Developing a Methodology for Model Intercomparison and Its Application to Improve Simulated Streamflow by Land Surface Models

AULIA FEBIANDA ANWAR TINUMBANG,[a] KAZUAKI YOROZU,[a] YASUTO TACHIKAWA,[a] AND YUTAKA ICHIKAWA[a]

[a] *Graduate School of Engineering, Kyoto University, Kyoto, Japan*

(Manuscript received 5 October 2022, in final form 20 January 2023, accepted 21 February 2023)

ABSTRACT: Runoff generated by land surface models (LSMs) is extensively used to predict future river discharge under global warming. However, the structural bias of LSMs, the precipitation bias of the climate model, and other factors could cause the runoff to be biased. A model intercomparison study can help understand LSM behavior. Traditional model intercomparison can discover output variation and evaluate performance, but explaining the reason for the difference is challenging. This study developed a novel method to identify the reasons for disparities and suggest improvements. Consequently, we investigated the impacts of model settings by adopting the settings of another model in one model until it can mimic similar features in its output. Hence, we developed a process called the "emulation model." We employed two LSMs [Simple Biosphere with Urban Canopy (SiBUC) and Meteorological Research Institute Simple Biosphere model (MRI-SiB)] in the Thai River basin. SiBUC produced a higher surface runoff than MRI-SiB, and the development of the MRI-SiB emulation revealed the cause of this variation. The differences in runoff characteristics affected streamflow estimation. For instance, the SiBUC peak discharge was faster and higher than observed in the dry year. Conversely, there was a tendency to underestimate the flow estimated by MRI-SiB runoff during the transition from dry to wet seasons. Incorporating other model settings can alleviate the shortcomings of each model. Overall, the proposed method can identify the strengths and weaknesses of a model and enhance the reproducibility of the hydrological characteristics of the observed discharge in the basin.

SIGNIFICANCE STATEMENT: This study aims to develop a new methodology for model intercomparison to identify the reasons for model output variation. Understanding why models behave differently is important to enhancing the reliability of model prediction. Our findings guide what affects disparities in land surface model runoff-based streamflow estimation, which will help reduce the uncertainty of future flood and drought predictions.

KEYWORDS: Streamflow; Runoff; Water budget/balance; Hydrologic models; Land surface model; Model comparison

1. Introduction

The risks associated with water availability and water-related hazards will continue to increase due to the intensification of the hydrological cycle because of climate change (IPCC 2022). Climate models are important tools for projecting future changes in the climate and their effects. These numerical models use a three-dimensional grid to simulate physical processes in the atmosphere, ocean, ice, and land surface across the globe. The land surface model (LSM), a part of the climate model that represents the land surface, is an important component for simulating hydrological cycle. The LSM divides the available energy into sensible and latent heat, and separates precipitation into evaporation, runoff, and water storage components.

For climate change studies, runoff output from the LSM has shown broad applications in predicting changes in future river discharge under global warming, such as in the studies by Nohara et al. (2006), Hirabayashi et al. (2013), and Dottori et al. (2018). These studies report runoff bias when simulating river flow, which might be attributed to precipitation bias from the climate models when simulating precipitation or structural bias owing to missing key processes in the runoff generation schemes of the LSM.

Furthermore, different LSMs have shown large variability in runoff output (Nohara et al. 2006; Hirabayashi et al. 2013), leading to uncertainty in streamflow prediction. Differences in parameterization, representation of physical processes in the model, numerical schemes, and other factors could cause these disparities. A method called "model intercomparison" makes it possible to analyze the variation in model output by driving multiple models with a common dataset and subjecting the output to a mutual comparison.

Extensive studies on comparing LSMs have been undertaken, often to determine how different the LSMs are in simulating the energy budget and/or water budget, including the Project for the Intercomparison of Land-Surface Parameterization Schemes (PILPS; Henderson-Sellers et al. 1995), the Rhône-Aggregation Land Surface Scheme Intercomparison Project (Boone et al. 2004), the African Monsoon Multidisciplinary Analysis Land Surface Model Intercomparison Project (ALMIP; Boone et al. 2009), and the Asian Dryland Model Intercomparison Project (ADMIP; Asanuma et al. 2013).

These projects have helped improve the understanding of the abilities of the models to simulate the land surface processes. Furthermore, comparing the performance of one model

Denotes content that is immediately available upon publication as open access.

Corresponding author: Aulia Febianda Anwar Tinumbang, tinumbang.auliafebiandaanwar.6c@kyoto-u.ac.jp

DOI: 10.1175/JHM-D-22-0183.1

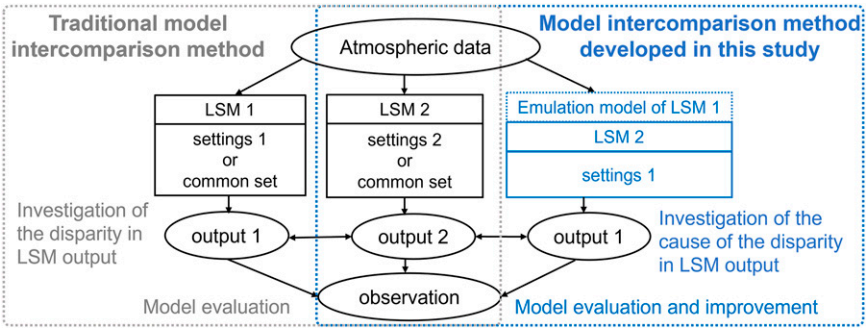

FIG. 1. Comparison of traditional model intercomparison methods vs the method developed in this study.

with that of other models helps identify which model performs better than the other models while also providing a direction of improvement for the lower-performing models. Some reasons for differences in model output could be due to complex interactions between model components (Koster and Milly 1997), or human errors in the modeling, for example, inadequate documentation during model development (Menard et al. 2021). However, analyzing why LSMs behave differently is challenging, and intercomparison results do not always explain why LSM simulations differ from one another or observations (Best et al. 2015). Therefore, this study aims to develop a new model intercomparison methodology to provide a better understanding of the reasons for output variation among LSMs.

The rest of the paper has six sections. Section 2 describes the developed model intercomparison approach, followed by the design framework in section 3. Section 4 provides the results of the intercomparison analysis, and section 5 addresses the model evaluation and improvement. Finally, sections 6 and 7 describe discussion and conclusions, respectively.

2. Model intercomparison methodology

a. *Past research on model intercomparison*

Previous research on model intercomparison has mainly focused on model comparison and evaluation, performed by comparing the performance of one model to that of other models and observations. Conventional methods of model intercomparison typically involved running multiple models with the same meteorological forcing (Materia et al. 2010; Getirana et al. 2017). It also occasionally includes a common set of land surface parameters (e.g., soil and vegetation parameters) into the participating models (Henderson-Sellers et al. 1995; Boone et al. 2004).

The intercomparison results revealed significant differences in how the models simulated the energy and/or water balances (Pitman et al. 1999; Asanuma et al. 2013; Boone et al. 2004; Materia et al. 2010; Getirana et al. 2017). These differences can be attributed to different types of flux partitioning (Boone et al. 2004, 2009), different complexities of the land surface schemes (Materia et al. 2010), and insufficient or missing representation of key processes in some models (Getirana et al. 2017).

Even though it can be challenging to explain why the model outputs vary, it is crucial to comprehend the reasons behind the scatter because large uncertainties in model prediction reduce its reliability (Pitman et al. 1999). Furthermore, without understanding the underlying causes of such disparities, developed models may become more similar to other models, but not always closer to the observations (Best et al. 2015).

b. *Model intercomparison approach developed in this study*

In this study, we developed a new methodology for model intercomparison to address the limitations of conventional methods, which have difficulty explaining the reasons for discrepancies among model outputs. Figure 1 compares the traditional methods of model intercomparison with the proposed approach. Figure 2 shows a detailed flowchart of the proposed method.

The approach in this study seeks to investigate the causes of the spread of LSM output and propose improvements. First, various settings for LSMs (e.g., land surface parameters, representative of physical processes, and numerical schemes) were examined. To find a direction for improvement, one must understand the impact of the model settings on the output estimation. This study analyzed those impacts by adopting the settings of one model in another.

As an illustration, we consider the intercomparison of two LSMs: LSM 1 and LSM 2. To suggest improvement strategies for LSM 2, for example, the impacts of model settings on output estimation are investigated by incorporating some settings from LSM 1 into LSM 2. If this process is repeated until LSM 2 can mimic the features of LSM 1 output, the settings that have a major impact on the LSM behavior and the source of their output discrepancy can be identified. This iterative process represents developing an emulation model. In this case, the LSM 1 emulation model is constructed by adopting its settings into LSM 2 until it produces an output comparable to LSM 1. First, one setting of LSM 1 is adopted to LSM 2 and evaluate its results. If one LSM 1 setting does not produce similar results to LSM 1, a different model setting should be tested. When a single setting fails to produce a desired outcome, a combination of two or more settings should be examined. In contrast to traditional intercomparison methods, in which the simulation of each model is undertaken by its corresponding modelers or

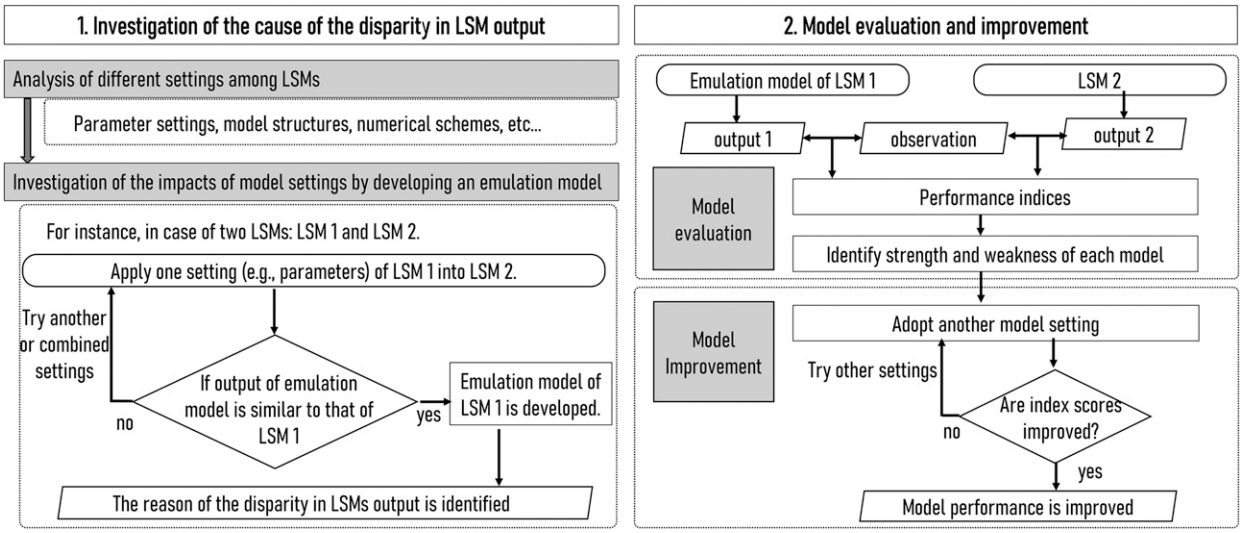

FIG. 2. Flowchart of the model intercomparison method in this study.

experts, constructing the emulation of one model is advantageous because it makes an intercomparison analysis possible without directly carrying out the simulation using another LSM.

