterogeneity than the lumped one. In addition, the

352 percentage differences between our modeled ET and the VIC ET product further confirm that the

353 distributed scheme significantly outperforms the lumped one (Fig. S8), with much lower

354 differences from the VIC ET product, although discrepancies still exist in some extremely cold

355 (e.g., Greenland) or dry regions (e.g., North Africa), which is because small differences in ET

356 will lead to large percentage difference in those regions with low ET. Further, given the changes

357 in basin-scale monthly soil moisture is relatively small, precipitation should approximate the sum

358 of ET and runoff according to the water mass balance, the good predictability of seasonality in

359 runoff as illustrated in Fig. 4 also reflects similar performance for ET.

360   The distributed scheme appears to outperform the lumped scheme in term of goodness-

361 of-fit, especially in some cold (e.g., Arctic, Northern European, Interior Tibet) and in some dry

362 (e.g., North Africa) regions (Fig. 6). This is possibly because distributed inputs can reflect basin-

363 level heterogeneity, and thus better capture the characteristic of the hydrological conditions in

364 those regions. However, both schemes do not perform well in the southern end of the Andes

365 Mountains (Fig. 6). This may be attributed to the complex land surface characteristics in that

366 mountainous area, which cannot be resolved due to the coarse spatial resolution. Moreover, the

367 distributed scheme seems not performing very well in some cold regions (Fig. 6), which is

368 possibly due to lack of representation for permafrost in the model.

369   Previous studies investigating the credibility of lumped and distributed hydrological

370 models indicate that, in many cases, lumped models perform comparably or just as well as

371 distributed models (Asadi, 2013; Brirhet and Benaabidate, 2016; Ghavidelfar et al., 2011;

372 Michaud and Sorooshian, 1994; Obled et al., 1994; Reed et al., 2004; Refsgaard and Knudsen,

373 1996; YAO et al., 1998). However, distributed models may have advantages for predicting

374 runoff in ungauged watersheds (Reed et al., 2004; Refsgaard and Knudsen, 1996), for capturing

375 spatial distribution of runoff due to heterogeneity in rainfall patterns or in land surface (Downer

376 et al., 2002; Paudel et al., 2011; YAO et al., 1998). Our results on the predictability of lumped

377 and distributed "*abcd*" model are in line with previous findings in the literature.

378   The good agreement between our modelled water fluxes, including total runoff, direct

379   runoff, baseflow and ET, and the benchmark products provides confidence in the capability of

380   both the lumped and distributed schemes in estimating temporal and spatial variations in major

381   water fluxes across the globe. In addition, to identify a suitable HE, the required computation

382   cost is another key factor as detailed below.

383

384   **3.3 Evaluation of computational efficiency**

385   While the performance of model predictability is comparable for the lumped and

386   distributed schemes as elucidated above, great disparities still exist for runtime of the two

387   schemes and the VIC model (Table S1). Take the Amazon basin that covers a total number of

388   2002 0.5-degree grid cells as an example, it takes 11.05 minutes for model calibration via the GA

389   method for the distributed scheme but only 0.16 minute for the lumped one. Similar disparity is

390   also found for model simulation with calibrated parameters, with runtime of 0.03 and 3.20

391   seconds for a 1000-year simulation of the Amazon basin for the lumped and distributed schemes,

392   respectively. However, according to the authors' experience, it will take ~1 week for the VIC

393   model to accomplish the same job, which is far more computationally expensive. In general, the

394   computational efficiency of the lumped scheme is two orders of magnitudes higher than the

395   distributed one, although that of the distributed one is still much higher than the VIC (~five

396   orders of magnitude) and many other GHMs and LSMs. Note that all of the simulations here are

397   conducted on the Pacific Northwest National Laboratory (PNNL)'s Institutional Computing (PIC)

398   Constance cluster using 1 core (Intel Xeon 2.3 GHz CPU) with the same configuration.

399

400   **3.4 Potential application of the "*abcd*" model as a hydrological emulator**

401   The good predictability and computational efficiency of both the distributed or lumped

402   schemes as elucidated in Sections 3.2 and 3.3 suggest its suitability for serving as HEs that can

403   efficiently emulate complex GHMs (e.g., the VIC or others). The source codes, input data, basin-

404   specific parameters across the globe for both the lumped and distributed schemes are open-

405   source and well-documented, which will make the HE ready to use and facilitate their wide and

406   easy use with minimal efforts.

407   The choice of either the distributed or lumped scheme as HE depends on the user's

408   specific needs. There is a tradeoff between the model predictability and computational efficiency.

409   While the distributed scheme tends to better capture the spatial heterogeneity of water fluxes and

410   can produce grid-level outputs that lumped scheme cannot, it incurs higher computational cost

411   than the lumped scheme. For applications that aim to strike a balance between predictability and

412   computation cost, such as practical assessment of water resources, or estimation of water supply

413   for integrated assessment models (IAMs), or quantification of uncertainty and sensitivity

414   analyses, it would be reasonable to employ the lumped scheme as a HE. The lumped scheme is

415   especially advantageous due to its minimal calibration and computational cost, parsimonious

416   efforts for model implementation, and reasonable fidelity in estimating major water fluxes (e.g.,

417   runoff, ET). For users from the IAM community, the lumped scheme might be sufficiently

418   suitable for their needs since 1) the lumped scheme can operate at the same spatial resolution at

419   which IAMs typically balance water demands and supplies (Edmonds et al., 1997; Kim et al.,

420   2006; Kim et al., 2016), and 2) the inherent uncertainty of the lumped scheme is likely

421   comparable or even overshadowed by the intrinsic uncertainty of IAMs (Kraucunas et al., 2015;

422   O'Neill et al., 2014). Similarly, for users who aim to conduct uncertainty and sensitivity analyses,

423   the high computational efficiency of the lumped scheme allow the users to emulate the

424   hydrological model of interest (e.g., GHMs, LSMs) and then run a large number of simulations

425   to conduct their uncertainty and sensitivity analysis (Scott et al., 2016). Therefore, the high

426   computational efficiency makes the lumped scheme more appealing as a HE in these cases.

427   However, if the research questions hinge on the gridded estimates, or emphasize the spatial

428   heterogeneity of the water fluxes or pools, it would be more desirable to deploy the distributed

429   scheme as a HE instead. For example, a follow-up work is coupling the distributed scheme of the

430   HE with a widely-used IAM, the Global Change Assessment Model (GCAM, Edmonds et al.,

431   1997), and then using the coupled model to investigate the impacts of a variety of land use

432   policies on global water scarcity, where the HE is used to estimate grid-level runoff globally

433   under different land use policies.

