Using a Surrogate-Assisted Bayesian Framework to Calibrate the Runoff Generation Scheme in E3SM Land Model V1

Donghui Xu[1], Gautam Bisht[1], Khachik Sargsyan[2], Chang Liao[1], L. Ruby Leung[1]

[1]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA
[2]Sandia National Laboratories, Livermore, CA, United States

Correspondence to: Donghui Xu (donghui.xu@pnnl.gov)

Abstract. Runoff is a critical component of the terrestrial water cycle and Earth System Models (ESMs) are essential tools to study its spatio-temporal variability. Runoff schemes in ESMs typically include many parameters so model calibration is necessary to improve the accuracy of simulated runoff. However, runoff calibration at global scale is challenging because of the high computational cost and the lack of reliable observational datasets. In this study, we calibrated 11 runoff relevant parameters in the Energy Exascale Earth System Model (E3SM) Land Model (ELM) using a surrogate-assisted Bayesian framework. First, the Polynomial Chaos Expansion machinery with Bayesian Compressed Sensing is used to construct computationally inexpensive surrogate models for ELM-simulated runoff at 0.5° × 0.5° for 1991-2010. Error metric between the ELM simulations and the benchmark data is selected to construct the surrogates, which facilitates efficient calibration and avoids the more conventional, but challenging, construction of high-dimensional surrogates for the ELM simulated runoff. Second, the Sobol index sensitivity analysis is performed using the surrogate models to identify the most sensitive parameters, and our results show that in most regions ELM-simulated runoff is strongly sensitive to 3 of the 11 uncertain parameters. Third, a Bayesian method is used to infer the optimal values of the most sensitive parameters using an observation-based global runoff dataset as the benchmark. Our results show that model performance is significantly improved with the inferred parameter values. Although the parametric uncertainty of simulated runoff is reduced after the parameter inference, it remains comparable to the multi-model ensemble uncertainty represented by the global hydrological models in ISMIP2a. Additionally, the annual global runoff trend during the simulation period is not well constrained by the inferred parameter values, suggesting the importance of including parametric uncertainty in future runoff projections.

1 Introduction

Runoff is an essential source of freshwater resource, and its variability has profound socio-economic impacts (Hall et al., 2014; Vörösmarty et al., 2000). Flooding in wet regions during peak streamflow is among the most impactful natural hazards of all weather-related events in terms of fatalities and material costs (Doocy et al., 2013). However, higher streamflow replenishes reservoirs that help provide water for agriculture and hydropower generation, and transports nutrients to the floodplain. Drought is a form of hydrological extreme that can also result in immense damages to the ecosystem and agriculture

(Mishra and Singh, 2010). It is associated with abnormally low runoff, especially in arid and semi-arid regions. Therefore, understanding the spatial and temporal patterns of runoff is crucial for flood control, water management, crop yield, ecosystem services, etc. The runoff variability has been impacted by human-induced land use and climate change (Milly et al., 2008; Fischer and Knutti, 2016; Bosmans et al., 2017; Dai, 2013; Xu et al., 2021a), and the changes are projected to be more significant towards the end of this century (Xu et al., 2021a).

The spatial and temporal patterns of runoff and its response to climate change for water security assessments and water management are commonly studied using Earth System Models (ESMs) (Milly et al., 2002; Hirabayashi et al., 2013; Schewe et al., 2014). Current generation ESMs have large uncertainty in the simulation of runoff and its changes under future scenarios. However statistical methods have been applied recently to reduce uncertainty in model predictions (Yang et al., 2017; Gosling and Arnell, 2011; Lehner et al., 2019; Xu et al., 2021a). Uncertainties in ESMs simulation of runoff stem from uncertain model inputs, model structural uncertainty, and parametric uncertainty (Sun et al., 2013; Giuntoli et al., 2018). Input uncertainties consist of uncertainties in atmospheric forcing and land surface cover data that can be reduced by improving observation quality as more data become available. Model structural uncertainty is due to knowledge gaps or simplifications of the physical processes of the earth system. Specifically, the typical coarse resolution (~100 km) of ESMs cannot capture a few of the key physical factors that control runoff generation process such as terrain and soil variations. Downscaling methods have been developed to reduce model bias when projecting the changes of hydrological variables from the coarse resolution ESM simulation to a fine resolution (Tebaldi et al., 2005; Knutti et al., 2010; Xu et al., 2019). Recent development in the Energy Exascale Earth System Model (E3SM) has introduced a sub-grid topography based downscaling of precipitation (Tesfa et al., 2020) to understand the role of topography in hydrological processes. Over the past few decades, the land component of ESMs has continuously been improved by developing new representations of physical processes, such as implementing variable soil thickness (Brunke et al., 2016), solving the variably saturated flow in groundwater dynamics (Bisht et al., 2018), including land-river interactions (Decharme et al., 2019; Xu et al., 2021b), representing lateral subsurface flow (Swenson et al., 2019), and increasing spatial resolution (Haarsma et al., 2016). While these advances improve our understanding of the earth system, they may not lead to reduced uncertainties in future projections (Knutti and Sedláček, 2012; Lehner et al., 2020). This is because parametric uncertainty may increase as new processes are included in the model. The uncertainty in ESM simulated runoff must be reduced before reliable conclusions can be drawn regarding ESM projections of future changes in the runoff characteristics.

The parametric uncertainties in simulated runoff can be reduced by model calibration (Gupta et al., 1998). Previous studies have shown that it is possible to constrain the uncertainty of runoff by calibrating the relevant model parameters at regional scale (Ray et al., 2015; Sun et al., 2013; Sheng et al., 2017; Xie et al., 2007; Troy et al., 2008; Hou et al., 2012; Huang et al., 2013). Hou et al. (2012); Huang et al. (2013) identified the most sensitive hydrologic parameters of the Community Land Model (CLM) for simulating runoff and surface energy fluxes at a few selected watersheds and flux tower sites in the US. They found that reducing the dimensionality of uncertain parameters using sensitivity analysis speeds up the calibration processes (Huang et al., 2013). Consequently, Sun et al. (2013) successfully applied a Bayesian inversion approach to estimate

the optimal parameters to improve the performance of runoff generation in CLM. Troy et al. (2008) proposed an efficient framework to calibrate the Variable Infiltration Capacity (VIC) model for the contiguous US by interpolating the calibrated parameters from small gauged basins. While previous studies performed comprehensive model calibration of runoff at regional scales, it remains challenging to calibrate land components of ESM at global scales due to (1) the lack of runoff observations and (2) the high computational cost of running a large ensemble of global land model simulations. For (1), it is common to validate land models with streamflow (i.e., flow rate accumulated from runoff within a drainage area) observation (Li et al., 2015; Krysanova et al., 2020; Beck et al., 2017; Zhang et al., 2016), as runoff is not directly measured. However, routing the runoff to simulate streamflow at coarse resolution introduces additional uncertainties due to the representation of stream network (Wu et al., 2011; Liao et al., 2022) and river channel geometry (Andreadis et al., 2013). A recent observation-based global runoff dataset (GRUN; Ghiggi et al., 2019) provides a good benchmark for calibrating runoff generation related parameters without the needs of coupling the land model with a river routing model. For (2), tens of thousands of simulations are typically needed for parameter calibration when the parameter dimension is high, but it is not computationally feasible to run a large ensemble of ESM simulations at global scale.

The computational cost of model calibration can be significantly reduced by using an uncertainty quantification (UQ) framework that develops surrogate models of complex physical models. UQ frameworks include several steps: 1) Construction of a surrogate model that can mimic the behaviour of a physical model; 2) Identification of sensitive parameters to reduce the dimensionality of uncertain parameters; 3) Use of the parameter inference process to constrain the parametric uncertainty by comparing surrogate model prediction against observation. The surrogate modelling approach has received wide attention in hydrological applications (Razavi et al., 2012; Ivanov et al., 2021; Wang et al., 2014) to calibrate large-scale land models in terms of different hydrological processes (Gong et al., 2015; Lu et al., 2018; Müller et al., 2015; Huang et al., 2016; Ray et al., 2015; Sargsyan et al., 2014; Ricciuto et al., 2018). Multiple methods falling into the class of surrogate models include Gaussian process models, artificial neural networks, support vector machines, and polynomial chaos expansions (PCEs). In this study, we rely on PCEs as convenient machinery for uncertain input parameter representation and surrogate construction. The PCE surrogate captures the complex, non-linear behaviour of the physical model through a learned polynomial expansion. This method also provides convenient global sensitivity analysis (Dwelle et al., 2019). Further, we employ Bayesian compressive sensing (BCS) to arrive at sparse PCEs that include only polynomial terms relevant to the model, thus facilitating PCE surrogate construction in the presence of a large number of uncertain inputs and a relatively small number of model simulations (Sargsyan et al., 2014). Once the surrogate model is constructed, it replaces the expensive physical model in simulation-intensive studies such as global sensitivity analysis and parameter inference.

The objective of this work is to use the UQ framework to improve the performance of runoff generation at monthly scale and quantify the associated parametric uncertainty in the E3SM Land Model version 1 (ELM-v1) (E3SM; Golaz et al., 2019). This study is organized in the following structure. We briefly describe the runoff generation process in ELM-v1, the UQ framework, and the data used in this study in Sections 2, 3, and 4, respectively. In Section 5, we first present the validation of the surrogate models, sensitivity of simulated runoff to the uncertain parameters, dimensional reduction of uncertain

parameters, and estimation of optimal parameters. Then we evaluate the performance of ELM-simulated runoff with the optimal parameters, the runoff sensitivity to precipitation, and the changes due to the use of optimal parameters on ELM-simulated water- and energy-related variables against various benchmarks using the ILAMB package (Collier et al., 2018). Lastly, we present the simulated runoff uncertainty associated with parameters and theirs impacts on runoff trends at global scale. Section 6 discusses the limitations of this work, followed by the conclusions in Section 7.

2 E3SM Land Model

2.1 Runoff generation scheme in ELM-v1

The ELM-v1(hereafter, v1 is omitted) was developed based on the Community Land Model 4.5 (CLM4.5; Oleson et al., 2013) to understand the water availability and water cycle extremes (Leung et al., 2020). The new physical processes added in ELM to better represent the terrestrial water cycle include a variably saturated flow model (Bisht et al., 2018), a soil erosion model (Tan et al., 2020), dynamic roots (Drewniak, 2019), and a two-way coupled irrigation scheme (Zhou et al., 2020). The runoff generation in ELM is based on the simple TOPMODEL-based runoff parameterization (SIMTOP; Niu et al., 2005) in which the total runoff (R_{total}) consists of three components: surface runoff (R_{over}, e.g., saturation excess runoff), surface water runoff (R_{h2osfc}, e.g., surface water drainage from depressions/wetlands), and subsurface runoff (R_{drai}):

$$R_{total} = R_{over} + R_{h2osfc} + R_{drai} \quad \text{Eq. (1)}$$

A fraction of the flux of water reaching the soil surface (q_{liq}) generates surface runoff and the fraction is determined by the saturation fraction (f_{sat}) of the grid cell:

$$R_{over} = f_{sat} q_{liq} \quad \text{Eq. (2)}$$

$$f_{sat} = f_{max} \exp(-0.5 f_{over} z_{\nabla}) \quad \text{Eq. (3)}$$

where f_{max} represents the maximum saturation fraction for a given grid cell that is calculated with high-resolution compound topographic indices, f_{over} is a decay factor, and z_{∇} is the water table depth.

ELM includes surface water storage to represent inland/wetland surface water dynamics (Ekici et al., 2019). When the surface water storage is fully filled, surface water runoff is generated:

$$R_{h2osfc} = k_{h2osfc} f_{connected} \left(W_{sfc} - W_c\right) \frac{1}{\Delta t} \quad \text{Eq. (4)}$$

where k_{h2osfc} represents the linear storage coefficient, $f_{connected}$ is the interconnected fraction of the inundated areas, W_{sfc} is the mass of surface water, W_c is the mass of surface water when the storage is full, and Δt is the model time step. W_{sfc} is formulated as:

$$W_{sfc} = \frac{d}{2}\left(1 + \mathbf{erf}\left(\frac{d}{\sigma_{micro}\sqrt{2}}\right)\right) + \frac{\sigma_{micro}}{\sqrt{2\pi}} e^{\frac{-d^2}{2\sigma_{micro}^2}} \quad \text{Eq. (5)}$$

where **erf** represents the error function, d is the height of the surface water relative to the cell averaged elevation, and σ_{micro} is the standard deviation of the microtopographic distribution that characterizes sub-grid elevation variation. Given the surface water height from the previous equation, the surface water fraction (f_{h2osfc}) of a cell is estimated with:

$$f_{h2osfc} = \frac{1}{2}\left(1 + \mathbf{erf}\left(\frac{d}{\sigma_{micro}\sqrt{2}}\right)\right) \quad \text{Eq. (6)}$$

The inundation areas are assumed to be randomly distributed within the grid cell, and the interconnected fraction of the inundated areas can be estimated based on percolation theory:

$$f_{connected} = \begin{cases} \left(f_{h2osfc} - f_c\right)^{\mu} & if\ f_{h2osfc} > f_c \\ 0, & if\ f_{h2osfc} \leq f_c \end{cases} \quad \text{Eq. (7)}$$

where f_c is the threshold below which the inundated areas are not connected, and μ is a scaling exponent. The default parameter values in ELM of f_c and μ are 0.4 and 0.14 for all the global cells, respectively.

