corrected Proof

Hydrology Research

Hydrology Research Vol 00 No 0, 1 doi: 10.2166/nh.2025.021

Rainfall–runoff modelling in a tropical climate (Citarum basin, West Java, Indonesia): Model and climate data intercomparison

Gianni Vesuviano [a,*], Simon A. Mathias [b], Rahmawati Rahayu [c,d] and Sim M. Reaney [e,f]

[a] Water and Climate Science, UK Centre for Ecology & Hydrology, Wallingford, UK
[b] Department of Engineering, Durham University, Durham, UK
[c] Department of Meteorology, Institut Teknologi Bandung, Bandung, Indonesia
[d] Department of Earth Sciences, Durham University, Durham, UK
[e] Department of Geography, Durham University, Durham, UK
[f] Institute of Hazard, Risk and Resilience, Durham University, Durham, UK
*Corresponding author. E-mail: giaves@ceh.ac.uk

GV, 0000-0003-2157-8875; SAM, 0000-0003-3054-9056; SMR, 0000-0003-3063-2044

ABSTRACT

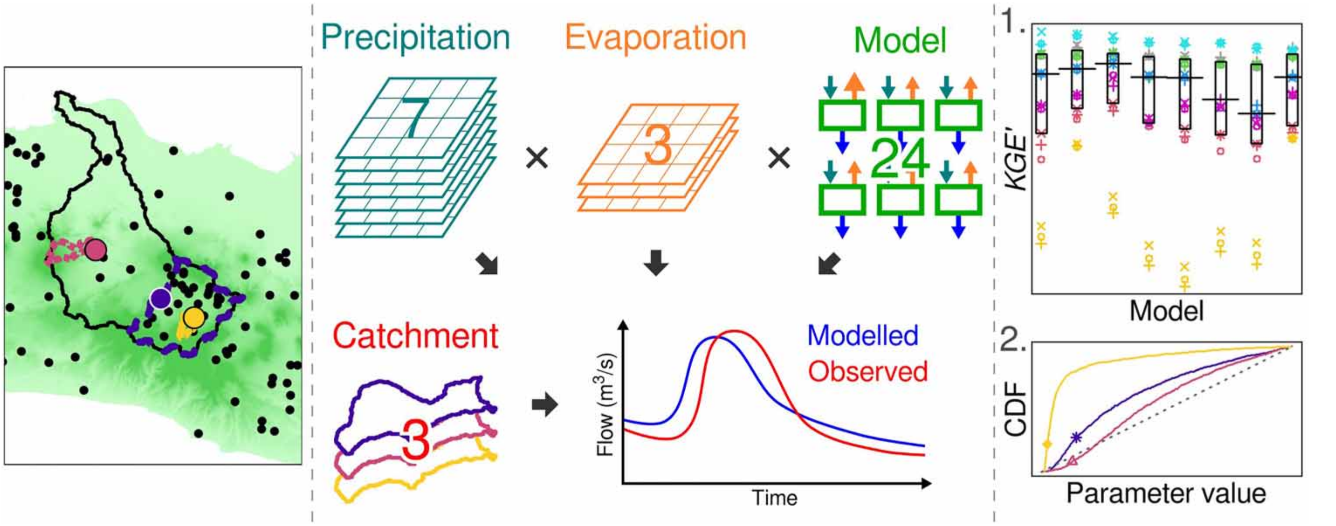
Hydrological research in tropical regions has been limited by a lack of high-quality precipitation and evapotranspiration input data, and high-quality observed runoff calibration data. Hence, most hydrological models have been developed for non-tropical regions and calibrated on non-tropical climatic data. To address these challenges, we assessed the performance of 24 rainfall–runoff models in combination with seven precipitation and three evapotranspiration datasets (504 combinations), in three catchments in the Citarum basin, Indonesia, to identify which models and input datasets were best suited to modelling daily mean flow in a volcanic region and tropical climate, and to assess the parameter sensitivity of the most suitable models. We found that model performance was most strongly affected by choice of input precipitation data and that the most suitable option was different for each of the three study catchments, varying over tens of kilometres. Hydrological model structure played a less significant role, and models developed explicitly for volcanic or tropical regions did not show a clear advantage over those designed for other regions or climates. However, models with two separate, parallel, fast and slow routing pathways had higher predictive skill than those with just one pathway in representing the year-round range of flow behaviours.

Key words: climate data, model intercomparison, parameter sensitivity, rainfall–runoff model, tropics, volcanic region

HIGHLIGHTS

- Large model and gridded climate data intercomparison in a tropical volcanic region.
- Catchment runoff simulation performance depends most on precipitation suitability.
- The suitability of precipitation datasets can change over tens of kilometres.
- Models developed for tropical or volcanic regions do not outperform other models.
- Minimum model complexity requires two separate, parallel internal flow paths.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Conceptual or lumped rainfall–runoff modelling is often used in situations where limited computing power and/or knowledge prohibit more detailed modelling of the catchment(s) of interest. These models simplify and conceptualize the main runoff-generating processes and can be used to provide input boundaries to more detailed hydrodynamic models, assess water resources and storage, and estimate the relationships between flood peaks and rainfall events under different land-use or climatic conditions (e.g. Mishra *et al.* 2017; Rickards *et al.* 2020; Zhang *et al.* 2023; Li *et al.* 2024). However, most hydrological models have been developed to model runoff generation in non-tropical regions and calibrated on non-tropical climatic data. For example, of the 46 rainfall–runoff models (RRMs) described and implemented by Knoben *et al.* (2019), only MODHYDROLOG (Chiew 1990; Chiew & McMahon 1994) and SIMHYD (Chiew *et al.* 2002) were originally developed using data recorded at a handful of gauging stations in northern Australia and the SMAR (Soil Moisture Analytical Relationship) model (Kachroo 1992) was developed using data from one catchment in Malaysia. In all three cases, non-tropical catchments were also used, and tropical catchments were outnumbered in the development dataset. Hence, the performance or suitability of RRMs in tropical regions is only very rarely considered during the development phase and often in conjunction with performance and suitability in other climate groups, despite tropical regions experiencing higher annual precipitation and evapotranspiration totals, more intense precipitation events, and often strongly seasonal precipitation, in comparison with other climate types. This bias against tropical regions exists despite approximately 40% of the world's population living in tropical countries (though not always under a tropical climate), a percentage expected to reach 50% by 2050 (State of the Tropics 2014).

The suitability of well-known hydrological models for tropical catchments has only received limited attention. Petheram *et al.* (2012) compared five models in 105 catchments in northern Australia for modelling daily mean flows, finding that RRMs designed for temperate climates may not be suited to the dramatic fluctuations in streamflow within and between wet seasons in tropical climates but that the same general modelling strategies applied to temperate climates are applicable to tropical (savanna) regions. Chiew *et al.* (2018) modelled daily mean flows in 780 Australian catchments, including a handful under a tropical monsoon climate in the northeast. Model performance was poorest in northern Australia, though this was attributed to the larger typical catchment size compared to other regions. All models were able to reproduce high flows but struggled with low flows – both occur annually in monsoon-driven climates. Desclaux *et al.* (2018) modelled four small (32–51 km^2) and mountainous catchments in New Caledonia, being able to model both flash floods at hourly timesteps and large annual variations. However, the models were less reliable in catchments where land cover had changed over the study period.

Despite the challenges posed by tropical catchments, numerous studies in Indonesia have been undertaken to explore the complexities of rainfall–runoff modelling in these environments. Gunawan (2021) compared flood peak magnitudes and timings for the Air Bengkulu watershed estimated using different synthetic unit hydrographs and HEC–HMS (Hydrologic Engineering Center-Hydrologic Modeling System) but did not recommend any one over any other. Yanto *et al.* (2017a) used the VIC model with gridded precipitation, temperature, and wind speed products (Yanto *et al.* 2017b) to estimate daily mean flows in five catchments in Java but had to reject 14 potential catchments due to poor observed flow data quality. Of the accepted catchments, calibration and validation Nash–Sutcliffe efficiency (Nash & Sutcliffe 1970) ranges were 0.31–0.89 and 0.07–0.79, respectively. Harlan *et al.* (2010) compared the GR4J and NRECA RRMs for calibration and validation periods in the Citarum Hulu basin. Nash–Sutcliffe efficiency was in the range of 0.73–0.78 for both models during calibration and validation periods. Van den Brink (2009) used HBV (Hydrologiska Byråns Vattenbalansavdelning) to model the daily discharge of the Cidanau River in Java. The model underestimated large peak flows and could not reproduce an observed constant baseflow. Jones *et al.* (2002) used PDM (Probability Distributed Model) to model 15-min river flow at several gauging stations in the Citanduy basin, but the ability to calibrate the model was strongly limited by the low density and unrepresentativeness of available rain gauge data and discontinuities in the 15-min gauged river flow records.

The low availability of high-quality observed runoff data for tropical (and arid and polar) catchments has limited the improvements that can be made in rainfall–runoff modelling outside temperate and continental climates (Beck *et al.* 2016). The reduced abundance of historical river gauging data can largely be attributed to the fact that tropical countries tend to have benefited from lower levels of historical economic development compared to temperate countries. Petheram *et al.* (2012) identified that, even within Australia, a more economically developed country, tropical regions suffer from more erroneous rainfall and evaporation data. Consequently, comparatively less research into rainfall–runoff modelling has focused on tropical climates. The availability of various gridded precipitation and evapotranspiration products can mitigate a lack of equivalent gauged data in sparsely gauged regions, but work to assess the relative suitability of different gridded products is ongoing. Suroso *et al.* (2023) assessed the suitability of TRMM rainfall data for hydrological modelling using three contrasting catchments in Java, Indonesia, and found rain gauge data to be systematically superior. Sekaranom *et al.* (2018) found TRMM data to underestimate extreme daily rainfall totals in Central Java, while Senjaya *et al.* (2020) found TRMM data to underestimate wet-season monthly totals in East Java.