434   While many studies indicate that basin runoff generation is sensitive to factors such as

435   physical characteristics, spatiotemporal variability in storage distribution and forcing input,

436   evidence also show that basin response can be captured using a handful of parameters (Hsu et al.,

437   1995; Young and Parkinson, 2002). In this study, the lumped scheme of the HE ignores the

438   spatiotemporal variability in basin characteristics by averaging the input forcing data;

439   consequently, the associated responses in within-basin runoff or ET variations cannot be

440   captured. In contrast, the distributed scheme presents a better performance in capturing

441   spatiotemporal variability of runoff and ET with use of the same input data, and without

442   increasing the number of parameters. Thus, the use of the distributed scheme is preferred when

443   the tradeoff in the computational efficiency is not a constraining factor.

444   Moreover, a combination of a top-down approach (Sivapalan et al., 2003) and a multi-

445   objective approach to model evaluation (Gupta et al., 1998) could be used to explore internal

446   basin behavior, wherein the top-down approach would start from a simple structure and then

447   progressively expand based on its caveats in reproducing overall basin behavior [e.g.,

448   Jothityangkoon et al., 2001]. In this study we adopt a similar framework, by starting from a

449   baseline model and then expanding to the "abcd" model with snow representation, also by

450   incorporating the baseflow index into the objective function to exert a multi-objective approach.

451   Our assessment indicates that a baseline model characterized by mean seasonal cycle still holds a

452   promise in predicting runoff at basins with small variability in basin characteristics, such as

453   basins of Ob, Lena, Yenisey, Siberia and Mackenzie in the Arctic area, where the baseline model

454   yields KGE values of greater than 0.90 from our evaluation. Further, while Martinez and Gupta

455   (2010) indicated that the incorporation of the snow component and an additional snow parameter

456   into the original "abcd" model has greatly improved model performance in snow-prevailed

457   regions, areas without prevailing snow (e.g., tropical zone) could still utilize the original version

458   of the "abcd" model to keep the model as parsimonious as possible without compromising model

459   predictability. In addition, although our results reveal that incorporation of baseflow index into

460   the objective function generally improves the model performance in partitioning of runoff

461   between direct runoff and baseflow, simply employing a single-objective approach (i.e., only

462   involving total runoff) also works well for some basins such as North Interior Africa and Interior

463   Australia. Thus, the single-objective approach is also acceptable for those basins with the

464   advantage of simplicity without compromise in performance. In short, according to specific basin

465   characteristics and the research needs, suitable model complexity and number of parameters

466   could be identified by following abovementioned scenarios, such that either the baseline model

467   or a reduced format of the HE (e.g., without snow representation or single-objective) could be

468   potentially utilized with the merits of simplicity, reasonable predictability and computational

469   efficiency, rather than adopting the full format of the HE. This HE could be used to emulate a

470   wide range of models with different spatial and temporal complexities, and its performance may

471   vary from model to model. Thus, examining and comparing the extent to which the HE could

472   mimic the behaviors of different GHMs and LSMs is of our future research interest. In addition,

473   future research can extend this work by systematically investigating the role of different levels of

474   inputs and parameters on model performance in different basins across the globe.

475   Based upon our open-source HE and the validated basin-specific parameters across the

476   globe, researchers can easily investigate the variations in water budgets at the basin/

477   regional/global scale of interest, with minimum requirements of input data, efficient computation

478   performance and reasonable model fidelity. Likewise, researchers can utilize the framework of

479   the HE with any alternative input data, or recalibrate the HE to emulate other complex GHMs or

480   LSMs of interest, to meet their own needs.

481

482   **3.5 Case study for uncertainty analysis**

483   To demonstrate the capability of the examined "*abcd*" model serving as a HE, we use the

484   lumped scheme to conduct parameter-induced uncertainty analysis for the runoff simulation at

485   the world's sixteen river basins with top annual flow (Dai et al. 2009). Specifically, for each of

486   the sixteen basins, we first apply ±10% change to each of the five calibrated parameters (a, b, c,

487   d, m) to compose varying ranges; note that we just truncate the range to those valid in Table 1 if

488   the ±10% change exceeds the valid range. Then we randomly sample the five parameters from

489   corresponding ranges for 100,000 times (i.e., 100,000 combinations of parameters). After that,

490   we run the lumped scheme 100,000 times for each basin with the 100,000 combinations of

491   parameters to examine the parameter-induced uncertainty in total runoff. The uncertainty

492   analysis indicates that most basins are robust to changes in parameters, other than the Tocantins,

493   Congo and La Plata (Fig. 7). In other words, for basins Congo and La Plata, slight changes in

494   parameters may lead to large changes in runoff estimates. Then the uncertainty in the calibrated

495   parameters for the two basins may lead to large bias in the simulated runoff, which may more or

496   less explain why modeled runoff for the two basins tend to have higher biases than other basins

497   (Fig. 4). Notably, the 100,000 times of simulations only takes ~80 seconds on a Dell Workstation

498   T5810 with one Intel Xeon 3.5 GHz CPU, which demonstrates the extraordinary computational

499   efficiency of the lumped scheme and its advantage for serving as a HE.

500

501   **4 Conclusions**

502   Toward addressing the issue that many global hydrological models (GHMs) are

503   computationally expensive and thus users cannot afford to conduct a large number of simulations

504   for various tasks, we firstly construct a hydrological emulator (HE) that possesses both

505   reasonable predictability and computation efficiency for global applications in this work. Built

506   upon the widely-used *"abcd"* model, we have adopted two snow-related parameters from

507   literature rather than tuning them for parameter parsimony, and also have improved the partition

508   of total runoff between the direct runoff and baseflow by introducing baseflow index into the

509   objective function of the parameter optimization. We then evaluate the appropriateness of the

510   model serving as an emulator for a complex GHM – the VIC, for both the lumped and distributed

511   model schemes, by examining their predictability and computational efficiency.

512   In general, both distributed and lumped schemes have comparably good capability in

513   simulating spatial and temporal variations of the water balance components (i.e., total runoff,

514   direct runoff, baseflow, evapotranspiration). Meanwhile, the distributed scheme has slightly

515   better performance than the lumped one (e.g., capturing spatial heterogeneity), with mean Kling-

516   Gupta efficiency of 0.79 vs. 0.75 across global 235 basins, and also it provides grid-level

517   estimates that the lumped one incapable of. Additionally, the distributed scheme performs better

518   in extreme climate regimes (e.g., Arctic, North Africa) and Europe. However, the distributed one

519   incurs two more orders of magnitudes of computation cost than the lumped one. A case study of

520   uncertainty analysis with 100, 000 simulations for each of the world's sixteen basins with top

521   annual streamflow further demonstrates the lumped scheme's extraordinary advantage in terms

522   of computational efficiency. Therefore, the lumped scheme could be an appropriate HE –

523   reasonable predictability and high computational efficiency. At the same time, the distributed

524   scheme could be a suitable alternative for research questions that hinge on grid-level spatial

525   heterogeneity. Finally, upon open-sourcing and well-documentation, the HE is ready to use and it

526   provides researchers an easy way to investigate the variations in water budgets at a variety of

527   spatial scales of interest (e.g., basin, region or globe), with minimum requirements of efforts,

528   reasonable model predictability and extraordinary 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 user's manual are available. In addition, the HE documented here has been translated into Python and is being incorporated into Xanthos (Li et al., 2017), which is an open-source global hydrologic model that allows users to run different combinations of evapotranspiration, runoff, and routing models. The HE will be the default runoff model used in Xanthos 2.0 and will be available on GitHub (https://github.com/JGCRI/xanthos).