In the second step, we evaluated the performance of each model by comparing its output to observation. We investigated the strengths and weaknesses of the model in reproducing observed data and suggested some strategies for overcoming this limitation. For example, the disadvantages of the LSM 2 can be alleviated by considering some settings of LSM 1. Simultaneously, proposing solutions for improving LSM 1 is feasible when the emulation model of LSM 1 can properly replicate the behavior of LSM 1, for instance, by incorporating LSM 2 settings.

The developed methodology applies to all output variables. This study focuses its constructed method on the ability of the LSM to simulate runoff and its performance in reproducing observed streamflow.

3. Design framework

a. *Land surface models*

This study employed two LSMs: Simple Biosphere model including Urban Canopy (SiBUC; Tanaka 2005) and Meteorological Research Institute Simple Biosphere model (MRI-SiB; Hirai et al. 2007). The SiBUC LSM uses a mosaic scheme to account for a mix of land use, paddy fields, and irrigation systems. In this model, land use is separated into three categories: green area, urban area, and water body. In contrast, MRI-SiB is a land surface model developed by the Meteorological Research Institute and Japan Meteorological Agency. The two LSMs have been developed based on the Simple Biosphere model (SiB; Sellers et al. 1986). Therefore, the basic structures are similar, for example, soil discretization into three layers, and using Richard's equation to describe the governing equations for soil moisture. However, the detailed structures, including the number of vegetation layers, surface and subsurface runoff calculation, and vertical soil water flux estimation, differ because different institutions developed them independently. The following section discusses the differences between them.

b. *Hydrometeorological forcing data*

The first part of this study used hydrometeorological output data (including precipitation, air temperature, surface pressure, humidity, shortwave and longwave radiation in the downward direction, and wind speed) from the atmospheric general circulation model MRI-AGCM 3.2S (Mizuta et al. 2012) to conduct an intercomparison analysis. The MRI-SiB was embedded in this climate model. The simulation by the MRI-AGCM 3.2S model was applied to the entire globe with 20-km spatial resolution. The output data were generated for 25 years of historical climate (1979–2003) and future climate under the RCP8.5 scenario (2080–99). In this study, only the historical data were utilized for the simulation.

The second part used satellite rainfall from Climate Hazards Group Infrared Precipitation with Station data (CHIRPS; Funk et al. 2015) and the atmospheric reanalysis dataset JRA-55 (Kobayashi and Iwasaki 2016) as forcing for both LSMs. The datasets included air temperature, humidity, shortwave and longwave radiation in the downward direction, wind speed, and surface pressure. The CHIRPS dataset has a high spatial resolution of 0.05°, whereas the JRA-55 output has a 60-km spatial resolution. We interpolated the JRA-55 data from approximately 60 to 5 km using the weighted average method. The CHIRPS dataset had a temporal resolution of 1 day, and that of the JRA-55 was 3 h. We linearly interpolated the JRA-55 data from 3 to 1 h to create hourly input forcing. We then used the JRA-55 hourly rainfall to interpolate the CHRIPS daily data.

c. *Flow routing model*

We used 1K-FRM, a distributed flow routing model developed based on the one-dimensional kinematic wave theory, to

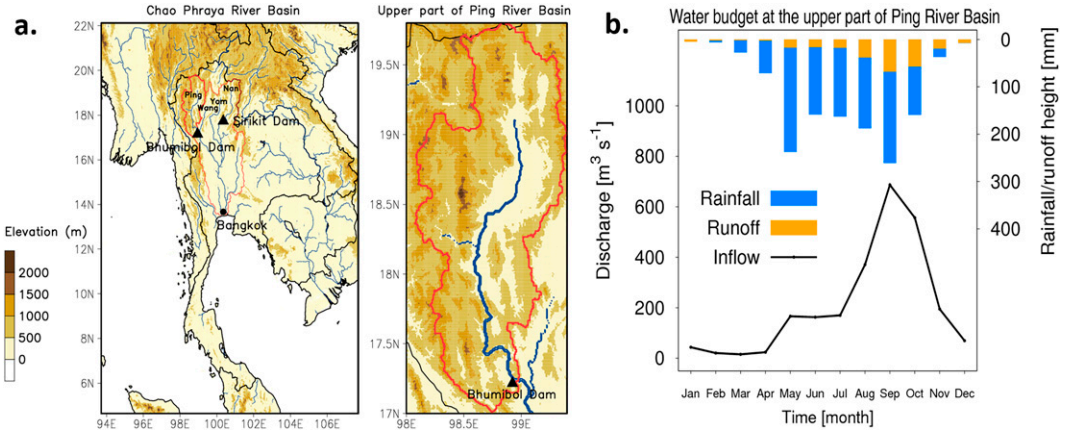

FIG. 3. (a) Topography (m) of Thailand and its surrounding countries (shown at left) and the upper part of Ping River basin (shown at right). The catchment boundary is shown by the red line. Major rivers and dams are shown in blue lines and black triangles, respectively. The location of the capital city of Bangkok is shown with a black circle. (b) The 10-yr-mean (2003–12) monthly observed rainfall (blue bars), runoff height (orange bars), and inflow (black line) at the Bhumibol Dam. The runoff height is the inflow volume per unit area of the catchment.

convert LSM runoff into river discharge. The source codes and manual operations are available on the Hydrology and Water Resources Research Laboratory Kyoto University website (http://hywr.kuciv.kyoto-u.ac.jp/products/1K-DHM/1K-DHM.html). This study used SiBUC coupled with 1K-FRM (Yorozu and Tachikawa 2015). This coupled model can investigate the interaction effect between flow routing and land surface processes. The 1K-FRM was also used to estimate streamflow using MRI-SiB runoff.

d. Study area

Figure 3a shows the topography of the Chao Phraya River basin in Thailand. The Chao Phraya River basin consists of the confluence of the Ping, Wang, Yom, and Nam River basins. The catchment area is approximately 159 000 km^2, which accounts for about 30% of the land surface area in Thailand. There are two major dams in this basin: Bhumibol Dam, built in the Ping River basin, and Sirikit Dam in the Nan River basin.

Thailand experiences strong interannual variability of rainfall. Flood disasters frequently occur during the rainy season and drought in the dry or early rainy season. Thailand recently experienced severe flood and drought cycles. For instance, in 2011, the country suffered severe flood damage due to unprecedented rainfall, in contrast to the low rainfall in the previous year (Pavelic et al. 2012). Predictions indicate that these water-related hazards will worsen due to global warming (Hunukumbura and Tachikawa 2012).

We conducted numerical experiments for this study in the upper part of the Ping River basin. The basin covers an area of approximately 26 100 km^2. We used the observation inflow at the outlet of the Bhumibol Dam, located at 17.2425°N, 98.9722°E, and obtained daily discharge data from the Electricity Generating Authority of Thailand (EGAT) available from June 1964, when the Bhumibol Dam began operation. Several smaller dams (e.g., Mae Ngat Somboon Chon Dam) located in the upper part of this basin may influence runoff observations.

Figure 3b shows the climatological mean (from 2003 to 2012) basin average monthly rainfall, runoff height, and inflow at the Bhumibol Dam. Runoff height is the volume of inflow per unit area of the catchment. The average annual rainfall in this basin, estimated from the CHIRPS dataset, is approximately 1300 mm.

This basin has distinct rainy and dry seasons, with roughly 90% of the precipitation falling during the rainy season and the rest in the dry season. The wet season is from May to October, and the dry season is from November to April. The rainy season can be divided into an early rainy season from May to July and a late rainy season from August to October. Despite comparable precipitation, the observed discharge in the early rainy season was much lower than in the late rainy season. In this basin, more than 80% of the total rainfall is evapotranspiration; thus, the runoff ratio is low. It is crucial to develop a hydrological model that can reproduce such basin characteristics to minimize the impacts and manage the risks of future flood and drought disasters in this area.

4. Investigation of the cause of the disparity in the LSM output

a. Analysis of different settings among LSMs

Earlier research (Tinumbang et al. 2020, 2021) examined different settings related to runoff generation schemes among SiBUC and MRI-SiB. Table 1 presents a summary of the comparison, while Fig. 4 provides a schematic image of each LSM. The outline below provides detailed descriptions.

1) LAND SURFACE PARAMETERS

LSMs generally require numerous input parameters, including soil parameters (e.g., saturated hydraulic conductivity K_s, saturated water content θ_s, soil depth), and vegetation

TABLE 1. Comparison of parameterization and land surface schemes between SiBUC and MRI-SiB. One asterisk indicates the dominant value in the target basin corresponds to clay loam soil; two asterisks indicate the dominant value in the target basin corresponds to the forest vegetation type.

Settings	SiBUC	MRI-SiB
1) Land surface parameters	User defined	User defined
Soil parameters	Based on soil types	Based on vegetation types
Saturated hydraulic conductivity K_s (m s^{-1})	1.44×10^{-6}*	2.00×10^{-5}**
Matric potential of saturation ψ_s (m)	−0.63*	−0.086**
Saturated water content θ_s	0.478*	0.42**
Soil wetness parameter B	8.41*	7.12**
Depth of surface layer, root zone, and recharge layer (m)	0.02, 1–5, 2.7–12.5	0.02, 0.5–1.5, 1.5–3.5
Vegetation parameters	For canopy layer	For canopy, ground-cover layer
Greenness fraction	0.6–0.99	0.075–0.98, 0.075–1.0
Leaf area index (LAI)	0.2–6.5	0.04–6.5, 1.0×10^{-4}–4.8
Fraction of photosynthetically active radiation (FPAR)	0.1–0.9	0.02–0.93, 1.0×10^{-4}–0.73
2) Structure for direct infiltration into root zone	—	Incorporated
Saturation–excess surface runoff generation	Surface soil is saturated	Surface soil and root zone are saturated
3) Soil–water flux between adjacent soil layers	Eq. (1a)	Eq. (1b)
4) Subsurface runoff calculation	Eq. (2a)	Eq. (2b)
5) Numerical scheme for updating soil moisture	Explicit–midpoint	Semi-implicit, Eq. (3)

parameters (e.g., greenness fraction, leaf area index, the fraction of photosynthetically active radiation). In SiBUC and MRI-SiB, users can define the land surface parameters. The soil hydraulic properties in SiBUC were set for each soil class and its parameter value was following Cosby et al. (1984). We used the Ecoclimap dataset (Masson et al. 2003) to derive the soil class in SiBUC. In MRI-SiB, all soil parameters were assigned to each vegetation class. Table 1 compares soil and vegetation parameters between the two LSMs for the upper part of the Ping River basin. The soil characteristics parameters displayed in this table correspond to the basin's major soil or vegetation. In SiBUC, clay loam soil is the primary soil type, while forest vegetation predominates in MRI-SiB.

