Inference of parameters for a global hydrological model by applying Approximate Bayesian Computation: Identifiability of climate-based parameters

T. Yoshida[1], N. Hanasaki[2], K. Nishina[2], J. Boulange[2], M. Okada[2], and P. A. Troch[3]

[1]. Institute for Rural Engineering, National Agriculture and Food Research Organization, Tsukuba, Japan.

[2]. National Institute for Environmental Studies, Tsukuba, Japan.

[3]. University of Arizona, Tucson, AZ.

Corresponding author: first and last name (takeoys@affrc.go.jp)

†Additional author notes should be indicated with symbols (current addresses, for example).

Key Points:

- We tested identifiability of parameters of a global hydrological model based on climate properties using Approximate Bayesian Computation.
- NSE scores with the identified parameters for 11 Köppen climate classes outperformed than those with the default and global parameters.
- The identified parameters showed consistency with the physical interpretation of soil formation and efficiencies in vapor transfer.

Abstract

The calibration of global hydrological models has been attempted for over two decades, but an effective and generic calibration method has not been proposed. In this study, we investigated the application of Approximate Bayesian Computation (ABC) to calibrate the H08 global hydrological model by running global simulations with 5000 randomly generated sets of four sensitive parameters. This yielded satisfactory results for 777 gauged watersheds, indicating that ABC can be used to optimize H08 parameters to calibrate individual watersheds. We tested the identifiability of the parameters to yield satisfactory representations of hydrological functions based on Köppen's climate classification ("climate-based" calibrations hereafter) We aggregated 5000 simulation results per catchment based on the 11 Köppen climate classes, then selected the parameters that exceeded the Nash–Sutcliffe efficiency (NSE) scores predefined by the acceptance ratio for each climate class. Our results indicate that the number of stations showing satisfactory (NSE > 0.0) and good (NSE>0.5) performances were 480 and 234 (61.7% and 30.1% of total stations, respectively), demonstrating the effectiveness of climate-based calibration. We also showed that the climate-based parameters outperformed the default and global parameters in terms of representativeness (global-scale differences of hydrological properties among climate classes) and robustness (consistency in yielding satisfactory results for watersheds in the same climate class). The identified parameters for 11 Köppen climate classes showed consistency with the physical interpretation of soil formation and efficiencies in vapor transfer with a wide variety of vegetation types, corroborating the strong influence of climate on hydrological properties.

Plain Language Summary

This is optional but will help expand the reach of your paper. Information on writing a good plain language summary is available here.

1 Introduction

Global hydrological models are essential tools to analyze Earth's hydrological cycle and water resources (Bierkens, 2015; Pokhrel et al., 2016). Over the past two decades, there have been numerous efforts to develop and use such models (Döll et al., 2003; Döll et al., 2014; Gerten et al., 2004; Hanasaki et al., 2008a, 2018; Rost et al., 2008; Sutanudjaja et al., 2018; Wada et al., 2014). Their applications include assessing the impact of climate change on water resources (Haddeland et al., 2014; Schewe et al., 2014), environmental footprint analyses (Dalin et al., 2012; Gleeson et al., 2012), and historical drought analyses (Schewe et al., 2019).

Further work is needed to improve the overall skill scores of estimations of basic hydrological variables, particularly streamflow (Oki et al., 1999). Comparative studies of models have shown that streamflow simulations sometimes deviate considerably from observation records (Gudmundsson et al., 2012; Haddeland et al., 2011; Zaherpour et al., 2018). Most global hydrological models adopt empirical a priori model parameters, limiting the effectiveness of simulations. There are two main obstacles to calibrate global model parameters accurately: 1. The difficult and computationally expensive calibration of parameters at numerous worldwide stations; 2. Inference of parameter values for watersheds having no observation records (hereafter, ungauged watersheds). The spatio-temporal distribution of streamflow observations is uneven,

with data unavailable for ~50% of the global land surface over substantial periods (Döll et al., 2003; Fekete & Vörösmarty, 2007).

Various studies have tackled these challenges. Nijssen et al. (2001a, b) developed the global Variable Infiltration Capacity (VIC) hydrological model by manually calibrating six hydrological parameters for nine river watersheds, each in different climatic zones. These parameters were then used in simulations for 17 other watersheds, ensuring that climate zones of the original calibration and subsequent simulation were the same. They found no reduction in bias and root-mean-square error for individual watersheds, although the transfer of climate-specific calibrated parameters between watersheds improved overall simulation performance. Döll et al. (2003) developed the Water-Global Assessment and Prognosis (WaterGAP) global hydrological model, and manually calibrated one hydrological parameter for 724 gauged watersheds. For ungauged watersheds, it was estimated by multiple linear regression using air temperature, area of open freshwater, and the length of non-perennial river stretches within each watershed as explanatory variables. Validation of streamflow simulations for nine watersheds in comparison to gauge data showed reasonable accuracy at all stations. Widén-Nilsson et al. (2007) developed a simple global water balance model (Water And Snow Modeling System; WASMOD-M) and generated 1680 parameter combinations. They identified the "best" parameter combination that maximized the skill score of streamflow simulation for gauged watersheds. For ungauged watersheds, they transferred the best combination of parameters from the nearest gauged watershed within 19.5° (latitude)/8.5° (longitude). The simulation employing transferred best parameters outperformed those with spatially uniform parameters. Beck et al. (2016) applied the Hydrologiska Byråns Vattenbalansavdelning (HBV) hydrologic model globally. They calibrated 14 parameters for 1787 catchments using an evolutionary algorithm and selected 674 of these whose simulation performance exceeded a particular threshold (donor catchments). For each ungauged watershed, they ran simulations using parameters from 10 donor catchments most similar to the watershed (devised by Beck et al., 2015). The ensemble means of these 10 simulations outperformed those with spatially uniform parameters for 79% of the watersheds. The influences of climate properties on hydrological parameters were tested using the Budyko framework (Greve et al., 2020). They calibrated the additional parameter to account for the residuals from the Budyko equation based on the empirical relationship obtained in the contiguous US and showed that the long-term water and energy balance could be improved without any additional data.

These studies imply that parameter calibrations for gauged watersheds are effective if the models are reasonably simple with a limited number of parameters. This implication is a constraint for state-of-the-art models because their formulations and structures are becoming increasingly complex. The integrity of the parameter transfer technique to simulate ungauged watersheds is not yet established. Several improvements are reported using parameter transfer in ungauged watersheds (e.g., Nijssen et al., 2001b; Widén-Nilsson et al., 2007) with a limited number of validation stations (e.g., Döll et al., 2003), or using the necessity of ensemble technique (e.g., Beck et al., 2016).

Approximate Bayesian Computation (ABC) is a promising new technique in the field of biology for inferring complex models (Beaumont et al., 2002; Sisson et al., 2018). Avoiding explicit evaluation of the likelihood function, it uses a set of summary statistics to extract information from observations to approximate target distributions. Hydrological modeling dealing with complex water flow processes through watersheds can benefit from the strengths of ABC (Sadegh & Vrugt, 2014; Sadegh et al., 2015). In ABC, a candidate parameter set (proposal) is first

sampled from some prior distribution, which is then used to simulate the output of the model. Instead of a likelihood evaluation in Bayesian approaches, a distance function is used to determine the acceptance of the proposal. Accepted samples are then used to summarize target posterior distributions. Applying ABC directly to an individual watershed would yield an optimized parameter set for that watershed (Sadegh & Vrugt, 2014; Vrugt & Sadegh, 2013). However, those parameters are not guaranteed to behave like other watersheds because of the overfitting of model parameters to epistemic errors associated with a model's structural and climate forcing errors (Beven and Freer, 2001).

In this study, we applied ABC to calibrate hydrological parameters of the H08 global hydrological model (Hanasaki et al., 2018) and to identify representative parameter sets based on climate properties. Our study aims to identify parameter sets that effectively reproduce a satisfactory performance for groups composed of watersheds, rather than "optimal" parameters for a single watershed. Hydrological similarities are divided into similarities in climate and watershed properties (Wagener et al., 2012; Troch et al., 2017). Based on previous studies, we hypothesize that on a global scale, similarity in climate properties is a dominant control on hydrological properties (Beck et al., 2016; Nijissen et al., 2001ab). We address two key research questions in this study: 1. Do climate properties exert dominant controls on hydrological properties on a global scale? 2. How can we identify representative parameters for watersheds under specific climate systems using the ABC technique?

2 Materials and Methods

2.1 H08 global hydrological model

This model comprises six sub-models: land-surface hydrology, river routing, crop growth, reservoir operation, water abstraction, and environmental flow. Here, we used the land-surface hydrology and river routing sub-models. Hanasaki et al. (2008a, 2008b, 2010, 2018) provides descriptions of the sub-models.

Land surface hydrology is based on a single-layer bucket model (Manabe, 1969; Robock et al., 1995). It resolves the surface-energy and water-budget (including snow) at daily intervals and has a single soil moisture layer. Storage capacity (S_{max}) is expressed as

$$S_{\mathrm{max}} = SD \times (f_{\mathrm{FC}} - f_{\mathrm{WP}}), \quad (1)$$

where SD is soil depth (m), f_{FC} is soil moisture fraction at field capacity (unitless parameter), and f_{WP} is soil moisture fraction at wilting point. The default (global) settings (Hanasaki et al., 2018) are 1 m for SD, 0.30 for f_{FC}, and 0.15 for f_{WP} (Robock et al., 1995). Water balance of the soil moisture layer is expressed as

$$\frac{dS}{dt} = R + Q_{\mathrm{sm}} - E - Q_{\mathrm{s}} - Q_{\mathrm{sb}}, \quad (2)$$

where S is soil moisture (kg m^{-2}), R is rainfall (kg m^{-2} s^{-1}), Q_{sm} is snowmelt (kg m^{-2} s^{-1}), E is evapotranspiration (kg m^{-2} s^{-1}), Q_{s} is surface runoff (kg m^{-2} s^{-1}), and Q_{sb} is subsurface runoff (kg m^{-2} s^{-1}). Evapotranspiration (E) is expressed as

$$E = \beta \rho C_D U (q_{\mathrm{SAT}}(T_{\mathrm{s}}) - q), \quad (3)$$

where β is the evaporation coefficient (unitless), ρ is the density of air (kg m^{-3}), C_D is the bulk transfer coefficient (unitless), U is wind speed (m s^{-1}), $q_{SAT}(T_s)$ is saturation specific humidity at surface temperature T_s (kg kg^{-1}), and q is specific humidity of air (kg kg^{-1}). Surface runoff (Q_s) occurs when soil moisture exceeds storage capacity (S_{max}). Subsurface runoff (Q_{sb}) occurs under the condition

$$Q_{sb} = \frac{S_{max}}{\tau * 86400} \times \left(\frac{S}{S_{max}}\right)^{\gamma}, \quad (4)$$

where τ is a time constant (days) and γ is a shape parameter (unitless). Surface and subsurface runoff are divided into two components: direct runoff to rivers and groundwater recharge. The proportions of these two flows are determined by a function of indexes representing topographic relief, soil texture, geology, permafrost, and glacier (Döll & Fiedler, 2008). Recharged water is stored in the groundwater reservoir and is formulated using Eq (4).