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

560   in the model, in order to alleviate the computation load during the parameter optimization

561   process.

562   The model defines two state variables "available water" and "evapotranspiration

563   opportunity", denoted as W_i and Y_i, respectively. The W_i is defined as:

564    $W_i = SM_{i-1} + Rain_i + SNM_i$ (4)

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

566   snowmelt during period i.

567   Y_i stands for the maximum water that can leave the soil as evapotranspiration (ET) at

568   period i, and it is defined as below:

569    $Y_i = ET_i + SM_i$ (5)

570   where ET_i is the actual ET at time period i and SM_i is soil moisture at the end of time step i.

571   Further, Y_i has a non-linear relationship with W_i as:

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

573   where a and b are parameters detailed in Section 2.1.2.

574   Allocation of W_i between ET_i and SM_i is estimated by assuming that the loss of soil

575   moisture by ET will be proportional to potential evapotranspiration (*PET)* as:

576    $\dfrac{dS}{dt} = -PET \times \dfrac{SM}{b}$ (7)

577   where *PET* is calculated by using the Hargreaves-Samani method (Hargreaves and Samani,

578   1982).

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

580    $SM_i = Y_i \times \exp(-PET_i / b)$ (8)

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

582   In the model framework, $W_i - Y_i$ is the sum of the groundwater recharge (RE) and direct

583   runoff (Q_d), and the allocation is determined by the parameter c:

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

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

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

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

588   where d is a parameter reflecting the release rate of groundwater to baseflow. Then the total

589   runoff (Q_t) is the sum of the direct runoff and baseflow:

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

591   The GW_i is the sum of groundwater storage at the end of last time step and the groundwater

592   recharge minus the baseflow, and GW_i is derived as:

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

594   Then, all the water fluxes and pools are solved.

595   Author contribution

596   Yaling Liu and Mohamad Hejazi designed this work, and all co-authors offered help through
597   discussions. Yaling Liu developed the hydrological emulator and conducted the simulations and
598   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.

601   **Acknowledgement:** This research was supported by the Office of Science of the U.S.

602   Department of Energy through the Integrated Assessment Research Program. PNNL is operated

603   for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. We thank Chris

604   Vernon for his help in maintaining the Github repository for the hydrological emulator

605   (https://github.com/JGCRI/hydro-emulator/).

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.

652 Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic
653   algorithm: NSGA-II. IEEE transactions on evolutionary computation, 6(2): 182-197.
654 Döll, P., Fiedler, K., Zhang, J., 2009. Global-scale analysis of river flow alterations due to water
655   withdrawals and reservoirs. Hydrology and Earth System Sciences, 13(12): 2413-2432.
656 Downer, C.W., Ogden, F.L., Martin, W.D., Harmon, R.S., 2002. Theory, development, and
657   applicability of the surface water hydrologic model CASC2D. Hydrol. Processes, 16(2):
658   255-275.
659 Edmonds, J. et al., 1997. An integrated assessment of climate change and the accelerated
660   introduction of advanced energy technologies-an application of MiniCAM 1.0.
661   Mitigation and adaptation strategies for global change, 1(4): 311-339.
662 FAO, 1998. Digital soil map of the world and derived soil properties. In: Organization, F.a.A.
663   (Ed.). Land and 410 Water Digital Media Series 1.
664 Fekete, B., Vorosmarty, C., 2011. ISLSCP II UNH/GRDC Composite Monthly Runoff. ISLSCP
665   Initiative II Collection, edited by: Hall, FG, Collatz, G., Meeson, B., Los, S., Brown de
666   Colstoun, E., and Landis, D., Data set, available at: http://daac.ornl.gov/, from Oak Ridge
667   National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, doi,
668   10.
669 Fekete, B.M., Vörösmarty, C.J., Grabs, W., 2002. High-resolution fields of global runoff
670   combining observed river discharge and simulated water balances. Global Biogeochem.
671   Cycles, 16(3).
672 Fernandez, W., Vogel, R., Sankarasubramanian, A., 2000. Regional calibration of a watershed
673   model. Hydrol. Sci. J., 45(5): 689-707.
674 Gerten, D. et al., 2005. Contemporary "green" water flows: Simulations with a dynamic global
675   vegetation and water balance model. Physics and Chemistry of the Earth, Parts A/B/C,
676   30(6): 334-338.
677 Ghavidelfar, S., Alvankar, S.R., Razmkhah, A., 2011. Comparison of the lumped and quasi-
678   distributed Clark runoff models in simulating flood hydrographs on a semi-arid
679   watershed. Water Resour. Manage., 25(6): 1775-1790.
680 GRDC, BfG The GRDC - Global Runoff Database. Available at:
681   http://www.bafg.de/GRDC/EN/01_GRDC/13_dtbse/database_node.html. Accessed
682   09/13/2017
683 Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared
684   error and NSE performance criteria: Implications for improving hydrological modelling.
685   J. Hydrol., 377(1): 80-91.
686 Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic
687   models: Multiple and noncommensurable mea- sures of information, Water Resour. Res.,
688   34, 751–763.
689 Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M.,
690   Ludwig, F., Masaki, Y., Schewe, J. and Stacke, T., 2014. Global water resources affected
691   by human interventions and climate change. Proc. Natl. Acad. Sci., 111(9), 3251-3256.
692 Hanasaki, N. et al., 2008. An integrated model for the assessment of global water resources–Part
693   2: Applications and assessments. Hydrology and Earth System Sciences, 12(4): 1027-
694   1037.
695 Hansen, M., DeFries, R., Townshend, J.R., Sohlberg, R., 2000. Global land cover classification
696   at 1 km spatial resolution using a classification tree approach. Int. J. Remote Sens., 21(6-
697   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.

Hsu, K., H. V. Gupta, and S. Sorooshian (1995), Artificial neural network modeling of the rainfall-runoff process, Water Resour. Res., 31(10), 2517 – 2530.

Jothityangkoon, C., M. Sivapalan, and D. Farmer (2001), Process controls of water balance variability in a large semi-arid catchment: Downward approach to hydrological model development, J. Hydrol., 254, 174 – 198.

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.

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.

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.