The subsurface runoff is parameterized as an exponential function of water table depth and includes an ice impedance factor (Θ_{ice}) to account for the reduction in the bottom drainage when ice is present in the soil (Swenson et al., 2012):

$$R_{drai} = \Theta_{ice} q_{drai,max} \exp(-f_{drai} z_{\nabla}) \quad \text{Eq. (8)}$$

$$\Theta_{ice} = 10^{-\Omega \frac{\theta_{ice}}{\theta_{sat}}} \quad \text{Eq. (9)}$$

where $q_{drai,max}$ is the maximum drainage rate, f_{drai} is the decay factor, $\frac{\theta_{ice}}{\theta_{sat}}$ represents the ice-filled fraction of the pore space for the soil under the water table, and Ω is an adjustable parameter.

We follow the work of Huang et al. (2013) in selecting uncertain parameters and their corresponding ranges (Table 1). Three additional parameters are included in this study for surface water storage drainage and impacts of ice to subsurface runoff and soil water dynamics, which represent new features in ELM compared to CLM4.0 used in Huang et al. (2013). All the parameter prior distributions are assumed to be a uniform distribution.

Table 1. Uncertain parameters' information.

Parameter	Definition	Default value	Priors
f_{max}	Maximum saturated fraction for a grid cell [-]	Derived from high-resolution DEM	$U(0.01,0.907)$
f_{over}	Decay factor for surface runoff $[m^{-1}]$	0.5	$U(0.1,5)$
f_{drai}	Decay factor for subsurface runoff $[m^{-1}]$	2.5	$U(0.1,5)$
$q_{drai,max}$	Maximum subsurface drainage rate $[kg \cdot m^{-2} \cdot s^{-1}]$	5.5×10^{-3}	$U(1 \times 10^{-6}, 1 \times 10^{-1})$

b	Clapp and Hornberger exponent [-]	Determined by plugging the soil type into the equations of means from Table 5 of Cosby et al. (1984).	Uniform distributions with $\pm 50\%$ of the means as the lower and upper bounds.
ψ_s	Saturated soil matrix potential [mm]		
K_s	Hydraulic conductivity $[mm \cdot s^{-1}]$		
θ_s	Porosity [-]		
f_c	Surface water fraction threshold for outflow [-]	0.4	$U(0.1, 0.7)$
μ	Scaling exponent for estimating connected surface water fraction [-]	0.14	$U(0.04, 0.24)$
Ω	Adjustable parameter for ice impedance factor [-]	6	$U(0.6, 60)$

2.2 Model configuration

We ran ELM globally at a spatial resolution of 0.5°×0.5° driven by the Global Soil Wetness Project forcing data set (GSWP3v1) from 1991 to 2010, featuring 3-hourly, 0.5°×0.5° global atmosphere forcing. GSWP3v1 has been dynamically downscaled and bias-corrected based on the reanalysis data of Compo et al. (2011). The default configuration of ELM was used with a 30 min time step. With the default configuration, the hydrologic representations of ELM are the same as those in CLM4.5, as new model features such as the variably saturated flow model and subgrid topography are not included. Except the uncertain parameters listed in Table 1, the default values of all other ELM parameters were used in this study.

3 Uncertainty quantification framework

A detailed derivation of the PCE-based uncertainty quantification framework and BCS method used in this work is presented in Sargsyan et al. (2014); Debusschere et al. (2016). In this study, we used the Uncertainty Quantification Toolkit (UQTk; Debusschere et al., 2004; Debusschere et al., 2016) that includes implementations of PCE construction with BCS and subsequent global sensitivity analysis. Only a brief description of constructing the PCE-based surrogate for the ELM simulations is summarized below.

3.1 Polynomial Chaos Expansion

Let $\mathcal{M}$ denote a physical model (e.g., ELM) with uncertain parameters $\boldsymbol{X}$, where $\boldsymbol{X} = [X_1, X_2, \ldots, X_D]$ and D represents the total number of uncertain parameters. In this study, the uncertain parameters $\boldsymbol{X}$ are listed in Table 1 and D is 11. A scalar Quantity of Interest (QoI), $\boldsymbol{y}$ (e.g. runoff at a specified time from a specified location), obtained using a sample of random parameters, $\boldsymbol{x}$, can be expressed as a polynomial expansion:

$$\boldsymbol{y} = \mathcal{M}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(x) \qquad \text{Eq. (10)}$$

where $\Psi_{\boldsymbol{\alpha}}$ is a polynomial and $c_{\boldsymbol{\alpha}}$ is the corresponding coefficient. In practice, $\boldsymbol{x}$ is scaled to [-1 1] from the original uncertainty input range. The polynomial expansion in Eq. (10) is written with respect to multivariate orthogonal polynomials:

$$\Psi_{\boldsymbol{\alpha}}(x) = \prod_{i=1}^{D} \Psi_{\alpha_i}(x_i) \qquad \text{Eq. (11)}$$

where $\Psi_{\alpha_i}(x_i)$ is a univariate polynomial, whose form is associated with the prior distribution of uncertain input variable X_i (e.g., Legendre polynomials are used when the input variable follows a uniform distribution), and α_i is a member of the multi-index $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_D]$, which represents the degrees of the univariate polynomial terms. Readers should refer to Dwelle et al. (2019) for details about the selection of polynomial terms and an illustration of how the multi-index is used to construct a PCE-based surrogate. In practice, Eq (11) is approximated with a truncated PCE by only selecting terms with a total degree of polynomials smaller than a certain value p (Xiu and Karniadakis, 2002; Lin and Karniadakis, 2009). This leads to a finite set $\mathcal{A}_p = (\boldsymbol{\alpha}: \sum_{i=1}^{D} \alpha_i \leq p)$ for the multi-index $\boldsymbol{\alpha}$ to take:

$$\mathcal{M}(\boldsymbol{x}) \approx \mathcal{M}^{PC}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}_p} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(x) = \sum_{j=0}^{P} c_j \Psi_j(x) \qquad \text{Eq. (12)}$$

where j represents the counter index of any possible multi-index $\boldsymbol{\alpha}$ in $\mathcal{A}_p$ in a predefined order (see details in Appendix B of Dwelle et al. (2019)). The coefficients (c_j) for the $P+1$ polynomial bases are computed using training simulations of $\mathcal{M}(\boldsymbol{x})$ (e.g., ELM) to construct the truncated PCE approximation in Eq (12). The number of the polynomial basis is determined by both the input dimension D and the total degree for truncation p (Xiu and Karniadakis, 2002):

$$P + 1 = \frac{(D+p)!}{D!\,p!}, \qquad \text{Eq. (13)}$$

The value P increases rapidly as the number of uncertainty input variables increases. For example, 11 uncertain parameters (e.g., $D = 11$) with a truncated PCE order of $p = 4$ leads to **1,365** coefficients to solve in Eq (12). It is computationally prohibitive to run 1,365 global ELM simulations, so we adopted the BCS method of Sargsyan et al. (2014) that requires a much smaller number of ELM simulations to construct a PCE-based surrogate. The BCS method computes only a sparse set of c_j to construct the surrogate of a form given by Eq (12) because not all $\Psi_j(\boldsymbol{x})$ are relevant for the given QoI (Sargsyan et al., 2014).

3.2 Global sensitivity analysis

In this study, we performed variance-based, global sensitivity analysis using Sobol indices (Sobol′, 2001). For PCE-based surrogate model, the main Sobol index, S_i, for the uncertain parameter X_i can be estimated as:

$$S_i = \frac{\sum_{j \in \Pi_i} c_j^2 \, ||\Psi_j||^2}{\sum_{j=0}^{P} c_j^2 \, ||\Psi_j||^2}, \quad \text{Eq. (14)}$$

where Π_i denotes all the indices of polynomial basis terms in Eq (12) that only involve parameter X_i, and $||\Psi_j||$ is the norm of the polynomial $\Psi_j(x)$. The main Sobol index S_i can be interpreted as the fraction of variance in the output that is associated with the uncertainty model parameter X_i only when other parameters are fixed at constant values. Similarly, one can estimate the Sobol index for any pair of parameters X_i and $X_{i'}$ to represent parameter interaction sensitivity with the coefficients c_j (Sargsyan et al., 2014).

3.3 Parameter inference

Parameter inference is used to determine a set of model parameters that reduces the error between observation and model prediction. The model inverse problem can be solved with the Bayes theorem:

$$p(\boldsymbol{X}|\boldsymbol{y}) = \frac{L(\boldsymbol{y}|\boldsymbol{X})p(\boldsymbol{X})}{p(\boldsymbol{y})}, \quad \text{Eq. (15)}$$

where $p(\boldsymbol{X}|\boldsymbol{y})$ is the posterior distribution of parameter $\boldsymbol{X}$ given observation $\boldsymbol{y}$, $L(\boldsymbol{y}|\boldsymbol{X})$ is the likelihood function, $p(\boldsymbol{X})$ represents the prior distribution of $\boldsymbol{X}$, and $p(\boldsymbol{y})$ is merely a normalizing constant for the purposes of parameter calibration. The discrepancy between the model and observations, $\boldsymbol{\epsilon} = \boldsymbol{y} - \mathcal{M}(\boldsymbol{X})$, should be included in the likelihood function. It is common to assume the error term (e.g., $\boldsymbol{\epsilon}$) follows a Gaussian distribution with vanishing mean:

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2), i = 1, 2, \dots, N \quad \text{Eq. (16)}$$

where N is the number of observations used to infer the parameters (e.g., time series of monthly runoff), and the standard deviation, σ, can be inferred from the data (see Sec 3.4). Then, the likelihood function can be written as:

$$L(\boldsymbol{y}|\boldsymbol{X}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(y_i - \mathcal{M}_i(\boldsymbol{X})\right)^2}{2\sigma^2}\right], \quad \text{Eq. (17)}$$

The logarithm of Eq (17) leads to the least-squares objective function that is used for deterministic parameter estimation in practice (Sargsyan et al., 2015):

$$\log L(\boldsymbol{y}|\boldsymbol{X}) = -\sum_{i=1}^{N} \frac{\left(y_i - \mathcal{M}_i(\boldsymbol{X})\right)^2}{2\sigma^2} - \frac{N}{2}\log(2\pi\sigma^2), \quad \text{Eq. (18)}$$

The posterior distribution in Eq (18) is difficult to compute in practice, hence we estimate it through samples obtained by the Markov Chain Monte Carlo (MCMC) method. Specifically, 1,000 iterations are used as the "burn-in" period in this study and the sampling of the posterior distribution is saved every 10 iterations. We run MCMC for 10,000 steps, resulting in 1,800

samples to construct the posterior distribution. We have employed adaptive MCMC method of Haario et al. (2001), in which the parameter space is searched according to proposal steps with a covariance that is updated on-the-fly.

3.4 Quantity of interest

In this study, the physical model $\mathcal{M}$ and the QoI $\mathcal{Y}$ correspond to ELM and runoff, respectively. The development of a surrogate model for the simulated runoff for each grid cell for each month of a 20-year simulation would require 240 (= 12 months × 20 years) PCE-based surrogates. Although developing a PCE-based surrogate is not expensive, it is computationally expensive to train 240 PCEs for each of the 70302 grid cells in the global domain. The parameter inference process for 240 PCEs for each grid cell will further increase the computational cost. We reduce the number of QoIs by training the surrogate model for the root mean square error (RMSE) between simulated runoff and observations instead of training the surrogate model to predict monthly runoff. The RMSE is given as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(R_i^{sim} - R_i^{obs}\right)^2}, \quad \text{Eq. (19)}$$

where R_i^{sim} and R_i^{obs} represent grid-level simulated total runoff and observed total runoff, respectively, for i-th month in the simulated period, and N represents the number of simulation months. Consequently, only one surrogate model is needed for each grid cell to quantify the performance of ELM in capturing the monthly runoff variation for a given uncertain parameter set. The selection of RMSE as QoI in constructing surrogate models significantly reduce the computational burden of surrogates' construction and parameter inference. We performed ELM simulations using 200 parameter sets that were randomly sampled from the range specified in Table 1. A set of 175 ELM simulations were used for training the surrogate models and the other 25 simulations were used for validating their performances. The performance of the PCE-based surrogate model can be affected by the truncated order (Dwelle et al., 2019). For each grid cell, we train the surrogate with $p = 1, 2, \ldots, 7$ separately, and picked the order that minimizes the relative norm-2 error (RE) of validation simulations:

$$RE = \frac{||RMSE_{val}^{PC} - RMSE_{val}^{\mathcal{M}}||_2}{||RMSE_{val}^{\mathcal{M}}||_2}, \quad \text{Eq. (20)}$$

where $RMSE_{val}^{PC}$ and $RMSE_{val}^{\mathcal{M}}$ represent the PCE-simulated and ELM-simulated vector of $RMSE$ of the 25 validation simulations, respectively. Then, the trained surrogate models, $RMSE^{PC}$, can be plugged into the likelihood function of Eq (18) seamlessly:

$$\log L(\boldsymbol{y}|\boldsymbol{X}) = -\frac{N \cdot (0 - RMSE^{PC})^2}{2\sigma^2} - \frac{N}{2}\log(2\pi\sigma^2), \quad \text{Eq. (21)}$$

The standard deviation of error between model simulated runoff and observation exhibits significant monthly variation. To provide a reasonable value of σ, we further assume σ in Eq (21) has a different meaning than that in Eq (18) by taking RMSE as model simulation, and 0 to as the target. Therefore, σ is approximated as the standard deviation of the difference between 0 and RMSEs, where each RMSE was calculated using simulated runoff and observation for a given training simulation. Our

estimation of σ leads to a reasonable posterior (see Sec 5.4), though other methods can also be used to estimate σ. We acknowledge that the value of σ may have an impact on the parameter posteriors, but investigating the sensitivity of σ on the posteriors is beyond the scope of this study.