The river routing model routes runoff through the global digital river network, with a spatial resolution of 0.5° (lat) × 0.5° (long)（Döll & Lehner, 2002）, at a constant flow velocity of 0.5 m s^{-1}.

Default values of τ and γ were determined empirically for four distinct climate zones: tropical, monsoon and dry, temperate, and polar (Hanasaki et al. 2008a). Previous studies which applied the H08 model to specific basins suggested that the calibration of four sensitive parameters *SD*, C_D, γ, and τ, improved the representation of the observed long-term variations of streamflow (Hanasaki et al., 2014; Masood et al., 2015; Mateo et al., 2014; Yoo, 2016).

2.2 Global meteorological data

We used WATCH Forcing Data (WFD; Weedon et al., (2011), which provides global land coverage (excluding Antarctica) at a spatial resolution of 0.5° (lat) × 0.5° (long) at daily intervals for the period 1901–2001. WFD is derived from a global-grid of monthly ground observations (CRU TS2.1; New et al., 2000) and six-hourly global reanalysis data (ERA-40; Uppala et al., 2005) including seven variables: air temperature, specific humidity, wind speed, surface air pressure, downward shortwave radiation, downward longwave radiation, and precipitation. Using mean monthly temperature and precipitation from WFD, we constructed a global map using 11 Köppen climate zones (Table 2; Figure S1; Köppen, 2011).

2.3 Hydrological data and simulation

Using monthly (3045 stations) streamflow data collected by the Global Runoff Data Center (GRDC; https://www.bafg.de/GRDC), we identified records suitable for calibration and validation by applying two thresholds with: 1) Catchment areas >10,000 km^2; and 2) Continuous records for the period 1961–1970. We set the first limit because the H08 model was configured at a spatial resolution of 0.5° (lat) × 0.5° (long) (~55 km × 55 km at the equator, equivalent to ~3000 km^2). The second limit was selected the period 1961–1970 provided the most extensive global coverage of streamflow data. In total, 777 stations in 500 basins met these criteria.

All stations were geo-referenced to the global digital river-network of the H08 model so that the errors of the modeled catchment areas with respect to the observed catchment areas were

<20%. The most common climate classification within the catchment was assigned as the representative climate zone for each catchment (Table 2).

H08 simulations were conducted at daily intervals and the land surface and river sub-models were set up following the boundary conditions and model parameters described by Hanasaki et al. (2018).

2.4 Calibration of H08 parameters by ABC

2.4.1 Implementation of ABC in H08 framework

We selected four hydrological parameters (**θ**) in the H08 hydrological model (i.e., *SD*, C_D, γ, and τ) as inference parameters in ABC, all of which have physical meaning. However, identifying their "true" value for each grid cell is challenging due to the heterogeneity within the grid cells and the simplification of physical processes inherent to the model.

We applied a simple rejection algorithm in ABC to infer the parameters. The priors "q (θ)" of **θ** are summarized in Table 1. The protocols of ABC were:

1. Generate N samples of **θ′**, according to q (θ).
2. Simulate runoff y′ using sampled **θ′** by H08 model and extract the monthly streamflow simulation time-series y at 777 stations for the period 1961–1970.
3. Calculate the Nash–Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970) using y and y′ for each simulation and obtain N number of NSEs in each watershed.
4. For the calibration of a group of watersheds, aggregate NSE scores from target M watersheds and compile N*M number of NSEs in each group.
5. In each group or watershed, evaluate the X_{th} percentile of NSE as tolerance ε_i.
6. Then, if NSE > ε, store $\boldsymbol{\theta}'_i$ as a posterior. We conducted this procedure for each group or watershed.

We set the number of simulations "N" to 5000 with "i" indicating the watershed identifier. This number was determined because of the restrictions of computational resources. Instead of the fixed tolerance, ϵ, which is commonly used in ABC procedures, we selected parameter sets yielding NSE scores exceeding a specific quantile value (X_{th} percentile) of the NSE distribution for each watershed. This was done because the NSE scores generated by the 5000 simulation runs showed completely different ranges for each watershed, presumably reflecting the inaccuracy of the climate forcing and oversimplification of the hydrological systems in the model. Thus, the fixed tolerance would lead to different acceptance ratios (ratio of accepted samples to the total simulation runs) for all the climate classes, making it difficult to interpret the results if the posteriors were too few and their distributions were discrete.

224 2.4.2 Representative parameters for the watershed groups

225 When one considers the representative parameters for the groups of watersheds, which is
226 the scope of this study, there is presumably a trade-off between the acceptance ratio used and the
227 representativeness of the parameter sets. If a low acceptance ratio is used, the number of
228 watersheds included in the posterior may be too few to capture the hydrological behavior for the
229 entire group. Conversely, the higher the acceptance ratio, the lower the range of NSE values; thus,
230 the parameters may converge to the mean of the priors and may not reflect the hydrological
231 properties of each watershed. The statistical measure used (e.g., mode, median, or mean) for
232 parameter identification may also affect the representative parameters' veracity. It seems
233 reasonable to use frequently occurring values of parameters (the mode) in the posterior
234 distribution, but the mode may reflect local optima not representative of the entire climate class.

235 Thus, it is essential to use a suitable acceptance ratio for sampling the posterior distribution,
236 and a suitable statistical measure (mode, median, and mean) to derive representative parameters
237 from the posterior distribution. We considered acceptance ratios of 0.1, 1, 5, 10, and 20% and an
238 appropriate statistical measure to provide the best determination of representative parameters for
239 an entire group of watersheds. In total, we tested 15 options to select representative parameter sets;
240 a combination of multiple acceptance ratio of the samples in procedure 5 (5 options) and statistical
241 measures, which were mode, median, and mean (3 options).

242 We implemented ABC to calibrate H08 model parameters for individual watersheds
243 (hereafter, "individual" calibration). Then, we tested two categorizations of groups of watersheds.
244 First, we implemented ABC to calibrate H08 model parameters based on the Köppen climate
245 classes (hereafter referred to as "climate-based" calibration). This categorization assumes that
246 climate properties dominate global-scale differences in hydrological behaviors (Nijssen et al.,
247 2001a, b; Beck et al., 2016). We aggregated 5000 simulation runs (samples) of all watersheds in
248 each climate class as prior distributions and applied the ABC technique to derive posterior
249 distribution. We attempted 15 combinations of the acceptance ratio and statistical measures to
250 determine the representative parameter sets satisfying the two criteria mentioned in the next
251 section. We also tested the transferability of the representative parameter sets for each of the
252 climate classes to the watersheds in the same climate class. Second, to compare the effectiveness
253 of the climate-based calibration, we determined the parameter sets at a global scale by applying
254 the same procedures as the climate-based calibration to the aggregated samples of all the gauged
255 777 watersheds (hereafter, "global" calibration).

256 2.4.3 Evaluation of procedures

257 The inferred parameters required the fulfillment of two criteria: 1) Consistent yield of
258 satisfactory results for ungauged watersheds (robustness); and 2) Reflect differences of climate
259 and catchment properties at a global scale (representativeness). To examine the robustness and
260 representativeness, we divided the samples of each climate class into calibration and validation
261 datasets by randomly selecting half of the watersheds. We then checked whether the representative
262 parameters derived from the calibration dataset were consistently representative of the validation
263 dataset (Repeated two-fold cross-vlidation). We repeated this process 100 times to check the
264 robustness and representativeness of the calibration and transfer processes by comparing the range
265 of NSE scores. We conducted the repeated two-fold cross-vlidation for each of the Köppen climate
266 classes in which the number of watersheds exceeded 50 (i.e., classes Aw: tropical monsoon, Ca:

hot summer temperate, Cb: warm summer temperate, Db: warm summer continental, and Dc: subarctic, Table 2), which included 91% of the observed stations.

We evaluated the effectiveness of the process from the following points of view: 1) The improvement of the NSE values obtained compared to those of the default parameters, and 2) The number of stations for which satisfactory and good NSE values were obtained. NSE thresholds used were initially based on the recommendations of Moriasi et al. (2007, 2015) but also that of Krysanova et al. (2018) thereafter. Here, we adjusted the thresholds (made them less strict) for the global models, e.g., NSE thresholds used by Moriasi et al. (2015) for a satisfactory and good performance of monthly runoff simulations were $0.55 < \text{NSE} < 0.70$ and $0.70 \leq \text{NSE} \leq 0.85$, respectively. Here, we used $0.0 < \text{NSE} < 0.50$ for a satisfactory and $\text{NSE} \geq 0.50$ for a good performance, respectively.

3 Results

3.1 Calibration of individual watershed

Figure 1 showed the cross plots of the posterior distributions at one of the watersheds (ID: 4362600, at Boca Del Cerro station in Rio Usumacinta River). The acceptance ratio was 30% for Figure 1. The color of the plots signified the NSE values, and the crosses represented the values corresponding to the mode (yellow), mean (light blue), and median (white) of the posterior distribution. Among the six combinations of parameters, the *SD*-C_D plot (Figure 1 (a)) showed the most constrained posterior distributions, and parameter sets that yielded higher NSE values were consistently clustered in a specific region of the search domain (near the upper-right corner). The other plots did not show such explicit constraints, and the NSE scores did not appear to correlate with the selected parameter values. This was particularly evident for the cross plot of γ and τ (Figure 1 (f)), which show widely scattered plots in the search domain and the totally randomized NSE scores.

The modal values of the posterior distributions (yellow crosses) was plotted centrally in the accumulated plots with higher NSE scored (blue crosses). In contrast, the mean and median values (crosses with light blue and white, respectively) deviated from the plots with higher NSE scores, plotting near the center of the search domain. These posterior distribution patterns were typically observed for the other watersheds or when the varied acceptance ratios were used. These results suggested that using the modal values of the posterior distribution was suitable for identifying the optimal parameters for the individual watersheds. Here, we decided to use an acceptance ratio of 10% because the posterior distributions obtained with acceptance ratios lower than 10% were discrete.

The spatial distribution of the parameters identified for individual watersheds were shown in Figure 2. The NSE threshold range showed higher NSE scores for the temperate and continental climate classes (Figure S2), and relatively low scores for watersheds in the arid (BW), semi-arid (BS), and tropical rainforest (Af) climate classes. Two possible explanations exist:

1. The structure of the model is based on a simple bucket model wherein all precipitation reaching the land surface infiltrates to the subsurface, with subsurface drainage continuing until empty. This would differ for watersheds in arid and semi-arid climate zones, where excess infiltration plays a critical role in runoff generation (Goodrich et al., 1994; Nicolau et al. 1996).

2. The WFD dataset is an integrated compilation of reanalyzed atmospheric conditions and meteorological data observed at the land surface. Thus, the sparser observation networks in tropical-rainforests, arid, and semi-arid zones (Schneider et al., 2014) result in less reliable WFD forcing data than that obtained in temperate and continental climate zones.