744 Nijssen, B., Lettenmaier, D.P., Liang, X., Wetzel, S.W., Wood, E.F., 1997. Streamflow
745   simulation for continental-scale river basins. Water Resour. Res., 33(4): 711-724.
746 Nijssen, B., O'donnell, G.M., Hamlet, A.F., Lettenmaier, D.P., 2001a. Hydrologic sensitivity of
747   global rivers to climate change. Clim. Change, 50(1-2): 143-175.
748 Nijssen, B., O'Donnell, G.M., Lettenmaier, D.P., Lohmann, D., Wood, E.F., 2001b. Predicting
749   the discharge of global rivers. J. Clim., 14(15): 3307-3323.
750 O'Neill, B.C. et al., 2014. A new scenario framework for climate change research: the concept of
751   shared socioeconomic pathways. Clim. Change, 122(3): 387-400.
752 Obled, C., Wendling, J., Beven, K., 1994. The sensitivity of hydrological models to spatial
753   rainfall patterns: an evaluation using observed data. J. Hydrol., 159(1-4): 305-333.
754 Oubeidillah, A.A., Kao, S.-C., Ashfaq, M., Naz, B.S., Tootle, G., 2014. A large-scale, high-
755   resolution hydrological model parameter data set for climate change impact assessment
756   for the conterminous US. Hydrology and Earth System Sciences, 18(1): 67-84.
757 Paudel, M., Nelson, E.J., Downer, C.W., Hotchkiss, R., 2011. Comparing the capability of
758   distributed and lumped hydrologic models for analyzing the effects of land use change.
759   Journal of Hydroinformatics, 13(3): 461-473.
760 Reed, S. et al., 2004. Overall distributed model intercomparison project results. J. Hydrol.,
761   298(1): 27-60.
762 Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types
763   of hydrological models. Water Resour. Res., 32(7): 2189-2202.
764 Rost, S., Gerten, D., Heyder, U., 2008. Human alterations of the terrestrial water cycle through
765   land management. Advances in Geosciences, 18: 43-50.
766 Sankarasubramanian, A., Vogel, R.M., 2002. Annual hydroclimatology of the United States.
767   Water Resour. Res., 38(6).
768 Sankarasubramanian, A., Vogel, R.M., 2003. Hydroclimatology of the continental United States.
769   Geophys. Res. Lett., 30(7).
770 Scott, M.J. et al., 2016. Sensitivity of future US Water shortages to socioeconomic and climate
771   drivers: a case study in Georgia using an integrated human-earth system modeling
772   framework. Clim. Change, 136(2): 233-246.
773 Sivapalan, M., G. Blöschl, L. Zhang, and R. Vertessy (2003), Downward approach to
774   hydrological prediction, Hydrol. Processes, 17, 2101–2111.
775 Tang, Q. et al., 2010. Dynamics of terrestrial water storage change from satellite and surface
776   observations and modeling. J. Hydrometeorol., 11(1): 156-170.
777 Tebaldi, C., Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate
778   projections. Philosophical Transactions of the Royal Society of London A: Mathematical,
779   Physical and Engineering Sciences, 365(1857): 2053-2075.
780 Thomas, H., 1981. Improved methods for national water assessment. Report WR15249270, US
781   Water Resource Council, Washington, DC.
782 Van Beek, L., Bierkens, M.F., 2009. The global hydrological model PCR-GLOBWB:
783   conceptualization, parameterization and verification. Utrecht University, Utrecht, The
784   Netherlands.
785 Vandewiele, G., Xu, C.-Y., 1992. Methodology and comparative study of monthly water balance
786   models in Belgium, China and Burma. J. Hydrol., 134(1-4): 315-347.
787 Vogel, R.M., Sankarasubramanian, A., 2003. Validation of a watershed model without
788   calibration. Water Resour. Res., 39(10).

789 Vörösmarty, C.J., Federer, C.A., Schloss, A.L., 1998. Potential evaporation functions compared
790   on US watersheds: Possible implications for global-scale water balance and terrestrial
791   ecosystem modeling. J. Hydrol., 207(3-4): 147-169.
792 Vörösmarty, C.J., Green, P., Salisbury, J., Lammers, R.B., 2000. Global water resources:
793   vulnerability from climate change and population growth. science, 289(5477): 284-288.
794 Wang, D., Tang, Y., 2014. A one-parameter Budyko model for water balance captures emergent
795   behavior in darwinian hydrologic models. Geophys. Res. Lett., 41(13): 4569-4577.
796 Weedon, G. et al., 2011. Creation of the WATCH forcing data and its use to assess global and
797   regional reference crop evaporation over land during the twentieth century. J.
798   Hydrometeorol., 12(5): 823-848.
799 Wen, L., Nagabhatla, N., Lü, S., Wang, S.-Y., 2013. Impact of rain snow threshold temperature
800   on snow depth simulation in land surface and regional atmospheric models. Adv. Atmos.
801   Sci., 30(5): 1449-1460.
802 Widén-Nilsson, E., Halldin, S., Xu, C.-y., 2007. Global water-balance modelling with
803   WASMOD-M: Parameter estimation and regionalisation. J. Hydrol., 340(1): 105-118.
804 Wilkinson, K., von Zabern, M., Scherzer, J., 2014. Global Freshwater Fluxes into the World
805   Oceans, Tech. Report prepared for the GRDC. Koblenz, Federal Institute of Hydrology
806   (BfG),(GRDC Report No. 44. doi: 10.5675/GRDC_Report_44, 23pp.[Available from h
807   ttp://www. bafg. de/GRDC/EN/02_srvcs/24_rprtsrs/report_44. pdf].
808 Young, P. C., and S. Parkinson (2002), Simplicity out of complexity, in Environmental Foresight
809   and Models: A Manifesto, edited by M. B. Beck, Elsevier Science, The Netherlands,
810   251–294.
811 YAO, H., HASHINO, M., TERAKAWA, A., SUZUKI, T., 1998. Comparison of distributed and
812   lumped hydrological models. PROCEEDINGS OF HYDRAULIC ENGINEERING, 42:
813   163-168.
814 Zhang, X.-J., Tang, Q., Pan, M., Tang, Y., 2014. A long-term land surface hydrologic fluxes and
815   states dataset for China. J. Hydrometeorol., 15(5): 2067-2084.

816 **Figure Caption**

817 **Figure 1** Schematic diagram of the "*abcd*" model, with enhancements of snow and partition of

818 total runoff between direct runoff and baseflow.

819 **Figure 2** Kling-Gupta efficiency of the simulated basin-level total runoff across the global 235

820 basins (lump = lumped, dist = distributed, cal = calibration, the x-axis labels of "lump_cal" or

821 "dist_cal" represent lumped/distributed scheme during calibration period).