Deleted: Note here σ has a different meaning than that of Eq (18), since RMSE is selected as QoI, and the objective is to minimize RMSE (e.g., 0 is regarded as the target in Eq (21)). Therefore, σ refers to the standard deviation of difference between RMSE and 0. We estimated σ as the standard deviation of RMSEs between simulated runoff and GRUN from all the training simulations

3.5 Calibration procedure

In summary, the following procedures were implemented to determine the optimal parameter values and their joint probability distribution:

1. Run ELM with 200 parameter sets randomly sampled with the range specified in Table 1. Construct PCE-based surrogate models to mimic the RMSE between the ELM and GRUN runoff dataset with 175 simulations and validate the performance of the surrogate models with the other 25 simulations.
2. Implement sensitivity analysis with the surrogate models to reduce parameter dimensionality for calibration by ignoring the parameters with negligible Sobol index (e.g., less than 0.05).
3. Estimate the Bayesian posterior of the most sensitive parameters for each grid through MCMC process with the runoff dataset of Ghiggi et al. (2019).
4. It has shown small surrogate error can result in significant deviation of the inferred parameter (Laloy and Jacques, 2019). To further search the optimal parameters and construct the runoff posterior uncertainty, we ran ELM simulations with additional 100 samples from the posteriors of the 3 most sensitive parameters for all global grid cells and default values were used for less sensitive parameters.
5. The parameters with the minimum RMSE between simulations and reference runoff data from the 100 ELM simulations were determined as the optimal parameter value for each grid cell.

4 Data

4.1 Observation-based runoff data

The $0.5° \times 0.5°$ global observed-based runoff (GRUN) dataset of Ghiggi et al. (2019) was used in this study as the observation within the calibration framework for parameter inference. The GRUN dataset was generated from a trained random forests (RF) model (Breiman, 2001) that used precipitation and near-surface temperature to predict monthly runoff. The training runoff data were derived from Global Streamflow Indices and Metadata Archive (GSIM; Gudmundsson et al., 2018; Do et al., 2018), and only the gauges with contributing area comparable to cell area of $0.5° \times 0.5°$ were used. GSWP3 atmospheric forcing was used for training and reconstruction of the monthly global runoff.

4.2 Model benchmarks

The ILAMB package (Collier et al., 2018) was used to evaluate the simulated water and energy cycles from the calibrated ELM against various benchmarks. Specifically, a gridded energy flux data (FLUXCOM; Jung et al., 2019) that was

generated by machine learning with flux tower measurements was used to evaluate latent and sensible heat fluxes; Global Land Evaporation Amsterdam Model version 3 (GLEAMv3; Martens et al., 2017) product was used to evaluate global ET; Gravity Recovery And Climate Experiment (GRACE; Kim et al., 2009) data were used to evaluate terrestrial water storage anomaly (TWSA). Details about ILAMB can be found at https://www.ilamb.org.

The inter-Sectoral Impact Model Intercomparison Project (ISIMIP) archived simulations from multiple global hydrological models and land surface model forced by the same atmosphere forcings (Warszawski et al., 2014). We used 13 available models from the second phase water sector (ISIMIP2a; Gosling et al., 2019) to provide a benchmark for the uncertainty of annual runoff magnitude and trend. Only the models in ISIMPI2a that were driven by the GSWP3 forcing without accounting for human activity impacts were selected here to be consistent with ELM's configuration.

4.3 Evaluation metrics

Two metrics were used to evaluate ELM's performance of simulating runoff at monthly scale with calibrated parameters, including the Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970) and the Kling-Gupta Efficiency (KGE; Gupta et al., 2009) which are computed as

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(R_i^{sim} - R_i^{obs}\right)^2}{\sum_{i=1}^{N}\left(R_i^{obs} - \mu_{obs}\right)^2}, \quad \text{Eq. (22)}$$

$$KGE = 1 - \sqrt{(\rho - 1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2}, \quad \text{Eq. (23)}$$

where R_i^{sim} and R_i^{obs} represent cell-level simulated total runoff and observed total runoff, respectively, for the i-th month, μ_{obs} is the corresponding averaged observation, ρ is the correlation coefficient between simulation and observation, σ_{sim} is the standard deviation in simulations, σ_{obs} is the standard deviation in observations, and μ_{sim} is the simulation mean. Both NSE and KGE vary from $-\infty$ to 1, and a perfect model performance is indicated by NSE = 1 and KGE = 1. NSE < 0 and KGE < -0.41 mean the simulations are worse estimates than the mean of observations, indicating a bad model performance (Knoben et al., 2019).

The sensitivity of runoff to precipitation is a critical aspect for runoff simulation evaluation, considering changes in precipitation will continue in the future (Trenberth, 2011). Therefore, we evaluated the sensitivity of runoff to the precipitation anomalies with the calibrated parameters. The sensitivity was quantified by the slope of linear regression (β) between runoff anomalies (ΔR) and precipitation anomalies (ΔP):

$$\Delta R = \beta \Delta P + \epsilon, \quad \text{Eq. (24)}$$

The interception $\epsilon \approx 0$, implies the mean runoff is related to mean precipitation.

We also evaluated the impacts of parameters on the runoff trend. Specifically, the magnitude of runoff trend was calculated with Sen's slope (Sen, 1968), which is nonparametric and not sensitive to the outliers. Then, Mann-Kendall test was used to determine if the trend is significant or not at confidence level $\alpha = 0.05$.

5 Results

5.1 Refinement of f_{drai} for arid region

The proposed prior for f_{drai} is not suitable for all the climate regions, such as simulations with the full the range of f_{drai} defined in Table 1 results in unrealistic runoff for arid regions. For example, the simulated runoff from an example grid cell with $f_{drai} < 0.4$ shows higher magnitudes and lower variabilities compared to simulations with $f_{drai} \geq 0.4$ (Figure 1a). Lower f_{drai} can lead to unrealistically high subsurface runoff according to the exponential function of baseflow drainage (Eq (8)) for the arid regions, where the precipitation is not enough to maintain the water table at a reasonable level. Such simulations with $f_{drai} < 0.4$ result in high nonlinearity in the simulated runoff, and hence the PCE-based surrogate model cannot capture the model behaviours (Figure 1b). The performance of surrogate models is improved by constraining the lower bound of f_{drai} to 0.4 (Figure 1c). Therefore, f_{drai} is refined as [0.4 5] for areas that are identified as arid climate in the Köppen climate classification (Figure S1), and [0.1 5] is used in other regions.

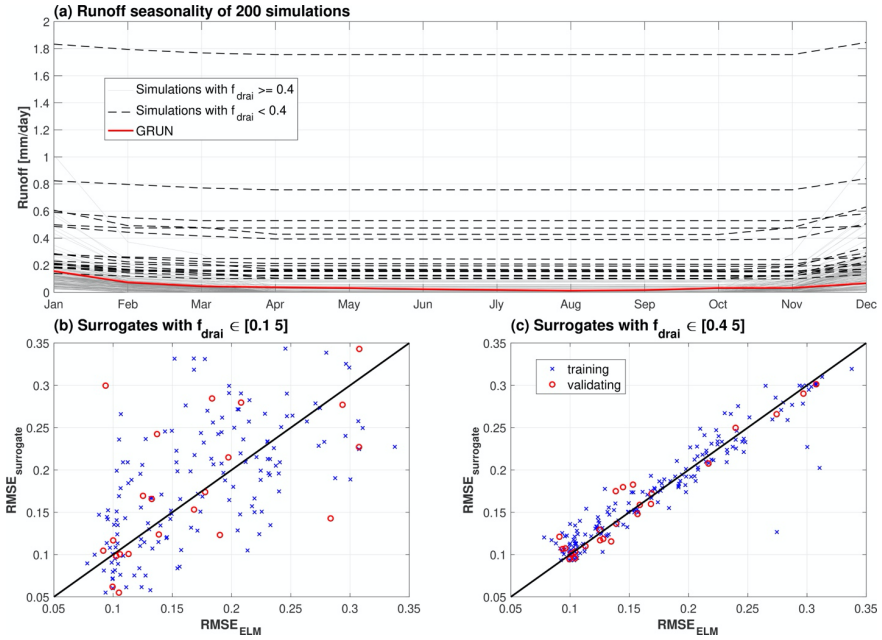
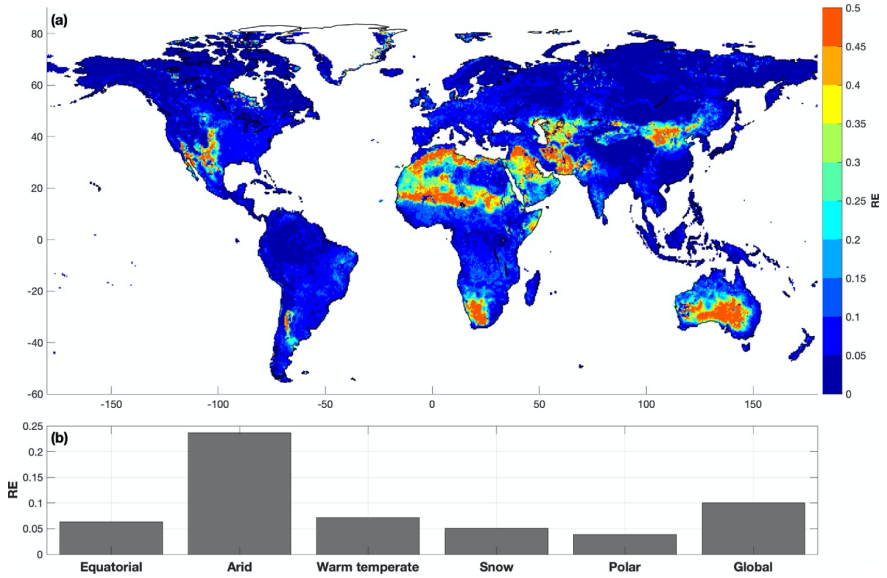


Figure 1. Validation of surrogate performance for an example grid cell from arid region. Subplot (a) shows runoff seasonality from all the 200 simulations with samples from parameter priors. Subplot (b) shows the validation of the surrogate model trained with original ranges of f_{drai} given in Table 1 in main text. Subplot (c) shows the validation of the surrogate model trained with constrained f_{drai}.

305

5.2 Validation of surrogate models

The PCE-based surrogate models can mimic the variations of RMSE between ELM-simulated runoff and the GRUN runoff with the truncated order determined in Figure S2. Specifically, the surrogate models exhibit good performance for the validation simulations with $RE < 0.1$ for 70% of the global domain (Figure 2a). The global averaged RE of surrogate models 310 for the validation simulations is around 0.1, with the largest error over the arid regions (Figure 2b). While 41% of the arid region shows an acceptable performance in the surrogate models when narrowing the range of f_{drai} with RE less than 0.15, the RE of other arid areas remain high (Figure 2a). Additional simulations were performed to investigate if the lower performance of surrogate models for arid regions is due to insufficient number of training simulations. We randomly selected 20 grid cells from the arid region and ran 2,000 ELM simulations with random samples from the parameter priors as 315 summarized in Table 1. The RE of surrogate models for the 20 grid cells remained large (e.g., Re > 0.2) even as the number of training simulations were increased (Figure S3). Thus, the lower performance of surrogate models over the arid regions is not dependent on the number of training simulations.



320 **Figure 2.** Relative norm-2 error of the surrogate models for the validation simulations. Subplot (a) shows the spatial distribution of the errors, and subplot (b) shows the average errors for the grid cells in each climate defined by Köppen climate classification.

Most surrogate models with large RE are in extremely dry arid region; for example, RE > 0.2 are mainly from grid 325 cells with annual runoff < 0.05 mm/day (Figure 3). The RE of surrogate models tends to decrease for areas with relatively higher annual runoff that are still from arid region (annual runoff < 0.5 mm/day in Figure 3). However, the runoff uncertainty in extremely dry areas will have negligible impact on the global water cycle. Surrogate model with RE > 0.15 is considered as not sufficiently accurate and such grid cells are excluded in the sensitivity analysis presented next.

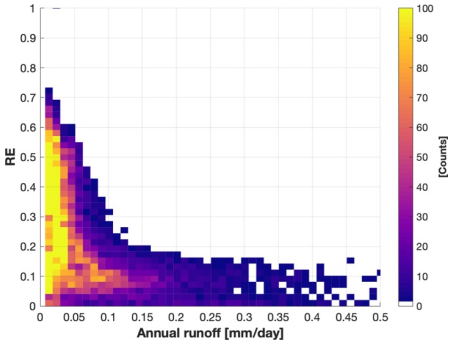


Figure 3. Plot of relative norm-2 error (RE) of surrogate models for the validation simulations vs. averaged annual runoff magnitude with all the grid cells from arid region.