There has been limited research on the integrated problem of weather data and hydrological model structure. Therefore, this study evaluates the suitability of common hydrological models and a range of different gridded precipitation and evapotranspiration datasets in tropical climates, using high-quality river flow data obtained from three gauging stations in the Citarum basin, West Java, Indonesia, for calibration. Seven rainfall and three evapotranspiration products are combined with 24 RRMs. Our aim is to identify suitable RRMs and provide an empirical basis for their use in this tropical volcanic region. The Citarum basin has been selected because of its relatively high density of hydrological monitoring, with one river flow gauge per approximately 290 km^2. Furthermore, land uses in the gauged catchments of the Citarum basin are common in many other tropical monsoon regions around the world, consisting mainly of forest and cropland (Tanaka *et al.* 2021; Rahayu *et al.* 2023). Additionally, relatively low levels of urban development make these catchments more suitable for conceptual hydrological modelling.

The aims of this study were to identify the most appropriate hydrological models and input datasets for modelling daily mean flow in a volcanic region under a tropical climate and to assess the importance of individual model parameters in the most suitable models.

2. STUDY AREA

The Citarum basin is located in West Java province on Java in Indonesia, approximately between 107° and 108° east and 6 and 7¼° south (Figure 1). The total catchment area is 6,600 km^2 (Fulazzaky 2010; Yoshida *et al.* 2017). The southern part of the basin is ringed by several volcanoes, the highest of which, Mandalawangi, peaks at 3,019 m above sea level (Lehner *et al.* 2008). These volcanoes surround a plateau at approximately 650–700 m, on which sits the largest metropolitan area in the catchment, Bandung, with an estimated population of 8.9 million as of mid-2022. The Citarum descends through three reservoirs north and east of Bandung, then flows across a large alluvial plain at near sea level before reaching the eastern edge of Jakarta Bay. Land cover in the Citarum basin is typical of West Java, with urban and wetland farming (mostly rice paddies) in the lowlands, dryland farming (non-irrigated seasonal crops) in the adjacent lower slopes, tea plantations in the higher slopes and primary forest on the mountain peaks (Rahayu *et al.* 2023).

The basin's climate is predominantly tropical rainforest and tropical monsoon (Köppen groups Af and Am), with small areas of tropical savannah (Aw) at the coast and temperate climate (Cfb) in the highest mountains (Beck *et al.* 2018). The majority of annual rainfall is driven by the Indo-Australian monsoon, which causes streamflow to fluctuate dramatically during the year. Haylock & McBride (2001) suggest that wet-season rainfall is especially, and perhaps inherently,

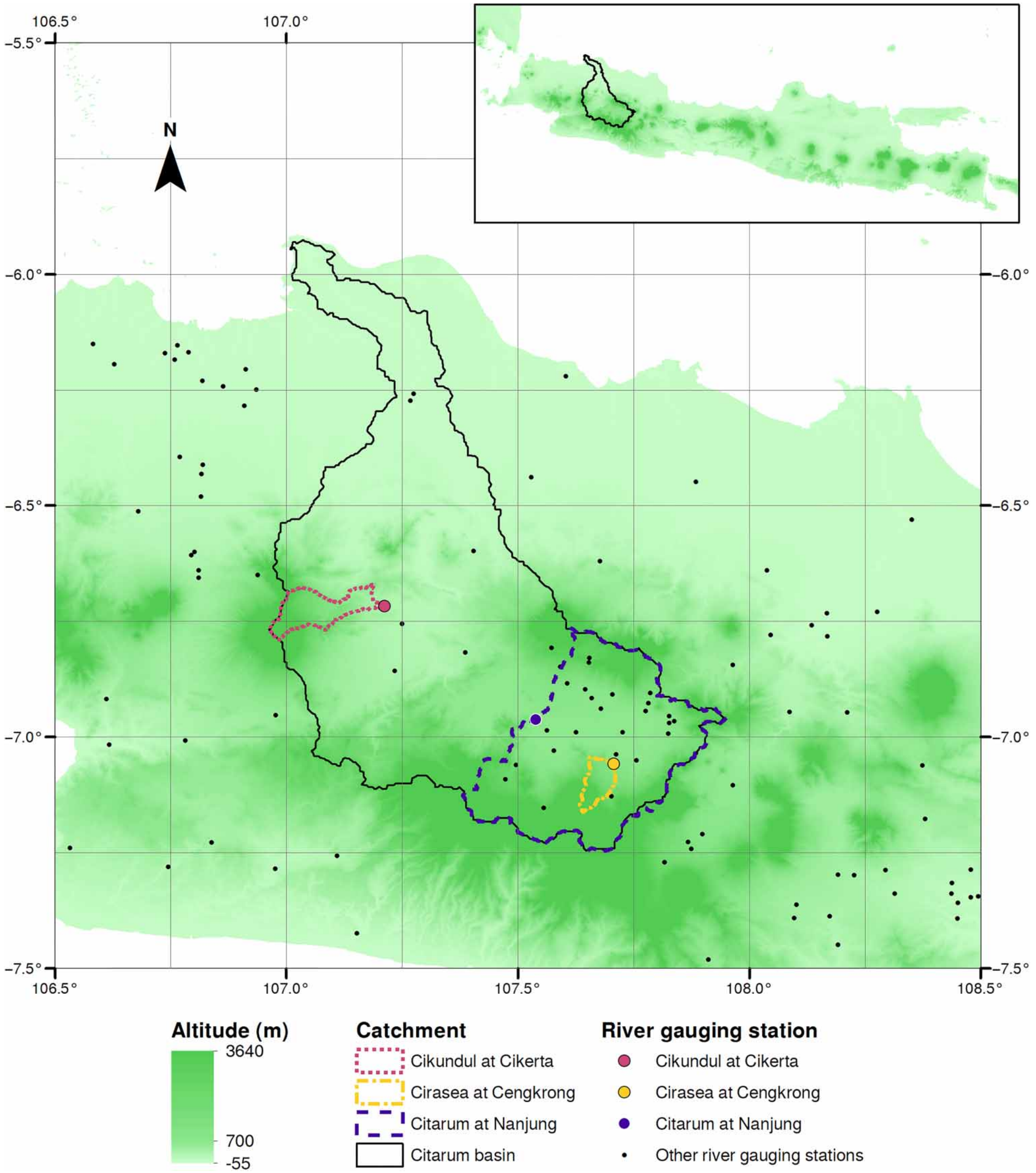


Figure 1 | Map of Citarum basin and gauging stations: Citarum at Nanjung (purple outline), Cikundul at Cikerta (pink outline), and Cirasea at Cengkrong (yellow outline).

unpredictable in Indonesia, and global climate models disagree on how climate change is expected to affect Indo-Australian Monsoonal rainfall over Indonesia (Jourdain *et al.* 2013; Narsey *et al.* 2020). However, a significant international focus on improving the understanding and prediction of atmospheric and oceanic multiscale variability in the Maritime Continent

(Indonesia, Borneo, New Guinea, the Philippines, Malay Peninsula and surrounding seas) has developed in recent years (Yoneyama & Zhang 2020), and a more recent regional climate model has predicted increases in mean and extreme precipitation over land in the Maritime Continent (Argüeso *et al.* 2022), with decreases in both over the sea. Regardless, Indonesia already experiences regular flooding, with 6,098 damaging flood events reported over 2020–2023 (BNPB 2024), a mean of over 1,500 per year.

3. DATA

3.1. Gauged streamflow

Gauged daily mean flows were acquired from the Center for Hydrology and Water Management (Indonesia) for three gauging stations in the Citarum basin: Citarum at Nanjung (58.8 years during 1918–2016), Cirasea at Cengkrong (24.3 years during 1984–2012) and Cikundul at Cikerta (10.1 years during 1998–2011). There are no apparent changes to the flow regime in any specific period of record, despite the recent and rapid urbanization of the Citarum basin. This stationarity could indicate that the study catchments remained rural, that they remained at the same urbanization level, or that their flow regimes are heavily managed. The third possibility presents a challenge to rainfall–runoff modelling, as heavy management of flows implies that discharge varies in ways that do not relate to measurable natural processes (Vu *et al.* 2023). However, non-stationary flow regimes also present a challenge, as they imply that the model parameters must vary over time to account for changes in local climate or land use/land cover. Time-varying parameter values were not considered in this study due to additional challenges in deciding which parameters should vary with time, the nature of each variation, and the additional variables needed to specify time-varying (versus constant) parameters.

3.2. Digital elevation model

The HydroSHEDS 15-arcsecond digital elevation model (DEM) (Lehner *et al.* 2008) was used to generate boundaries for the catchment area upstream of each gauging station, after first snapping the location of each station to the nearest location on the mainstream network defined by the HydroSHEDS grids. Nanjung is the largest catchment, with an area of 1,817 km^2. It has been the subject of several previous hydrological and flood studies of the Citarum basin, often alone (e.g. Harlan *et al.* 2010; Nastiti *et al.* 2015; Julian *et al.* 2019; Hatmoko *et al.* 2020; Rusli *et al.* 2021). Cikerta and Cengkrong are considerably smaller, with areas of 160.6 and 67.7 km^2, respectively.

3.3. Precipitation and evapotranspiration data

All RRMs require input rainfall data from which to generate modelled runoff hydrographs. Continuous-simulation models also require evapotranspiration data to model the drying of water storage elements between precipitation events. Many also require temperature data to model snowfall, thawing, and freezing, and temperature data may also be used to estimate evapotranspiration if it is not supplied.

Precipitation data were obtained from seven gridded precipitation products with daily temporal resolution and various spatial resolutions and start/end dates (Table 1).