The two LSMs have different vegetation schemes. In SiBUC, canopy and ground cover are modeled as a single canopy layer, as opposed to MRI-SiB, which treats each separately. Therefore, the parameters related to vegetation in both LSMs were assigned differently, depending on the vegetation layer.

2) STRUCTURE FOR DIRECT INFILTRATION

In MRI-SiB, there is a structure that enables rainwater to flow directly from the ground surface into the root zone (second soil layer). This scheme, P_2, was employed to alleviate the defect in the conventional infiltration scheme resulting from insufficient vertical resolutions and an integration time step (Nakaegawa and Sugi 2001). In SiBUC, such a scheme does not exist. This structure has an impact on the saturation–excess runoff calculation. In MRI-SiB, it is computed when the surface soil layer and the root zone have saturated, whereas it is determined in SiBUC when the surface layer reaches saturation.

3) SOIL WATER FLUX BETWEEN ADJACENT SOIL LAYERS

In SiBUC, the transfer of water between soil layers, $Q_{i,i+1}$, is calculated using Darcy's law by considering hydraulic diffusion and gravitational drainage of water in soil, as expressed in Eq. (1a). MRI-SiB neglects gravitational drainage and estimates the soil water flux based on hydraulic diffusion, as

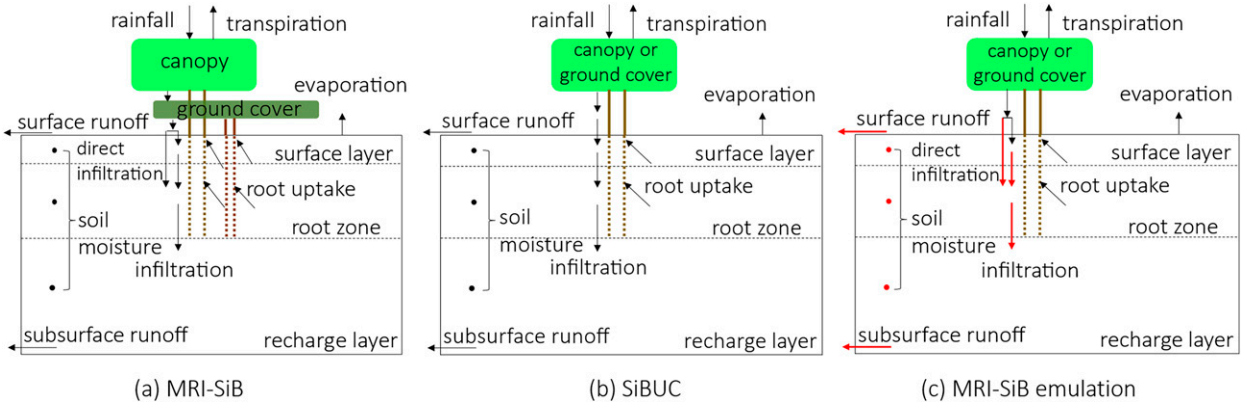

FIG. 4. Schematic image of (a) MRI-SiB, (b) SiBUC, and (c) MRI-SiB emulation. The red lines highlight processes in SiBUC modified by MRI-SiB settings.

expressed in Eq. (1b). The method of calculating vertical soil water flux without considering gravitational flow is peculiar, but explaining why is outside the scope of this study. In both LSMs, hydraulic conductivity and matric potential are functions of soil moisture based on the formulae of Clapp and Hornberger (1978):

$$Q_{i,i+1} = K\left(\frac{\partial \psi}{\partial z} + 1\right) \quad (i = 1, 2), \tag{1a}$$

$$Q_{i,i+1} = K\left(\frac{\partial \psi}{\partial z}\right) \quad (i = 1, 2), \tag{1b}$$

where K, ψ, i, and z indicate hydraulic conductivity, matric potential, soil layer, and soil depth, respectively.

4) Subsurface runoff calculation

SiBUC computes the subsurface runoff, Q_3, only based on gravitational drainage, expressed in Eq. (2a). However, the calculation in MRI-SiB considers the gravitational drainage and potential gradients between the second and third soil layers, as expressed in Eq. (2b):

$$Q_3 = \sin\oslash_s K_s W_3^{2B+3}, \tag{2a}$$

$$Q_3 = \sin\oslash_s K_s W_3^{2B+3}\left(1 + \frac{\psi_2 - \psi_3}{D_3}\right), \tag{2b}$$

where $\sin\oslash_s$, B, W_3, and D_3 express the topographic slope, Clapp and Hornberger equation shape parameter, soil wetness, and soil depth of the third soil layer, respectively.

5) Numerical scheme for updating soil moisture

In SiBUC, an explicit midpoint (modified Euler) method was used to update the soil moisture in each time step. This method was employed to solve the differential equation for soil moisture, matric potential, and hydraulic conductivity at the midpoint between the current and the next step. In MRI-SiB, a semi-implicit method was implemented for matric potential by linearizing the soil moisture using a Taylor expansion, as expressed in Eq. (3):

$$\psi_i^{t+\Delta t} = \psi_i^t + \frac{\partial \psi_i^t}{\partial W_i^t}\frac{\partial W_i^t}{\partial t}\Delta t. \tag{3}$$

b. *Investigation of the impacts of model settings by developing an emulation model*

By incorporating MRI-SiB settings into SiBUC, the effects of the model settings were examined. Trial and error were used to iterate this process until SiBUC replicated similar features of MRI-SiB runoff. We started by investigating the effect of each setting and identifying which variables most affected the runoff. We also examined combinations of land surface parameters with another setting because their impacts on runoff could be large. Then, all variations related to runoff generation were combined to evaluate their impacts. A version of SiBUC that adopts the MRI-SiB settings and can reproduce runoff its features is the MRI-SiB emulation model.

It was not possible to run the simulation by MRI-SiB directly; therefore, constructing the MRI-SiB emulation was necessary to perform the intercomparison analysis. Appendix A describes the advantages and disadvantages of each model in reproducing observed river discharge in this basin. As a forcing for SiBUC, the atmospheric output from the MRI-AGCM 3.2S datasets, simulated under the historical climate from 1979 to 2003, was employed. This dataset also contained the runoff output produced by MRI-SiB. A 25-yr numerical simulation was run, with the period from 1994 to 2003 being used to analyze the long-term trends of hydrological characteristics of this basin, and the rests as spinup. Because this basin is dry (more than 80% of rainfall becomes evapotranspiration), we set the spinup period rather long based on the assumption that the soil subsystem takes a long time to reach equilibrium. The spatial and temporal resolutions of both LSMs were approximately 20 km and 1 h, respectively.

Since both LSMs used the same atmospheric data, we assumed the cause of the difference in runoff output was the various settings between the two LSMs. However, the MRI-SiB was simulated by considering the interaction with the atmospheric model in the MRI-AGCM 3.2S (commonly known as an "online simulation"). SiBUC, nevertheless, was run without considering such interaction, which is generally referred to as an "offline simulation." Although such differences may influence the output estimation, they are outside the scope of this investigation. The process of constructing the MRI-SiB emulation model also ignored the land–atmosphere interaction in MRI-AGCM 3.2, which was largely impacted by soil moisture, soil temperature, and other factors. Thus, it might have affected the amount of evapotranspiration and subsurface runoff. Therefore, this study focused on the reason for the differences in runoff characteristics between the two LSMs instead of water budget's reproducibility.

Table 2 lists the experimental settings of the MRI-SiB emulation model. The control experiment was a simulation using the default settings of SiBUC. We conducted six experiments to investigate the impact of the individual or combined MRI-SiB settings on SiBUC. Experiments a and b employed SiBUC and MRI-SiB parameters, respectively. We based the design of experiments 1–4 and 6 on previous studies (Tinumbang et al. 2020, 2021).

1) Effects of model settings on water budget

Figure 5 compares the 10-yr-mean water budgets calculated by MRI-SiB, SiBUC with default settings (control), and each experiment to develop the emulation model. Figures B1a and B1b in appendix B show the spatial pattern of the total soil depths of MRI-SiB and SiBUC, respectively. SiBUC with default settings estimates 4% less evapotranspiration and 30% more runoff than MRI-SiB. In terms of runoff components, it generated approximately 15-fold higher surface runoff than MRI-SiB. Approximately one-third of the total runoff of SiBUC is surface runoff, as opposed to predominantly subsurface runoff for MRI-SiB. SiBUC estimated approximately 8%

TABLE 2. Experimental settings to develop MRI-SiB emulation by SiBUC. The applied settings are indicated by open circles. Experiments 1–4 followed Tinumbang et al. (2020), and experiment 6 was according to Tinumbang et al. (2021).

	Experiments										
Changing SiBUC to MRI-SiB settings	Control	1	2a	2b	3a	3b	4a	4b	5a	5b	6
Land surface parameters	—	○	—	○	—	○	—	○	—	○	○
Direct infiltration structure	—	—	○	○	—	—	—	—	—	—	○
Calculation for soil-water movement	—	—	—	—	○	○	—	—	—	—	○
Calculation for subsurface runoff	—	—	—	—	—	—	○	○	—	—	○
Numerical scheme for updating soil moisture	—	—	—	—	—	—	—	—	○	○	○

higher evaporation and 14% lower transpiration than MRI-SiB did for the evapotranspiration components. Appendix C briefly discusses the effects of changing model settings on the energy budget.

The experimental results reveal that altering the model settings affected the runoff characteristics. For instance, when we used MRI-SiB parameters (experiment 1), the surface runoff increased by 2.6-fold, whereas the subsurface runoff decreased by one-third compared with the control. The increase in surface runoff was probably because of the thinner soil depth, which reduced the capacity of soil to store rainwater. In contrast, surface runoff decreased when the P_2 structure was adopted (experiment 2) because rainwater could infiltrate directly into the second soil layer.

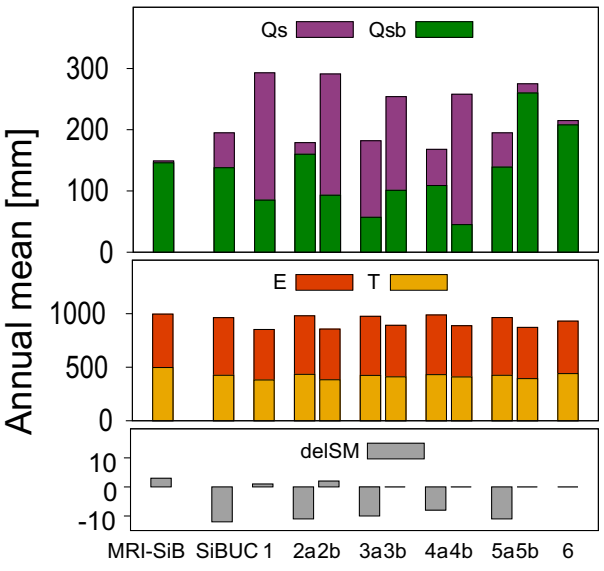

FIG. 5. The 10-yr-mean annual water budget (mm). Surface runoff (Qs), subsurface runoff (Qsb), evaporation (E), transpiration (T), and soil moisture changes (delSM) are represented by purple, green, red, orange, and gray bars, respectively. The x and y axes represent each simulation and the annual mean of each variable (mm), respectively. The total soil depth of (a) MRI-SiB and (b) SiBUC parameters are 1.5–3.5 and 2.7–12.5 m, respectively, with the spatial pattern shown in Fig. B1. SiBUC is the experiment with default settings (control). Raw data from 20 years of results of experiments 1–4 and 6 have previously been published (Tinumbang et al. 2021, 2022).