5.3 Global sensitivity analysis

The most significant ELM parameters identified for runoff generation are $f_{over}, f_{drai}, \psi_s, f_c$, and Ω based on the spatial distribution of the main Sobol indices (Figure 4), while the other 6 parameters have negligible contributions to the runoff variations (Figure S4). In equatorial regions, f_{drai} and f_{over} are equally sensitive and account for 39% and 36% of the average runoff variations, respectively, as indicated by the size of circles in Figure 5a; while ψ_s is the secondary sensitive parameter. For the arid regions, f_{drai} is the most sensitive parameter, and f_{over}, f_c, K_s, and ψ_s are secondary sensitive parameters with a similar value for the main Sobol indices (Figure 5b). Although other parameters show negligible main Sobol indices for arid regions, they have shown sensitivities when interacting with each other as denoted by the thickness of the lines between each pair of parameters in Figure 5b. The complex joint sensitivity results in high nonlinearity in the runoff variations, representing a possible reason for the poor performance of PCE for arid regions. The most significant uncertain parameters for the warm temperate region are the same as those for the equatorial region (Figure 5c). Snow and polar climates have similar sensitivity pattern, with f_c and Ω are the two most important uncertain parameters (Figure 5d and e). In colder region, the contribution of surface water storage drainage, which is controlled by f_c, is large to the total runoff because of prominent surface water areas (Pekel et al., 2016). The hydraulic conductivity and groundwater drainage when ice is present in the soil is controlled by Ω, which has a significant impact on runoff generation process when the soil is partial or fully frozen. The surface water storage and ice impedance factor, which were not included in the version of the model used in previous study (Hou et al., 2012; Huang et al., 2013), are found to be the most sensitive parameters in cold regions. Besides arid region, other regions show smaller sensitivities to parameter interactions.

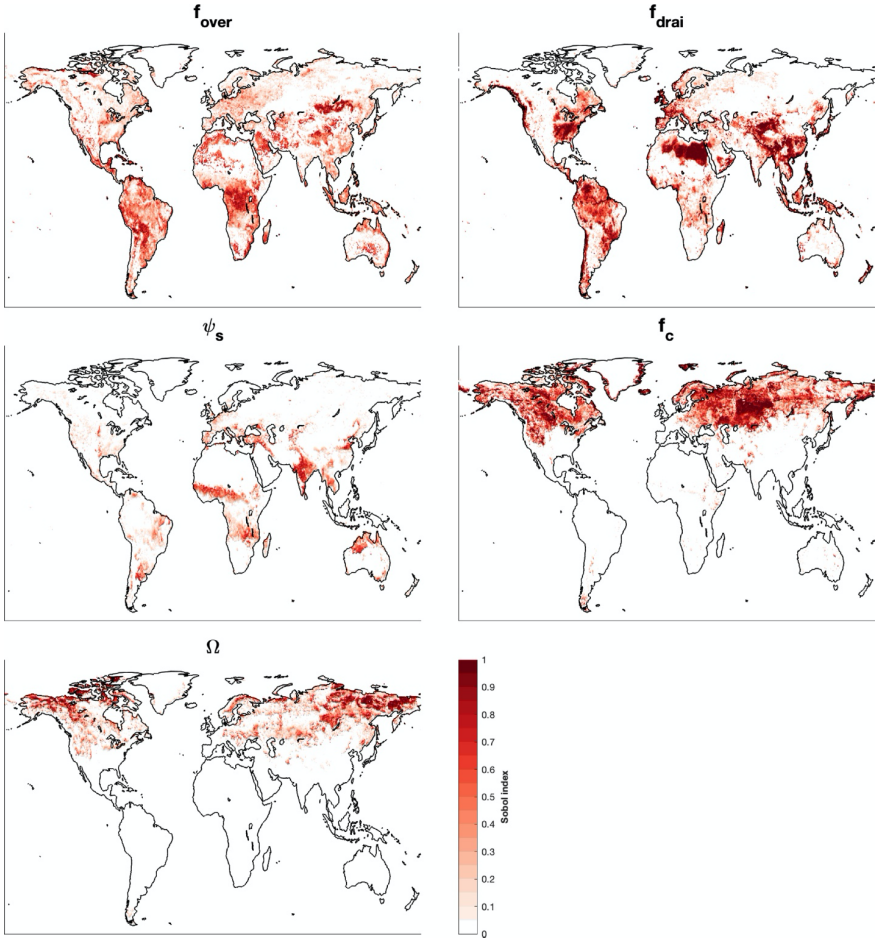


Figure 4. Spatial distribution of main Sobol index for the sensitive parameters.

355

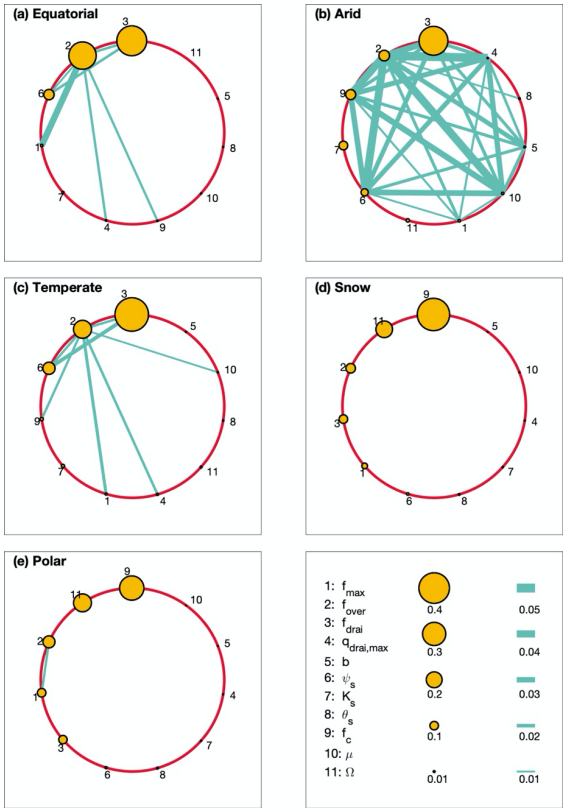


Figure 5. Averaged main Sobol index and joint Sobol index for different climates defined by Köppen climate classification. Only the cells with the relative norm-2 errors of PCE-based surrogate models for validating simulations less than 0.15 are used in estimating the averaged sensitivity for each climate region. The size of the circles and thickness of the lines are proportional to main Sobol index and joint Sobol index, respectively. The legend in the right bottom subplot shows the Sobol index for the corresponding size of circle and thickness of line.

5.4 Parameter dimensionality reduction

The ELM simulated runoff is significantly sensitive to three or fewer parameters with Sobol index > 0.05 for 81.3% of the total grid cells (Figure S5). Therefore, we sampled only the three most sensitive parameters in each grid cell in the MCMC process to perform parameter inference as mentioned in Sec 3.3. The posteriors of the three calibrated parameters ($f_{drai}, f_{over}, \psi_s$) at an example grid cell ($56.75°W, 11.25°S$) are much more constrained than the priors after the MCMC simulation with the surrogate model (Figure 6a, b, and c). The third parameter, ψ_s, has a relatively wider posterior than the first two parameters because its sensitivity is much smaller (e.g., Sobol index = 0.08). The Gelman-Rubin R statistic of Gelman and Rubin (1992) computed with 5 MCMC chains (after burn-in period) is 1.002, 1.004, 1.003 for f_{drai}, f_{over}, and ψ_s, respectively, suggesting our MCMC simulation has converged (see convergence curve in Figure S6). ELM simulations with a large number of samples from parameter priors are needed to identify the optimal parameter that minimizes RMSE, for example, 10,000 surrogate simulations are used to find the parameters that yield $RMSE = 1$ (Figure 6d). In contrast, due to the reduced parameter dimensionality and narrowed range, much fewer samples (e.g., 100) are needed to find the better parameter values (e.g., corresponding to $RMSE < 1$) when they are sampled from the parameter posteriors (Figure 6d). The spatial distribution of the parameter values at 5% and 95% of the posteriors is shown in Figure S7 and Figure S8, respectively.

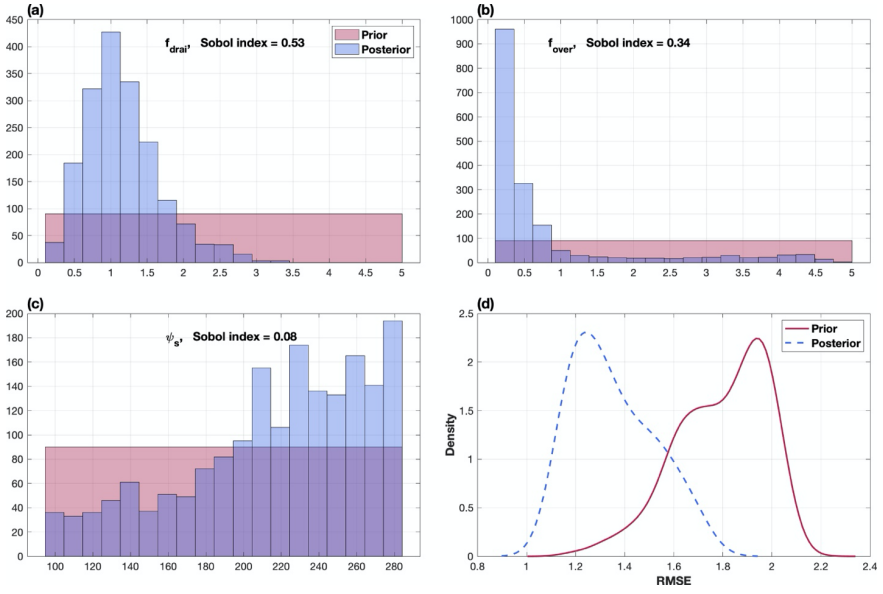


Figure 6. Posteriors of (a). f_{drai}, (b). f_{over}, and (c). ψ_s from parameter inference process at an example grid cell. Subplot (d) shows the probability density function (PDF) of RMSE evaluated with surrogate models forced by 100 samples from parameter posteriors and 10,000 samples from parameter priors.

5.5 Optimal parameter values

The procedure described in Sec 3.5 is used to find the optimal parameter values for the three most sensitive parameters for each grid cell. For the grid cells with RE > 0.15 for surrogate models, the optimal parameter value is determined from the training and validation simulations (e.g., 200 simulations with random parameter values from priors) that yield minimum RMSE. The optimal parameter values show clear regional patterns (Figure 7). Specifically, the optimal f_{over} tends to be lower than the default value for the equatorial and partial snow areas (Figure 7a). The optimal f_{over} is found to be higher than the default value for the arid areas, while it is around the default value on average for the warm temperate areas (Figure 7a). For the same water table depth, lower f_{over} leads to higher saturation fraction (Eq. 3), that in turn leads to larger surface runoff (Eq. 2). The calibrated f_{drai} is lower than the default value for both equatorial and arid regions (Figure 7b). The optimal f_{drai} for warm temperate areas show different patterns, with higher values over eastern US and Europe, but lower values over South-eastern China. The generation of subsurface runoff depends on f_{drai} (Eq. 8) with lower f_{drai} leading to larger subsurface runoff. ψ_s affects the runoff generation through its impact on soil water movement, such as the soil water flux is larger at saturation with higher ψ_s. As shown in Figure 7c, higher ψ_s are needed to minimize the RMSE for all regions that show sensitivity to this parameter, except some grid cells from polar area. Over the high latitudes of Northern Hemisphere, higher f_c and lower Ω are found in the optimal parameters (Figure 7d, e). The surface water storage can store more water at higher f_c by reducing surface water runoff (Eq. 4, 7), thus leading to a lower and delay peak runoff than the default values. Further, the lower Ω values have less impacts of ice on hydraulic conductivity (Eq. 7.89 in Olson et al. (2016)) and drainage (Eq. 9), which leads to higher runoff for the winter seasons.

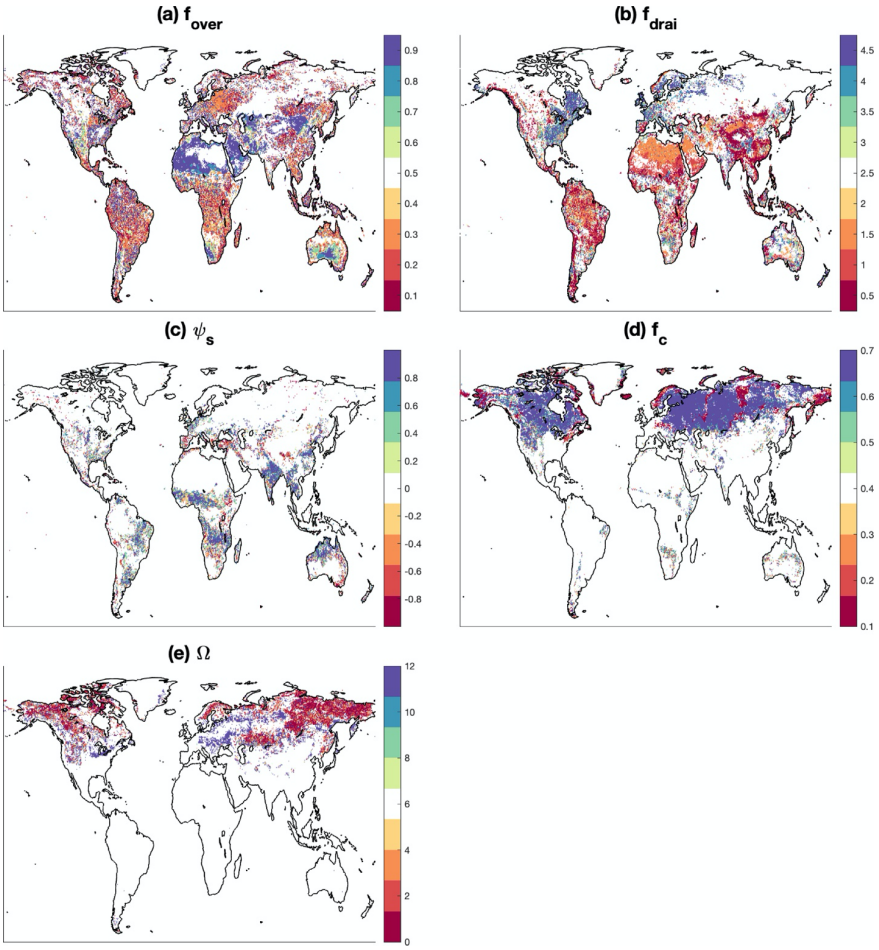


Figure 7. Optimal values for the sensitive parameters. The default values for the parameters are defined at the midpoint of the colormap. There are no certainty bounds for ψ_s from different grid cells because it is determined by the soil properties. Therefore, the values of ψ_s are scaled to [-1, 1] in subplot (c) for each grid cell with the corresponding upper bound ($\psi_{s,max}$) and lower bound ($\psi_{s,min}$): $\frac{2}{\psi_{s,max}-\psi_{s,min}}\psi_s - \frac{\psi_{s,max}+\psi_{s,min}}{\psi_{s,max}-\psi_{s,min}}$.