AgMERRA and Yanto also contained gridded temperature and wind speed data at the same resolution as their precipitation data, while SA-OBS contained gridded temperature (but not wind speed) data; each was used separately to provide three estimated evapotranspiration time series for this study using methods as described by Mathias (2023, p. 423). The gridded

Table 1 | Precipitation datasets used in this study

Dataset	Spatial resolution	Temporal range	Reference
AgMERRA	0.25	1980–2010	Ruane *et al.* (2015)
CHIRPS	0.05	1981–near present	Funk *et al.* (2015)
CPC	0.5	1979–near present	Xie *et al.* (2010)
MSWEP v2	0.1	1979–near present	Beck *et al.* (2019)
PERSIANN-CDR	0.25	1983–near present	Ashouri *et al.* (2015)
SA-OBS	0.25	1981–2014	van den Besselaar *et al.* (2017)
Yanto	0.125	1985–2014	Yanto *et al.* (2017b)

products were averaged inside the three study catchment boundaries on each timestep to give catchment-averaged precipitation (seven products) and catchment-average evapotranspiration (three products). The seven precipitation and three evapotranspiration datasets have different start and end dates, but all were trimmed to the longest common period of 01-01-1985 to 31-12-2010. Representative precipitation and evapotranspiration years for the full Citarum basin, consisting of the mean of monthly totals for 1985–2010 inclusive, are shown in Figure 2 and Supplementary material, Table S1. Several studies exist comparing different gridded precipitation products in Indonesia (e.g. Vernimmen *et al.* 2012; As-syakur *et al.* 2013; Rahmawati *et al.* 2021; Wiwoho *et al.* 2021; Wati *et al.* 2022), so only the comparisons specific to the Citarum basin are discussed here.

While precipitation throughout the representative year follows the same overall 'monsoonal' pattern in all seven precipitation datasets, there are differences between them. Mean annual rainfall varies from 1,962 mm (CPC) to 2,811 mm (CHIRPS). The mean of all seven yearly totals is 2,359 mm, which is most closely matched by Yanto (2,374 mm), a dataset focused specifically on Java Island, based entirely on rain gauge data without satellite or reanalysis products, and incorporating over 750 gauges, although with relatively sparse coverage in the middle of the Citarum basin. The only other dataset to incorporate many Javanese rain gauges is SA-OBS, which matches Yanto more closely than any other dataset but still shows differences, including a wetter January–March and drier and longer dry season. Correspondence between all seven rainfall datasets is best in the drier months – four agree on August totals of 55–62 mm, while two agree on 42–43 mm – although CHIRPS estimates double this, at 87 mm. Correspondence is worse during the wettest months – AgMERRA and PERSIANN-CDR both estimate a lower total in February than either January or March; this unusual agreement is possibly a consequence of AgMERRA using PERSIANN as one-third of its ensemble 'target' for average monthly precipitation patterns (Ruane *et al.* 2015). CHIRPS estimates the shortest wet season, while CPC estimates the longest. AgMERRA estimates a wet season onset near the average of all seven datasets but still has the second lowest annual total rainfall as its estimated totals for December, January, and February are the lowest of any dataset. Three models, CPC, MSWEP, and SA-OBS, estimate a $>20\%$ increase in monthly totals from December to January, while CHIRPS estimates an 8% decrease. Unlike rainfall, all three evapotranspiration datasets estimate similar totals for each month and the year as a whole, although Yanto estimates over 20 mm more potential evapotranspiration per month than AgMERRA during July–October. Evapotranspiration is relatively stable throughout the representative year, ranging from 107 to 163 mm per month in all datasets and being more than (less than) representative rainfall in the dry (wet) season, respectively.

Figure 3(a) shows the 100 largest catchment-average daily rainfalls in each dataset during 1985–2010. CHIRPS, despite being the wettest dataset overall, is below average in terms of the mean and standard deviation (SD) of its 100 largest

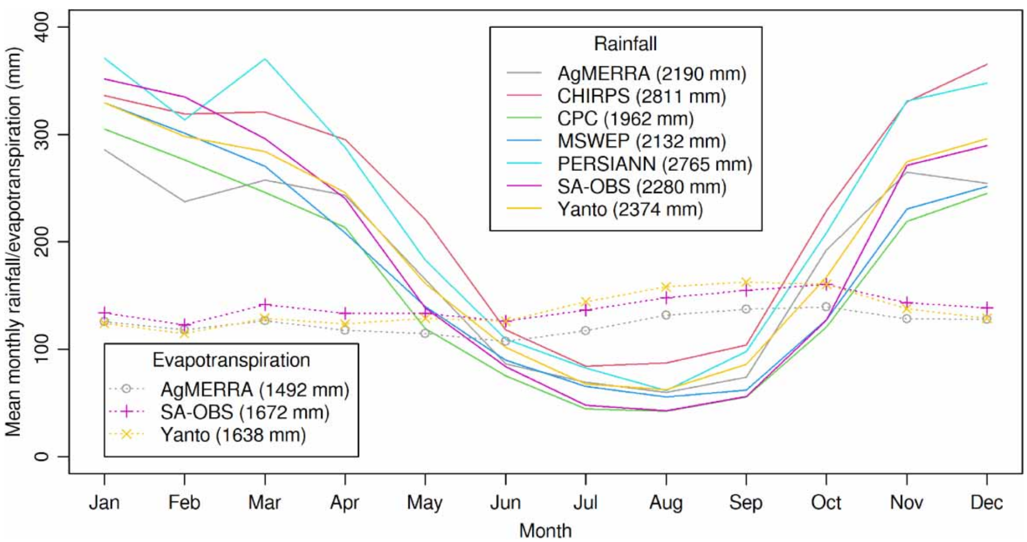


Figure 2 | Mean monthly precipitation and evapotranspiration in the Citarum basin during 1985–2010. Bracketed values in legends indicate the mean annual total during 1985–2010.

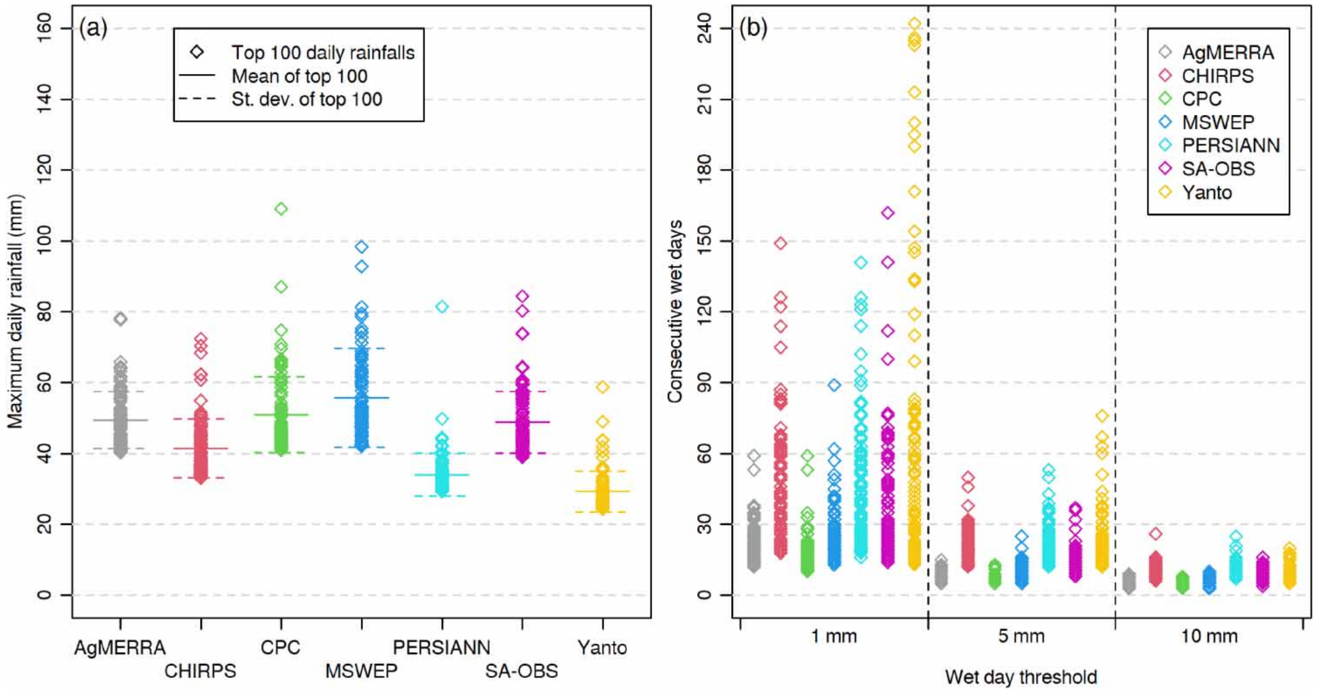


Figure 3 | The 100 largest daily accumulations in each precipitation dataset from 1985 to 2010, including the mean and SD of these accumulations (a). The longest continuous run of wet days per hydrological year in each precipitation dataset for wet day thresholds of 1, 5, and 10 mm (b).

daily rainfalls. AgMERRA, CPC, and SA-OBS have similar means and standard deviations to each other, as do PERSIANN and Yanto. MSWEP has, on average, the largest daily rainfall values of all datasets. In all seven datasets, the mean of the 100 largest events is about one SD larger than the 100th-largest event but several SDs below the largest event.

Figure 3(b) estimates the continuity of the wet season according to each rainfall model by plotting the longest continuous run of days per hydrological year (01 September–31 August) where rainfall exceeds a threshold of either 1, 5, or 10 mm on each day. This shows some similarity with Figure 3(a). For example, Yanto has many more runs of days with >1 mm or >5 mm rainfall than any other model, but this matches its middling annual rainfall total because it also has the smallest top 100 events. This result implies that wet-season rainfall in Yanto is distributed more evenly across days, while in other models, particularly MSWEP, wet-season rainfall is more concentrated in fewer days. In general, there is an inverse relationship between a model's top 100 event sizes and its longest run of days with >1 mm or >5 mm rainfall. This relationship is subdued when a dry/wet threshold of 10 mm is considered because 10 mm/day is approximately the mean daily rainfall for the wet season, and many daily rainfalls must be below the mean by definition. However, the pattern seen using a 10 mm threshold is still similar to those observed using a 1 or 5 mm threshold.

It is important to note that a continuous run of 242 wet days, the maximum seen in Yanto, does not imply 242 days of continuous rain and does not even imply as many rainfall minutes over a whole wet season as a shorter run of consecutive wet days. Due to the monsoonal nature of the rain, it is likely that no rain actually occurred during most hours of each day in any wet season (the storms are likely concentrated within short sub-daily durations). Also, due to the definition of 'consecutive', a single dry day in the middle of any period would be sufficient to halve the number of consecutive wet days. Conversely, a longer-than-24-h dry period would not result in a dry day if it occurred across 2 days that both had rain outside of the dry period.