The effects of neglecting gravitational drainage for calculating the soil water flux (experiment 3) varied depending on the parameter settings. When we utilized SiBUC settings, surface runoff increased, whereas the MRI-SiB parameters caused a decrease. However, there was a reduction in the subsurface runoff when the potential gradient between the root zone and the recharge layer was considered when estimating the subsurface runoff (experiment 4). This outcome was mainly because the soil moisture in the root zone was lower than the recharge layer.

The impacts of the numerical schemes for updating soil moisture (experiment 5) on runoff estimation also differed depending on the soil parameters. When SiBUC parameters were used from a different scheme, the runoff components showed no apparent change. However, they significantly varied when we employed different numerical schemes with MRI-SiB parameters. A detailed investigation (results not provided) demonstrated that soil moisture predicted by the two schemes differed greatly during rainfall events, mostly due to the saturated hydraulic conductivity (K_s) settings.

These results demonstrate that when we used only a single setup (e.g., parameters, structures, or time integration method), we could not reproduce similar runoff features by MRI-SiB. However, when all MRI-SiB variables were considered (experiment 6), SiBUC mimicked characteristics of runoff by MRI-SiB, indicating a negligible amount of surface runoff and predominance of subsurface runoff. SiBUC adopting the settings of experiment 6 is the "MRI-SiB emulation model." Figure 4c shows the MRI-SiB emulation model schematic. Overall, the runoff and evapotranspiration predicted by the emulation model were approximately 43% greater and 7% lower than MRI-SiB, respectively. This outcome was primarily due to a lower estimate of transpiration, influenced by the difference in canopy and ground cover layer treatments between the two LSMs. In the emulation model, only MRI-SiB parameters for the canopy were employed, leading to significantly less transpiration in the grid cells dominated by ground cover.

2) IMPACTS OF MODEL SETTINGS ON STREAMFLOW ESTIMATION

We investigated the impacts of model settings on streamflow simulated by runoff from each LSM by comparing the hydrographs and examining the variation in lag time and volume of peak discharge.

Figure 6 shows the mean of daily discharge simulated by the runoff from SiBUC, MRI-SiB, and the MRI-SiB emulation

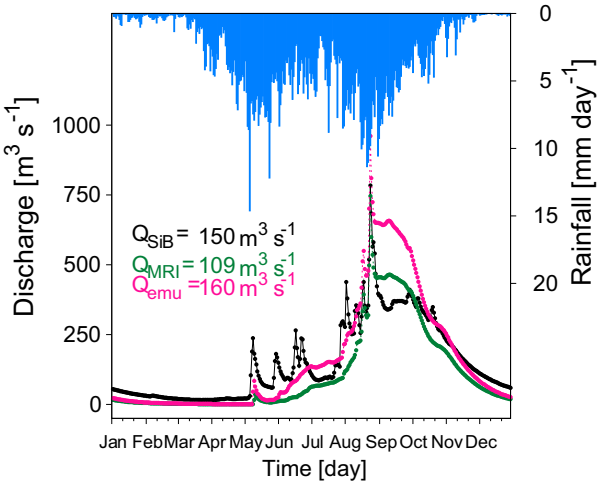

FIG. 6. The 10-yr-mean daily discharge using MRI-AGCM 3.2S forcing. Streamflow simulated using runoff by iBUC (control), MRI-SiB, and MRI-SiB emulation (experiment 6) are shown by black, green, and pink lines, respectively. The *x* and *y* axes represent the day of the year and the mean daily discharge ($m^3\ s^{-1}$), respectively.

model using MRI-AGCM 3.2S forcing. The comparison of the climatological mean of daily discharge was based on the method used by Getirana et al. (2017). In general, MRI-AGCM 3.2S rainfall (Fig. 6) could reproduce seasonal patterns of observed rainfall in this catchment (Fig. 3), with approximately 84% of the rainfall falling during the wet season (May–October) and the rest falling during the dry season (November–April). In addition, the early rainy season (May–July) and the late rainy season (August–October) have comparable rainfall ratios. The discharge simulated by SiBUC runoff was approximately 1.4-fold higher than that of MRI-SiB, owing to higher runoff and lower evapotranspiration. In particular, the estimated discharge by SiBUC in the dry, early, and late rainy seasons were 2.0-, 3.2-, and 1.1-fold higher than those of MRI-SiB, respectively. Thus, the disparity between the two LSMs is noticeable during the early rainy season.

Conversely, the streamflow estimated by MRI-SiB emulation runoff was 47% and 7% higher than MRI-SiB and SiBUC, respectively. Although the emulation model overpredicted the discharge by MRI-SiB runoff, it closely reproduced its features, particularly during the early rainy season when it was considerably lower than the SiBUC discharge.

Figure 7a depicts a time series of daily discharge, and Figs. 7b and 7c indicate the time lag of the simulated discharge in 1994 and 1998, respectively. The timings of peak discharge in response to peak rainfall of SiBUC and MRI-SiB are sometimes similar, as in 1994 and 1996, but significantly different in others, such as 1998 and 2000–02. The difference in peak discharge timing can be investigated by calculating the lag time. It is defined as the time elapsed between peak rainfall and peak discharge. Here, the peak rainfall and discharge are the annual maximum daily rainfall and discharge.

In 1994, the peak rainfall occurred on 22 August, and the peak discharge by both SiBUC and MRI-SiB were produced 3 days later. This occurrence was because of the enormous amount of antecedent precipitation that saturated the drainage basin in this case, which reduced infiltration and increased

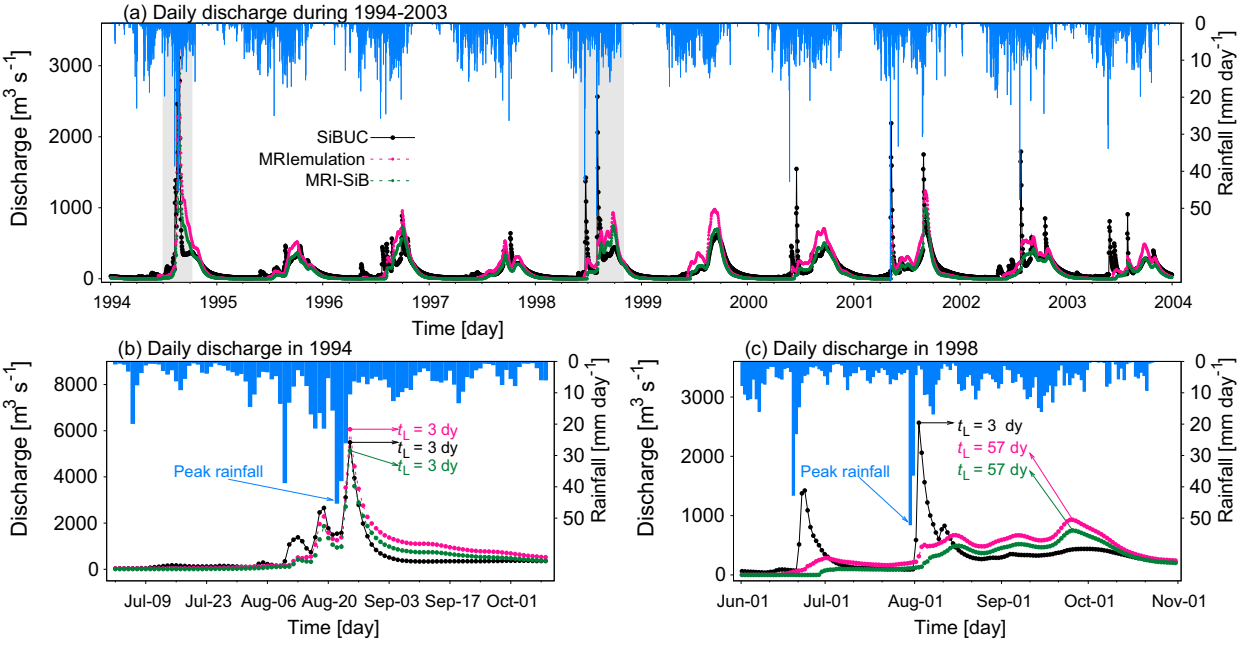

FIG. 7. (a)–(c) Time series of daily discharge. The *x* and *y* axes represent the day of the year and daily discharge ($m^3\ s^{-1}$), respectively. The gray area in (a) indicates daily discharge during heavy rainfall events in (b) 1994 and (c) 1998. The t_L represents the lag time of each simulation (day).

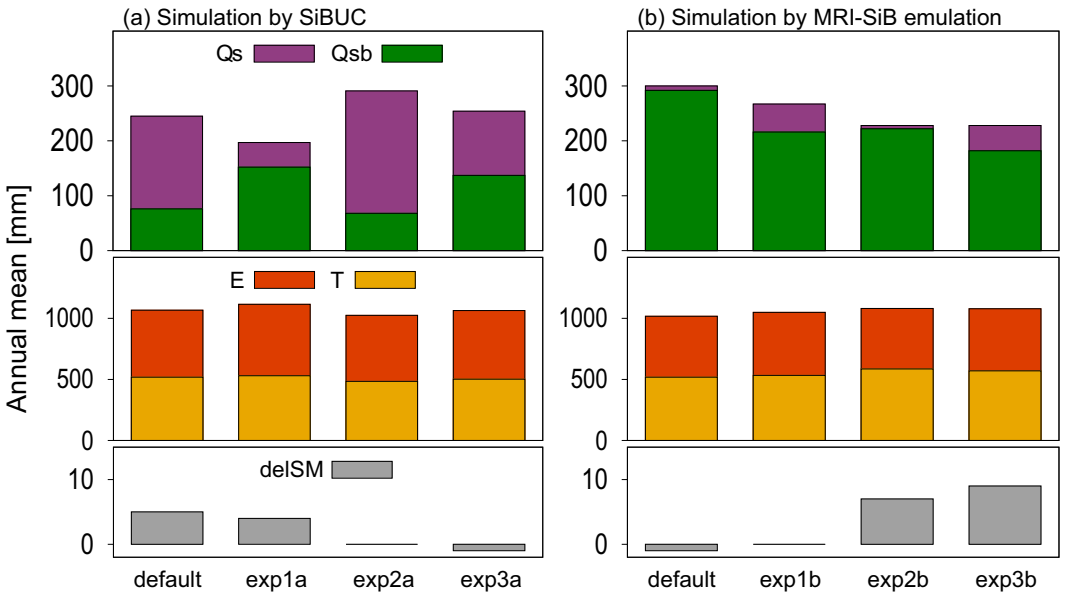

FIG. 8. As in Fig. 5, but for water budget using observed forcing data. The average annual rainfall is 1316 mm. In experiments 2 and 3, the total soil depth of both models is 1.5–5 m, with the spatial pattern shown in Fig. B1c.

surface runoff. Rainfall reaches the river more quickly and produces a short lag time because surface runoff is quicker than subsurface runoff. The peak volumes of the two LSMs were comparable, with a difference of only 7%.