5.6 Evaluation of ELM with the optimal parameters

The ELM-simulated runoff with the optimal parameter values shows improved skills of capturing the spatiotemporal variation of monthly runoff at global scale with higher NSE and KGE compared to the simulation with default parameter values (Figure 8). Specifically, the median of NSE and KGE from all global grid cells increases from -0.88 and -0.05 to 0.06 and 0.31, respectively. Over the western US coast, southeast and Midwest of US, western Europe, equatorial areas, the performance of the calibrated ELM is better with NSE > 0.5 and KGE > 0.7. While the performance of other areas (e.g., western US, Sahara and Arabian desert, central and eastern Asia, and partial high latitude regions) is improved compared to simulations with the default parameter values, the NSE and KGE still have negative values. The higher model errors in those regions cannot be resolved by calibration as 1) the simulation resolution is too coarse to resolve the topographic impacts (Chegwidden et al., 2020); 2) the snow melting processes are not calibrated in this study, and the onset of snowmelt in ELM is poorly represented (Toure et al., 2018); and 3) hydrology of arid areas is not well understood (Pilgrim et al., 1988). Except for the calibration period, ELM with the optimal parameters also shows an improved performance in runoff simulation for another period (2011-2013) (Figure S9).

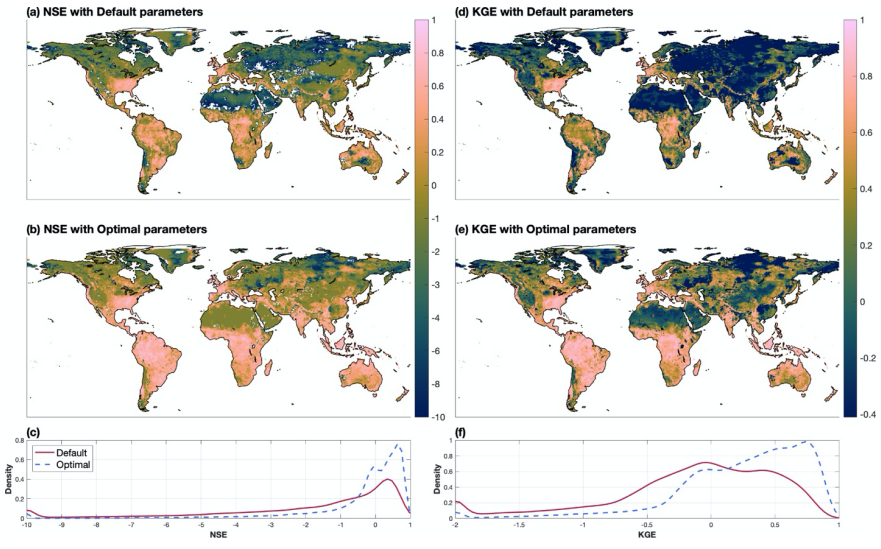


Figure 8. Evaluation of simulated monthly runoff for 1991-2010 at grid level with default and optimal parameters. Subplot (a) and (b) show the NSE metrics between the GRUN runoff and simulated runoff with default and optimal parameter, respectively. Subplot (c) shows the comparison of the probability density function (PDF) of NSE metrics from all the global grid cells. Subplot (d), (e), and (f) illustrate the evolution with KGE metric.

Compared to the reference runoff (Figure 9a), the ELM simulation with default parameter values tends to overestimate the sensitivity of runoff to precipitation (β in Eq (24)) for the equatorial and arid regions, but underestimates β in the warm temperate regions, such as eastern US, China, and eastern coasts of Australia (Figure 9b). The simulation with optimal parameter values is able to more accurately estimate β than the simulation with default parameter values with improved spatial correlation coefficient from 0.22 to 0.56, and lower RMSE from 1.22 to 0.65 (Figure 9c). However, some significant discrepancy of β still exists in the simulation with optimal parameter values (e.g., eastern China), implying the sensitivity is not well constrained in ELM for certain regions even after model calibration.

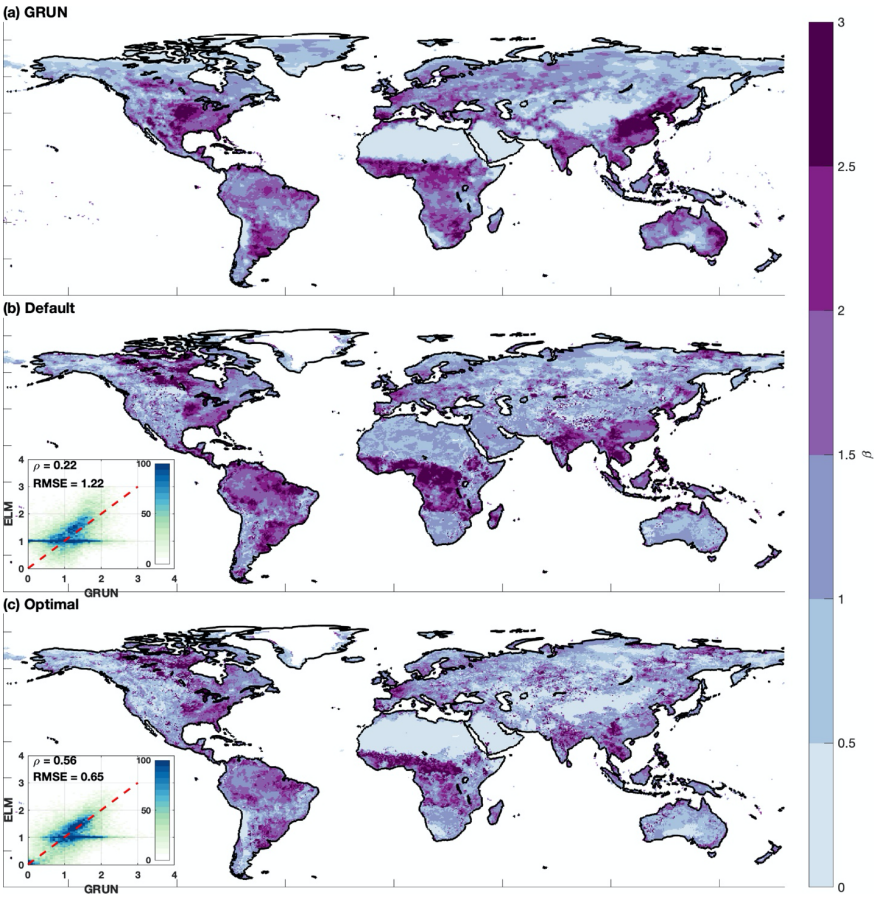


Figure 9. Sensitivity of runoff to precipitation (β) estimated from (a). GRUN runoff dataset, (b). ELM simulation with default parameter, and (c) ELM simulation with optimal parameter. The inserts show the scatter plots with density for cell-to-cell comparison of β between GRUN and ELM simulations.

440

According to the evaluation with the ILAMB package, ELM shows similar performance in simulating other variables (e.g., latent heat flux, sensible heat flux, ET, and TWSA) with optimal parameter values compared to use of default parameter values

(Table 2). However, both the default and optimal simulations fail to capture the spatial variation of TWSA with the spatial distribution score less than 0.05. This is because the coarse resolution (e.g., several hundred km) of GRACE product (Seyoum et al., 2019) cannot resolve the spatial variability of TWSA for our model resolution.

Table 2. ILAMB benchmark scores for latent heat flux, sensible heat flux, evapotranspiration, and terrestrial water storage anomaly with default and optimized parameters in ELM. Description of each score metric can be found in http://redwood.ess.uci.edu/CMIP6_bnchmrk1_9_8/.

Variable	Data source	Parameter	Bias Score	RMSE Score	Seasonal Cycle Score	Spatial Distribution Score	Overall Score
Latent Heat Flux	FLUXCOM	Default	0.740	0.680	0.910	0.993	0.800
		Optimal	0.730	0.677	0.909	0.992	0.797
Sensible Heat Flux	FLUXCOM	Default	0.682	0.643	0.932	0.940	0.768
		Optimal	0.680	0.636	0.932	0.933	0.763
ET	GLEAM3.3	Default	0.714	0.675	0.870	0.971	0.781
		Optimal	0.705	0.672	0.873	0.967	0.778
TWSA	GRACE	Default	0.901	0.554	0.818	0.003	0.566
		Optimal	0.900	0.545	0.817	0.004	0.562

5.7 Parametric uncertainty

The parameter priors listed in Table 1 result in significant uncertainties in the total runoff, with global average annual runoff for 1991-2010 varying from 30,999 - 76,496 $[km^3/yr]$ (Figure 10a). After parameter inference, the uncertainty of the runoff constructed using simulations with parameter posteriors is constrained to 35,389 - 49,741 $[km^3/yr]$. The constrained annual runoff uncertainty captures the reference runoff (38,443 $[km^3/yr]$) and is consistent with previous global runoff studies (Schellekens et al., 2017; Rodell et al., 2015; Clark et al., 2015; Haddeland et al., 2011). The simulation with the optimal parameter values yields an averaged global annual runoff of 42,156 $[km^3/yr]$, overestimating the reference runoff by 9.6%. The overestimation is mainly from Amazon, Asia, and Eastern Europe (Figure 10b), and Ghiggi et al. (2019) reported a similar spatial bias pattern between global hydrological model simulations in ISIMP2a. The simulation with the default parameter values shows smaller biases in terms of annual runoff magnitude as compared to the reference runoff data, with an overestimation of 5.3% on average. However, the smaller biases of annual runoff with the default parameters are because of cancelling out of the monthly errors to some extent. For example, the default parameters tend to overestimate the runoff during

the wet periods but underestimate the runoff during the dry periods in Amazon basin (Figure S10a). While the default simulation shows higher RMSE and lower NSE at monthly scale, it yields smaller biases at annual scale than the optimal parameter (Figure S10b). Therefore, the calibrated simulation shows better performance in capturing the spatiotemporal variability (higher NSE and KGE in Figure 8b and e), but it doesn't lead to a reduced bias at annual scale. We further acknowledge that 200 simulations with 11 random parameters may not be sufficient to capture the full variations of simulated runoff.

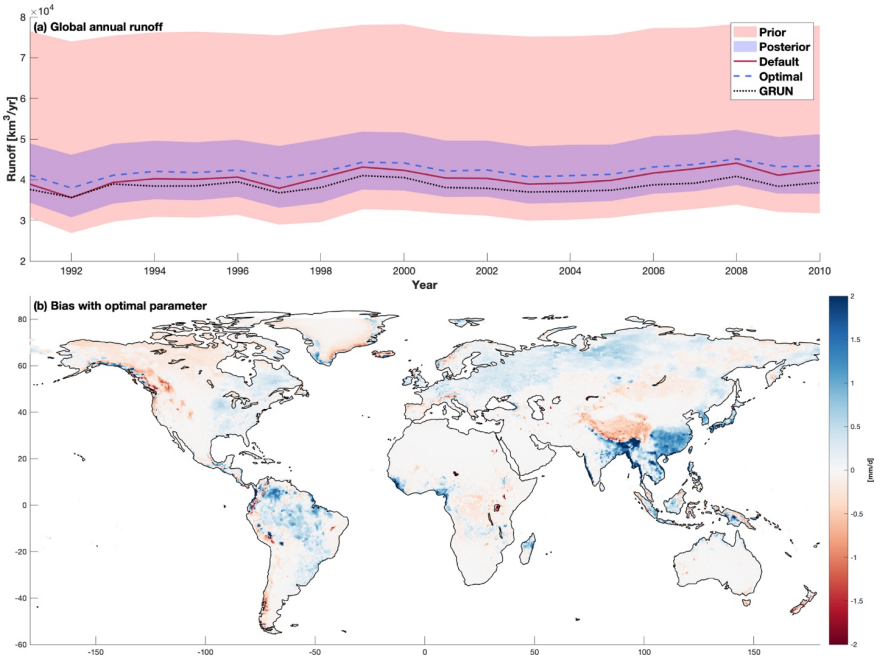


Figure 10. (a). Annual global runoff from default ELM simulation, optimal ELM simulation, and GRUN runoff dataset for the simulation period (1991-2010). The red and blue shade areas represent the uncertainties constructed from the simulations with parameter sampled on priors and posteriors, respectively. Subplot (b) shows the absolute difference of annual average runoff between ELM simulation with optimal parameter and GRUN runoff data.