3.2 Climate-based calibration

3.2.1 Optimal method to identify climate-based parameters

Figure 3 showed the two-dimensional posterior distributions for the combinations of parameters for the climate class, Db, obtained with acceptance ratios of 5%. The crosses indicated the locations of the mode (yellow), median (white), and mean (light blue), the same notation used in Figure 1. As suggested by the individual calibration, the *SD*-C_D plot showed the most constrained distribution compared to the other five distributions. Ideally, the posterior distributions should exhibit a clear peak and concentrated in a specific region of the search domain to ensure that the choice of statistical measures does not affect the identified representative parameter values. However, the posterior distribution for the climate-based calibration suggested that the representative parameters differed depending on the choice of statistical measures.

We calculated the NSE scores of the validation data sets for the climate classes Db and Dc based on the parameters selected from the calibration data sets with different acceptance ratios and statistical measures. Figure 4 showed the distribution of NSE values for the validation groups obtained from the 100 repetitions of the repeated two-fold cross-vlidation. We then considered the influence of the choice of statistical measure on the NSE scores. For both climate classes, there was a larger spread of data when the mode was used to determine the representative parameters than when the median or mean were used. This was particularly noticeable for the Dc climate class for which the parameter values obtained from the posterior mode varied substantially (Figure 5(b)). However, despite unimodal parameter distributions obtained for Db, this climate class yielded lower first quartiles for the mode than for the median and mean (Figure 5(a)). Compared to the mode, the mean and median values were stable and yielded narrower ranges of NSE scores, implying the robustness of the method. These results indicate the effectiveness of the posterior-mean or -median for identifying robust parameters for climate-based calibration. Note that different statistical measures were used for determining the representative parameters in the individual and climate-based calibrations: the mode for the individual calibration and the median or mean for the climate-based calibration.

The variations in the tolerance, ϵ, which depend on the acceptance ratios, are summarized for each climate class (Table 3). The samples were selected as posterior distributions if the NSE scores exceeded the tolerance, ϵ. For acceptance ratios not exceeding 10%, tolerance exceeded 0.0 for all the climate classes, suggesting that the selected samples can be used to provide donors for at least satisfying the criteria of 'good' performance. The only exception was the climate class BW (ϵ =-0.873 for 5% of the acceptance ratio), and care should be taken when interpreting the identified parameters for BW. We also found that a narrow acceptance ratio (e.g., 0.1%) did not necessarily provide an improvement, probably owing to the overfitting of only a few watersheds, which are not representative of an entire climate class.

Based on our analyses, we provisionally postulate that the posterior-median from climate classes with acceptance ratios of 5% is the optimal method to identify climate-based parameters

based on the number of stations yielding good or satisfactory simulations (Table 4). The only exception to this is the inferred parameter sets in the climate class Dc, which performed inadequately compared with the other climate classes (see the row 'Dc unif.' in Table 4). This can probably be attributed to the many watersheds (261) in this climate class, which resulted in lower NSE scores and a significant deterioration of the overall scores.

3.2.2 Rigorous investigation of inadequate performance in the Dc climate class

Figure 5 displayed the cross plots of the parameters identified from the individual calibration of the Dc climate class. We found that there were multiple clusters of identified parameters compared to Figure 3, which presented the posterior distributions of the climate-based calibration for Db. We also observed regional patterns for each parameter (Figure 6), indicating that geographic regions can determine the parameters. Encouraged by the spatial smoothness of the individually calibrated parameters, we divided the Dc climate class into six subareas depending on the longitude (Table 5). We then conducted ABC for each subarea and inferred the representative parameter sets. The combinations of the acceptance ratios and statistical measures were also tested, as for the climate-based calibration.

Figure 7 summarized how the divisions of the Dc climate class on the changes in the NSE values of each watershed. The initial climate-based calibration in Dc resulted in NSE values lower than 0.25 for most of the stations (shown as red in Figure 7(a)), except for the stations in the subarea IV. The higher NSE scores in the subarea IV indicated that the initial sampling of the posterior was selective and that the posterior predominantly consisted of the stations in the area. Next, we conducted a simulation with the representative parameters that were determined for each subarea from the median (Figure 7 (b)) and mode (Figure 7 (c)) of the posterior distributions. Our simulations revealed that modal values of the posterior distributions yielded better NSE scores compared to those of the median, especially in subareas I and V. This indicates that the posterior distributions for each area were well constrained in the search domain. We also indicated in Figure 7 the differences in NSE from the initial climate-based calibration (i.e., uniform sampling throughout Dc; Figure 7(a)) to the calibration for the divided subareas with the posterior-mode (Figure 7(c)). This shows improvement with the division method for 87% of the stations in the Dc climate class (Figure 7 (d)).

3.3 Effectiveness of calibrated parameters

The NSE values obtained from the four parameter sets were compared (Figure 8). The four parameter sets are the H08 model default parameters, the optimized parameters for an individual watershed, the representative parameters obtained for each climate class, and the parameters optimized for the entire global data set. As expected, the boxplots for the individual calibrations outperformed for all the climate classes, exhibiting the best median values and the narrowest ranges between the first and third quartiles. Using the default parameters as a reference, we compared the gained improvement of the other two calibrated parameters. The NSE distributions with the climate-based parameter substantially improved from those with global calibrations and default parameters in the climate classes Af, Aw, BS, and BW, highlighting the representativeness of the climate-based parameters. A notable feature of the NSE distributions of the climate-based calibrations was the narrower ranges between the first and third quartiles compared to those with default parameters, exhibiting the robustness of the calibration procedures. The climate-based and

global calibrations produced similar NSE distributions in the temperate and continental climate classes (i.e., Ca, Cb, Da, and Db) suggesting that a large number of stations in these climate classes probably contributed to the parameters selected in the uniform calibration because: (1) Relatively higher NSE scores were obtained for these climate classes, and (2) The number of stations listed in the four climate classes totaled 400 (51.2% of total stations).

Overall, using ABC in the climate-based calibration procedure markedly improved the representativeness and robustness of the parameter sets used in the H08 global hydrological model. Table 5 summarized the number of stations with "satisfactory" and "good" performances (see section 2.4.3). For the Dc climate class, two cases of the climate-based calibrations were also presented: the differences between the initial attempt (Dc (unif.): treating the whole area uniformly) and the second attempt (Dc (div.): dividing the whole into six subareas). The number of total stations with good and satisfactory performances sequentially increased from the default to the global, reaching a maximum for the climate-based parameters. In total, 61.7% and 30.1% of the stations exhibited "satisfactory" and "good" performances, respectively. For the climate-based calibration, the NSE scores improved at 72.2% of all the stations from the default, and 35.6% from the global parameters. Note that the global and climate-based parameters for each of the climate classes Ca, Cb, and Db were identical; thus, no improvement from the global parameters was achieved in these climate classes. However, the effects of dividing Dc were remarkable, significantly increasing the number of stations with a "satisfactory" performance (from 93 to 158) and "good" performance (from 18 to 70).

Figure 9 compared the default parameters and identified values of *SD*, C_D, γ, and τ obtained from the three calibration methods (individual, climate-based, and global). Note that the number of stations per catchment varied (Table 2). The blue-shaded boxplots, representing the distributions of individual calibrations for Af, Am, and BS showed quite narrow ranges because the number of the stations in those classes was too few (3, 4 and 6, respectively). Moreover, the posterior distribution of the climate-based calibration for BW was indistinct because the posteriors included the parameter sets that yielded NSE scores lower than 0.0. Except for the climate classes Af, Am, and BS, the blue-shaded boxplots of the two parameters, *SD* and C_D, showed relatively narrow ranges for the Aw, Ca, Cb, Da, and Db classes. Because *SD* and C_D were sensitive parameters and explain more effectively variabilities in the NSE scores than γ and τ (Figure 1), the narrower ranges of the individual parameters for these climate classes justify our hypothesis to identify representative parameters based on their climate class. Conversely, the ranges of the individual parameter *SD* in the class Dc and ET showed broad distributions in the search domain. This was consistent with the lower performance of the initial attempt to search climate-based parameters in these classes.

The climate-based parameters (red dots in Fig. 9) were mostly identified in the ranges of the first and third quartiles of the individual parameter distributions. This suggests that they successfully captured the differences in the hydrological properties on a global scale. The six climate-based parameters of the Dc climate class also showed consistency with the individual calibration. While the individual parameters of *SD* demonstrated wide distributions in the search domain, four out of the six climate-based parameters identified for each subarea were within the box. The individual parameters of C_D displayed a relatively narrow range, and the six parameters were concentrated near the median values of the individual parameter.

In contrast, the global calibration (green dots) and default parameter (blue dots) values were both set constant for *SD* and C_D and deviated from the distributions of the individual

parameters, particularly in the tropical (Af, Am, and Aw) and arid (BS and BW) climate classes (in Fig. 9). This fact highlighted the significance of the gained improvement from the default and global parameters in these regions (Table 5). Because the number of stations in the tropical and arid regions accounts for only 11.7% (91 stations), the improvement will be more prominent if the number of stations per climate system increases.

4 Discussion

4.1 Does climate exert a dominant control on hydrological properties on a global scale?

In this study, we tested the identifiability of representative parameters determined for the Köppen climate classes. The representative parameters provided improved streamflow simulations compared with those of the default and the global calibration. Moreover, the climate-based parameters showed remarkable improvement in four out of the five climate classes composed of more than 50 watersheds (i.e., Aw, Ca, Cb, and Db). This supports our initial hypothesis that similarity in climate properties is a dominant control on hydrological properties on a global scale. Our results are also consistent with the previous findings of the validities of transferring parameters of global hydrological models based on climate properties (Nijssen et al., 2001a, b; Beck et al., 2016).

We highlighted the importance of the direct or indirect connections of the effects of climate on the hydrological function of watersheds. First, we revealed the most critical relationship between the climate and modeled values of the bulk transfer coefficient C_D, which depends on the roughness of the canopy surface (Stull 1991). C_D showed a decreasing trend from the warmer to the cooler climate classes (Figure 9 (b)). The parameter values obtained for the individual and climate-based calibrations are consistent with the notion that tropical regions typically have a high evapotranspiration efficiency due to dense vegetation, unlike cooler regions, which generally have a low evapotranspiration efficiency. In the first generations of land surface models (LSMs), C_D was set to the standard value for grassland vegetation (Manabe 1969, Hartmann 1994). Milly and Shmakin (2002) developed the Land Dynamics (LaD) model, which calculates the water and energy balance with parameters based on the vegetation and soil types. The global simulation of LaD showed an improved annual water balance, justifying the parameterization based on land surface attributes. Most of the operational LSMs today employ detailed and complex parameterization of the land surface scheme but use a priori parameters with look-up tables, limiting their abilities for model improvement via sensitivity analysis (Samaniego et al., 2017). Our approach employed a simple land surface scheme, but the identifiability of the "effective" heat-flux parameterization will benefit further improvement.