In contrast, the hydrograph responses of the two LSMs during the heavy rainfalls of 1998 varied significantly. We also observed a substantial difference in the peak volume, with the SiBUC peak discharge being approximately 3.4-fold higher than the MRI-SiB. The amount of antecedent rainfall in 1998 was far lower than that in 1994. Peak rainfall occurred on 31 July and SiBUC peak discharge was produced 3 days later. The MRI-SiB peak discharge formed 54 days after SiBUC. The varied lag times of the two models, and the difference in the early rainy season (Fig. 6), resulted from different runoff characteristics caused by the variation in infiltration capacity. SiBUC had lower infiltration rates than MRI-SiB and generated a surface runoff thin surface soil was saturated. As a result, the peak discharge was produced predominantly by surface runoff, resulting in a short lag time. Conversely, the infiltration rate in MRI-SiB was higher and rainwater infiltrated directly into the second soil layer. Therefore, the impact of catchment wetness could be observed because subsurface runoff dominated the runoff component in MRI-SiB. Precipitation was used to saturate the soil, and the discharge began to rise as the soil became saturated. Thus, the time required for the simulated discharge by MRI-SiB runoff to reach the stream was substantially longer than that for SiBUC.

Although the emulation model tended to have a higher discharge than the MRI-SiB, it successfully mimicked this feature of the MRI-SiB discharge. The correlation coefficient score of the time series of daily discharge between MRI-SiB and the emulation model is 0.989, indicating a nearly perfect correlation.

5. Model evaluation and improvement

a. *Introduction*

In this section, we evaluated the ability of SiBUC and MRI-SiB emulation model to reproduce observed river discharge and proposed strategies to improve the performance of each model. Both models were driven by the CHIRPS rainfall and JRA-55 reanalysis datasets. The spatial and temporal resolutions of both LSMs are 0.05° and 1 h, respectively. The numerical simulations were run from 1981 to 2012, with the results from 1981 to 2002 discarded as spinup and those from 2003 to 2012 were used to analyze the long-term hydrological features of this basin. As stated in the previous section, we set the spinup time to be rather long for the soil subsystem to achieve an equilibrium state in this dry basin.

The performance indices used to evaluate the simulated daily discharge over 10 years are the Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe 1970), root-mean-square coefficient–observations standard deviation ratio (RSR), bias, coefficient of variation (CV), Pearson correlation coefficient (r), and Kling–Gupta efficiency (KGE; Gupta et al. 2009). Appendix D shows the equation of each index.

b. *Model evaluation*

Figure 8 compares the water budget simulated by SiBUC (default settings) and MRI-SiB emulation. Figure B1c shows the spatial patterns of the total soil depth in experiments 2 and 3. SiBUC estimated 4% higher evapotranspiration and 20% lower runoff than the MRI-SiB emulation. The ratio of evaporation to transpiration was similar in both LSMs, but the runoff characteristics were opposite. The surface runoff by SiBUC accounted for approximately 70% of the total runoff, with the remainder as subsurface runoff. Subsurface

TABLE 3. Evaluation scores of simulated daily discharge from 2003 to 2012.

LSM	SiBUC				MRI-SiB emulation			
Experiments	Default	1a	2a	3a	Default	1b	2b	3b
NSE	0.54	0.44	0.48	0.62	0.34	0.52	0.57	0.65
RSR	0.68	0.75	0.72	0.62	0.81	0.70	0.66	0.59
r	0.78	0.68	0.83	0.82	0.77	0.82	0.77	0.82
CV	0.97	0.99	0.99	0.99	1.06	1.11	0.99	1.04
Bias (%)	7.62	−19.88	26.63	5.19	19.00	9.26	−9.79	−6.51
KGE	0.77	0.63	0.68	0.81	0.70	0.77	0.75	0.80

runoff was the dominant runoff component in the MRI-SiB emulation default settings, accounting for approximately 97% of the runoff. Appendix C briefly presents the impacts of modifying model settings on the energy budget.

Table 3 shows the evaluation scores for each simulation. Figures 9 and 10 show a time series of daily discharge from 2003 to 2012 and a 10-yr mean of daily discharge between observation and each simulation, respectively. Overall, the simulated flow with default SiBUC settings outperforms the emulation model in terms of NSE and bias scores. However, because the emulation model produced a higher runoff than the original MRI-SiB, this contributed to an overestimated bias by the emulation model. Other performance indices, including the correlation coefficient and KGE scores, were comparable, indicating that both model settings could reproduce the observed discharge.

In terms of seasonal patterns, SiBUC typically reproduced the actual discharge well. However, it underestimated the peak flow in September while significantly exceeding the observed flow in October. We thought that the underestimation in September resulted from insufficient subsurface runoff contribution. In contrast, the overestimation in October was due to excessive surface runoff generation in response to intense rainfall events over this period.

Similarly, the discharge simulated by the MRI-SiB emulation showed seasonal patterns close to the observed inflow. However, it significantly underestimated the observed inflow, particularly from April to May. Because the subsurface runoff dominated the runoff by MRI-SiB emulation, the precipitation during this transition period between the dry and rainy seasons was mainly used to saturate the soil, resulting in a considerably low discharge. Similar to SiBUC, the discharge by the emulation model also overestimated the observed flow in October. However, the reproducibility of peak flow in September was better than that of SiBUC.

We also evaluated the discharge simulated by the two LSMs by comparing the lag time and peak volume based on observation. We investigated the performance of each model for each year to comprehend the model behavior based on the difference in rainfall amount because this catchment has a strong interannual variability in rainfall. Using a 10-yr mean annual rainfall (1316 mm) set as the threshold, we classified the annual rainfall into wet and dry years. We defined wet years as those with annual precipitation that exceeded the climatological mean rainfall, whereas dry years were those with less than the threshold. In this study, we selected 2004 and 2011 as the representative dry (annual rainfall of 1205 mm) and wet years (annual rainfall of 1645 mm), respectively. During the 10-yr analysis period, 2011 experienced the highest rainfall. However, we considered 2004 for the dry-year analysis even though 2003 had less annual rainfall (988 mm) because both LSMs performed significantly better.

Figure 11 depicts a time series of daily discharge during peak rainfall events in 2004 (dry year) and 2011 (wet year). It also presents the difference in lag time and the peak volume between the observed and simulated discharge. During the dry year of 2004, the peak rainfall occurred on 7 September, and the peak of observed discharge occurred 13 days later. SiBUC peak flow happened 6 days after the peak rainfall, a week earlier than the observed peak discharge. This duration was because SiBUC settings had a limited infiltration capacity, leading to a fast-rising limb of the hydrograph owing to excessive surface runoff generation. However, the lag time of the peak discharge by the emulation model was closer to the observation. The longer lag time for the MRI-SiB emulation peak discharge compared with SiBUC resulted from predominant subsurface runoff. Nevertheless, the volume of peak discharge by both LSMs considerably exceeded the observed peak flow.

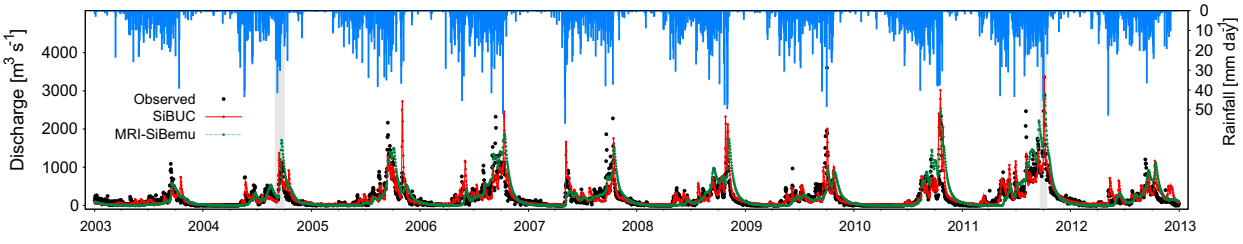

FIG. 9. Time series of daily discharge of observation (black dots) and simulation from SiBUC (red lines) and MRI-SiB emulation (green lines) during 2003–12. The *x* and *y* axes represent the day of the year and mean daily discharge ($m^3 s^{-1}$), respectively. The gray areas indicate daily discharge during heavy rainfall events in 2004 (dry year) and 2011 (wet year).

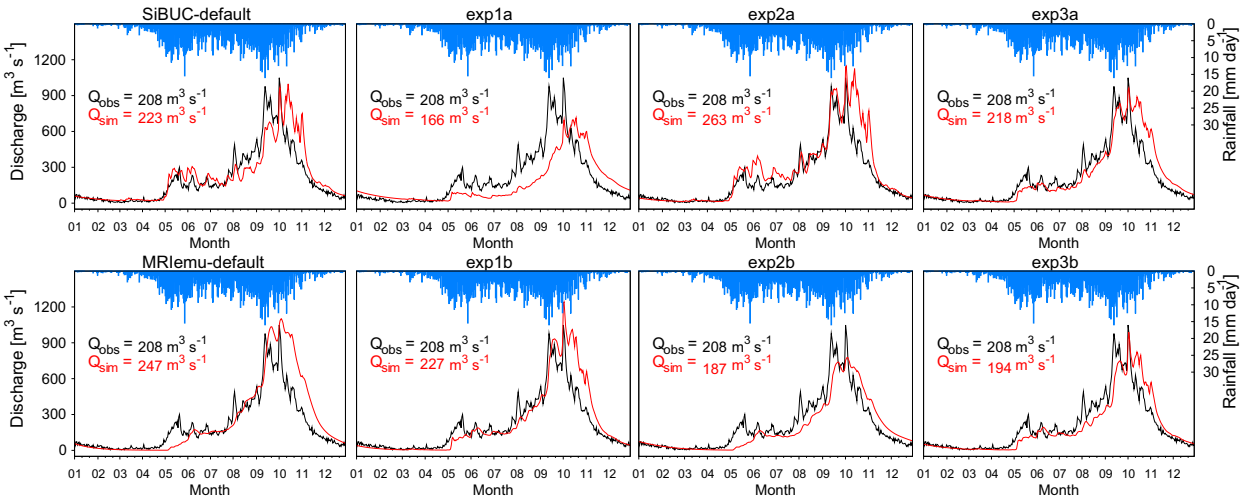

FIG. 10. As in Fig. 6, but for 10-yr-mean (2003–12) daily discharge using observation forcing.