The runoff uncertainties associated with parameters are constrained significantly with the parameter posteriors at basin scale as well (Figure 11). Noticeably, the posterior uncertainty of annual runoff is larger over the equatorial regions (e.g., Parana, Amazon, Godavari, Congo) than other regions. The simulation with optimal parameter values yields larger overestimation of total runoff compared with the simulation using the default parameter values for the selected basins, except Mississippi, Godavari, and Loire basin (Table S1). The reason for the overestimations is that the optimal parameters are determined by maximizing NSE at monthly scale, which cannot ensure the annual runoff to be appropriately constrained. There exist significant discrepancies between simulations and GRUN for basins located at high latitudes (e.g., Mackenzie, Volga, Ob, Yenisey, and Lena) even when the posterior uncertainties are considered (Figure 11), highlighting the importance of snow-melting processes in snow-dominated regions. However, the large difference between ELM and the reference runoff in Yangtze river basin may be caused by the bias of the reference runoff since previous study reported annual discharge to be around 900 $[km^3/yr]$ (Yang et al., 2015).

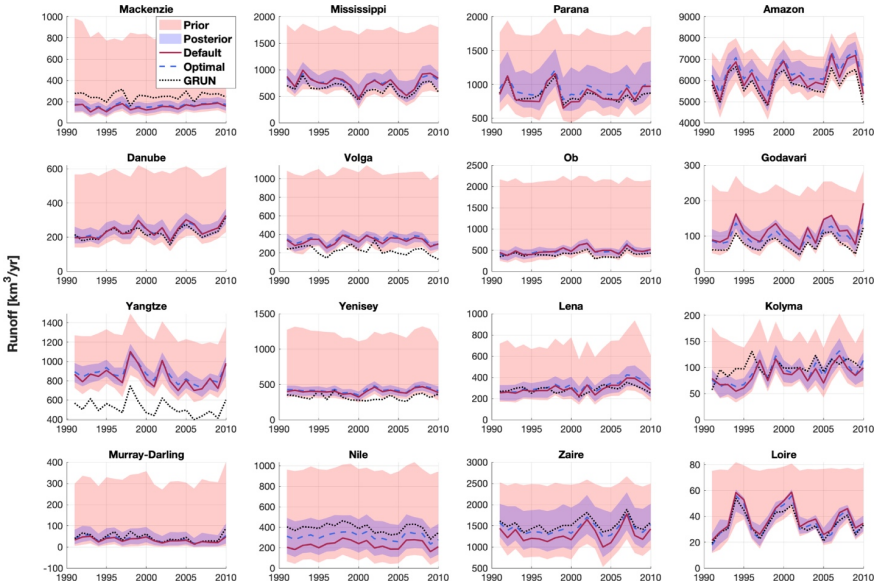


Figure 11. Annual runoff at basin scale from default ELM simulation, optimal ELM simulation, and GRUN runoff dataset for the simulation period (1991-2010). The red and blue shade areas represent the uncertainties constructed from the simulations with parameter sampled on priors and posteriors, respectively.

Despite being constrained by the parameter inference process, the parametric uncertainty of ELM-simulated annual runoff is considerable. Specifically, the posterior uncertainty of global runoff simulated by ELM is comparable to that of the multi-model ensemble constructed with the 13 global hydrological models from ISIMIP2a (Figure 12a). The parametric uncertainty affects not only the magnitude of global runoff but also the trend for the simulation period, during which a rapid increase of temperature has occurred (Figure S11a). The Sen's slope (Sen, 1968) for the reference runoff data is found to be 54.7 $[km^3/yr]$, but this increasing trend is not significant according to the Mann-Kendall test (Figure 12b). Other studies also reported no significant changes in the global runoff with observed streamflow data (Alkama et al., 2013; Dai et al., 2009; Milliman et al., 2008; Alkama et al., 2011). However, the default and calibrated ELM simulations yielded the Sen's slope to be 188.9 $[km^3/yr]$ and 133.8 $[km^3/yr]$, respectively. Although the Sen's slope is reduced with the optimal parameters, the increasing trend remains significant. Likewise, all the other global hydrological models of ISIMIP2a exhibit significant increasing trend in the annual runoff, with the Sens' slope varying from 93 $[km^3/yr]$ to 272 $[km^3/yr]$ (Figure 12b). Considering the GRUN dataset and all model simulations are forced by the same atmosphere forcing (i.e., GSWP3), the differences of the global runoff trends can be attributed to the model structural/parametric uncertainty. We note that there exists a significant trend in GSWP3 precipitation at global scale, with an increase of 246.1 $[km^3/yr]$ during the simulation period (Figure S11b). But it remains unclear how the runoff responds to the increase of precipitation at global scale because the concurrent increased temperature (Figure S11a) leads to more ET, which can potentially balance the increased precipitation to some extent. The inconsistency of the global runoff trend between the model simulations and observation-based data can be caused by uncertainties of different sources. For example, the accuracy of GRUN is limited by the coverage of the streamflow gauges, as over half of the global areas are ungagged (Alkama et al., 2013). The model parametric uncertainty is significant, as ELM simulations with parameters posteriors show a wide range of annual runoff trend, from no trend to significant increasing trend (Figure 12b). This highlights the necessities of including parametric uncertainty in future runoff projections since runoff trend is not well constrained even if the model performance in the control period is improved.

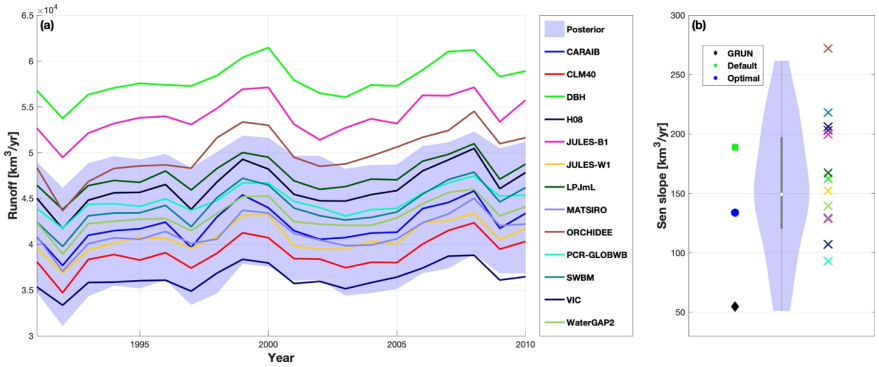


Figure 12. (a) Annual global runoff from 13 global hydrological models participated in ISIMIP2a, and ELM simulated runoff uncertainty constructed using simulations with parameter posteriors. (b) Sen's slope for the global annual runoff for the GRUN runoff dataset and simulations. The voilinplot (Hintze and Nelson, 1998) are generated with Sen's slope of ELM simulations with parameter posteriors, and the white point is the median values and the grey line represents range of the 25% - 75% percentile. The matlab fucntion of Bechtold (2016) was used to create the voilinplot. The cross signs are the Sen's slopes estimated from the ISIMIP2a model simulations.

6 Limitations

We note there can be other better choices of priors for the parameter whose range covers several orders of magnitude. For example, sampling $q_{drai,max}$ on a uniform distribution (e.g., $U[10^{-6}, 10^{-1}]$ results in fewer prior samples with values less than 10^{-2}. A log-transformed uniform distribution can be a good alternative to guarantee enough samples over each range of the desired values. Using a log-transformed uniform distribution for $q_{drai,max}$ prior doesn't impact our results significantly because the simulated runoff is not sensitive to $q_{drai,max}$ (Figure S4), and more samples over smaller values of $q_{drai,max}$ will not lead to more variation in runoff. However, careful selection of prior distributions can be important for sensitive parameters in future application of surrogate-assisted calibration framework.

By using RMSE instead of the simulated runoff as the QoI, only one PCE-based surrogate model is constructed for each grid cell to represent the ELM performance of simulating monthly runoff time series. Although selecting RMSE as QoI significantly reduces the computational burden of surrogates' construction and parameter inference, the corresponding surrogates cannot be used to estimate posterior uncertainty of physical model outputs. For example, we still need to run ELM simulations after the parameter inference to construct the runoff posterior uncertainty (Sec 3.5). Additionally, the objective of this study is to minimize RMSE at monthly scale, hence an improved model performance at annual scale is not guaranteed.

Including both monthly and annual performance metrices in objective function may balance the performance at different temporal scales. However, only one objective is accepted in the uncertainty quantification framed used in this study.

We further acknowledge the poor performance of PCE-based surrogate model in capturing the ELM-simulated runoff over extremely arid regions (Figure 2 and 3). This can be attributed to the limitation of polynomial-based surrogate models in capturing highly non-smooth or strongly nonlinear relationships. Machine learning algorithms (Dagon et al., 2020) and deep neural network (Tsai et al., 2021) are alternative techniques for surrogate modelling, which are better at capturing non-smooth or nonlinear functions, but future research is needed to investigate the capability.

The calibrated parameters have a significant impact on baseflow index, which is the ratio between subsurface runoff and total runoff. For example, the baseflow of Amazon basin with default and optimal parameters are 0.53 and 0.70, respectively (Figure S12). Mortatti et al. (1997) reported the baseflow index of Amazon basin to be 0.70 with isotopic tracer method, which is consistent with our simulation with optimal parameter values. However, accurate separation of surface runoff and subsurface runoff over other regions is not guaranteed, though the total runoff has been calibrated to match with the reference runoff dataset. The global baseflow index dataset of Beck et al. (2013) that derived from observed streamflow provides us the benchmark for evaluating the baseflow index simulated in ELM. Constraining the baseflow index during the ELM validation and calibration study will be investigated in the future.

We further note that uncertainty in the reference runoff data of GRUN used in the parameter inference is inevitable. While Ghiggi et al. (2019) found that GRUN outperformed other global hydrological models and multi-model ensemble, lower accuracy over mountainous regions due to the coarse resolution has been reported. Additionally, the irrigation and water management impacts on streamflow was included for some regions during the training process of GRUN (Ghiggi et al., 2019), but irrigation and water management are not active in ELM configuration used in this study. This inconsistency may explain the significant overestimation of ELM simulated runoff compared to GRUN for certain regions, for example, Yangtze River basin (Figure 10).

Another limitation of this study is that the snow melting processes were not calibrated. A poor representation of snow melting process can result in poor skill of runoff generation in snow-dominant regions, where snowmelt is an important contribution to runoff (Jenicek and Ledvinka, 2020). This could explain the low performance (i.e., negative NSE) of calibrated ELM over the Northern Hemisphere high latitudes and mountainous regions. However, including parameterizations of snow processes such as snow albedo, solar absorption, and snow aging (Lawrence et al., 2011) can introduce more uncertain parameters, which will make calibration more challenging (Huang et al., 2013). In the future, a dedicated calibration on the snow melting process is needed to improve the runoff generation in snow-dominated regions.

7 Conclusion

In this study, we applied an UQ framework to calibrate the runoff generation relevant parameters in the ELM-v1 using an observation-based runoff dataset as benchmark. The parameters with higher sensitivity are identified through the

sensitivity analysis with the PCE-based surrogate models. While different sensitivity patterns are found for different regions, 81.3% of the global cells show significant sensitivities to three or fewer parameters of the 11 selected parameters. The results of our sensitivity analysis are consistent with those of previous studies over the US continent (Huang et al., 2013; Sun et al., 2013), with runoff showing the largest sensitivity to the subsurface runoff parameter. The Bayesian posterior distribution of the highly sensitive parameters at each grid cells is estimated with MCMC simulations, using the surrogate model to construct the likelihood function. Additional ELM simulations with parameter samples from the posterior run to estimate the optimal parameter values and construct the parametric uncertainty for the simulated runoff. While the optimal parameter values improve the model performance of runoff significantly, the parametric uncertainty is comparable to the uncertainty in a multi-model ensemble in ISIMIP2a, which is appreciable. Furthermore, the parameters are found to impact the annual global runoff trend for our simulation period. Specifically, the simulations with parameter posteriors show a wide range of the annual runoff trends at global scale, from no trend to significant increasing trend. In summary, parameter calibration is necessary to improve model performance and parametric uncertainties should be considered for comprehensive analysis of runoff and its projections.

Code and Data Availability

The current version of ELM is available from E3SM project (https://github.com/E3SM-Project/E3SM/releases/tag/v1.1.0). The UQTk code and documentation are available from https://www.sandia.gov/uqtoolkit/. The exact version of ELM, exact version of UQTk source code, and scripts to produce the plots in this study is archived on Zenodo (https://doi.org/10.5281/zenodo.5815500). Matlab version R2019b Update 4 was used to run the processing and plotting scripts. ILAMB version 2 was used in this study, and the package can be accessed at 10.18139/ILAMB.v002.00/1251621. The domain file and surface data file that used to run ELMv1, and processed ISIMP2a runoff data are archived on Zenodo (https://doi.org/10.5281/zenodo.5815730). The GRUN runoff dataset was downloaded from https://doi.org/10.6084/m9.figshare.9228176.

Author Contribution

DX and GB designed the study. DX run the simulations, performed the analysis, visualized the results, and prepared the first draft of manuscript. GB mentored DX through this study. KS helped with the Uncertainty Quantification methodology. CL investigated the results. All authors contributed to the discussion and review of the results and manuscript writing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the Earth System Model Development program area of the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research as part of the multi-program, collaborative integrated Coastal Modeling (ICoM) project. The Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RLO1830. CL was supported through Next Generation Ecosystem Experiments-Tropics, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research at Pacific Northwest National Laboratory. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

References

Alkama, R., Decharme, B., Douville, H., and Ribes, A.: Trends in Global and Basin-Scale Runoff over the Late Twentieth Century: Methodological Issues and Sources of Uncertainty, J Climate, 24, 3000-3014, 10.1175/2010JCLI3921.1, 2011.