822 **Figure 3** Comparison of basin-specific long-term annual total runoff, direct runoff and baseflow

823 estimates from both the lumped and distributed "*abcd*" model schemes against VIC products,

824 across global 235 basins and for the calibration period of 1971-1990 and validation period of

825 1991-2010. The labels are denoted as combination of model scheme and period, where lump and

826 dist stand for lumped and distributed model scheme, cal and val represent the calibration and

827 validation period, respectively. These denotations remain the same for all figures in this work.

828 Note that the basin-level VIC baseflow is derived by multiplying the gridded VIC long-term

829 annual total runoff and the mean of the four gridded baseflow index products from Beck et al.

830 (2014), and then aggregating from grid-level to basin-level. The basin-level VIC direct runoff is

831 then calculated by subtracting baseflow from the total runoff.

832 **Figure 4** Time series of basin-specific total runoff (Q_{total}) from the VIC product, the lumped and

833 distributed "*abcd*" schemes for the world's sixteen river basins with top annual flow (Dai et al.

834 2009) during 1981-1990 (part of the calibration period 1971-1990). KGE_l and KGE_d stand for

835 KGE value for the lumped and distributed scheme, respectively.

836 **Figure 5** Spatial patterns of long-term annual total runoff (mm yr^{-1}) during 1971-1990 across

837 global 235 basins: a) VIC runoff product; b) total runoff estimates from the lumped "*abcd*"

838   scheme (lump = lumped); and c) total runoff estimates from the distributed "*abcd*" scheme (dist

839   = distributed).

840   **Figure 6** The spatial pattern of Kling-Gupta efficiency (KGE) for the total runoff estimates of

841   the global 235 basins for the calibration period of 1971-1990: a) the lumped "*abcd*" scheme; and

842   b) the distributed "*abcd*" scheme.

843   **Figure 7** Parameter-induced uncertainty in total runoff for the world's sixteen river basins with

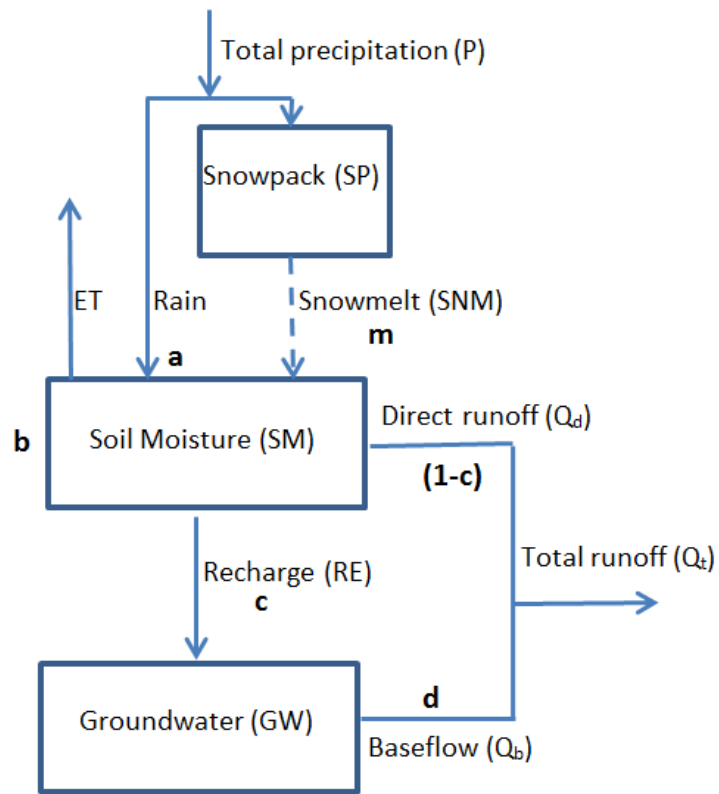
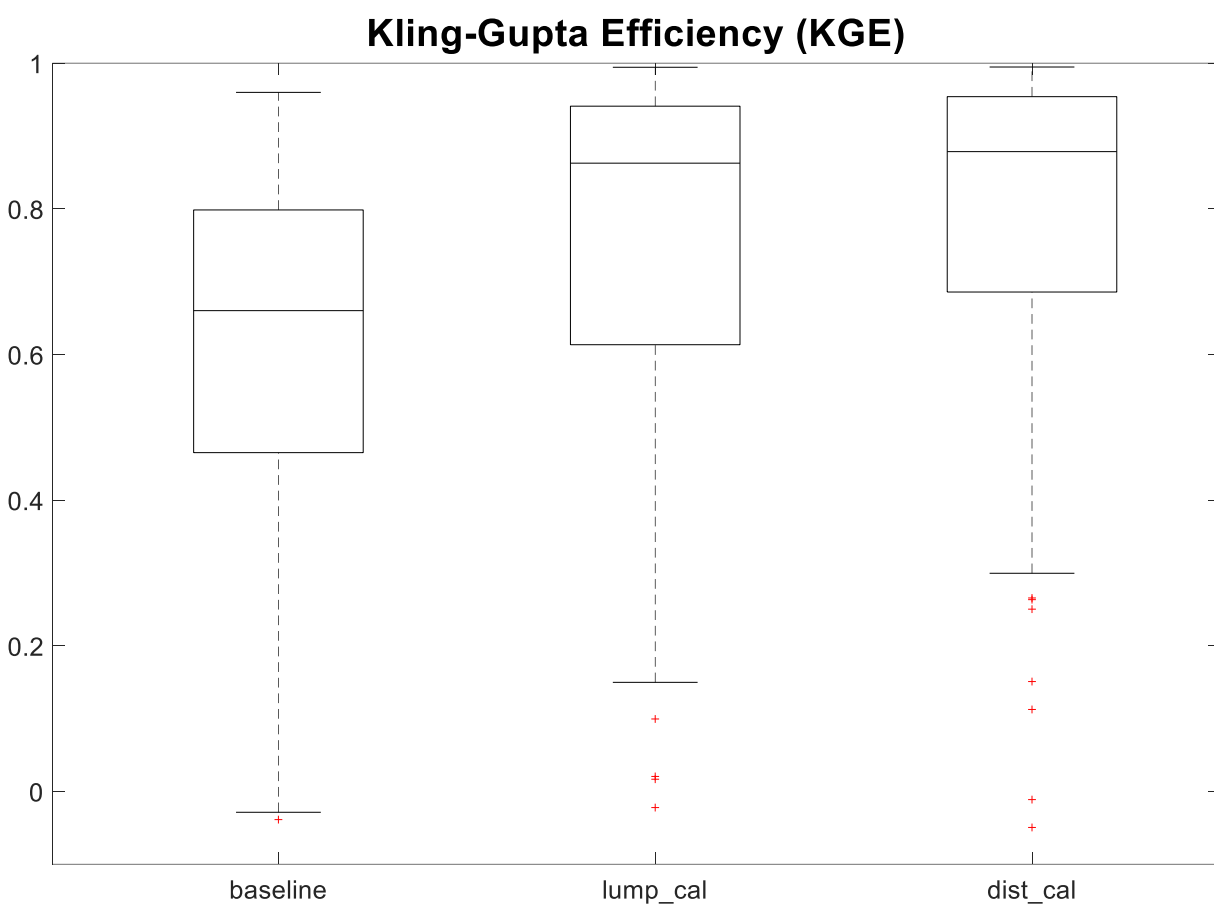
844   top annual flow. The green line stands for simulated total runoff using the calibrated parameters,