Table 2 presents the correlation between daily catchment-average precipitation and gauged daily mean flow for all precipitation datasets and all study catchments. In all catchments, precipitation-flow correlation is highest with Yanto, followed by PERSIANN and SA-OBS in Cikerta and Cengkrong; the order of second and third is reversed in Nanjung. In all catchments, precipitation-flow correlation is lowest with AgMERRA and CPC. Overall, correlation ranges from a minimum of 25.5% to a maximum of 52.6%, being under 46% in all except one case.

Table 2 | Correlation between daily catchment-average precipitation and gauged daily mean flow for three catchments and seven gridded precipitation datasets

Catchment	AgMERRA	CHIRPS	CPC	MSWEP	PERSIANN	SA-OBS	Yanto
Nanjung	0.315	0.393	0.308	0.336	0.416	0.459	0.526
Cikerta	0.308	0.364	0.318	0.353	0.389	0.380	0.409
Cengkrong	0.255	0.316	0.255	0.273	0.385	0.338	0.442

4. MODELS AND METHODS

This study compared 24 RRMs, all of which were coded in R (R Core Team 2019). Eighteen of the models followed the diagrams and equations in the supplement of Knoben *et al.* (2019) (Table 3).

As well as the 18 models in Table 3, two additional variants of HyMOD and four variants of the PDM were implemented (Table 4).

Snow and freeze/thaw routines were not implemented for any models as snow is not a realistic occurrence in the Citarum basin (note that MOPEX-2 was not considered as it is equivalent to MOPEX-1 with a snow store). Due to near-constant (i.e. non-seasonal) evapotranspiration rates, MOPEX-4 was slightly modified, with the equation for interception (I) replaced by a constant to be optimized. According to the classifications of Paul *et al.* (2021), all models are conceptual, continuous simulation, catchment-scale, lumped, and deterministic. They are all solved sequentially.

All the hydrological models tested here follow similar overall concepts, consisting of stores that are charged either by rainfall or previous stores and depleted by evapotranspiration, inter-store flow, and/or model outflow, which are all either related to the current volume or depth of water in the store, or are constant. The main differences between the models are the number of stores, their arrangement, and each store's relationship between storage and inter-store flow, model outflow, and/or evapotranspiration.

Table 3 | RRMs following Knoben *et al.* (2019)

Model	No. parameters	Reference
FLEX-I	10	Fenicia *et al.* (2008)
GR4J	4	Perrin *et al.* (2003); Santos *et al.* (2018)
HBV-96	9	Lindström *et al.* (1997)
Hillslope	7	Savenije (2010)
HYCYMODEL	12	Fukushima (1988)
HyMOD	6	Wagener *et al.* (2001); Boyle (2001)
IHACRES	7	Jakeman *et al.* (1990)
MODHYDROLOG	15	Chiew (1990); Chiew & McMahon (1994)
MOPEX-1	5	Ye *et al.* (2012)
MOPEX-3	6	Ye *et al.* (2012)
MOPEX-4	7	Ye *et al.* (2012)
NAM	10	Nielsen & Hansen (1973)
New Zealand v2	8	Atkinson *et al.* (2003)
SIMHYD	7	Chiew *et al.* (2002)
SMAR	8	Tan & O'Connor (1996)
Susannah Brook v1-5	6	Son & Sivapalan (2007)
Thames Catchment Model	6	Greenfield (1984); Moore & Bell (2001)
Xinanjiang	12	Zhao (1992)

Table 4 | Additional RRMs

Model	No. parameters	Notes
HyMOD (exp)	4	As HyMOD, replacing Pareto-distributed soil store with exponentially distributed soil store
HyMOD (1 res)	4	As HyMOD, replacing Pareto-distributed soil store with exponentially distributed soil store, and three serial 'fast' linear reservoirs with one linear reservoir
PDM	9	As Moore (2007)
PDM (exp)	7	As Moore (2007), replacing Pareto-distributed soil store with exponentially distributed soil store
PDM (3 par)	3	As Mathias *et al.* (2016), Mathias (2023, p. 447)
PDM (2 par)	2	As Mathias *et al.* (2016), Mathias (2023, p. 447), replacing a nonlinear routing reservoir with a linear routing reservoir

Jansen *et al.* (2021) noted that several models with identical names may differ internally in their numerical and mathematical formulations (e.g. implicit or explicit solution, sequential or simultaneous fluxes) and that parameter values are not transferrable between different models with the same name. However, it is possible for these different models to achieve very similar performance scores using different parameter sets.

Some of the compared models have been used previously in tropical climates, including PDM by Jones *et al.* (2002), HBV by van den Brink (2009), SMAR, IHACRES and SIMHYD by Petheram *et al.* (2012), GR4J, SIMHYD, and Xinanjiang by Chiew *et al.* (2018), GR4J by Harlan *et al.* (2010), and GR4H, an hourly version of GR4J, by Desclaux *et al.* (2018). Additionally, three of the models (MODHYDROLOG, SIMHYD, and SMAR) included some tropical catchments in their development datasets. However, no model was designed specifically for tropical catchments, and most were developed exclusively or almost exclusively for temperate and/or continental climates.

Model performance was quantified in this study by modified Kling–Gupta efficiency (*KGE*': Kling *et al.* 2012) and its components, corresponding to correlation, bias, and variation between modelled and observed daily flows. Despite its clear potential, few studies consider how the individual components of *KGE*' contribute towards the final score. Missing data periods and the period corresponding to the first three years covered by each input dataset, which were used as the spin-up period for the models' internal reservoirs, were excluded from the calculation.

Models were assessed through a two-stage procedure to allow the identification of the ensemble of behavioural model parameter sets. First, each model's parameterization was optimized via the shuffled complex evolution algorithm (Duan *et al.* 1992, 1994) to find the models with the potential to perform best in the tropical volcanic study region. This was implemented through the R function SCEoptim in the package 'hydromad' (Guillaume 2013). Four optimizations, each using 12 complexes, but with elitism set to 2, 1, 0.5, and 0.1 respectively, were performed for each combination of catchment, model, rainfall, and evapotranspiration data. In every case, we assumed no prior knowledge of the hydrological processes active in the catchments, so no model parameters were fixed and all were optimized. In each case, the parameter values corresponding to the highest *KGE*' were kept. Higher elitism gives more weight to more optimal parent parameter sets when evolving toward optimal parameter values.

The second stage of the assessment followed the GLUE methodology (Beven & Binley 1992) to account for the equifinality in the different parameter sets. A subset of models that performed well during the first stage were selected, and then each was input with one of 100,000 random parameter sets, where each parameter value was sampled randomly from a uniform distribution with limits suggested by Knoben *et al.* (2019). For each case, the subset of 1,000 parameter sets giving the top 1% of *KGE*' values was considered 'behavioural'. For selected models, parameter identifiability was measured using Kolmogorov–Smirnov (*KS*) tests (Massey 1951) to compare the complete set of 100,000 values against the subset of 1,000 behavioural values.

Since all input data were trimmed to the same 1985–2010 period, the first three years (1985–1987) were always used for model spin-up. Calibration and validation periods for each river gauge were defined according to the gauging period, with the first approximately two-thirds of data recorded between 1988 and 2010 used for calibration and the last third used for validation (Table 5). Note that missing data within a year may or may not form a continuous period. The results and discussion presented in this article consider the validation period only.

Table 5 | Period-of-record, calibration period, validation period, and missing data for the Citarum at Nanjung, Cikundul at Cikerta, and Cirasea at Cengkrong gauging stations

Gauge	Period-of-record	Calibration period (non-missing days)	Validation period (non-missing days)	Missing data from 1988 to 2010
Citarum at Nanjung	1918–2016 (58.8 years, 58 complete)	1988–2002 (5,114 days)	2003–2010 (2,556 days)	1989 (full year) 2004 (full year)
Cikundul at Cikerta	1997–2011 (10.1 years, 6 complete)	1998–2006 (2,294 days)	2008–2010 (1,048 days)	1988–1996 (full years) 1997 (215 days) 1998 (29 days) 2001 (36 days) 2002 (full year) 2003 (177 days) 2004 (171 days) 2005 (full year) 2007 (full year) 2009 (48 days)
Cirasea at Cengkrong	1984–2012 (24.3 years, 20 complete)	1988–2001 (4,263 days)	2002–2010 (2,701 days)	1995 (full year) 1996 (full year) 1997 (120 days) 2003 (103 days) 2004 (full year) 2008 (30 days) 2009 (48 days) 2010 (39 days)

5. RESULTS

5.1. Stage 1: Model optimization

Figure 4 plots the optimized *KGE*' achieved by each model with each combination of precipitation and evapotranspiration data, as well as the three component scores contributing to this maximum *KGE*', for the Citarum at Nanjung over the validation period. The point colours and shapes correspond to those used in Figure 2, and the models are ordered left-to-right by descending maximum optimized *KGE*'. Maximum *KGE*' scores per model were relatively similar, with all but the two lowest scores in the 0.65–0.75 range. The highest *KGE*' was most commonly achieved by the use of Yanto precipitation and AgMERRA evapotranspiration, though the model performance was less dependent on the choice of evapotranspiration dataset. Yanto precipitation achieved the highest or near-highest correlations between the observed and modelled daily flows for all models. With few exceptions, the high *KGE*' values resulted mainly from high correlations between the observed and modelled daily flows. Hence, as suggested previously, the Yanto precipitation dataset appears to most accurately represent the timing of actual rainfall events in this catchment (Table 2). Variation in daily flows was generally overestimated, with only SA-OBS precipitation able to match observed variation across all models and evapotranspiration datasets. Yanto precipitation data typically overestimated variation by less than other datasets, except SA-OBS. CHIRPS precipitation resulted in the lowest *KGE*' for most models, primarily because the use of CHIRPS resulted in overly high variability of daily mean flows. This finding is in direct contrast to what would be expected given Rusli *et al.* (2021), which concluded from an assessment of river flow and groundwater changes that CHIRPS had lower uncertainty than TRMM, SA-OBS, and local rain gauge estimates. Almost all combinations of model and dataset resulted in an underestimation of total flow following optimization (bias <1). PERSIANN and CHIRPS typically resulted in both the greatest underestimation in total flow and overestimation in variability, as well as low *KGE*'. Wang *et al.* (2023) reported that hydrological models were more able to adapt to consistent biases in rainfall products than to biases restricted to wet days.