Peak precipitation in the wet year of 2011 occurred on 30 September and observed peak flow was recorded 3 days later. Peak discharge simulated by SiBUC runoff reproduced this behavior well, whereas MRI-SiB emulation occurred 2 days later than observed. The runoff characteristics of the two LSMs influenced these variations. The SiBUC peak discharge was close to the observed peak flow in terms of volume, as opposed to MRI-SiB emulation, which underestimated the observation.

c. *Model improvement for SiBUC*

Based on the evaluation results from the previous section, we identified the strengths and weaknesses of each LSM in reproducing the observed river flow. Because the limitations of one model are the strengths of another model, it is possible to devise improvement strategies by adopting one LSM setting for another LSM. We performed three trial experiments to improve the simulated discharge by SiBUC runoff, as indicated in Table 4, experiment a. Table 3 and Figs. 8, 10, and 11 display the results of the performance scores, water budget, time series of the mean of daily discharge, and simulated discharge in the dry (2004) and wet (2011) years, respectively.

The evaluation showed that SiBUC runoff could reproduce the seasonal patterns of observed discharge well; however, the September peak flow was underestimated. SiBUC matched the

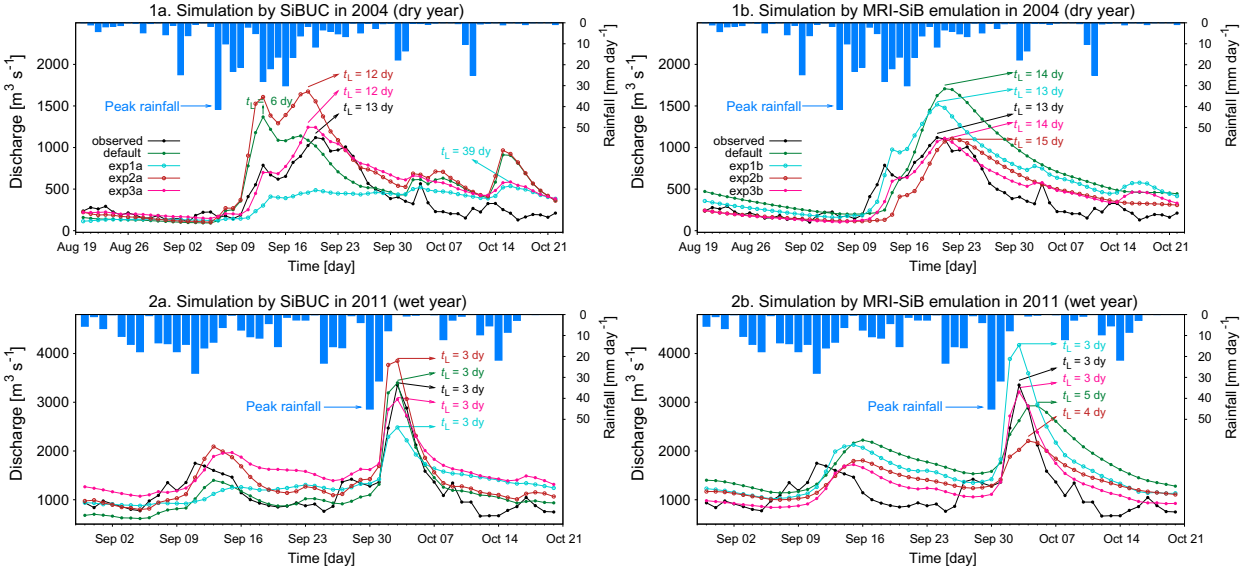

FIG. 11. (top) Daily discharge from 2 to 30 Sep 2004 (dry year) and (bottom) daily discharge from 20 Sep to 12 Oct 2011 (wet year) for simulations by (left) SiBUC and (right) MRI-SiB. The *x* and *y* axes represent the day of the year and mean daily discharge ($m^3\ s^{-1}$), respectively. Observation; simulation with default settings; and experiments 1, 2, and 3 are represented by black dots and light blue, green, yellow, and pink lines, respectively.

TABLE 4. Simulation settings by each LSM.

Experiments	Changing settings for SiBUC	Experiments	Changing settings for MRI-SiB emulation
Default	No change	Default	No change
1a	Incorporating P_2 structure	1b	Adopting soil characteristics parameters of SiBUC
2a	Reducing soil depth	2b	Increasing soil depth
3a	Combining 1a and 2a	3b	Combining 1b and 2b

observed flow in the wet year of 2011, whereas in the dry year of 2004, it produced a substantially faster and higher peak flow than observed flow. The results suggest that one of the main reasons for the shortcomings of SiBUC is its low infiltration capacity, leading to excessive generation of surface runoff.

The SiBUC trial experiments were performed using MRI-SiB settings to alleviate its deficiencies by increasing infiltration capacity. The first experiment was conducted by incorporating the MRI-SiB direct infiltration structure P_2. Compared with the default SiBUC, leveraging this scheme reduced surface runoff while increasing subsurface runoff. However, the SiBUC performance did not improve by adopting this structure because the predicted discharge was far below the observation. SiBUC soil depth was set rather deep (as much as 12.5 m); hence, saturating the soil took a long time, leading to increased water loss through evaporation and decreased runoff.

In experiment 2a, we attempted to increase the runoff by reducing the root zone depth from 1–5 to 0.5–2.5 m and the recharge layer from 2.7–12.5 to 1.5–5 m. Surface runoff increased due to a lower capacity to store rainwater in thinner soil, and subsurface runoff increased because the thinner depth resulted in faster saturation. Overall, this setting improved the peak flow in September while increasing the overestimated discharge in October because of excessive surface runoff generation in response to heavy rainfall events. In this regard, increasing the runoff by lowering the soil depth provides a way to enhance the outcome of experiment 1a.

The two prior settings were combined to conduct experiment 3a. Surface runoff decreased by two-thirds, and subsurface runoff increased by 2.5-fold when compared to the default settings. Overall, this setting slightly improved the bias of the simulated discharge and increased the peak volume in September by 20%. Furthermore, the lag time of peak flow in the dry year of 2004 became closer to the observation. For the wet year of 2011, the lag time remains unchanged, despite the peak flow decreasing by 9% relative to the observed peak flow.

d. Improvement direction for MRI-SiB

Because the MRI-SiB emulation model could mimic similar features of the MRI-SiB output, it is possible to suggest improvements for MRI-SiB. For instance, the simulated discharge by MRI-SiB emulation significantly underestimated the observed flow during the transition between the dry and rainy seasons. The predominant subsurface runoff generation by the MRI-SiB settings appears to be the cause of this limitation.

Three tests were done, as shown in Table 4, experiment b, by adopting some SiBUC settings in the MRI-SiB emulation model to increase the considerably low discharge in the transition between dry to rainy seasons and reduce the overestimated mean discharge. Table 3 and Figs. 8, 10, and 11 show all the results.

In the first experiment, we changed the soil characteristics parameters (K_s, ψ_s, θ_s, and B) for the forest vegetation used in MRI-SiB to those for clay loam soil used in SiBUC. As a result of the lower infiltration rates, overall runoff fell by 10%, with surface runoff increasing 6.6-fold and subsurface runoff decreasing by one-fourth. Despite being lower than the observation in May, which marks the beginning of the early rainy season, the discharge by this setting has increased by 30% compared with the default level. The lag time of the peak discharge in the wet year of 2011 also improved.

In the second trial, the depths of the root zone and recharge layer for the forest vegetation type were increased from 1.5 to 2.5 m and from 3.5 to 5 m, respectively. However, the discharge by the MRI-SiB emulation model tended to overpredict the observation. This outcome may also result from overestimating the runoff by the emulation model compared with the original MRI-SiB. Owing to the greater capacity of soil to hold water and delay soil saturation, increasing the soil depth in this context reduced subsurface runoff by one-fourth, despite no substantial change in surface runoff. Although the discharge in the early rainy season did not improve, it lowered the overestimated discharge in October by 50%.

We performed the third experiment by combining the previous two experiments. Similar to the outcome of the first test, the discharge in May increased by 30%, and improvements were made to the peak flow bias and the lag times in 2004 and 2011.

6. Discussion

In this study, we developed an approach for model comparison to help identify the reason for the model output disparity. This strategy is beneficial for identifying model behavior and the potential for structural bias in each model, recognizing the strengths and weaknesses of the model, and determining how to enhance it. Currently, SiBUC is used to construct an MRI-SiB emulation model by incorporating MRI-SiB settings to generate similar features of MRI-SiB runoff, which has a low surface runoff and a predominant subsurface runoff. Although we have not validated this process, we believe it is possible to develop a SiBUC emulation model by integrating the SiBUC settings into MRI-SiB. By adopting SiBUC parameters, which have a low saturated hydraulic conductivity, and removing a direct infiltration structure in MRI-SiB, we can produce the characteristics of SiBUC runoff with high surface runoff and low subsurface runoff.

The developed method has been evaluated only for LSMs using similar schemes. However, future work would require testing this method using more diverse LSM schemes.

In addition, emulation model development helps identify model settings impacts and understand the causes of disparities among LSMs. Although the emulation model could mimic similar features of the MRI-SiB runoff, it could not precisely reproduce the MRI-SiB water budget. Consequently, the performance of the emulation model, which tended to overestimate the original MRI-SiB runoff, influenced the MRI-SiB emulation's findings in reproducing the actual river flow. Further studies would need to clarify the assessment results and the proposed improvement to the original MRI-SiB embedded in MRI-AGCM 3.2S.

Moreover, the enhancement strategies proposed in this study are based only on the existing schemes (parameters or structures) of the model involved. Future research should test some new schemes, such as integrating groundwater flow, incorporating lateral flow, and addressing the heterogeneity of the soil profile, to obtain a more robust performance.

The bias of the simulated discharge may also be inherited from the satellite rainfall bias or the atmospheric reanalysis datasets. Several measures can reduce the uncertainties provided by the input data, such as bias correction of satellite rainfall or using rainfall data from multiple sources. However, such consideration is beyond the scope of this study.

Furthermore, we only tested this method in a selected basin with distinct dry and rainy seasons. When applied to another basin with different geography and climate features, such as snowfall or aridity, the performance of each LSM may vary. Applying this approach to other basins should be beneficial in further discovering model strengths and weaknesses. However, the strategy to develop an emulation model may differ because the key mechanisms driving runoff may vary. For instance, in a basin where snowmelt dominates the river discharge, the phase change of precipitation or other essential elements should be considered.

7. Conclusions

We developed and used a new methodology for model comparison to improve the accuracy of LSM runoff-based streamflow simulations. Unlike conventional intercomparison methods, which mainly focus on finding differences in model output and evaluating model performance, the proposed method can determine why the differences exist and provide feedback for improvement.