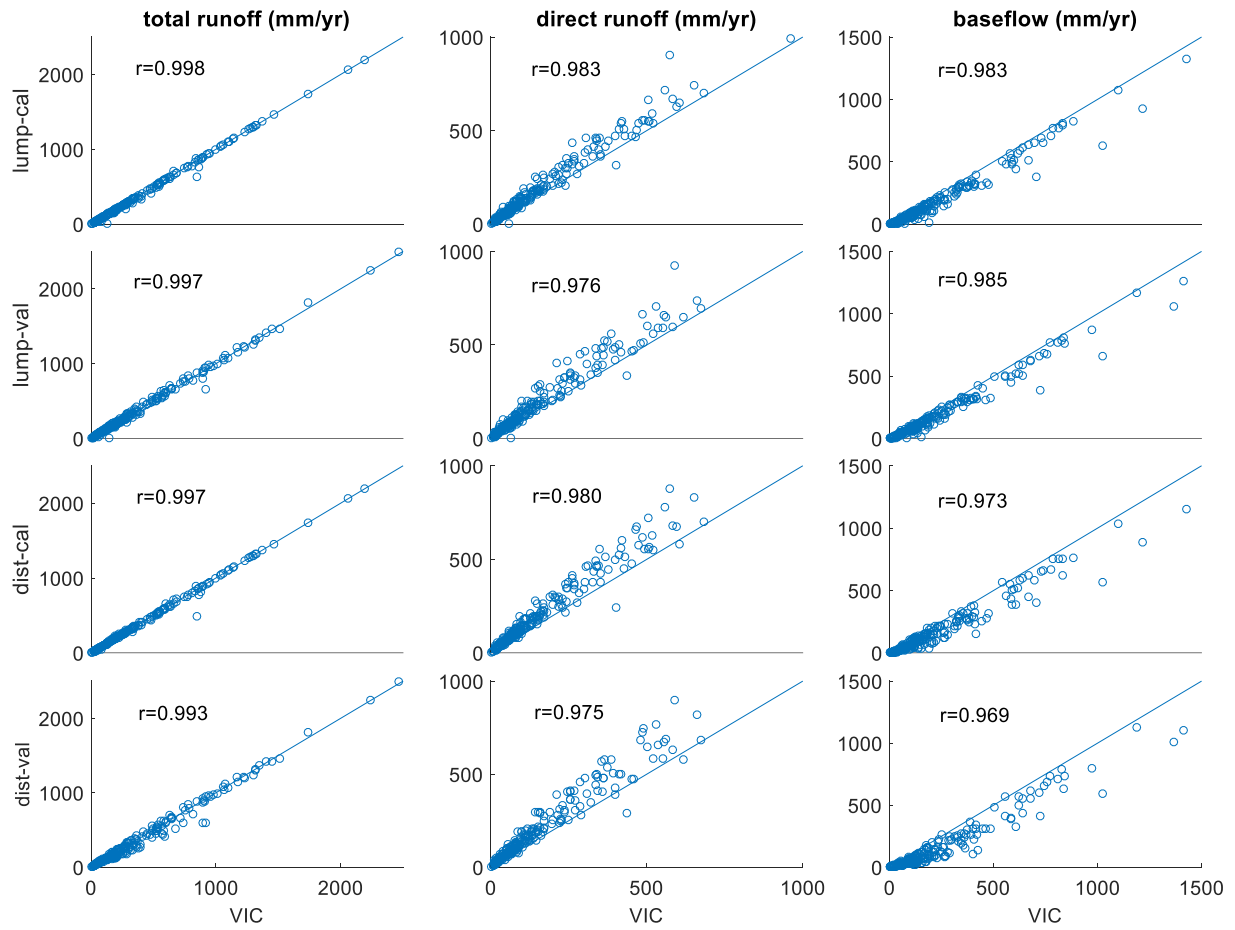
845   and the gray area represents the spread derived from variations in parameters.