Seven models were able to achieve *KGE*' scores above 0.7, exclusively with Yanto precipitation. The Citarum basin is a tropical and volcanic region, so it is notable that MODHYDROLOG, HYCYMODEL, and New Zealand v2 were within the top four performing models. MODHYDROLOG was developed purely on Australian catchments, some of which were in a tropical climate, while HYCYMODEL and New Zealand v2 were developed in Japan and New Zealand, both containing volcanic regions, though under non-tropical climates.

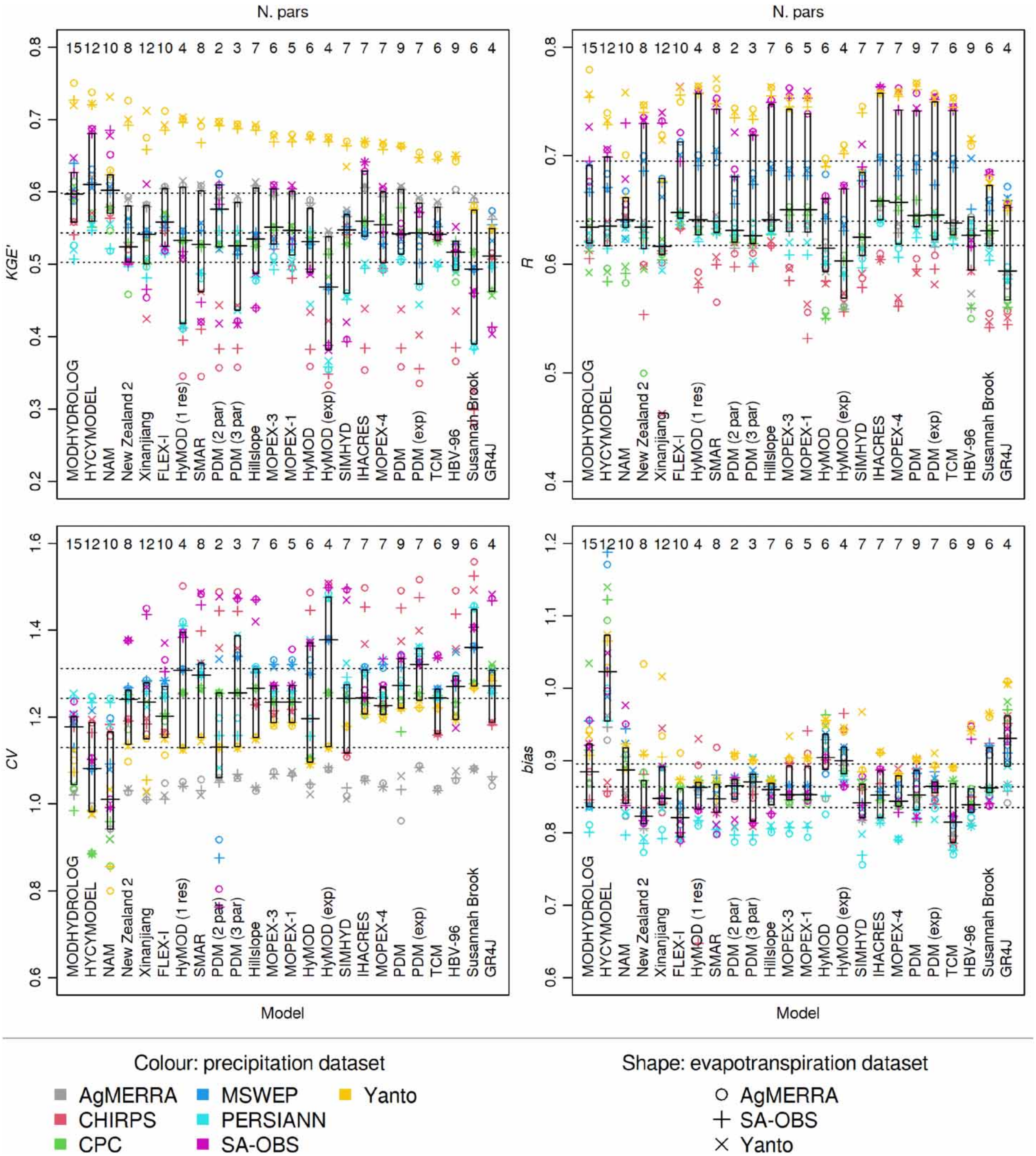


Figure 4 | Maximum *KGE*′ between observed and modelled daily flow achieved at Nanjung gauging station using all combinations of seven precipitation and three evapotranspiration input datasets through 24 models (top left). Components of the maximum *KGE*′ score (correlation coefficient, top right; coefficient of variation, bottom left; bias ratio, bottom right). Solid horizontal lines show the median value per model, boxes show the interquartile range. Dotted horizontal lines show median and interquartile range across all models. The numbers at the top of the bars show the number of model parameters.

Figure 5 is equivalent to Figure 4 but for the Cikerta gauging station. Here, *KGE*' scores were lower than in Nanjung, with only one model achieving a *KGE*' above 0.5. However, maximum *KGE*' scores per model were relatively similar, with all in the 0.42–0.51 range. There was once again a strong link between high *KGE*' and high correlation coefficient, but here,

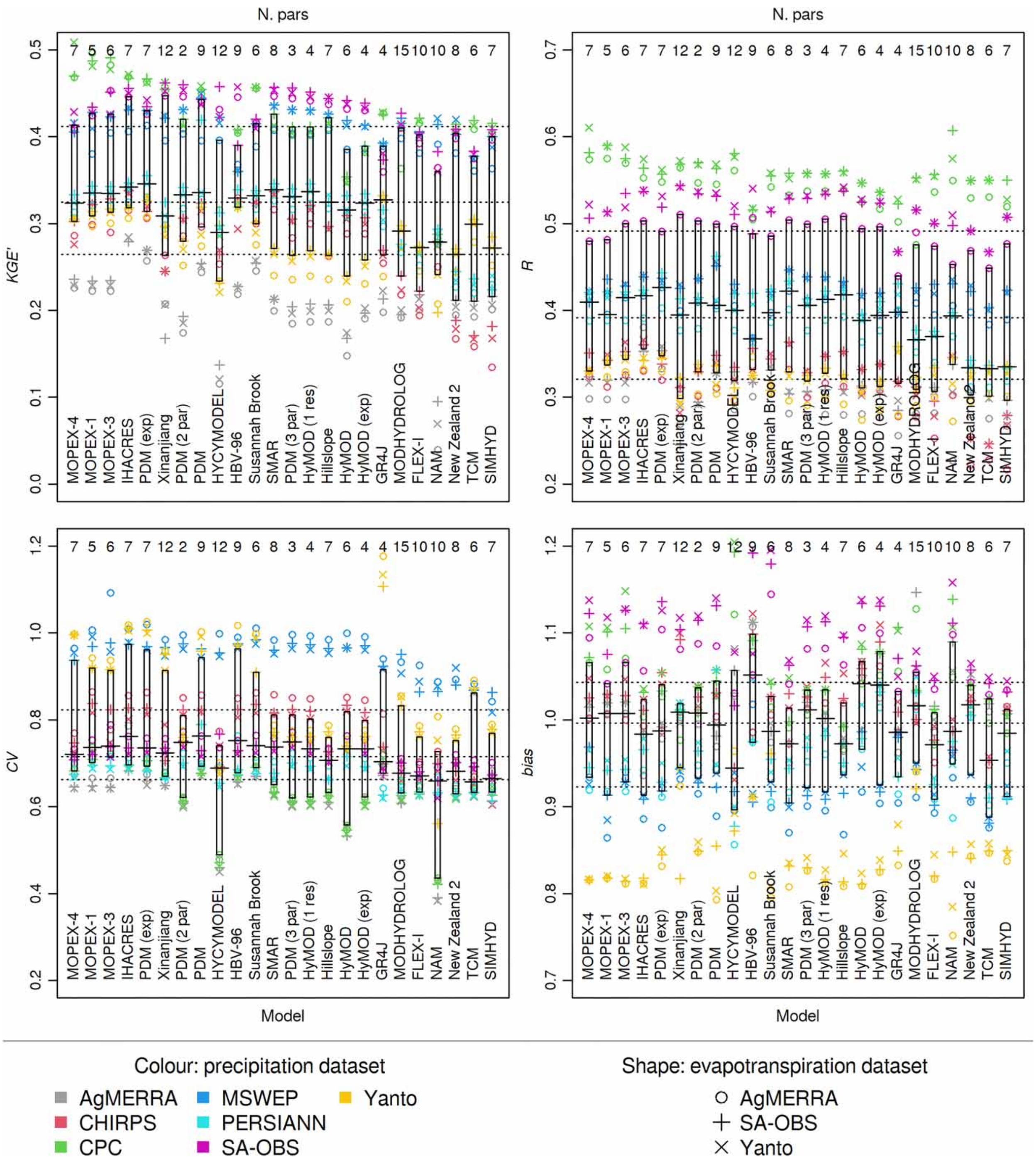


Figure 5 | Maximum *KGE*′ between observed and modelled daily flow achieved at Cikerta gauging station using all combinations of seven precipitation and three evapotranspiration input datasets through 24 models (top left). Components of the maximum *KGE*′ score (correlation coefficient, top right; coefficient of variation, bottom left; bias ratio, bottom right). Solid horizontal lines show the median value per model, boxes show the interquartile range. Dotted horizontal lines show median and interquartile range across all models. The numbers at the top of the bars show the number of model parameters.

SA-OBS and CPC precipitation data generally gave the highest performance, dependent on the model used. The Cikerta catchment is approximately 160 km^2. This is similar in area to (in some cases smaller than) a precipitation data grid cell, so the interpolation choices made by the dataset creators during the gridding procedure may have considerable effects on the extracted catchment-average values. Most combinations of model and climatic input dataset estimated total flow to within 10% but underestimated variation in flows, with only MSWEP regularly approaching coefficient of variation (CV) values of 1. Yanto achieved similarly high accuracy in matching flow variation when used with several of the most successful RRMs. Unlike in Nanjung, MODHYDROLOG and New Zealand v2 place in the bottom half of models ranked by maximum performance, while HYCYMODEL places ninth. Here, six of the eight most successful models are variants of MOPEX and the PDM. Jones *et al.* (2002) achieved inconsistent results using the PDM in Java, which they attributed to discontinuous gauged flow records and unrepresentative rainfall data. The gauged flow records and precipitation data used in this study are longer and more representative of the study area, which may explain the improved PDM performance.