Alkama, R., Marchand, L., Ribes, A., and Decharme, B.: Detection of global runoff changes: results from observations and CMIP5 experiments, Hydrol. Earth Syst. Sci., 17, 2967-2979, 10.5194/hess-17-2967-2013, 2013.

Andreadis, K. M., Schumann, G. J.-P., and Pavelsky, T.: A simple global river bankfull width and depth database, Water Resour Res, 49, 7164-7168, 10.1002/wrcr.20440, 2013.

Bechtold, B.: Violin Plots for Matlab, Github Project, 10.5281/zenodo.4559847, 2016.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art hydrological models, Hydrol. Earth Syst. Sci., 21, 2881-2903, 10.5194/hess-21-2881-2017, 2017.

Beck, H. E., van Dijk, A. I. J. M., Miralles, D. G., de Jeu, R. A. M., Bruijnzeel, L. A., McVicar, T. R., and Schellekens, J.: Global patterns in base flow index and recession based on streamflow observations from 3394 catchments, Water Resour Res, 49, 7843-7863, https://doi.org/10.1002/2013WR013918, 2013.

Bisht, G., Riley, W. J., Hammond, G. E., and Lorenzetti, D. M.: Development and evaluation of a variably saturated flow model in the global E3SM Land Model (ELM) version 1.0, Geosci. Model Dev., 11, 4085-4102, 10.5194/gmd-11-4085-2018, 2018.

Bosmans, J. H. C., van Beek, L. P. H., Sutanudjaja, E. H., and Bierkens, M. F. P.: Hydrological impacts of global land cover change and human water use, Hydrol. Earth Syst. Sci., 21, 5603-5626, 10.5194/hess-21-5603-2017, 2017.

Breiman, L.: Random Forests, Machine Learning, 45, 5-32, 10.1023/A:1010933404324, 2001.

Brunke, M. A., Broxton, P., Pelletier, J., Gochis, D., Hazenberg, P., Lawrence, D. M., Leung, L. R., Niu, G.-Y., Troch, P. A., and Zeng, X.: Implementing and evaluating variable soil thickness in the Community Land Model, version 4.5 (CLM4. 5), J Climate, 29, 3441-3461, 2016.

Chegwidden, O. S., Rupp, D. E., and Nijssen, B.: Climate change alters flood magnitudes and mechanisms in climatically-diverse headwaters across the northwestern United States, Environ Res Lett, 15, 094048, 10.1088/1748-9326/ab986f, 2020.

Clark, E. A., Sheffield, J., van Vliet, M. T., Nijssen, B., and Lettenmaier, D. P.: Continental runoff into the oceans (1950–2008), J Hydrometeorol, 16, 1502-1520, 2015.

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., Mu, M., and Randerson, J. T.: The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation, J Adv Model Earth Sy, 10, 2731-2754, https://doi.org/10.1029/2018MS001354, 2018.

Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils, Water Resour Res, 20, 682-690, https://doi.org/10.1029/WR020i006p00682, 1984.

Dagon, K., Sanderson, B. M., Fisher, R. A., and Lawrence, D. M.: A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5, Adv. Stat. Clim. Meteorol. Oceanogr., 6, 223-244, 10.5194/ascmo-6-223-2020, 2020.

Dai, A.: Increasing drought under global warming in observations and models, Nat Clim Change, 3, 52-58, 10.1038/nclimate1633, 2013.

Dai, A., Qian, T., Trenberth, K. E., and Milliman, J. D.: Changes in Continental Freshwater Discharge from 1948 to 2004, J Climate, 22, 2773-2792, 10.1175/2008JCLI2592.1, 2009.

Debusschere, B., Sargsyan, K., Safta, C., and Chowdhary, K.: Uncertainty Quantification Toolkit (UQTk), in: Handbook of Uncertainty Quantification, edited by: Ghanem, R., Higdon, D., and Owhadi, H., Springer International Publishing, Cham, 1-21, 10.1007/978-3-319-11259-6_56-1, 2016.

Debusschere, B. J., Najm, H. N., Pébay, P. P., Knio, O. M., Ghanem, R. G., and Maı¿tre, O. P. L.: Numerical Challenges in the Use of Polynomial Chaos Representations for Stochastic Processes, SIAM Journal on Scientific Computing, 26, 698-719, 10.1137/s1064827503427741, 2004.

Decharme, B., Delire, C., Minvielle, M., Colin, J., Vergnes, J.-P., Alias, A., Saint-Martin, D., Séférian, R., Sénési, S., and Voldoire, A.: Recent Changes in the ISBA-CTRIP Land Surface System for Use in the CNRM-CM6 Climate Model and in Global Off-Line Hydrological Applications, J Adv Model Earth Sy, 11, 1207-1252, https://doi.org/10.1029/2018MS001545, 2019.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata, Earth Syst. Sci. Data, 10, 765-785, 10.5194/essd-10-765-2018, 2018.

Doocy, S., Daniels, A., Murray, S., and Kirsch, T. D.: The human impact of floods: a historical review of events 1980-2009 and systematic literature review, PLoS Curr, 5, 10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a, 2013.

Drewniak, B. A.: Simulating Dynamic Roots in the Energy Exascale Earth System Land Model, J Adv Model Earth Sy, 11, 338-359, https://doi.org/10.1029/2018MS001334, 2019.

Dwelle, M. C., Kim, J., Sargsyan, K., and Ivanov, V. Y.: Streamflow, stomata, and soil pits: Sources of inference for complex models with fast, robust uncertainty quantification, Adv Water Resour, 125, 13-31, https://doi.org/10.1016/j.advwatres.2019.01.002, 2019.

Ekici, A., Lee, H., Lawrence, D. M., Swenson, S. C., and Prigent, C.: Ground subsidence effects on simulating dynamic high-latitude surface inundation under permafrost thaw using CLM5, Geosci. Model Dev., 12, 5291-5300, 10.5194/gmd-12-5291-2019, 2019.

Fischer, E. M. and Knutti, R.: Observed heavy precipitation increase confirms theory and early models, Nat Clim Change, 6, 986-991, 10.1038/nclimate3110, 2016.

Gelman, A. and Rubin, D. B.: Inference from Iterative Simulation Using Multiple Sequences, Statist. Sci., 7, 457-472, 10.1214/ss/1177011136, 1992.

Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014, Earth Syst. Sci. Data, 11, 1655-1674, 10.5194/essd-11-1655-2019, 2019.

Giuntoli, I., Villarini, G., Prudhomme, C., and Hannah, D. M.: Uncertainties in projected runoff over the conterminous United States, Climatic Change, 150, 149-162, 10.1007/s10584-018-2280-5, 2018.

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., Abeshu, G., Anantharaj, V., Asay-Davis, X. S., Bader, D. C., Baldwin, S. A., Bisht, G., Bogenschutz, P. A., Branstetter, M., Brunke, M. A., Brus, S. R., Burrows, S. M., Cameron-Smith, P. J., Donahue, A. S., Deakin, M., Easter, R. C., Evans, K. J., Feng, Y., Flanner, M., Foucar, J. G., Fyke, J. G., Griffin, B. M., Hannay, C., Harrop, B. E., Hoffman, M. J., Hunke, E. C., Jacob, R. L., Jacobsen, D. W., Jeffery, N., Jones, P. W., Keen, N. D., Klein, S. A., Larson, V. E., Leung, L. R., Li, H.-Y., Lin, W., Lipscomb, W. H., Ma, P.-L., Mahajan, S., Maltrud, M. E., Mametjanov, A., McClean, J. L., McCoy, R. B., Neale, R. B., Price, S. F., Qian, Y., Rasch, P. J., Reeves Eyre, J. E. J., Riley, W. J., Ringler, T. D., Roberts, A. F., Roesler, E. L., Salinger, A. G., Shaheen, Z., Shi, X., Singh, B., Tang, J., Taylor, M. A., Thornton, P. E., Turner, A. K., Veneziani, M., Wan, H., Wang, H., Wang, S., Williams, D. N., Wolfram, P. J., Worley, P. H., Xie, S., Yang, Y., Yoon, J.-H., Zelinka, M. D., Zender, C. S., Zeng, X., Zhang, C., Zhang, K., Zhang, Y., Zheng, X., Zhou, T., and Zhu, Q.: The DOE E3SM Coupled Model Version 1: Overview and Evaluation at Standard Resolution, J Adv Model Earth Sy, 11, 2089-2129, https://doi.org/10.1029/2018MS001603, 2019.

Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Dai, Y., Ye, A., and Miao, C.: Multi-objective parameter optimization of common land model using adaptive surrogate modeling, Hydrol. Earth Syst. Sci., 19, 2409-2425, 10.5194/hess-19-2409-2015, 2015.

Gosling, S., Müller Schmied, H., Betts, R. A., Chang, J., Ciais, P., Dankers, R., Döll, P., Eisner, S., Flörke, M., Gerten, D., Grillakis, M., Hanasaki, N., Hagemann, S., Huang, M., Huang, Z., Jerez, S., Kim, H., Koutroulis, A., Leng, G., Liu, X., Masaki, Y., Montavez, P., Morfopoulos, C., Oki, T., Papadimitriou, L., Pokhrel, Y., Portmann, F. T., Orth, R., Ostberg, S., Satoh, Y., Seneviratne, S., Sommer, P., Stacke, T., Tang, Q., Tsanis, I., Wada, Y., Zhou, T., Büchner, M., Schewe, J., and Zhao, F.: ISIMIP2a Simulation Data from Water (global) Sector (V. 1.1). GFZ Data Services [dataset], https://doi.org/10.5880/PIK.2019.003, 2019.

Gosling, S. N. and Arnell, N. W.: Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis, Hydrol Process, 25, 1129-1145, https://doi.org/10.1002/hyp.7727, 2011.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment, Earth Syst. Sci. Data, 10, 787-804, 10.5194/essd-10-787-2018, 2018.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour Res, 34, 751-763, https://doi.org/10.1029/97WR03495, 1998.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J Hydrol, 377, 80-91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.
Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, Bernoulli, 7, 223-242, 220, 2001.
Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J. J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J. S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185-4208, 10.5194/gmd-9-4185-2016, 2016.
Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., and Gerten, D.: Multimodel estimate of the global terrestrial water balance: setup and first results, J Hydrometeorol, 12, 869-884, 2011.
Hall, J. W., Grey, D., Garrick, D., Fung, F., Brown, C., Dadson, S. J., and Sadoff, C. W.: Coping with the curse of freshwater variability, Science, 346, 429, 10.1126/science.1257890, 2014.
Hintze, J. and Nelson, R.: Violin plots : A box plot-density trace synergism, The American Statistician, 52, 181-184, 1998.
Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S.: Global flood risk under climate change, Nat Clim Change, 3, 816-821, 10.1038/Nclimate1911, 2013.
Hou, Z., Huang, M., Leung, L. R., Lin, G., and Ricciuto, D. M.: Sensitivity of surface flux simulations to hydrologic parameters based on an uncertainty quantification framework applied to the Community Land Model, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2012JD017521, 2012.
Huang, M., Ray, J., Hou, Z., Ren, H., Liu, Y., and Swiler, L.: On the applicability of surrogate-based Markov chain Monte Carlo-Bayesian inversion to the Community Land Model: Case studies at flux tower sites, Journal of Geophysical Research: Atmospheres, 121, 7548-7563, https://doi.org/10.1002/2015JD024339, 2016.
Huang, M., Hou, Z., Leung, L. R., Ke, Y., Liu, Y., Fang, Z., and Sun, Y.: Uncertainty Analysis of Runoff Simulations and Parameter Identifiability in the Community Land Model: Evidence from MOPEX Basins, J Hydrometeorol, 14, 1754-1772, 10.1175/JHM-D-12-0138.1, 2013.
Ivanov, V. Y., Xu, D., Dwelle, M. C., Sargsyan, K., Wright, D. B., Katopodes, N., Kim, J., Tran, V. N., Warnock, A., Fatichi, S., Burlando, P., Caporali, E., Restrepo, P., Sanders, B. F., Chaney, M. M., Nunes, A. M. B., Nardi, F., Vivoni, E. R., Istanbulluoglu, E., Bisht, G., and Bras, R. L.: Breaking Down the Computational Barriers to Real-Time Urban Flood Forecasting, Geophys Res Lett, n/a, e2021GL093585, https://doi.org/10.1029/2021GL093585, 2021.
Jenicek, M. and Ledvinka, O.: Importance of snowmelt contribution to seasonal runoff and summer low flows in Czechia, Hydrol. Earth Syst. Sci., 24, 3475-3491, 10.5194/hess-24-3475-2020, 2020.
Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, Scientific data, 6, 1-14, 2019.
Kim, H., Yeh, P. J. F., Oki, T., and Kanae, S.: Role of rivers in the seasonal variations of terrestrial water storage over global basins, Geophys Res Lett, 36, https://doi.org/10.1029/2009GL039006, 2009.
Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrol. Earth Syst. Sci., 23, 4323-4331, 10.5194/hess-23-4323-2019, 2019.
Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, Nat Clim Change, 3, 369-373, 10.1038/nclimate1716, 2012.
Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in Combining Projections from Multiple Climate Models, J Climate, 23, 2739-2758, 2010.
Krysanova, V., Zaherpour, J., Didovets, I., Gosling, S. N., Gerten, D., Hanasaki, N., Müller Schmied, H., Pokhrel, Y., Satoh, Y., Tang, Q., and Wada, Y.: How evaluation of global hydrological models can help to improve credibility of river discharge projections under climate change, Climatic Change, 163, 1353-1377, 10.1007/s10584-020-02840-0, 2020.
Laloy, E. and Jacques, D.: Emulation of CPU-demanding reactive transport models: a comparison of Gaussian processes, polynomial chaos expansion, and deep neural networks, Computational Geosciences, 23, 1193-1215, 2019.
Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model, J Adv Model Earth Sy, 3, https://doi.org/10.1029/2011MS00045, 2011.
Lehner, F., Wood, A. W., Vano, J. A., Lawrence, D. M., Clark, M. P., and Mankin, J. S.: The potential to reduce uncertainty in regional runoff projections from climate models, Nat Clim Change, 9, 926-933, 10.1038/s41558-019-0639-x, 2019.
Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., and Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth Syst. Dynam., 11, 491-508, 10.5194/esd-11-491-2020, 2020.
Leung, L. R., Bader, D. C., Taylor, M. A., and McCoy, R. B.: An Introduction to the E3SM Special Collection: Goals, Science Drivers, Development, and Analysis, J Adv Model Earth Sy, 12, e2019MS001821, https://doi.org/10.1029/2019MS001821, 2020.
Li, H.-Y., Leung, L. R., Getirana, A., Huang, M., Wu, H., Xu, Y., Guo, J., and Voisin, N.: Evaluating Global Streamflow Simulations by a Physically Based Routing Model Coupled with the Community Land Model, J Hydrometeorol, 16, 948-971, 10.1175/JHM-D-14-0079.1, 2015.