The soil depths (*SD*) obtained from the climate-based calibration showed a decreasing trend with cooling climate, from the tropical (Af, Am, and Aw) through to the subarctic (Dc) and tundra (ET) (Figure 9 (a)), suggesting that the water-holding capacity in the subsurface has direct or indirect links with climate (Harman & Troch, 2014; Troch et al., 2015). Direct links may include the weathering of rock, which is related to the amount and temperature of water flowing through it, so the rate of soil-formation is higher in regions with ample rainfall and warmer temperatures (Rasmussen et al., 2005). This rationale corroborated the decrease in soil depth along the climate gradient. Studies involving direct comparisons of modeled *SD* with global maps of soil depth (e.g., Pelletier et al., 2016) are worthy of future research. However, as water can be stored in the soil

layer as well as in deeper bedrock (Sayama et al., 2009; Ajami et al., 2011), and the optimized parameters of subsurface storage may include such deep aquifer storage, a direct comparison may not be possible between parameterized soil depths (*SD*) and the actual volume of water stored in the soil layer. Understanding the global variance in subsurface storage capacity is more relevant (Harman and Troch, 2014). Moreover, it would help us understand the long-term water-balance or the baseflow characteristics of watersheds. Milly (1994) demonstrated that water storage in the soil was essential to explain the seasonal variation of water-balance over the Eastern United States. Yoshida & Troch (2016) showed that the estimated storage of the deep aquifers in volcanic watersheds varied with geological timescales under similar climatic conditions of the Western United States and Japan. As these studies were conducted in regions with similar climatic conditions, a comprehensive understanding between the subsurface storage and climate on the global scale was limited.

The Dc climate class showed diversified individual parameters, and the representative parameters for the entire area yielded poor results. This suggests that the Köppen climate classes are not the only measure that explains global hydrological differences. We found that the individual parameters in Dc did not exhibit spatially randomized patterns, but rather smoothness in space (Figure 6), which divided the entire area into sub-areas and improved the NSE score. This corroborates the findings of Addor et al. (2018) that the spatial smoothness of hydrological signatures can be satisfactorily regionalized when the signatures exhibit smoothness, most likely reflecting the climate.

It is out of the scope of this paper to elucidate why and how these differences in the subareas emerged; however, we will provide several possible explanations that can be investigated in the future. A highly constrained C_D suggests that the values reflect surface roughness's actual properties, showing low evaporative potential in the arctic climate. This is also consistent with the global differences in C_D (Figure 9 (b)). Conversely, the interpretations of variabilities in *SD* are not straightforward, ranging from 0.208 (subarea VI) to 3.027 (subarea II). We suggest that the variance may reflect the processes relevant to permafrost or lakes, which are not or poorly represented in the H08 model. The extremely small *SD* values for the subarea VI (Eastern Siberia) correspond to the areal extension of the 'continuous' permafrost, which means that 90-100% of the area was covered permafrost (Brown et al, 2002, Figure S3). As the permafrost decreases, *SD* values tend to increase (westward on the Eurasia continent). However, for subarea III (Eastern Canada), the *SD* value was the largest even though the area was covered with 'continuous' or 'discontinuous' permafrost. This might be explained by the existence of large lakes (Lehner and Döll, 2004, Figure S4). Relatively low NSE scores in this area, even after the subarea-based calibration (Figure 7 (c)), also suggests the influence of other factors that were not accounted for by the model on observed discharges (i.e., disturbance by the storage in lakes).

4.2 Practical guidance for the application of ABC to large scale hydrological modeling

The ABC algorithm is based on the rationale that one can approximate 'true' posterior distributions when sampling can be conducted an unlimited number of times (Beaumont et al., 2002; Sisson et al., 2018). However, due to the high demand for computational resources, the calibrations of global hydrological models would not be conducted in an ideal way. In this study, the number of prior samples was limited to 5000. The challenge was to derive useful information from the limited number of samples using the framework of ABC. The successful identification of

the parameters in this study may be attributed to the low dimensionality of the problem of four sensitive parameters, thanks to the previous attempt to calibrate the H08 at the individual watershed (Hanasaki et al., 2014; Masood et al., 2015; Mateo et al., 2014; Yoo, 2016). For the problems with higher dimensionality, the simple rejection algorithm requires substantial iterations to obtain useful posterior distributions (Sadegh and Vrugt, 2013).

The approximated posterior distribution accuracy depends on the choice of the summary-metrics and the effects of other sources of errors (e.g., climate forcing or model structures). The use of discharge-based metrics in ABC possesses similarity with the limits of the acceptability approach of generalized likelihood uncertainty estimation, GLUE (Beven & Binley, 1992). The theoretical connections between ABC and GLUE were thoroughly discussed in the references (Nott et al., 2012; Sadegh & Vrugt, 2013), but we point out the consequences of selecting NSE as the summary metrics. The low sensitivity of γ and τ might partly reflect our use of monthly-averaged streamflow for the calibration or NSE for summary metrics. Parameters γ and τ explain the groundwater recession rate, but the monthly streamflow was presumably insufficient to represent the recession rate, especially for large watersheds (i.e., with catchment areas >10,000 km^2). γ and τ might have shown more sensitivity to the overall results if we had used daily or weekly mean streamflow data to calibrate the smaller watersheds. Moreover, NSE is particularly sensitive to the timing of flood peaks because it calculates error residuals based on the ratio of the mean square error to the variance of observed streamflows (Nash & Suttcliffe, 1979). Using other hydrological signatures would provide more explicit contributions of these parameters, e.g., baseflow index (Vogel & Kroll, 1992; Kroll et al., 2004), or the slope of the flow duration curves (Yadav et al., 2007; Zhang et al., 2008).

Lastly, we determined the method to identify representative parameter sets for each climate class using median values of the posterior distribution obtained with an acceptance ratio of 5%. However, we do not argue that these criteria are conclusive, but rather that the application of this method to other datasets and with different purposes may result in different criteria.

4 Conclusion

In this study, we introduced the ABC technique to calibrate four sensitive parameters of the H08 global hydrological model for gauged watersheds and aggregated the 5000 simulated samples into 11 Köppen climate classes. We then tested the hypothesis that the parameters derived from the aggregated posterior distribution represent the hydrological properties in the same climate class, and are transferable to the watersheds in that climate classes (climate-based calibration). By randomly splitting watersheds into equal-sized calibration and validation datasets, we found that the representativeness and robustness of the climate-based parameters are satisfied with an NSE acceptance ratio of 5% and the median of the posterior distribution to define representative parameters. The simulation with the climate-based parameters yielded satisfactory (NSE > 0.0) and good (NSE > 0.5) performances at 480 and 234 stations (61.7% and 30.1% of 777 stations), respectively, demonstrating a significant improvement from those simulated with default parameters. Simulations using climate-based parameters also showed higher NSE scores than those with default parameters for 72.2% of the watersheds.

The 11 Köppen climate classes' identified parameters showed consistency with the physical interpretation of soil formation and efficiencies in vapor transfer with a wide variety of vegetation

types. The consistency of the defined parameter values with physical underpinnings indicates that the correct parameters were determined, ensuring the robustness of the parameters, particularly when transferred to ungauged watersheds. One of the significant advantages of applying ABC to a global hydrological model is that it can be easily implemented without complex code modifications and, as in this study, the results of the same calculations can be used to determine the ideal combination of parameters in an exploratory manner. Therefore, this technique is suitable for studies aiming to constrain model parameters that better predict watershed behaviors on a global scale.

Acknowledgments

This work was funded by the Environment Research and Technology Development Fund (JPMEERF20182R02) of the Environmental Restoration and Conservation Agency, Japan, and Ministry of Education, Culture, Sports, Science and Technology/Japan Society for the Promotion of Science KAKENHI grant (grant number 19K06304). The Global Runoff Data Centre (GRDC) is gratefully acknowledged for providing their valuable river discharge data. Data is available through Weedon et al. (2011) and Global Runoff Data Centre (https://www.bafg.de/GRDC).

References

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018), A ranking of hydrological signatures based on their predictability in space. Water Resources Research, 54(11), 8792-8812.

Ajami, H., Troch, P. A., Maddock III, T., Meixner, T., & Eastoe, C. (2011), Quantifying mountain block recharge by means of catchment-scale storage-discharge relationships. *Water Resources Research*, 47(4).

Beck, H. E., Roo, A. d., & Dijk, A. I. J. M. v. (2015), Global maps of streamflow characteristics based on observations from several thousand catchments, *Journal of Hydromet*., 16(4), 1478-1501, doi: 10.1175/jhm-d-14-0155.1.

Beck, H. E., Dijk, A. I. J. M. van, Roo, A. de, Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016), Global-scale regionalization of hydrologic model parameters, *Water Resour. Res*., 52(5), 3599-3622, doi: 10.1002/2015wr018247.

Beven, K., & Freer, J. (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of hydrology*, *249*(1-4), 11-29.

Bierkens, M. F. P. (2015), Global hydrology 2015: State, trends, and directions, *Water Resour. Res*., 51(7), 4923-4947, doi: 10.1002/2015WR017173.

Brown, J., Ferrians, O., Heginbottom, J. A., & Melnikov, E. (2002), Circum-Arctic map of permafrost and ground-ice conditions, version 2. Boulder, Colorado USA, National Snow and Ice Data Center. [Date Accessed 25 June 2020].

Dalin, C., Suweis, S., Konar, M., Hanasaki, N., & Rodriguez-Iturbe, I. (2012), Modeling past and future structure of the global virtual water trade network, *Geophys. Res. Lett*., 39(24), L24402, doi: 10.1029/2012gl053871.

Döll, P., & Lehner, B. (2002), Validation of a new global 30-min drainage direction map, *J. Hydrol*., 258(1-4), 214-231.

Döll, P., & Fiedler, K. (2008), Global-scale modeling of groundwater recharge, *Hydrol. Earth Syst. Sci*., 12(3), 863-885, doi: doi:10.5194/hess-12-863-2008.

Döll, P., Kaspar, F., & Lehner, B. (2003), A global hydrological model for deriving water availability indicators: model tuning and validation, *J. Hydrol*., 270(1–2), 105-134, doi: http://dx.doi.org/10.1016/S0022-1694(02)00283-4.

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., & Eicker, A. (2014), Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. Water *Resour. Res*., 50(7), 5698-5720.

Fekete, B. M., & Vörösmarty, C. J. (2007), The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *IAHS publ*, 309, 129-136.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., & Sitch, S. (2004), Terrestrial vegetation and water balance—hydrological evaluation of a dynamic global vegetation model. *J. Hydrol*., 286(1-4), 249-270, doi: 10.1016/j.jhydrol.2003.09.029.

Gleeson, T., Wada, Y., Bierkens, M. F., & Van Beek, L. P. (2012), Water balance of global aquifers revealed by groundwater footprint. *Nature*, 488(7410), 197-200, doi: 10.1038/nature11295.