Figure 6 is equivalent to Figures 4 and 5, but for Cengkrong. Maximum *KGE*' scores per model were relatively similar, with all but the one lowest and two highest scores in the 0.60–0.66 range. The use of AgMERRA precipitation generally gave the highest scores, except with the three best-performing models, where PERSIANN precipitation was optimal. Yanto gave the lowest optimized *KGE*' scores in Cengkrong with most models due to modelled flow with a lower correlation to observed flow and overestimated total flows. However, Yanto was the only precipitation dataset that did not result in a consistent overestimation of CV. In general, SA-OBS did not greatly overestimate CV, while PERSIANN did, except when used with the top three performing models. Performance differences between different input datasets in Cengkrong did not seem greater than in Cikerta or Nanjung, despite these being about 10 and 100 times larger than Cengkrong, respectively, and Cengkrong being smaller in area than one grid cell of any precipitation dataset.

The highest *KGE*' scores, both above 0.68, were achieved by two variants of the PDM, with MOPEX variants taking three of the top seven places. Hence, there is some agreement between Cikerta and Cengkrong regarding optimal models that is not shared strongly with Nanjung, even though Cengkrong is nested inside Nanjung. However, the only commonality between the three catchments in terms of optimal input datasets is that the choice of evapotranspiration dataset is unimportant. The varying suitability of precipitation datasets for the three nearby catchments indicates that the suitability of any particular precipitation dataset can change over short distances.

5.2. Stage 2: Parameter identification

The parameter identification study focused on three models that were top or near-top performers in at least two catchments: MOPEX-1, NAM, and two-parameter PDM. PDM parameter names are taken from Moore (2007), and MOPEX-1 and NAM parameter names are taken from Knoben *et al.* (2019).

Figure 7 shows *KS* estimates for the values of the five MOPEX-1 parameters corresponding to the top 1% of *KGE*' values when the model is driven by 100,000 randomly generated parameter sets per catchment. Each catchment model was driven by the precipitation and evapotranspiration datasets that gave the highest *KGE*' when that model was optimized in Stage 1. A low *KS* indicates that a top 1% *KGE*' can be obtained with that parameter at a wide range of values, while a high *KS* indicates that a top 1% *KGE*' can only be obtained with a narrower range of values. *KS* statistics with a p-value of 0.01 or less are italicized in Figure 7. If the empirical cumulative density function of parameter values is most vertical at different values for different catchment models, then that parameter varies between catchments, and performance is sensitive to that parameter.

The performance of MOPEX-1 is most sensitive to the value of S_e, the root zone storage capacity. While all three catchments require low S_e for high performance, the cumulative density function (CDF) and optimized values for Cikerta and Nanjung are more similar to each other than they are to those for Cengkrong. Performance in Cikerta and Nanjung is very sensitive to the value of S_{b1}, maximum soil moisture (although with different optimal values), while performance in Cengkrong is more sensitive to t_u, the time parameter for leakage from the root zone to slow runoff store. Considering the single optimized values of all time constants together, Nanjung has the fastest flow through both the 'fast' and 'slow' paths, while Cikerta has a relatively slow 'fast' path, and Cengkrong has a fast 'fast' path and slow 'slow' path. This suggests that MOPEX-1 behaves as intended in Cengkrong and Nanjung, with reasonably high performance ($KGE' = 0.654$ and 0.679, respectively). All three catchment models are less sensitive to other parameters, but *KS* values are significant at $p = 0.01$ in all cases.

Figure 8 is equivalent to Figure 7 but for NAM. With this model, U^* (maximum upper zone storage), K_1 (time coefficient for interflow routing), and C_{L1} (threshold for interflow routing) are the most important parameters for all three catchments. This implies that correct parameterization of the interflow path is most important for good model performance. The optimized values for these parameters suggest that Cikerta is driven more by flow in the groundwater path, while Nanjung and

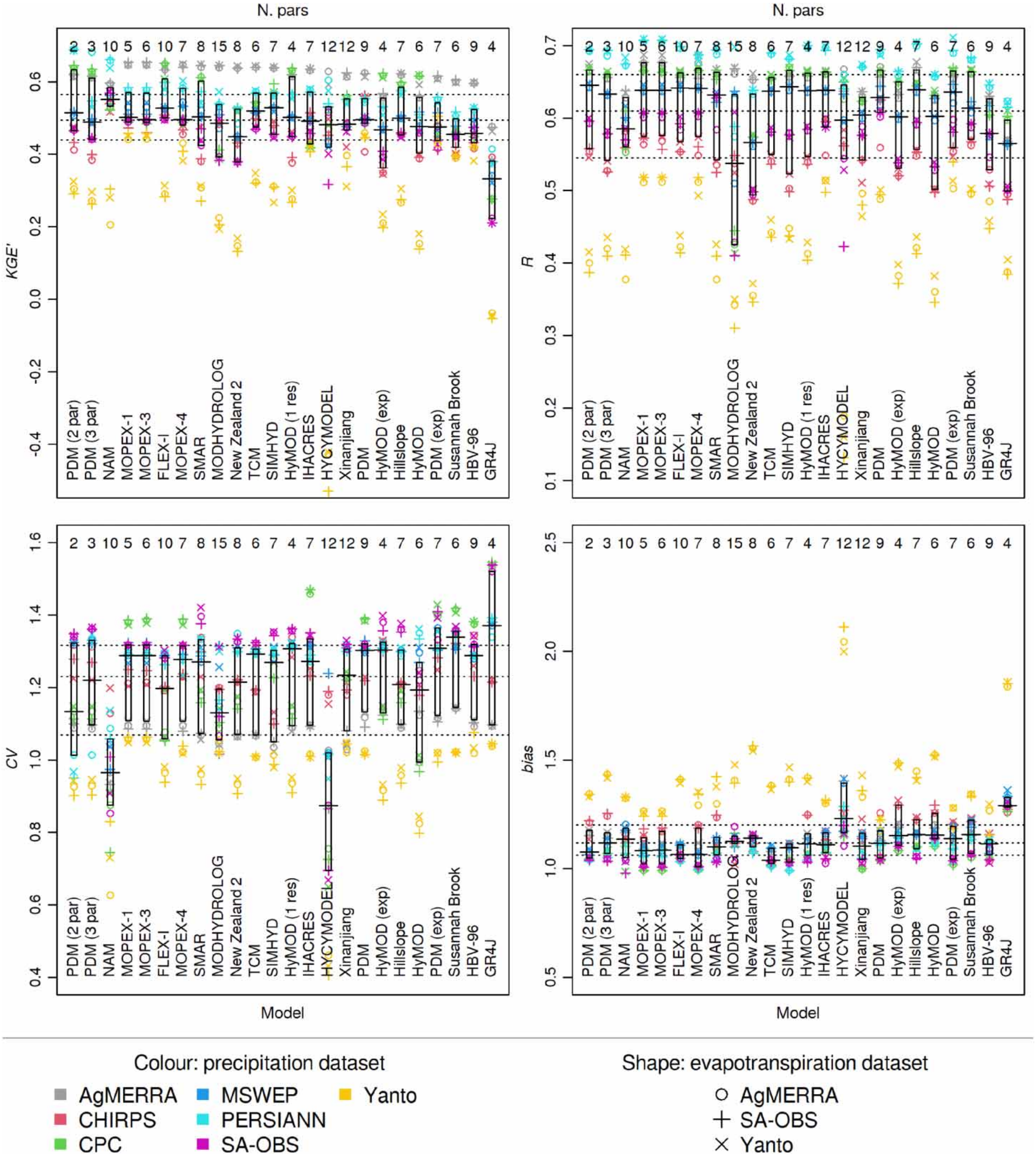


Figure 6 | Maximum *KGE*′ between observed and modelled daily flow achieved at Cengkrong gauging station using all combinations of seven precipitation and three evapotranspiration input datasets through 24 models (top left). Components of the maximum *KGE*′ score (correlation coefficient, top right; coefficient of variation, bottom left; bias ratio, bottom right). Solid horizontal lines show the median value per model, boxes show the interquartile range. Dotted horizontal lines show median and interquartile range across all models. The numbers at the top of the bars show the number of model parameters.

Cengkrong are driven more by interflow. NAM has 10 parameters and six stores. Given that values of C_s, C_{of}, and C_{L2} were not critical for high model performance in any catchment, and model performance strongly depended on low values for U^*, then NAM may be unnecessarily complex for these catchments. C_s is the degree-day factor for snowmelt, so it is unsurprising