In particular, we investigated the reasons for the output differences between LSMs by examining the impacts of model settings by transferring certain settings from one LSM to another. Once the model that incorporated the settings of another model was able to duplicate similar characteristics of certain target outputs of the other LSM, we could discover the model setting with substantial impacts on the LSM behavior and the reason for the output discrepancies. The iterative process of adopting and investigating model-settings effects is developing an emulation model.

We applied the developed approach using two Simple Biosphere model–based LSMs: SiBUC and MRI-SiB, to improve the accuracy of the runoff-based simulated streamflow in the upper part of the Ping River basin, Thailand. The variations in model settings between the two LSMs were investigated by constructing an MRI-SiB emulation model. This was done by adopting MRI-SiB settings in SiBUC until SiBUC could reproduce similar features to those of MRI-SiB runoff. The runoff features produced by the two LSMs varied: SiBUC tended to predict a larger surface runoff than MRI-SiB, whereas subsurface runoff was predominantly produced by MRI-SiB. One explanation for such variations is that MRI-SiB has a larger infiltration capacity owing to its higher hydraulic conductivity and a direct infiltration structure into deep soil, resulting in most runoff being subsurface runoff.

The differences in runoff characteristics affect the simulated discharge, particularly the lag time and peak volume. In the case of severe rainfall events with little antecedent rainfall, the peak discharge predicted by SiBUC runoff was generated rapidly and at a high volume because of surface runoff generation. In contrast, MRI-SiB peak discharge was produced with a substantially longer lag time and lower peak volume because most of the precipitation was used to saturate the soil.

Overall, the discharge simulated using both settings reproduced the seasonal patterns of the observed flow. However, the MRI-SiB emulation had a more significant bias, which may be affected by the tendency of the emulation model to overstate the runoff of the original MRI-SiB. Furthermore, we made some suggestions for enhancing their performance. For instance, SiBUC, which generated excessive surface runoff, produced a significantly faster and greater peak discharge than that was observed during the dry year. Therefore, integrating a structure for direct infiltration into a deep soil layer and reducing the soil thickness can reduce surface runoff while increasing subsurface runoff, resulting in a longer lag time and reducing the peak volume bias. Conversely, MRI-SiB simulated discharge, mainly from subsurface runoff, tended to underestimate the observation significantly during the transition from dry to wet seasons. This shortcoming could be alleviated by reducing infiltration rates by adjusting the soil parameters based on soil type to increase the discharge during this period.

The proposed method can adequately identify the disparity between the runoff outputs of LSMs and the cause. It can also improve the reproducibility of the hydrological characteristics of observed discharge in this basin. However, the method has only been tested on LSMs with similar schemes, and the proposed improvement follows the existing schemes of the participating models. Future research should apply this method to more diverse LSMs, consider new schemes for more robust performance, and implement this method in a different basin with diverse geography and climate factors.

Acknowledgments. This study was supported by MEXT Program for the Advanced Studies of climate change projection (SENTAN) Grant JPMXD0722680734, JST "Moonshot R&D Program" Grant Number JPMJMS2283, and Integrated Research Program for Advancing Climate Models (TOUGOU) Theme C and D Grants JPMXD0717935561 and JPMXD0717935498. There are no conflicts of interest to declare.

Data availability statement. The authors confirm that the data supporting the findings of this study are available within the article. Derived data supporting the findings of this study are available on request from the corresponding author.

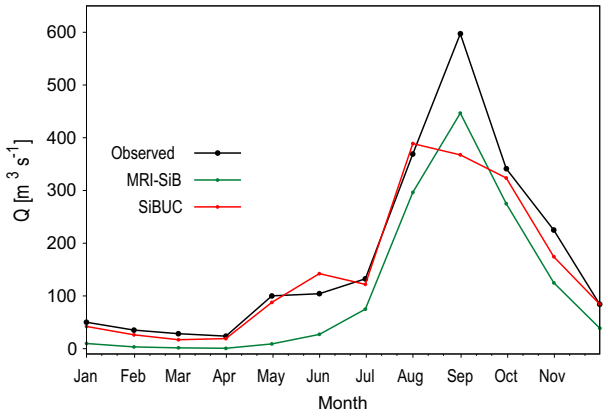

FIG. A1. Comparison of a 10-yr-mean (1994–2003) monthly observed inflow at the Bhumibol Dam outlet (black line) and discharge simulated by MRI-SiB runoff (green) and SiBUC runoff (red) driven by MRI-AGCM 3.2S atmospheric output.

APPENDIX A

LSM Performance in Reproducing Observed River Discharge

The models employed in this study have their advantages and limitations in reproducing observed river discharge. Figure A1 compares a 10-yr-mean (1994–2003) monthly observed inflow at the Bhumibol Dam outlet to discharge estimated by MRI-SiB and SiBUC runoff driven by MRI-AGCM 3.2S atmospheric output. Overall, the SiBUC runoff simulates streamflow well; however, there is a significant underestimation of the September peak flow. Compared to that of SiBUC, the MRI-SiB peak discharge is closer to observation despite mostly showing a lower discharge. To enhance the SiBUC performance, we selected SiBUC to emulate MRI-SiB to discover which runoff generation processes affect these differences and to alleviate SiBUC limitation by incorporating MRI-SiB advantages.

APPENDIX B

Spatial Pattern of the Total Soil Depth

Figure B1 shows a spatial pattern of total soil depth of MRI-SiB and SiBUC for simulations in section 4 and experiments 2 and 3 in section 5.

APPENDIX C

Effects of Model Settings on Energy Budget

Figure C1 shows the climatological mean of the energy budget driven by MRI-AGCM 3.2S (section 4) and observed forcing (section 5). When the model settings were modified, both latent and sensible heat fluxes changed slightly. The Bowen ratio (B), which represents the ratio of sensible to latent heat, expresses the partitioning of the net radiation at a surface. A study by Kim et al. (2014) reported that the mean monthly Bowen ratio in the northern Thailand varies with seasons: in wet seasons, the range of B is 0.25 ± 0.08, whereas, in dry seasons, it is 0.83 ± 0.28. Based on this report,

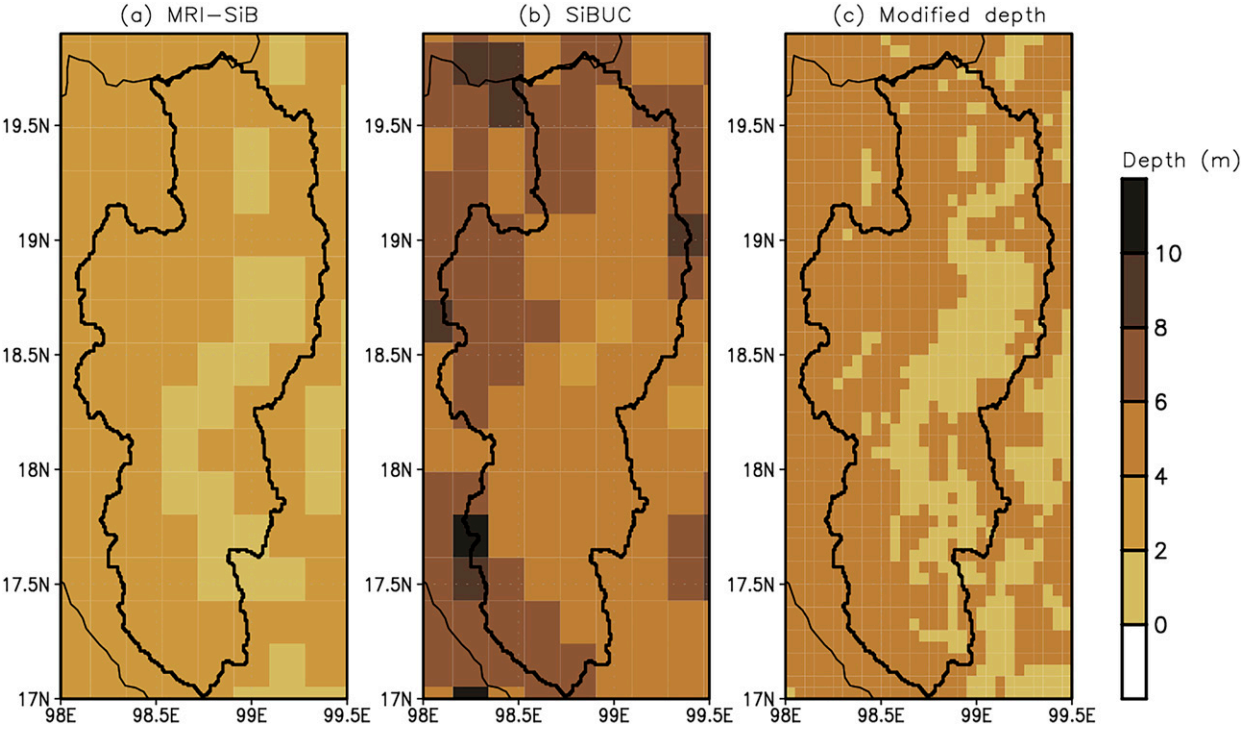

FIG. B1. Spatial pattern of total soil depth of (a) MRI-SiB, (b) SiBUC, and (c) modified depth of experiments 2 and 3 in section 5.

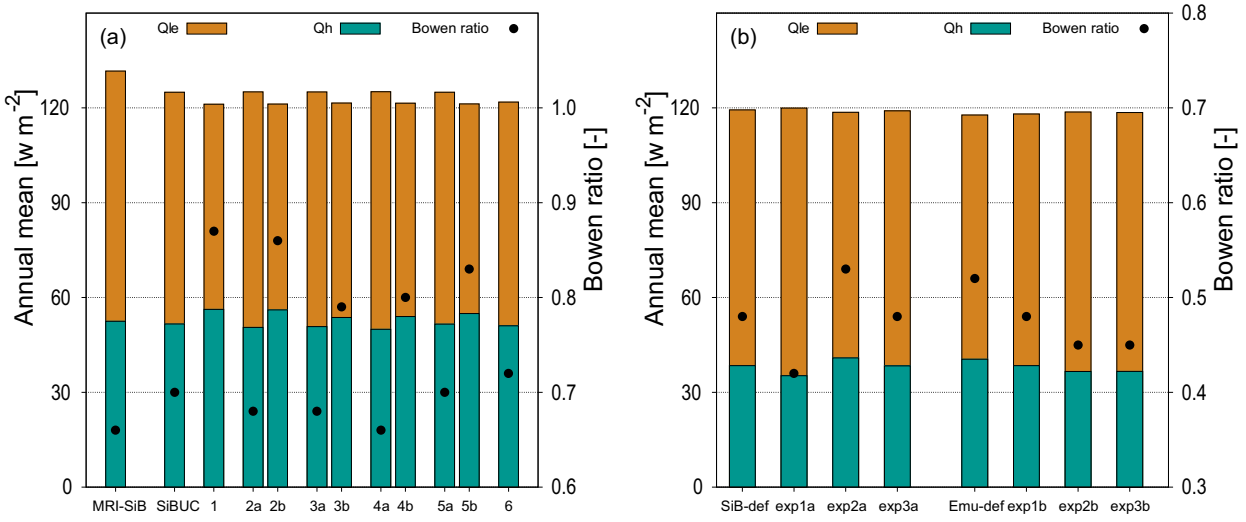

FIG. C1. The 10-yr-mean energy budget (W m^{-2}) driven by (a) MRI-AGCM 3.2S and (b) observed forcing. The brown and green bars represent the latent heat (Qle) and sensible heat (Qh), respectively. The points represent the Bowen ratio (sensible heat to latent heat ratio).

the energy fluxes estimated by changing the model settings in this study appear to be within a reasonable range.