Goodrich, D.C., T. J Schmugge, T. J. Jackson, C. L. Unkrich, T. O. Keefer, R. Parry, L. B. Bach, and S. A. Amer (1994), Runoff simulationsensitivity to remotelysensed initial soil water content, *Water Resour. Res*., 30(5), 1393-1405, doi/abs/10.1029/93WR03083.

Greve, P., Burek, P., & Wada, Y. (2020), Using the Budyko frameworkfor calibrating a global hydrological model. Water Resources Research, 56, e2019WR026280. https://doi.org/10.1029/2019WR026280

Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N. & Voss, F. (2012), *J. Hydromet*., 13(2), 604-620, doi: 10.1175/jhm-d-11-083.1.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J., Stacke, T., Tessler, Z. D., Wada, Y., & Stacke, T. (2014). Global water resources affected by human interventions and climate change. *P. Natl. Acad. Sci*., 111(9), 3251-3256, 3251-3256, doi: 10.1073/pnas.1222475110.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., & Yeh, P. (2011), Multimodel estimate of the global terrestrial water balance: setup and first results. *J. Hydromet*., 12(5), 869-884, doi: 10.1175/2011jhm1324.1.

Hanasaki, N., Inuzuka, T., Kanae, S., & Oki, T. (2010), An estimation of global virtual water flow and sources of water withdrawal for major crops and livestock products using a global hydrological model. *J. Hydrol*, 384(3-4), 232-244.

Hanasaki, N., Fujiwara, M., Maji, A., & Seto S. (2018), On the applicability of the H08 global water resources model to the kyusyu island, *JSCE Journal B1*, 74(4), I_109-I_114.

Hanasaki, N., Yoshikawa, S., Pokhrel, Y., & Kanae, S (2018), A global hydrological simulation to specify the sources of water used by humans, *Hydrol. Earth Syst. Sci*., 22(1), 789-817, doi: 10.5194/hess-22-789-2018.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., & Tanaka K. (2008a), An integrated model for the assessment of global water resources - Part 1: Model description and input meteorological forcing, *Hydrol. Earth Syst. Sci*., 12(4), 1007-1025, doi: 10.5194/hess-12-1007-2008.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., & Tanaka K. (2008b), An integrated model for the assessment of global water resources - Part 2: Applications and assessments, *Hydrol. Earth Syst. Sci*., 12(4), 1027-1037, doi: 10.5194/hess-12-1027-2008.

Hanasaki, N., Saito, Y., Chaiyasaen, C., Champathong, A., Ekkawatpanit, C., Saphaokham, S., Sukhapunnaphan, T., Sumdin, S., & Thongduang J. (2014), A quasi-real-time hydrological simulation of the Chao Phraya River using meteorological data from the Thai Meteorological Department Automatic Weather Stations, *Hydrological Research Letters*, 8(1), 9-14, doi: http://dx.doi.org/10.3178/hrl.8.9.

Harman, C., & Troch, P. A. (2014), What makes Darwinian hydrology" Darwinian"? Asking a different kind of question about landscapes. *Hydrol. Earth Syst. Sci.*, 18(2), 417.

Hartmann, D. L. (1994), Global Physical Climatology (International geophysics; v. 56). Academic Press.

Köppen, W. (2011), The thermal zones of the Earth according to the duration of hot, moderate and cold periods and to the impact of heat on the organic world, *Meteorologische Zeitschrift*, 20(3), 351-360, doi: 10.1127/0941-2948/2011/105.

Kroll, C., Luz, J., Allen, B., & Vogel, R. M. (2004). Developing a watershed characteristics database to improve low streamflow prediction. *J. Hydrol. Engineering*, 9(2), 116-125.

Krysanova, V., Donnelly, C., Gelfan, A., Gerten, D., Arheimer, B., Hattermann, F., & Kundzewicz, Z. W. (2018). How the performance of hydrological models relates to credibility of projections under climate change. *Hydrological Sciences Journal*, 63(5), 696-720.

Lehner, B. and Döll, P. (2004): Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology 296/1-4: 1-22.

Manabe, S. (1969), Climate and the ocean circulation 1. The atmospheric circulation and the hydrology of the earth's surface, *Mon. Weather Rev*., 97(11), 739-774, doi: 10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2.

Masood, M., Yeh, P. J. F., Hanasaki, N., and Takeuchi K. (2015), Model study of the impacts of future climate change on the hydrology of Ganges–Brahmaputra–Meghna basin, *Hydrol. Earth Syst. Sci*., 19(2), 747-770, doi: 10.5194/hess-19-747-2015.

Mateo, C. M., Hanasaki, N., Komori, D., Tanaka, K., Kiguchi, M., Champathong, A., Sukhapunnaphan, T., Yamazaki, D., & Oki T. (2014), Assessing the impacts of reservoir operation to floodplain inundation by combining hydrological, reservoir management, and hydrodynamic models, *Water Resour. Res*., 50(9), 7245-7266, doi: 10.1002/2013wr014845.

Milly, P. C. D. (1994), Climate, soil water storage, and the average annual water balance. Water Resources Research, 30(7), 2143-2156.

Milly, P. C. D., & Shmakin, A. B. (2002), Global modeling of land water and energy balances. Part I: The land dynamics (LaD) model. Journal of Hydrometeorology, 3(3), 283-299.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885-900.

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015), Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, 58(6), 1763-1785.

Nash, J. E., & Sutcliffe, J. V. (1970), River flow forecasting through conceptual models part I — A discussion of principles, *J. Hydrol*., 10(3), 282-290, doi: http://dx.doi.org/10.1016/0022-1694(70)90255-6.

New, M., Hulme, M. & Jones P. (2000), Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate, *J. Climate*, 13(13), 2217-2238.

Nicolau, J. M., Solé-Benet, A., Puigdefábregas, J., & Gutiérrez, L. (1996), Effects of soil and vegetation on runoff along a catena in semi-arid Spain. *Geomorphology*, 14(4), 297-309.

Nijssen, B., Schnur, R. & Lettenmaier D. P. (2001a), Global retrospective estimation of soil moisture using the variable infiltration capacity land surface model, 1980-93, *J. Climate*, 14(8), 1790-1808.

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

Nott

Oki, T., Nishimura, T., & Dirmeyer P. (1999), Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP), *J. Meteorol. Soc. Jpn*., 77(1B), 235-255, doi: 10.2151/jmsj1965.77.1B_235.

Pokhrel, Y. N., Hanasaki, N., Wada, Y., & Kim, H. (2016), Recent progresses in incorporating human land–water management into global land surface models toward their integration into Earth system models, *Wiley Interdisciplinary Reviews: Water*, 3(4), 548-574, doi: 10.1002/wat2.1150.

Rasmussen, C., Southard, R. J., & Horwath, W. R. (2005), Modeling energy inputs to predict pedogenic environments using regional environmental databases. *Soil Science Society of America Journal*, 69(4), 1266-1274.

Robock, A., Vinnikov, K. Y., Schlosser, C. A., Speranskaya, N. A., & Xue, Y. K. (1995), Use of Midlatitude Soil-Moisture and Meteorological Observations to Validate Soil-Moisture

729 Simulations with Biosphere and Bucket Models, *J. Climate*, 8(1), 15-35, doi: 10.1175/1520-
730 0442(1995)008<0015:UOMSMA>2.0.CO;2.

731 Rost, S., Gerten, D., Bondeau, A., Lucht, W., Rohwer, J., & Schaphoff, S. (2008), Agricultural
732 green and blue water consumption and its influence on the global water system, *Water Resour.*
733 *Res.*, 44, W09405, doi: doi:10.1029/2007WR006331.

734 Sadegh, M., & Vrugt, J. A. (2013), Bridging the gap between GLUE and formal statistical
735 approaches: approximate Bayesian computation. Hydrology and Earth System Sciences, 17(12),
736 4831-4850.

737 Sadegh, M., and Vrugt, J. A. (2014), Approximate Bayesian Computation using Markov Chain
738 Monte Carlo simulation: DREAM(ABC), *Water Resour. Res.*, 50(8), 6767-6787, doi:
739 10.1002/2014wr015386.

740 Sadegh, M., Vrugt, J. A., Xu, C., & Volpi, E. (2015), The stationarity paradigm revisited:
741 Hypothesis testing using diagnostics, summary metrics, and DREAM (ABC). *Water Resour. Res.*,
742 51(11), 9207-9231.

743 Schewe, J., et al. (2014), Multimodel assessment of water scarcity under climate change. *P. Natl.*
744 *Acad. Sci. USA*, 111(9), 3245-3250, doi: 10.1073/pnas.1222460110.

745 Schewe, J., et al. (2019), State-of-the-art global models underestimate impacts from climate
746 extremes. Nature communications, 10(1), 1-14.

747 Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., & Rudolf, B. (2014),
748 GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its
749 role in quantifying the global water cycle. *Theoretical and Applied Climatology*, 115(1-2), 15-40.

750 Sisson, S. A., Fan, Y., & Beaumont, M. (2018), *Handbook of approximate Bayesian computation*.
751 Chapman and Hall/CRC.

752 Stull, R. B. (2012), An introduction to boundary layer meteorology (Vol. 13). Springer Science &
753 Business Media.

754 Sutanudjaja, E. H., et al. (2018), PCR-GLOBWB 2: a 5 arcmin global hydrological and water
755 resources model. *Geoscientific Model Development*, 11(6), 2429-2453.

756 Troch, P. A., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J. W., Roy, T., & Valdés-Pineda,
757 R. (2015), Catchment coevolution: A useful framework for improving predictions of hydrological
758 change?. *Water Resour. Res.*, *51*(7), 4903-4922.

759 Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. Roy. Meteor. Soc.*, 131(612), 2961-
760 3012.

761 Vogel, R. M., & Kroll, C. N. (1992). Regional geohydrologic-geomorphic relationships for the
762 estimation of low-flow statistics. *Water Resour. Res.*, 28(9), 2451-2458.

763 Vrugt, J. A., & Sadegh, M. (2013), Toward diagnostic model calibration and evaluation:
764 Approximate Bayesian computation, *Water Resour. Res.*, 49(7), 4335-4345, doi:
765 10.1002/wrcr.20354.

766 Wada, Y., D. Wisser, and M. F. P. Bierkens (2014), Global modeling of withdrawal, allocation
767 and consumptive use of surface water and groundwater resources, *Earth Syst. Dynam.*, 5(1), 15-
768 40, doi: 10.5194/esd-5-15-2014.

Weedon, G. P., S. Gomes, P. Viterbo, W. J. Shuttleworth, E. Blyth, H. Österle, J. C. Adam, N. Bellouin, O. Boucher, and M. Best (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. Hydromet*., 12(5), 823-848, doi: 10.1175/2011jhm1369.1.

Widén-Nilsson, E., Halldin, E., S., & Xu, C.-y. (2007), Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation, *J. Hydrol*., 340(1), 105-118, doi: https://doi.org/10.1016/j.jhydrol.2007.04.002.

Yadav, M., Wagener, T., & Gupta, H. (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in water resources*, 30(8), 1756-1774.