846   Figure 1

847

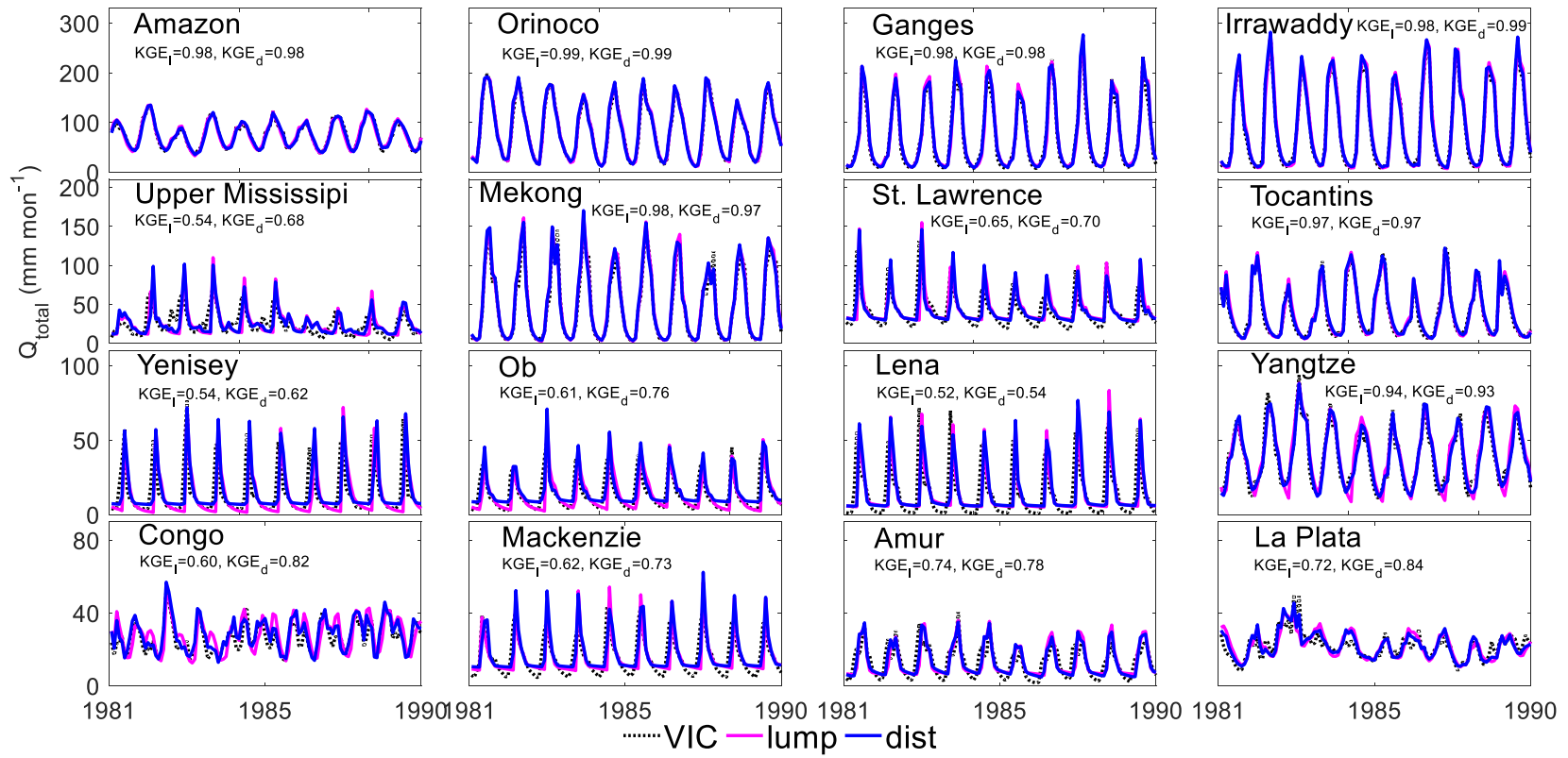
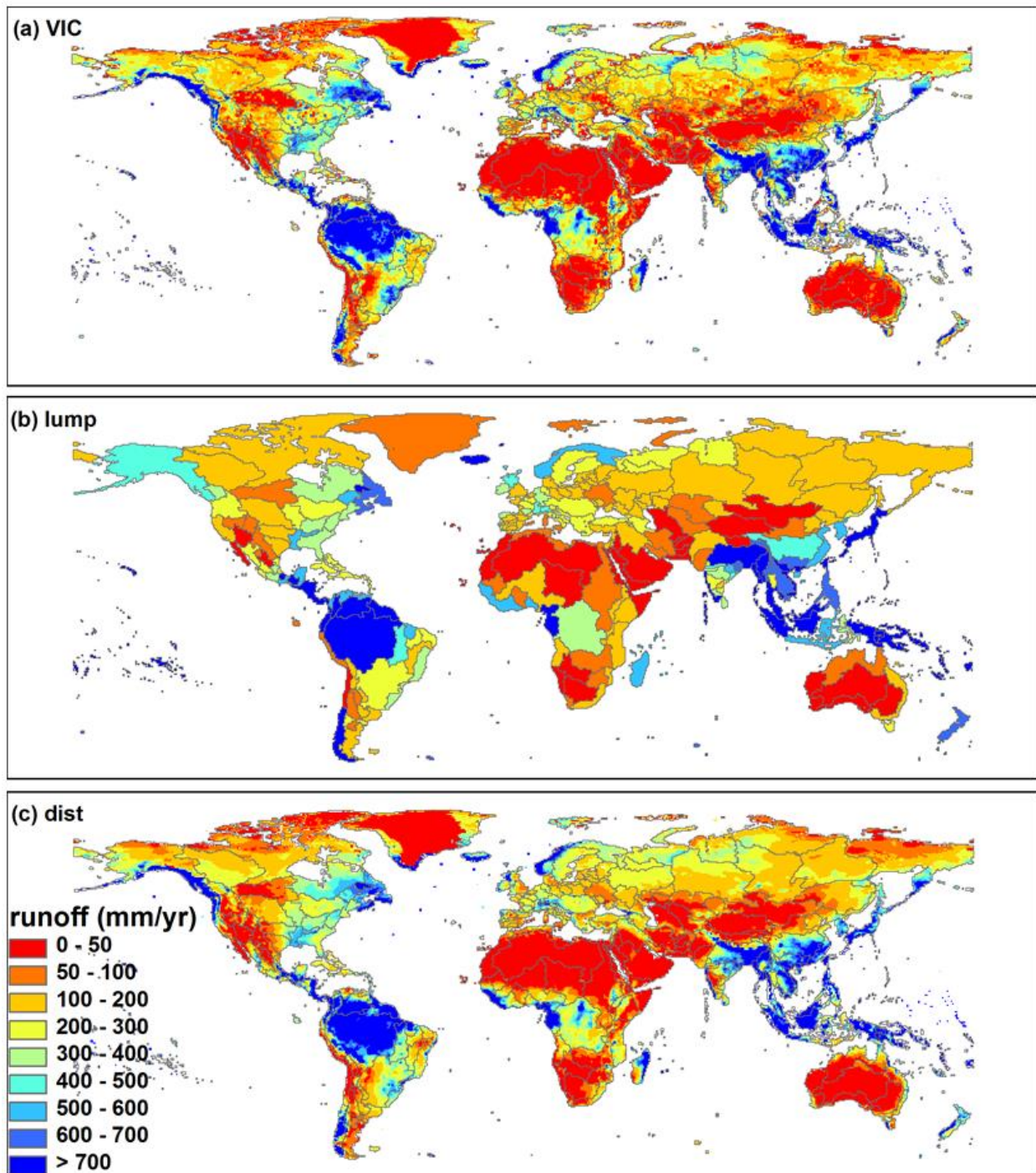
848