APPENDIX D

Performance Indices

We used the following indices to evaluate the simulated daily discharge over 10 years. In the following equations, M_i and O_i represent the simulated and observed discharges, respectively, and $\overline{M}$ and $\overline{O}$ represent the mean value of each.

a. Nash–Sutcliffe efficiency

The NSE is a popular metric for evaluating the performance of hydrological models, calculated as one minus the model's error variance divided by the variance of the observation. The NSE values range from $-\infty$ to 1, with one being the best fit and values less than zero indicating poor performance:

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}(O_i - M_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2}.$$

b. Root-mean-square coefficient–observations standard deviation ratio

The RSR was calculated as the ratio of the root-mean-square error and standard deviation of measured discharge. It had a best RSR value of 0, with lower scores indicating better model performance:

$$\mathrm{RSR} = \frac{\sqrt{\sum_{i=1}^{N}(O_i - M_i)^2}}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2}}.$$

c. Bias

Bias is the difference between the predicted and observed streamflow and indicates whether the simulated value is greater or less than the reference value. It had an optimal value of zero:

$$\mathrm{bias} = \frac{\sum_{i=1}^{N}(M_i - O_i)}{\sum_{i=1}^{N} O_i}.$$

d. Coefficient of variation

The CV of a time series is the ratio of the standard deviation δ to the mean of that time series. The equation below is the δ of the time series of simulated discharge (δ_M), and the observed discharge was calculated in the same manner:

$$\mathrm{CV}_M = \frac{\delta_M}{\overline{M}}, \qquad \delta_M = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(M_i - \overline{M})^2}.$$

e. Pearson correlation coefficient

The r calculates the degree of a linear relationship between observed and simulated data. If $r = 0$, no linear relationship exists, whereas if $r = 1$ or -1, there is a perfect positive or negative linear relationship:

$$r = \frac{1}{N-1}\frac{\sum_{i=1}^{N}(M_i - \overline{M})(O_i - \overline{O})}{\delta_M \delta_O}.$$

f. Kling–Gupta efficiency

The KGE assesses the agreement between the simulated and observed discharge. It divides the mean squared model error into three parts: r, the variability ratio of the simulated

to observed discharge, and the bias of the simulated discharge. Each of these elements has an ideal value of one:

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + \left(\frac{\mathrm{CV}_M}{\mathrm{CV}_O} - 1\right)^2 + \left(\frac{\overline{M}}{\overline{O}} - 1\right)^2}.$$

REFERENCES

Asanuma, J., D. Ojima, Ailikun, and K. Yorozu, 2013: Intercomparison of land surface process modelling in Asian drylands. *APN Sci. Bull.*, **3**, 36–44, https://doi.org/10.30852/sb.2013.36.

Best, M. J., and Coauthors, 2015: The plumbing of land surface models: Benchmarking model performance. *J. Hydrometeor.*, **16**, 1425–1442, https://doi.org/10.1175/JHM-D-14-0158.1.

Boone, A., and Coauthors, 2004: The Rhône-aggregation land surface scheme intercomparison project: An overview. *J. Climate*, **17**, 187–208, https://doi.org/10.1175/1520-0442(2004)017<0187:TRLSSI>2.0.CO;2.

——, and Coauthors, 2009: The AMMA Land Surface Model Intercomparison Project (ALMIP). *Bull. Amer. Meteor. Soc.*, **90**, 1865–1880, https://doi.org/10.1175/2009BAMS2786.1.

Clapp, R. B., and G. M. Hornberger, 1978: Empirical equations for some soil hydraulic properties. *Water Resour. Res.*, **14**, 601–604, https://doi.org/10.1029/WR014i004p00601.

Cosby, B. J., G. M. Hornberger, R. B. Clapp, and T. R. Ginn, 1984: 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.

Dottori, F., and Coauthors, 2018: Increased human and economic losses from river flooding with anthropogenic warming. *Nat. Climate Change*, **8**, 781–786, https://doi.org/10.1038/s41558-018-0257-z.

Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, https://doi.org/10.1038/sdata.2015.66.

Getirana, A., A. Boone, C. Peugeot, and ALMIP2 Working Group, 2017: Streamflows over a West African Basin from the ALMIP2 model ensemble. *J. Hydrometeor.*, **18**, 1831–1845, https://doi.org/10.1175/JHM-D-16-0233.1.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: 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.

Henderson-Sellers, A., A. J. Pitman, P. K. Love, P. Irannejad, and T. H. Chen, 1995: The Project for Intercomparison of Land Surface Parameterisation Schemes (PILPS): Phases 2 and 3. *Bull. Amer. Meteor. Soc.*, **76**, 489–504, https://doi.org/10.1175/1520-0477(1995)076<0489:TPFIOL>2.0.CO;2.

Hirabayashi, Y., R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae, 2013: Global flood risk under climate change. *Nat. Climate Change*, **3**, 816–821, https://doi.org/10.1038/nclimate1911.

Hirai, M., T. Sakashita, H. Kitagawa, T. Tsuyuki, M. Hosaka, and M. Oh'izumi, 2007: Development and validation of a new land surface model for JMA's operational global model using the CEOP observation dataset. *J. Meteor. Soc. Japan*, **85A**, 1–24, https://doi.org/10.2151/jmsj.85A.1.

Hunukumbura, P. B., and Y. Tachikawa, 2012: River discharge projection under climate change in the Chao Phraya River Basin, Thailand, using the MRI-GCM3.1S dataset. *J. Meteor. Soc. Japan*, **90A**, 137–150, https://doi.org/10.2151/jmsj.2012-A07.

IPCC, 2022: Summary for policymakers. *Climate Change 2022: Impacts, Adaptation, and Vulnerability*, H.-O. Pörtner et al., Eds., Cambridge University Press, 3–33.

Kim, W., D. Komori, J. Cho, S. Kanae, and T. Oki, 2014: Long–term analysis of evapotranspiration over a diverse land use area in northern Thailand. *Hydrol. Res. Lett.*, **8**, 45–50, https://doi.org/10.3178/hrl.8.45.

Kobayashi, C., and T. Iwasaki, 2016: Brewer–Dobson circulation diagnosed from JRA-55. *J. Geophys. Res. Atmos.*, **121**, 1493–1510, https://doi.org/10.1002/2015JD023476.

Koster, R. D., and P. C. D. Milly, 1997: The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models. *J. Climate*, **10**, 1578–1591, https://doi.org/10.1175/1520-0442(1997)010<1578:TIBTAR>2.0.CO;2.

Masson, V., J.-L. Champeaux, F. Chauvin, C. Meriguet, and R. Lacaze, 2003: A global database of land surface parameters at 1-km resolution in meteorological and climate models. *J. Climate*, **16**, 1261–1282, https://doi.org/10.1175/1520-0442-16.9.1261.

Materia, S., P. A. Dirmeyer, Z. Guo, A. Alessandri, and A. Navarra, 2010: The sensitivity of simulated river discharge to land surface representation and meteorological forcings. *J. Hydrometeor.*, **11**, 334–351, https://doi.org/10.1175/2009JHM1162.1.

Menard, C. B., and Coauthors, 2021: Scientific and human errors in a snow model intercomparison. *Bull. Amer. Meteor. Soc.*, **102**, E61–E79, https://doi.org/10.1175/BAMS-D-19-0329.1.

Mizuta, R., and Coauthors, 2012: Climate simulations using MRI–AGCM 3.2 with 20-km grid. *J. Meteor. Soc. Japan*, **90A**, 233–258, https://doi.org/10.2151/jmsj.2012-A12.

Nakaegawa, T., and M. Sugi, 2001: Impact of soil moisture movement schemes in a SVATS on global climate of AGCM. *IAHS Publ.*, **270**, 47–51.

Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual model part I - A discussion of principles. *J. Hydrol.*, **10**, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6.

Nohara, D., A. Kitoh, M. Hosaka, and T. Oki, 2006: Impact of climate change on river discharge projected by multimodel ensemble. *J. Hydrometeor.*, **7**, 1076–1089, https://doi.org/10.1175/JHM531.1.

Pavelic, P., and Coauthors, 2012: Balancing-out floods and droughts: Opportunities to utilize floodwater harvesting and groundwater storage for agricultural development in Thailand. *J. Hydrol.*, **470–471**, 55–64, https://doi.org/10.1016/j.jhydrol.2012.08.007.

Pitman, A. J., and Coauthors, 1999: Key results and implications from phase 1(c) of the Project for Intercomparison of Land-Surface Parametrization Schemes. *Climate Dyn.*, **15**, 673–684, https://doi.org/10.1007/s003820050309.

Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher, 1986: A Simple Biosphere model (SiB) for use within general circulation models. *J. Atmos. Sci.*, **43**, 505–531, https://doi.org/10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2.

Tanaka, K., 2005: Development of the new land surface scheme SiBUC commonly applicable to basin water management and numerical weather prediction model. Ph.D. thesis, Kyoto University, 292 pp.

Tinumbang, A. F. A., K. Yorozu, Y. Tachikawa, Y. Ichikawa, H. Sasaki, and T. Nakaegawa, 2020: Impacts of model structures and soil parameters on runoff characteristics in land surface

models. *J. Japan Soc. Civ. Eng.*, **76B1**, I_217–I_222, https://doi.org/10.2208/jscejhe.76.2_I_217.

——, ——, ——, ——, ——, and ——, 2021: Investigating the impacts of different time integration methods in land surface models on runoff estimation. *J. Japan Soc. Civ. Eng.*, **77B1**, I_253–I_258, https://doi.org/10.2208/jscejhe.77.2_I_253.

——, ——, ——, and ——, 2022: Sensitivity analysis of the runoff in the land surface models forced by the output of MRI–AGCM 3.2 climate model. *THA 2022 Int. Conf. on Moving towards a Sustainable Water and Climate Change Management after COVID-19*, Online, Chulalongkorn University, TA-105S, https://project-wre.eng.chula.ac.th/aseanacademicnetwork/files/THA2022/TA%20on%20web/TA-105S.pdf.

Yorozu, K., and Y. Tachikawa, 2015: The effect on river discharge estimation by considering an interaction between land surface process and river routing process. *Proc. Int. Assoc. Hydrol. Sci.*, **369**, 81–86, https://doi.org/10.5194/piahs-369-81-2015.