Yoo, S. (2016), Estimation of water availability in the Korean Peninsula considering climate change, 104 pp, Seoul National University.

Yoshida, T. and Troch, P. A. (2016), Coevolution of volcanic catchments in Japan, *Hydrol. Earth Syst. Sci*., 20, 1133–1150, https://doi.org/10.5194/hess-20-1133-2016.

Zaherpour, J., et al. (2018), Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts, *Environ. Res. Lett*., 13(6), 065015.

Zhang, X., Zhang, L., Zhao, J., Rustomji, P., & Hairsine, P. (2008). Responses of streamflow to changes in climate and land use/cover in the Loess Plateau, China. *Water Resour. Res.*, 44(7).

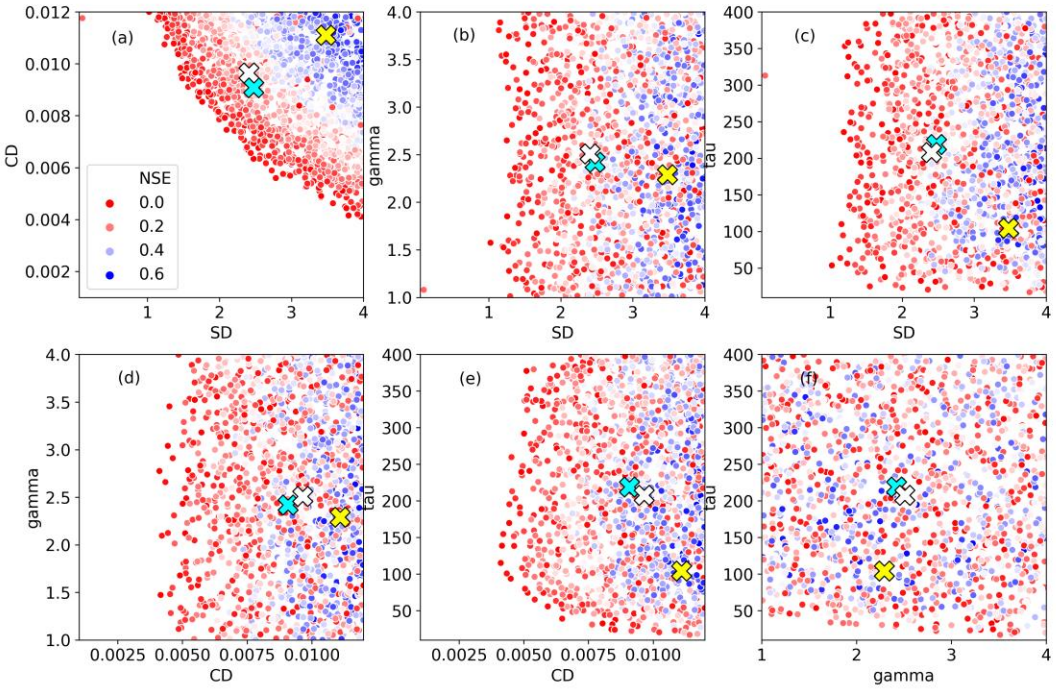


Figure 1. Cross plots of the posterior distributions of the individual calibration obtained with the acceptance ratio 30% (at watershed ID: 4362600). Each subplot represents the bivariate plots between (a): *SD*-C_D, (b): *SD*-γ, (c): *SD*-τ, (d): C_D-γ, (e): C_D-τ, and (f): γ-τ. The crosses indicate the mode (yellow), median (white) and mean (light blue) of the posterior distributions, respectively.

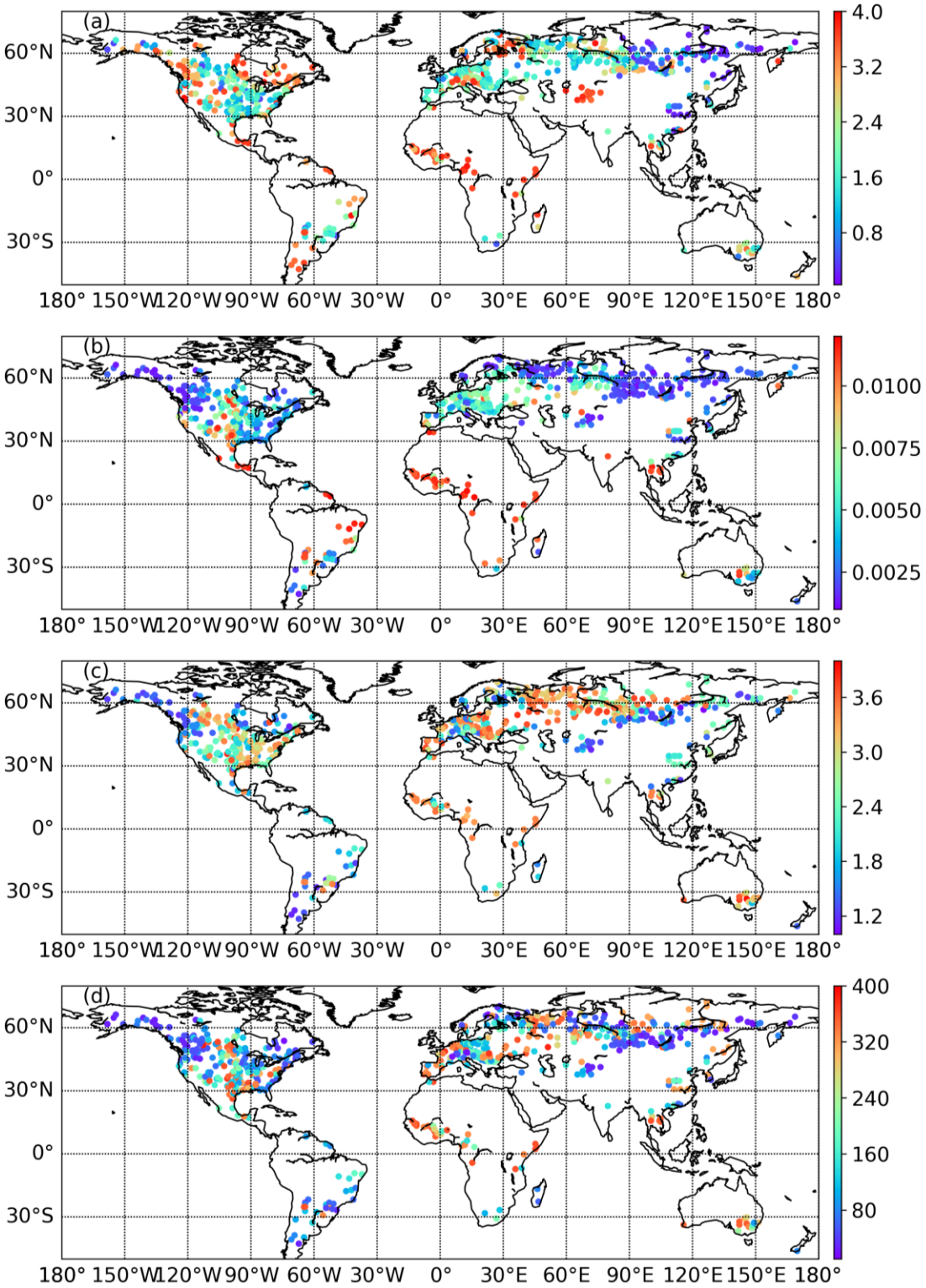


Figure 2. Spatial distributions of the parameter values obtained from the individual calibration. Each subplot represents the values of (a): *SD*, (b) C_D, (c) γ, and (d) τ.

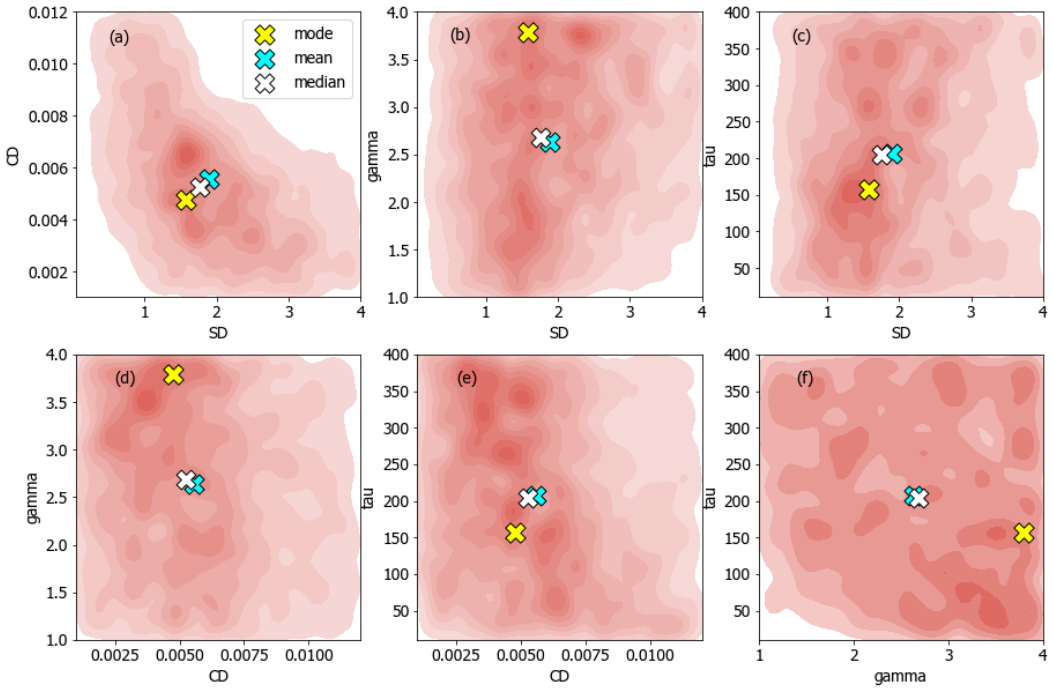


Figure 3. The cross plots of the posterior distributions for the climate class Db with the acceptance ratio of 10%. Each subplot represents the relations between (a): *SD*-C_D, (b): *SD*-γ, (c): *SD*-τ, (d): C_D-γ, (e): C_D-τ, and (f): γ-τ. The crosses depict the mode (yellow), mean (light blue), and median (white) of the posterior distributions, respectively.