849   Figure 2

850

851   Figure 3

852

853

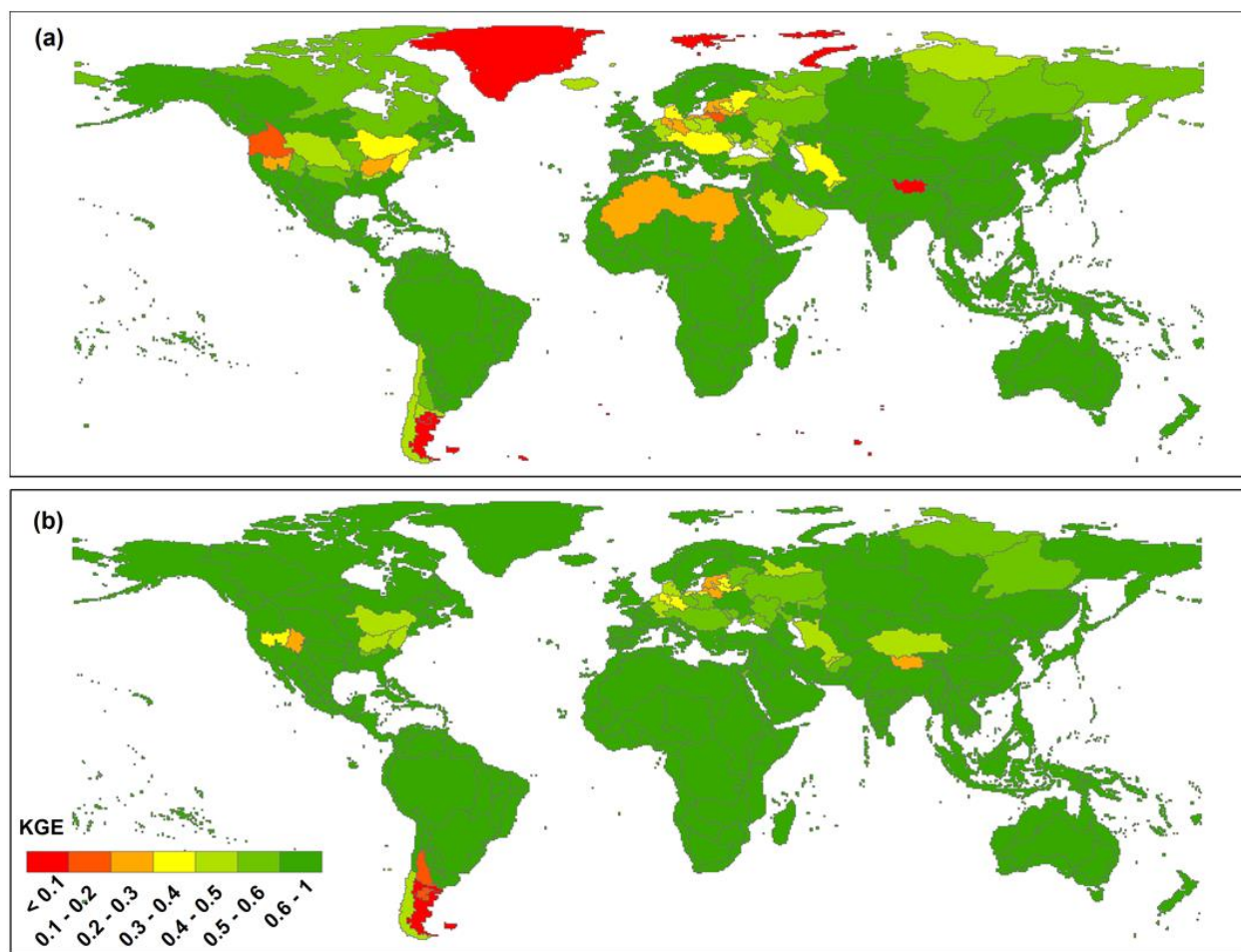
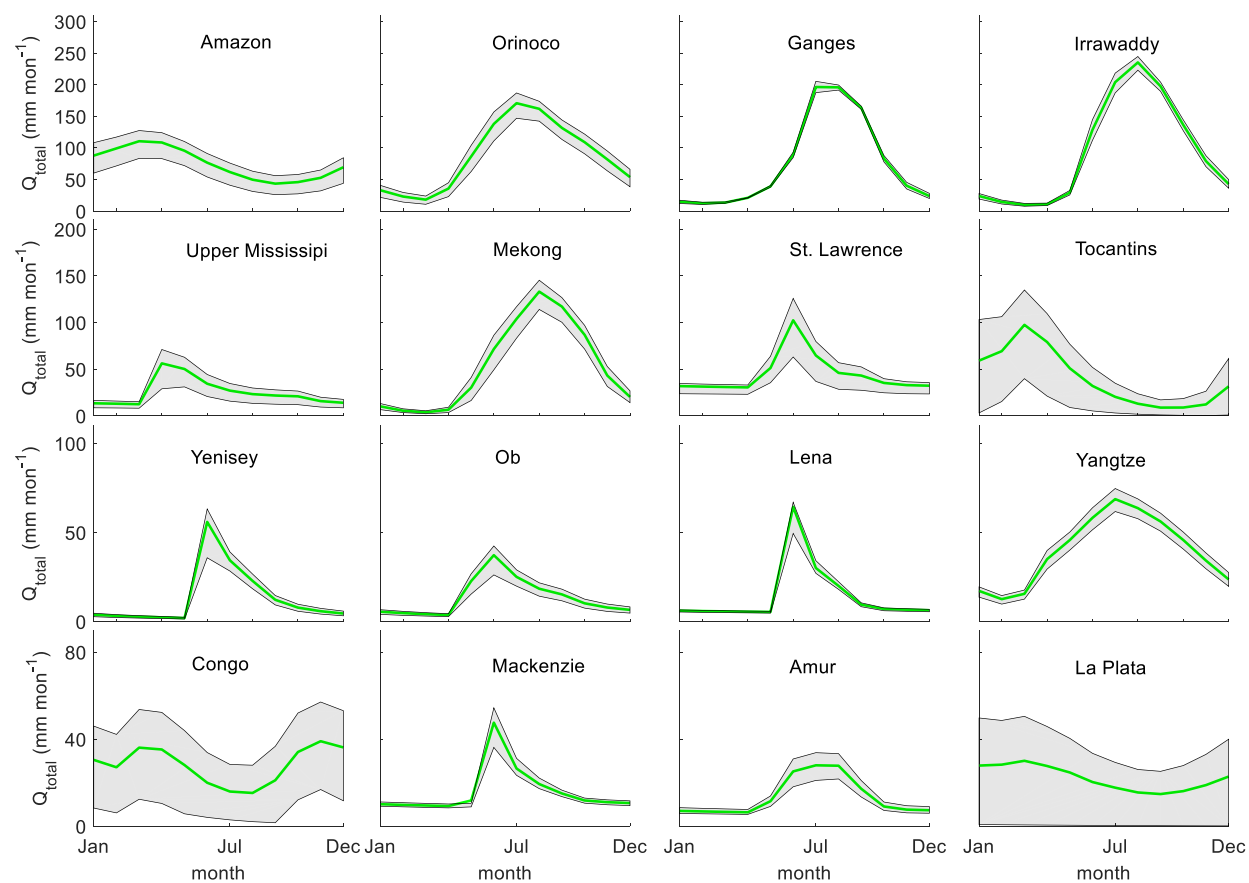
854   Figure 4

855

856   Figure 5

857

858

859   Figure 6

861   Figure 7

862

863   Table 1 Parameters description and ranges for the "*abcd*" model (the parameters *a,c,d* and *m* are

864   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;
b	Upper limit on the sum of evapotranspiration and soil moisture storage	0-4000	Sankarasubramanian and Vogel, 2002;
c	Degree of recharge to groundwater	0-1	Vandewiele and Xu,
d	Release rate of groundwater to baseflow	0-1	1992)
m	Snow melt coefficient	0-1	(Wen et al., 2013)

865