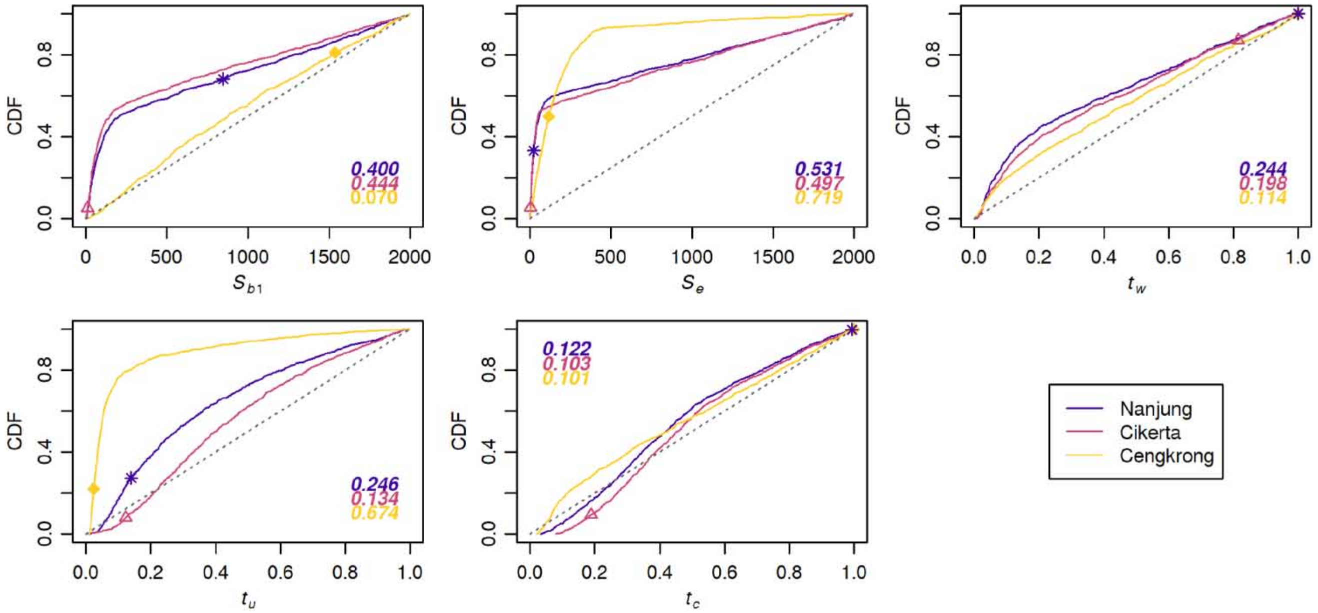


Figure 7 | Cumulative distributions for MOPEX-1 parameters, corresponding to 100,000 nominally uniform samples for each parameter (dotted grey lines), the 1,000 samples giving the highest *KGE'* scores compared to gauged flow data for each catchment (coloured lines), and the Stage 1 optimized values for each parameter (coloured symbols).

that its value is unimportant. C_{of} and C_{L2} are linked to overland flow, one of three parallel paths to outflow (with interflow and groundwater). Hence, the unimportance of these parameters and stores suggests that the NAM structure could be reduced to seven parameters and four stores with minimal impact on performance in these tropical catchments.

Figure 9 is equivalent to Figure 8 but for the two-parameter implementation of the PDM. With this model, $\bar{c}$ (the mean soil moisture store depth) is highly identifiable and varies between catchments. However, the linear routing time constant, k_1, is much less identifiable in all three catchments, and there are no significant (at $p = 0.01$) *KS* scores in Nanjung. In Nanjung, the optimal value $\bar{c}$ is 1, at the lower limit of the permitted range, while in Cengkrong, the optimal $\bar{c}$ is at the top of the permitted range and optimal $k1$ is low, suggesting significant attenuation. Parameterization at opposite extremes suggests that the model structure is unsuitable, despite the high *KGE*' values (0.694 in Nanjung and 0.681 in Cengkrong). This reinforces previous findings that high performance in wet catchments derives mainly from accurate driving data. It is only in Cikerta that the optimal parameter values are away from either extreme of the permitted range, yet only a moderate *KGE*' score (0.456) can be achieved with this arguably better-suited model structure.

6. DISCUSSION

This study confirms the Citarum at Nanjung as having sufficient flow data quality for hydrological modelling, giving confidence to previous studies in this area (e.g. Harlan *et al.* 2010; Nastiti *et al.* 2015; Julian *et al.* 2019; Hatmoko *et al.* 2020; Rusli *et al.* 2021). However, we also highlight the Cirasea at Cengkrong and the Cikundul at Cikerta as other stations in the same basin with potentially high-quality, long-duration flow records that could be used to broaden the scope of other hydrological studies focused on the Citarum basin, Indonesia, or volcanic or tropical regions generally.

Overall, considering both the broad model optimization study and the more focused study of individual catchment model parameterizations, it is apparent that performance, measured by modified *KGE*, is much more strongly linked to the availability of appropriate precipitation data than either hydrological model structure or evapotranspiration data. Overall, the greater influence of precipitation data than the model structure on high-flow estimates agrees with several previous studies (e.g. Bárdossy *et al.* 2022; Zhang *et al.* 2022; Aitken *et al.* 2023), particularly in wetter regions (Thober *et al.* 2018) and tropical climates (van Kempen *et al.* 2021). Furthermore, the strong relationship between precipitation characteristics and major runoff event characteristics in tropical climates agrees with, e.g. Birch *et al.* (2021). Perhaps surprisingly, given that it is so strongly based on gauged rainfall data from the study area, the use of Yanto precipitation data did not result in consistently

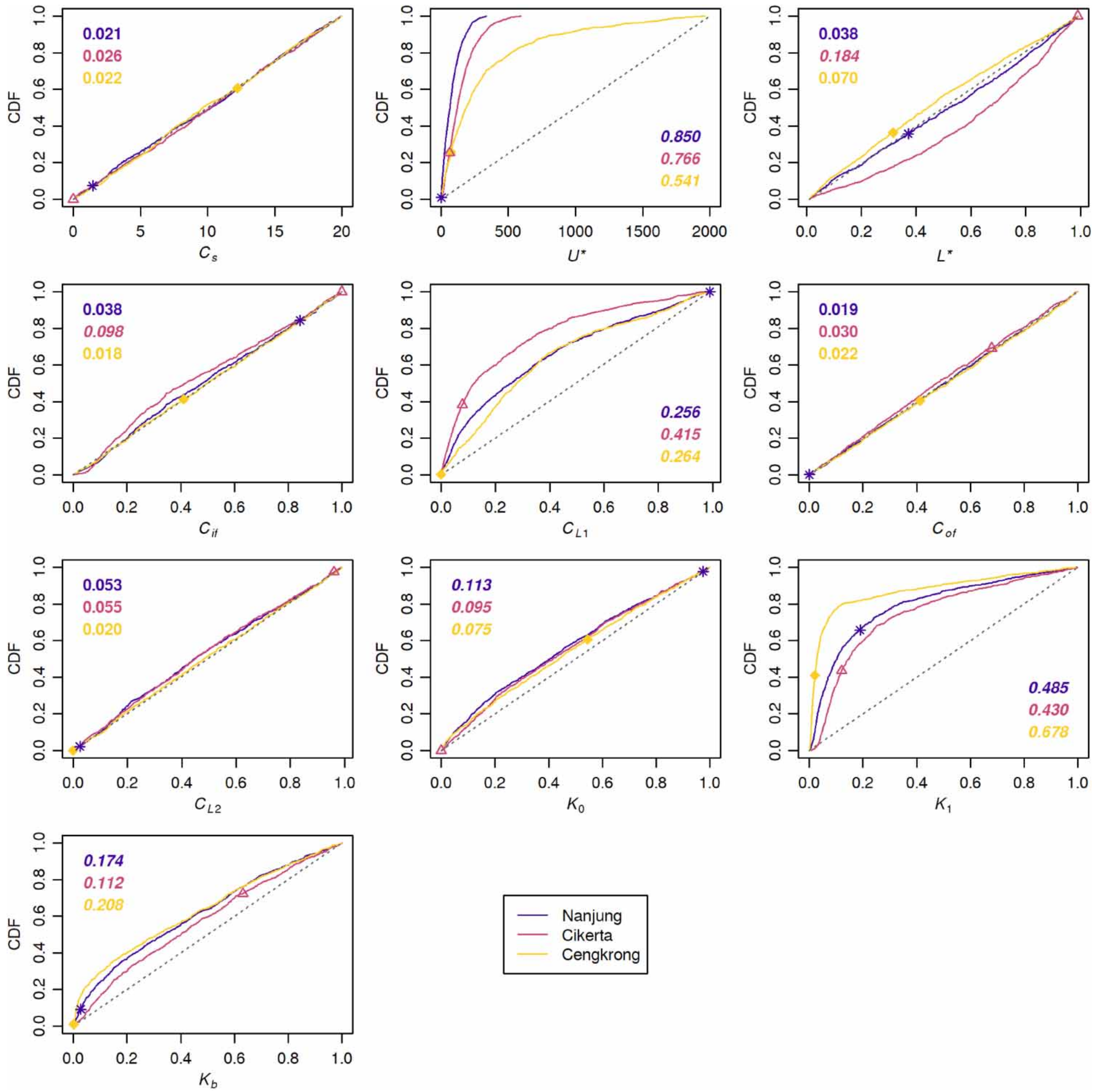


Figure 8 | Cumulative distributions for NAM parameters, corresponding to 100,000 nominally uniform samples for each parameter (dotted grey lines), the 1,000 samples giving the highest *KGE*' scores compared to gauged flow data for each catchment (coloured lines), and the Stage 1 optimized values for each parameter (coloured symbols).

high *KGE*' scores and even resulted in model outputs with greatly underestimated total flow (bias) and poor correlation with observed flows in Cikerta. Hence, it cannot be assumed that gauged rainfall within the study area is more 'true' than remotely sensed estimates.