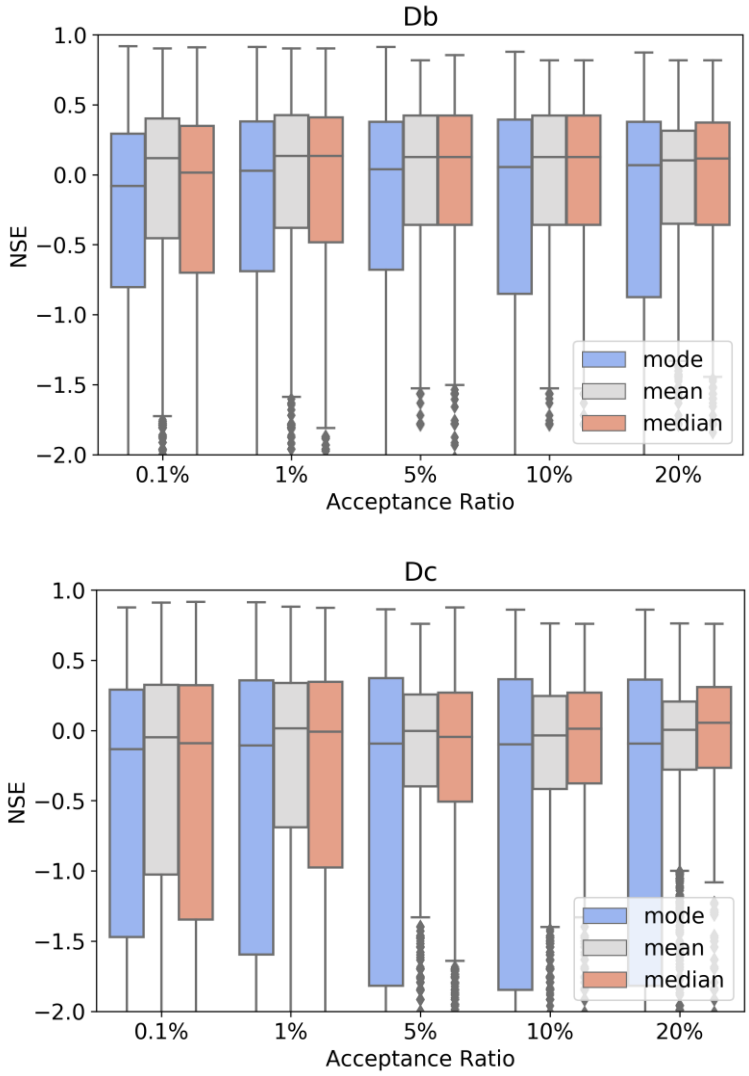


Figure 4. The ranges of NSE scores obtained from the validation dataset of the repeated two-fold cross-vlidation for two climate classes (Db and Dc) for 100 iterations.

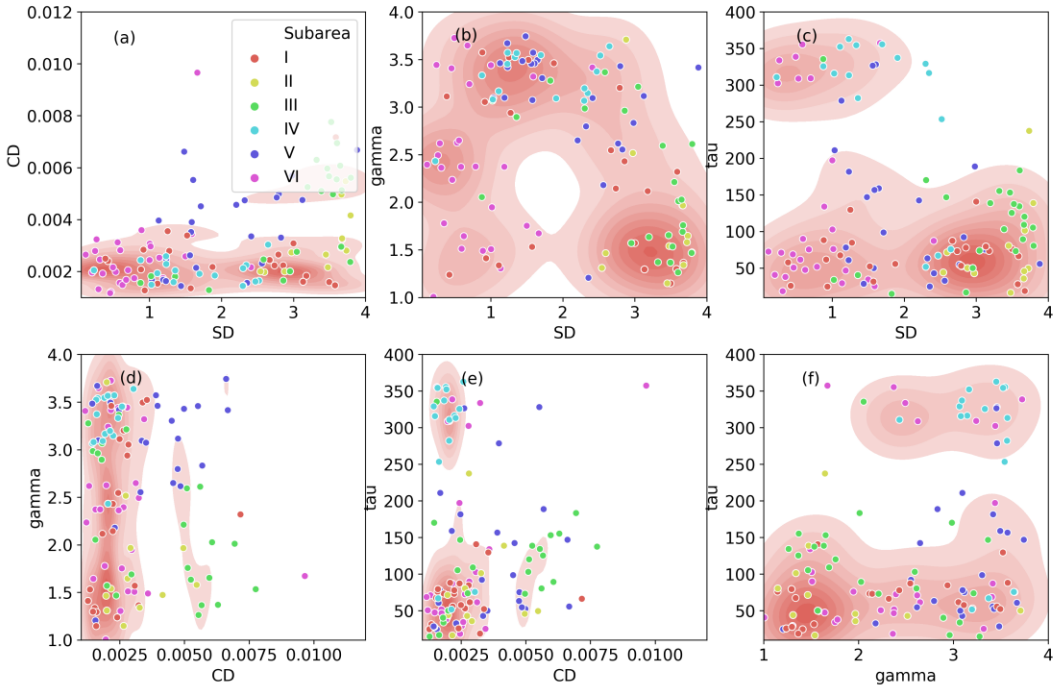
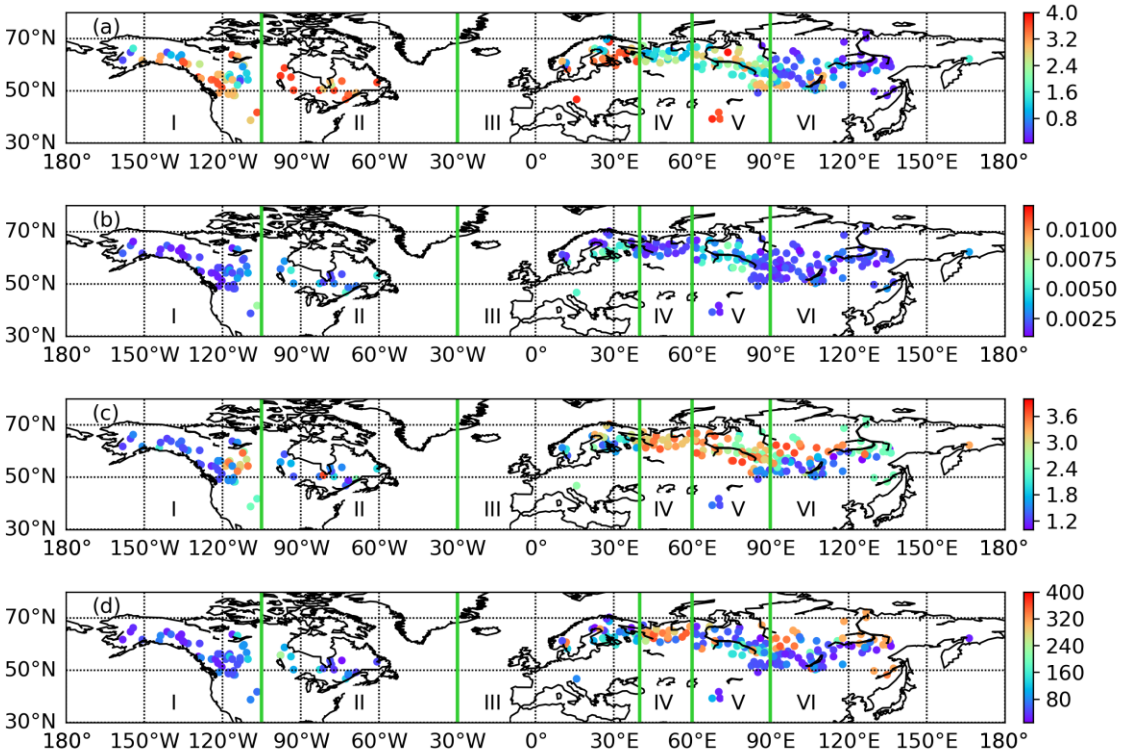
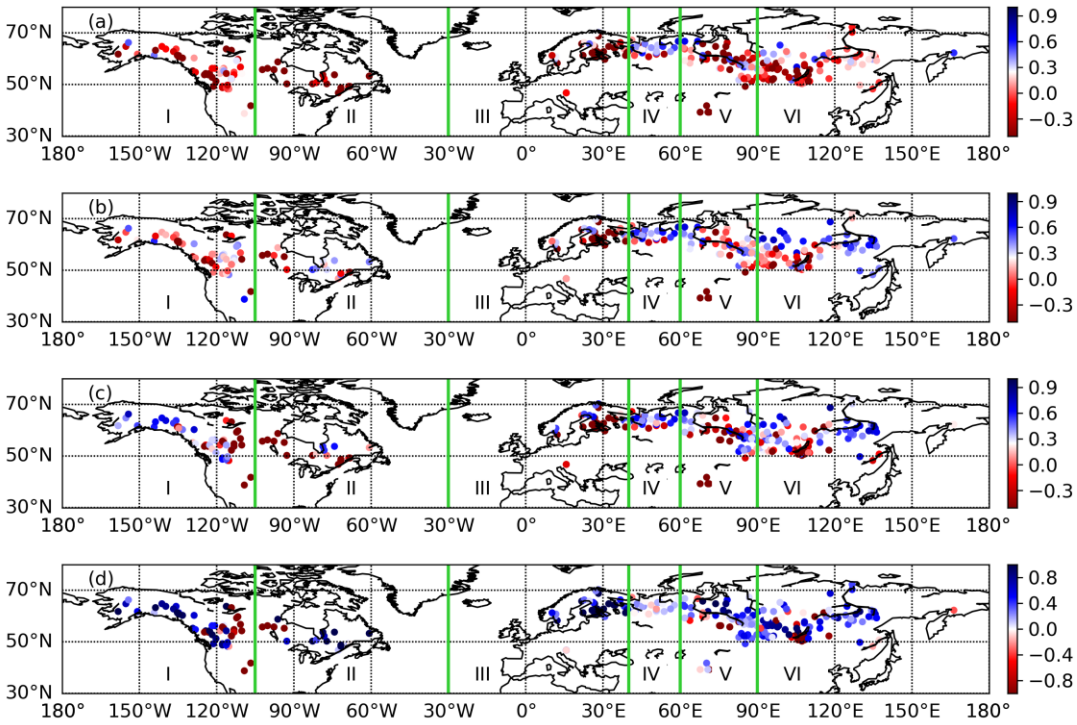


Figure 5. Cross plots of the parameters identified by the individual calibration for the climate class Dc. Each subplot represents the relations between (a): *SD*-C_D, (b): *SD*-γ, (c): *SD*-τ, (d): C_D-γ, (e): C_D-τ, and (f): γ-τ. The contour shows the kernel density of the individual parameters. The plots show the values of the individual parameters and their colors indicate the subareas shown in Figure 6 and Table 5.