Liao, C., Zhou, T., Xu, D., Barnes, R., Bisht, G., Li, H.-Y., Tan, Z., Tesfa, T., Duan, Z., Engwirda, D., and Leung, L. R.: Advances in hexagon mesh-based flow direction modeling, Adv Water Resour, 160, 104099, https://doi.org/10.1016/j.advwatres.2021.104099, 2022.
Lin, G. and Karniadakis, G. E.: Sensitivity analysis and stochastic simulations of non-equilibrium plasma flow, International Journal for Numerical Methods in Engineering, 80, 738-766, https://doi.org/10.1002/nme.2582, 2009.
Lu, D., Ricciuto, D., Stoyanov, M., and Gu, L.: Calibration of the E3SM Land Model Using Surrogate-Based Global Optimization, J Adv Model Earth Sy, 10, 1337-1356, https://doi.org/10.1002/2017MS001134, 2018.
Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev., 10, 1903-1925, 10.5194/gmd-10-1903-2017, 2017.
Milliman, J. D., Farnsworth, K. L., Jones, P. D., Xu, K. H., and Smith, L. C.: Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951-2000, Global and Planetary Change, 62, 187-194, 10.1016/j.gloplacha.2008.03.001, 2008.
Milly, P. C. D., Wetherald, R. T., Dunne, K. A., and Delworth, T. L.: Increasing risk of great floods in a changing climate, Nature, 415, 514-517, DOI 10.1038/415514a, 2002.
Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, Science, 319, 573-574, 10.1126/science.1151915, 2008.
Mishra, A. K. and Singh, V. P.: A review of drought concepts, J Hydrol, 391, 202-216, https://doi.org/10.1016/j.jhydrol.2010.07.012, 2010.
Mortatti, J., Moraes, J., RODRIGUES, J., Victoria, R., and Martinelli, L.: Hydrograph separation of the Amazon River using 18O as an isotopic tracer, Scientia Agricola, 54, 167-173, 1997.
Müller, J., Paudel, R., Shoemaker, C. A., Woodbury, J., Wang, Y., and Mahowald, N.: CH₄ parameter estimation in CLM4.5bgc using surrogate global optimization, Geosci. Model Dev., 8, 3285-3310, 10.5194/gmd-8-3285-2015, 2015.
Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J Hydrol, 10, 282-290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.
Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., and Gulden, L. E.: A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models, Journal of Geophysical Research: Atmospheres, 110, https://doi.org/10.1029/2005JD006111, 2005.
Oleson, K., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson, S., Thornton, P. E., Bozbiyik, A., Fisher, R., Heald, C. L., Kluzek, E., Lamarque, J.-F., Lawrence, P. J., Leung, L. R., Lipscomb, W., Muszala, S. P., Ricciuto, D. M., Sacks, W. J., Sun, Y., Tang, J., and Yang, Z.-L.: Technical description of version 4.5 of the Community Land Model (CLM), http://dx.doi.org/10.5065/D6RR1W7M, 2013.
Olson, R., Fan, Y. A., and Evans, J. P.: A simple method for Bayesian model averaging of regional climate model projections: Application to southeast Australian temperatures, Geophys Res Lett, 43, 7661-7669, 10.1002/2016gl069704, 2016.
Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, Nature, 540, 418-422, 10.1038/nature20584, 2016.
Pilgrim, D. H., Chapman, T. G., and Doran, D. G.: Problems of rainfall-runoff modelling in arid and semiarid regions, Hydrological Sciences Journal, 33, 379-400, 10.1080/02626668809491261, 1988.
Ray, J., Hou, Z., Huang, M., Sargsyan, K., and Swiler, L.: Bayesian Calibration of the Community Land Model Using Surrogates, SIAM/ASA Journal on Uncertainty Quantification, 3, 199-233, 10.1137/140957998, 2015.
Razavi, S., Tolson, B. A., and Burn, D. H.: Review of surrogate modeling in water resources, Water Resour Res, 48, https://doi.org/10.1029/2011WR011527, 2012.
Ricciuto, D., Sargsyan, K., and Thornton, P.: The Impact of Parametric Uncertainties on Biogeochemistry in the E3SM Land Model, J Adv Model Earth Sy, 10, 297-319, 10.1002/2017ms000962, 2018.
Rodell, M., Beaudoing, H. K., L'Ecuyer, T., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., and Chambers, D.: The observed state of the water cycle in the early twenty-first century, J Climate, 28, 8289-8318, 2015.
Sargsyan, K., Najm, H. N., and Ghanem, R.: On the Statistical Calibration of Physical Models, International Journal of Chemical Kinetics, 47, 246-276, https://doi.org/10.1002/kin.20906, 2015.
Sargsyan, K., Safta, C., Najm, H. N., Debusschere, B. J., Ricciuto, D., and Thornton, P.: Dimensionality Reduction for Complex Models Via Bayesian Compressive Sensing, Int J Uncertain Quan, 4, 63-93, 2014.
Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J. C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G. P.: A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset, Earth Syst. Sci. Data, 9, 389-413, 10.5194/essd-9-389-2017, 2017.
Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, Proceedings of the National Academy of Sciences, 111, 3245-3250, 10.1073/pnas.1222460110, 2014.
Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall's Tau, J Am Stat Assoc, 63, 1379-1389, 10.1080/01621459.1968.10480934, 1968.

Seyoum, W. M., Kwon, D., and Milewski, A. M.: Downscaling GRACE TWSA Data into High-Resolution Groundwater Level Anomaly Using Machine Learning-Based Models in a Glacial Aquifer System, Remote Sensing, 11, 824, 2019.
Sheng, M., Lei, H., Jiao, Y., and Yang, D.: Evaluation of the Runoff and River Routing Schemes in the Community Land Model of the Yellow River Basin, J Adv Model Earth Sy, 9, 2993-3018, https://doi.org/10.1002/2017MS001026, 2017.
Sobol′, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation, 55, 271-280, https://doi.org/10.1016/S0378-4754(00)00270-6, 2001.
Sun, Y., Hou, Z., Huang, M., Tian, F., and Ruby Leung, L.: Inverse modeling of hydrologic parameters using surface flux and runoff observations in the Community Land Model, Hydrol. Earth Syst. Sci., 17, 4995-5011, 10.5194/hess-17-4995-2013, 2013.
Swenson, S. C., Lawrence, D. M., and Lee, H.: Improved simulation of the terrestrial hydrological cycle in permafrost regions by the Community Land Model, J Adv Model Earth Sy, 4, https://doi.org/10.1029/2012MS000165, 2012.
Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M., and Perket, J.: Representing Intrahillslope Lateral Subsurface Flow in the Community Land Model, J Adv Model Earth Sy, 11, 4044-4065, https://doi.org/10.1029/2019MS001833, 2019.
Tan, Z., Leung, L. R., Li, H.-Y., Tesfa, T., Zhu, Q., and Huang, M.: A substantial role of soil erosion in the land carbon sink and its future changes, Global Change Biol, 26, 2642-2655, https://doi.org/10.1111/gcb.14982, 2020.
Tebaldi, C., Smith, R. L., Nychka, D., and Mearns, L. O.: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles, J Climate, 18, 1524-1540, 2005.
Tesfa, T. K., Leung, L. R., and Ghan, S. J.: Exploring Topography-Based Methods for Downscaling Subgrid Precipitation for Use in Earth System Models, Journal of Geophysical Research: Atmospheres, 125, e2019JD031456, https://doi.org/10.1029/2019JD031456, 2020.
Toure, A. M., Luojus, K., Rodell, M., Beaudoing, H., and Getirana, A.: Evaluation of Simulated Snow and Snowmelt Timing in the Community Land Model Using Satellite-Based Products and Streamflow Observations, J Adv Model Earth Sy, 10, 2933-2951, https://doi.org/10.1029/2018MS001389, 2018.
Trenberth, K. E.: Changes in precipitation with climate change, Clim Res, 47, 123-138, 2011.
Troy, T. J., Wood, E. F., and Sheffield, J.: An efficient calibration method for continental-scale land surface modeling, Water Resour Res, 44, https://doi.org/10.1029/2007WR006513, 2008.
Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., Liu, J., and Shen, C.: From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling, Nature Communications, 12, 5988, 10.1038/s41467-021-26107-z, 2021.
Vörösmarty, C. J., Green, P., Salisbury, J., and Lammers, R. B.: Global Water Resources: Vulnerability from Climate Change and Population Growth, Science, 289, 284, 10.1126/science.289.5477.284, 2000.
Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., and Miao, C.: An evaluation of adaptive surrogate modeling based optimization with two benchmark problems, Environmental Modelling & Software, 60, 167-179, https://doi.org/10.1016/j.envsoft.2014.05.026, 2014.
Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework, Proceedings of the National Academy of Sciences, 111, 3228, 10.1073/pnas.1312330110, 2014.
Wu, H., Kimball, J. S., Mantua, N., and Stanford, J.: Automated upscaling of river networks for macroscale hydrological modeling, Water Resour Res, 47, https://doi.org/10.1029/2009WR008871, 2011.
Xie, Z., Yuan, F., Duan, Q., Zheng, J., Liang, M., and Chen, F.: Regional Parameter Estimation of the VIC Land Surface Model: Methodology and Application to River Basins in China, J Hydrometeorol, 8, 447-468, 10.1175/JHM568.1, 2007.
Xiu, D. and Karniadakis, G. E.: The Wiener--Askey Polynomial Chaos for Stochastic Differential Equations, SIAM Journal on Scientific Computing, 24, 619-644, 10.1137/S1064827501387826, 2002.
Xu, D., Ivanov, V. Y., Kim, J., and Fatichi, S.: On the use of observations in assessment of multi-model climate ensemble, Stochastic Environmental Research and Risk Assessment, 33, 1923-1937, 10.1007/s00477-018-1621-2, 2019.
Xu, D., Ivanov, V. Y., Li, X., and Troy, T. J.: Peak Runoff Timing is Linked to Global Warming Trajectories, Earth's Future, n/a, e2021EF002083, https://doi.org/10.1029/2021EF002083, 2021a.
Xu, D., Bisht, G., Zhou, T., Leung, L. R., and Pan, M.: Development of Land-River Two-Way Coupling in the Energy Exascale Earth System Model, Earth and Space Science Open Archive, 41, doi:10.1002/essoar.10507802.1, 2021b.
Yang, H., Zhou, F., Piao, S. L., Huang, M. T., Chen, A. P., Ciais, P., Li, Y., Lian, X., Peng, S. S., and Zeng, Z. Z.: Regional patterns of future runoff changes from Earth system models constrained by observation, Geophys Res Lett, 44, 5540-5549, 10.1002/2017gl073454, 2017.
Yang, S. L., Xu, K. H., Milliman, J. D., Yang, H. F., and Wu, C. S.: Decline of Yangtze River water and sediment discharge: Impact from natural and anthropogenic changes, Scientific Reports, 5, 12581, 10.1038/srep12581, 2015.
Zhang, Y., Zheng, H., Chiew, F. H. S., Arancibia, J. P. a., and Zhou, X.: Evaluating Regional and Global Hydrological Models against Streamflow and Evapotranspiration Measurements, J Hydrometeorol, 17, 995-1010, 10.1175/JHM-D-15-0107.1, 2016.
Zhou, T., Leung, L. R., Leng, G., Voisin, N., Li, H.-Y., Craig, A. P., Tesfa, T., and Mao, Y.: Global Irrigation Characteristics and Effects Simulated by Fully Coupled Land Surface, River, and Water Management Models in E3SM, J Adv Model Earth Sy, 12, e2020MS002069, https://doi.org/10.1029/2020MS002069, 2020.