821

822 **Figure 6**. Parameter map for the climate class Dc. Each subplot represents the values of (a): *SD*,
823 (b): C_D, (c): γ, and (d): τ.
824
825



826

827 **Figure 7**. NSE scores for the climate class Dc. Each of the panel shows (a): NSE scores obtained
828 from the climate-based calibration (Dc unif., in which whole area was treated unifromly), (b):
829 NSE scores obtained from the climate-based calibration dividing the whole area into 6 subareas
830 (posterior-median to derive the representative parameters), (c): same as (b) but posterior-mode
831 was used, and (d): differences in NSE scores between (a) and (c).
832

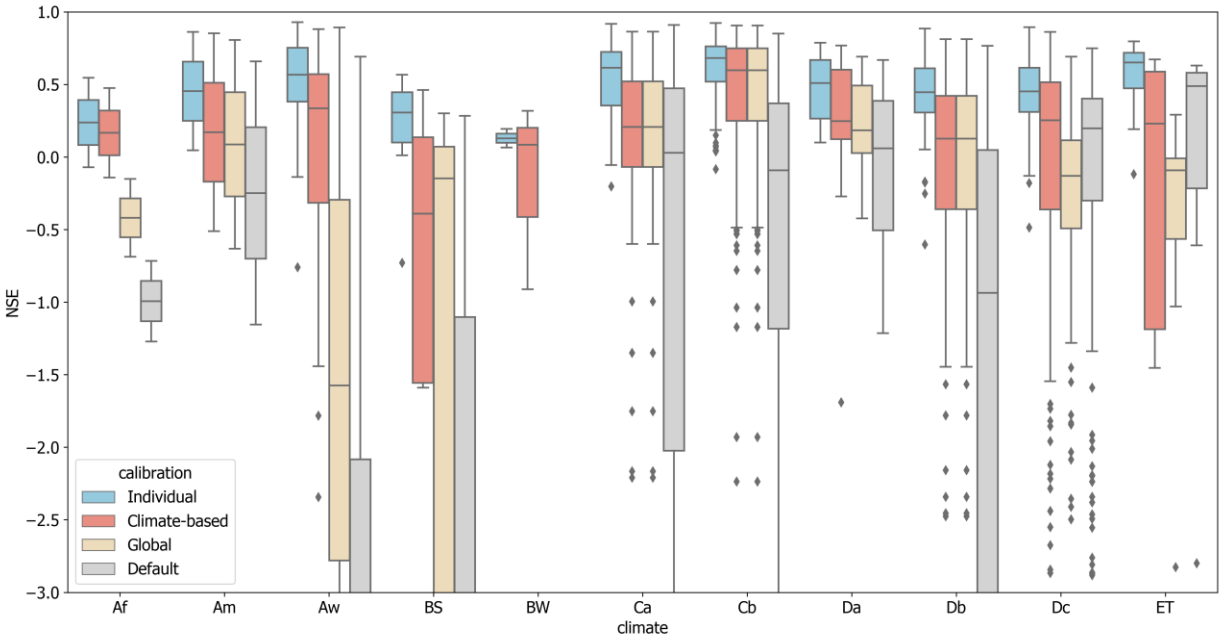


Figure 8. NSE scores obtained using the H08 model default parameters and calibrations using the individual watershed, climate-based, and global parameter sets for the 11 climate classes (statistical measure = median; acceptance ratio = 5%; Dc was divided into subareas). Note that NSE scores of BW for the default and global parameters were always below -3.0.

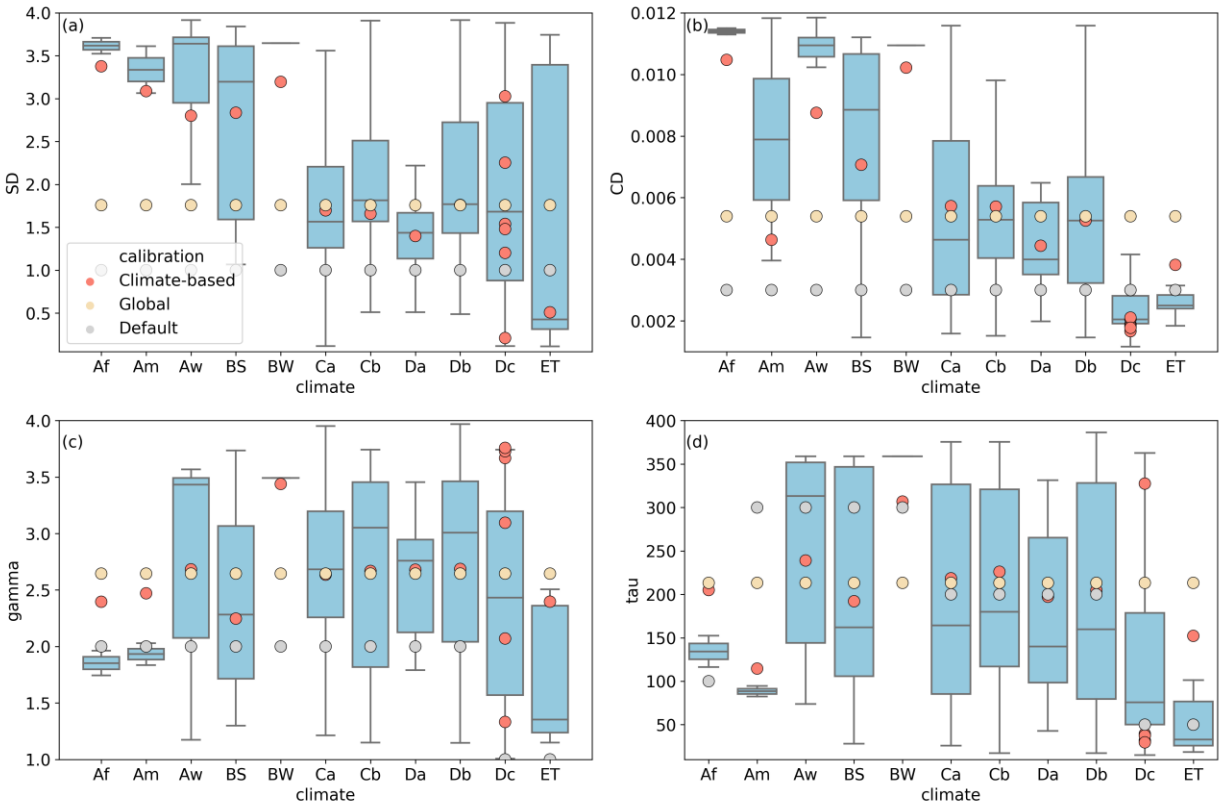


Figure 9. Comparison for all climate classes of optimized values of *SD*, C_D, γ, and τ (colored dots) with boxplots of the distributions of the individual calibrations (statistical measure = median; acceptance ratio = 5%).

847

848 **Table 1.** Default values for the H08 global hydrological model parameters and ranges of values
849 used for randomly generated parameter sets.

850

	Range	Default
Soil depth (*SD*)	0.05–4.0	1.0
Bulk transfer coefficient (C_D)	0.001–0.012	0.003
† Shape parameter for subsurface runoff (γ)	1–4	1–2†
† Time constant for subsurface runoff (τ)	10–400	50–300†

851 † Number differed by climatic zones. (γ, τ) is (2.0, 100) for tropical forest; (2.0, 300) for tropical
852 monsoon, savanna, and dry climates; (2.0, 200) for temperate and continental (warmer) climates;
853 and (1.0, 50.0) for continental (cooler) and polar climates (Hanasaki et al., 2008a).

854

855

856 **Table 2.** Köppen climate classes used in this study and number of stations for each climate class

Climate	Abbreviation	Number of stations
Tropical rain forest	Af	3
Tropical monsoon	Am	4
Tropical savanna	Aw	61
Arid	BW	17
Semi-arid	BS	6
Hot summer temperate	Ca	99
Warm summer temperate	Cb	129
Hot summer continental	Da	18
Warm summer continental	Db	164
Subarctic	Dc	262
Tundra	ET	14
Total		777

857

858

859 **Table 3.** NSE values as tolerance in ABC for each climate class according to the acceptance ratios

860

	Acceptance Ratio				
Climate	0.1%	1%	5%	10%	20%
Af	0.546	0.472	0.337	0.206	-0.033
Am	0.888	0.868	0.842	0.814	0.755
Aw	0.919	0.851	0.688	0.526	0.189
BS	0.598	0.455	0.255	0.113	-0.099
BW	0.396	0.240	-0.873	-4.527	-18.569
Ca	0.899	0.838	0.710	0.602	0.437
Cb	0.888	0.831	0.708	0.619	0.473
Da	0.811	0.718	0.602	0.516	0.355
Db	0.816	0.680	0.511	0.380	0.202
Dc	0.810	0.671	0.474	0.340	0.171
ET	0.797	0.684	0.548	0.401	0.212

861

862

863

864

865

866

867

868 **Table 4.** Number of stations for "satisfactory" and "good" performances for the climate-based, global, and default parameters.

869

		"Satisfactory" performance (0.5>NSE>0)			"Good" performance (NSE>0.5)			Gain of NSE Climate-based	
Climate	Num.	Climate-based	Global	Default	Climate-based	Global	Default	from Default	from Global
Af	3	1 (0.333)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	3 (1.000)	3 (1.000)
Am	4	1 (0.250)	1 (0.250)	1 (0.250)	1 (0.250)	1 (0.250)	1 (0.250)	4 (1.000)	4 (1.000)
Aw	61	24 (0.393)	9 (0.148)	2 (0.033)	14 (0.230)	3 (0.049)	1 (0.016)	61 (1.000)	57 (0.934)
BS	17	5 (0.294)	5 (0.294)	1 (0.059)	0 (0.000)	0 (0.000)	0 (0.000)	14 (0.824)	9 (0.529)
BW	6	2 (0.333)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	6 (1.000)	5 (0.833)
Ca	99	69 (0.697)	69 (0.697)	49 (0.495)	26 (0.263)	26 (0.263)	21 (0.212)	65 (0.657)	0 (0.000)
Cb	129	101 (0.783)	101 (0.783)	54 (0.419)	74 (0.574)	74 (0.574)	20 (0.155)	107 (0.829)	0 (0.000)
Da	18	16 (0.889)	14 (0.778)	10 (0.556)	7 (0.389)	4 (0.222)	3 (0.167)	17 (0.944)	11 (0.611)
Db	164	96 (0.585)	96 (0.585)	40 (0.244)	37 (0.226)	37 (0.226)	10 (0.061)	130 (0.793)	0 (0.000)
Dc (unif.)	262	93 (0.355)	91 (0.347)	166 (0.634)	18 (0.069)	8 (0.031)	23 (0.088)	71 (0.271)	160 (0.611)
Dc (div.)	262	157 (0.599)	91 (0.347)	166 (0.634)	70 (0.267)	8 (0.031)	23 (0.088)	150 (0.573)	178 (0.679)
ET	14	8 (0.571)	3 (0.214)	10 (0.714)	5 (0.357)	0 (0.000)	7 (0.500)	4 (0.285)	10 (0.714)
Total (Dc unif.)	777	416 (0.535)	389 (0.501)	333 (0.429)	182 (0.234)	153 (0.197)	86 (0.111)	482 (0.620)	259 (0.333)
Total (Dc div.)	777	480 (0.617)	389 (0.501)	333 (0.429)	234 (0.301)	153 (0.197)	86 (0.111)	561 (0.722)	277 (0.356)

870

871

872 **Table 5.** The geographical divisions of the climate class Dc.

873

Subarea	Range of longitude [degree]	Note (River Names, Area)
1:	180 W - 105 W	Alaska. Yukon River, East of McKenzie River
2:	105 W – 30 W	Eastern Canada, Canadian Prairies, Quebec
3:	30 W – 40 E	Scandinavia
4:	40 E – 60 E	Eastern Europe
5:	60 E -90 E	Western Siberia
6:	90 E – 180 E	Eastern Siberia

874

875

876