This study demonstrated that the suitability of many common, widely used, and easily available gridded precipitation datasets can change noticeably over short distances of tens of kilometres, extending the findings of Baez-Villanueva *et al.* (2018) and dos Santos Silva *et al.* (2023) to considerably smaller spatial scales. Our study also highlights that individual grid cells in gridded precipitation datasets may be larger than a study catchment, leaving calculated catchment-

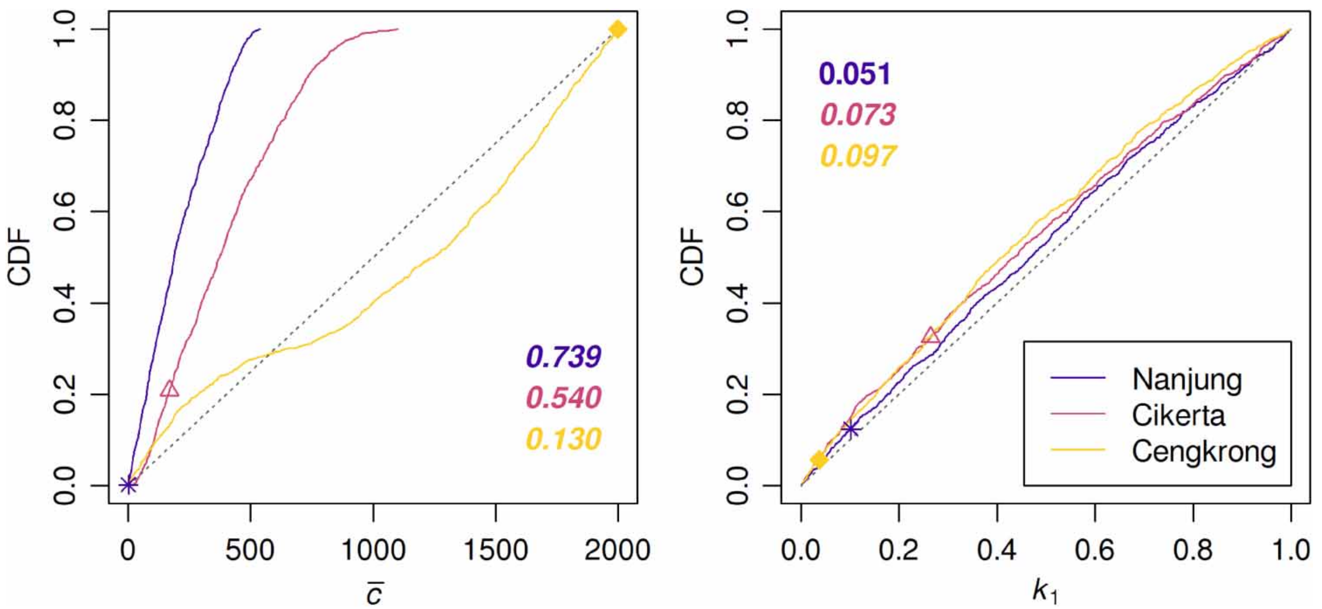


Figure 9 | Cumulative distributions for two-parameter PDM parameters, corresponding to 100,000 nominally uniform samples for each parameter (dotted grey lines), the 1,000 samples giving the highest *KGE*' scores compared to gauged flow data for each catchment (coloured lines), and the Stage 1 optimized values for each parameter (coloured symbols).

average rainfalls vulnerable to the interpolation procedures used to create the gridded datasets. Luo *et al.* (2019) found that a model of the lower Lancang-Mekong Basin driven by gridded (CHIRPS or TRMM) rainfall data outperformed the same model driven by gauged rainfall data, supporting this finding. For all these reasons, it is important that modelling of tropical catchments considers multiple precipitation datasets, and that assumptions about suitability are not made beforehand.

Within each catchment, the top 22–24 (of 24) models achieved *KGE*' values within a 0.1 range, meaning that almost all models performed similarly well. No model structure or structures ranked consistently highly in all three study catchments, nor was there a tendency for models developed in volcanic or tropical regions to perform particularly highly. However, a more detailed comparison of MOPEX-1, NAM, and the two-parameter PDM suggested that model structure did have some importance, as an optimal model structure seemed to need two separate, parallel internal flow paths (having more than two did not provide additional benefits). While these are nominally for 'fast' and 'slow' routing, the similar *KGE*' values achieved by almost all models (per-catchment) suggested that a range of internal structures, developed mainly for temperate climates, can be 'repurposed' to better suit other types of catchments, such as the surface and baseflow runoff paths of the standard PDM appearing to trade roles (consider for example the study of Vesuviano *et al.* (2022)).

It should be noted that, while *KGE*' is a suitable metric for high flow-focused studies, such as this one (Mizukami *et al.* 2019), alternative metrics should be considered for water resources studies. van Kempen *et al.* (2021) suggest that model structure may be relatively more important than input data for low-flow performance.

6.1. Practical implications for flood prediction and water management

Given our finding that modelled flow depends more strongly on input precipitation than model structure, future work to improve flood estimation, prediction, and risk management in tropical climates should focus more on improving precipitation estimates and forecasts than on changes to hydrological model structures. In this context, future work on flood modelling in tropical regions should prioritise the acquisition of more accurate precipitation data, while current studies should consider using an ensemble of precipitation inputs from varied sources to produce a range of flow estimates. It is important to note that locally gauged precipitation data cannot automatically be assumed to be more accurate than remotely sensed products. This is because a rain gauge represents a single point, compared to the spatially integrated remotely sensed products. This source of uncertainty should be carefully considered in flood prediction models.

To effectively improve flood prediction and water management, policymakers should focus on advancing the accuracy of precipitation data, emphasising the integration of ground-based, remotely sensed, and appropriate modelled estimates. This multi-source approach will enhance the reliability of flow predictions, providing a stronger foundation for flood risk management and water resource planning in tropical regions. The modelling approach generates an ensemble of river flow predictions that can be integrated into frameworks such as robust decision-making (Hall *et al.* 2012), enabling effective decision-making under uncertainty. These approaches allow the information contained within the different models and precipitation datasets to be effectively used to support water and flood risk management.

7. CONCLUSIONS

Flooding remains a significant problem for tropical regions due to their high annual and daily wet-season rainfalls and, in many cases, high population densities. However, most rainfall–runoff modelling tools were not developed for either tropical climates or, as particularly relevant in Indonesia, volcanic regions. Furthermore, while many different rainfall data sources are available for tropical regions, their accuracy is hard to assess as there are few rainfall gauging stations against which to compare them. In this study, we compared the performance of 24 existing hydrological model variants, with 21 combinations of seven precipitation and three evapotranspiration data sources, to model gauged daily flows at three river gauging stations in the Citarum basin, Java, Indonesia, first comparing overall performance per-catchment, per-climatic input and per-model via modified *KGE*', then identifying which specific parameters in which models had the greatest impact on performance.

We found that appropriate precipitation data had the greatest effect on the maximum *KGE*' achieved, but that the highest *KGE*' scores were achieved with different precipitation datasets in different study catchments. This highlights the need to consider multiple precipitation input datasets when modelling catchments with strong rainfall–runoff relationships in wet climates. The choice of evapotranspiration dataset had little effect on *KGE*' as the differences between datasets were relatively small. In comparison to precipitation, the hydrological model structure also had little effect on *KGE*', with the top 22–24 (of 24) models achieving *KGE*' values within a 0.1 range in each study catchment. Model structures developed for volcanic or tropical regions did not outperform other model structures in this tropical volcanic region. However, this study did imply a minimum model complexity required to successfully represent the range of flow behaviours seen in this volcanic tropical monsoon region, consisting of two separate, parallel, fast and slow routing pathways, with flexibility allowed in the details of how they are implemented. These findings demonstrate the need to focus on improving precipitation data and estimates to improve flood prediction and water management in tropical climates.

ACKNOWLEDGEMENTS

This work was supported by the Natural Environment Research Council (NE/S002790/1, NE/S002790/2 and NE/S00310X/1). Rahmawati Rahayu's PhD scholarship was funded by the Durham University Global Challenges Centre for Doctoral Training, which in turn was funded by the UK Global Challenges Research Fund (GCRF). The authors are also grateful for the river flow rate data provided by the Center for Hydrology and Water Management, which is part of the Indonesia Ministry of Public Works (Balai Besar Wilayah Sungai). We would like to thank *Hydrology Research*'s Subject Editor, Associate Editor and two anonymous reviewers for their helpful comments.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Aitken, G., Beevers, L., Parry, S. & Facer-Childs, K. (2023) Partitioning model uncertainty in multi-model ensemble river flow projections, *Climatic Change*, **176**, 153. https://doi.org/10.1007/s10584-023-03621-1.

Argüeso, D., Luca, A. D., Jourdain, N. C., Romero, R. & Homar, V. (2022) Mechanisms for extreme precipitation changes in a tropical archipelago, *Journal of Climate*, **35** (17), 5519–5536. https://doi.org/10.1175/JCLI-D-21-0224.1.

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, D., Nelson, B. R. & Prat, O. P. (2015) PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies, *Bulletin of the American Meteorological Society*, **96** (1), 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1.

As-syakur, A. R., Tanaka, T., Osawa, T. & Mahendra, M. S. (2013) Indonesian rainfall variability observation using TRMM multi-satellite data, *International Journal of Remote Sensing*, **34** (21), 7723–7738. https://doi.org/10.1080/01431161.2013.826837.

Atkinson, S. E., Sivapalan, M., Woods, R. A. & Viney, N. R. (2003) Dominant physical controls on hourly flow predictions and the role of spatial variability: mahurangi catchment, New Zealand, *Advances in Water Resources*, **26** (3), 219–235. https://doi.org/10.1016/S0309-1708(02)00183-5.

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Ribbe, L., Nauditt, A., Giraldo-Osorio, J. D. & Thinh, N. X. (2018) Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin-America, *Atmospheric Research*, **213**, 34–50. https://doi.org/10.1016/j.atmosres.2018.05.011.

Bárdossy, A., Kilsby, C., Birkinshaw, S., Wang, N. & Anwar, F. (2022) Is precipitation responsible for the most hydrological model uncertainty?, *Frontiers in Water*, **4**, 836554. https://doi.org/10.3389/frwa.2022.836554.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J. & Bruijnzeel, L. A. (2016) Global-scale regionalization of hydrologic model parameters, *Water Resources Research*, **52**, 3599–3622, https://doi.org/10.1002/2015WR018247.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A. & Wood, E. F. (2018) Present and future Köppen–Geiger climate classification maps at 1-km resolution, *Scientific Data*, **5**, 180214. https://doi.org/10.1038/sdata.2018.214.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. M., van Dijk, A. I. J. M., McVicar, T. R. & Adler, R. F. (2019) MSWEP v2 Global 3-Hourly 0.1° precipitation: methodology and quantitative assessment, *Bulletin of the American Meteorological Society*, **100** (3), 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1.

Beven, K. & Binley, A. (1992) The future of distributed models: model calibration and uncertainty prediction, *Hydrological Processes*, **6** (3), 279–298. https://doi.org/10.1002/hyp.3360060305.

Birch, A. L., Stallard, R. F. & Barnard, H. R. (2021) Precipitation characteristics and land cover control Wet season runoff source and rainfall partitioning in three humid tropical catchments in central Panama, *Water Resources Research*, **57**, e2020WR028058. https://doi.org/10.1029/2020WR028058.

BNPB (2024) *Infografis*. Jakarta: Badan Nasional Penanggulangan Bencana. Available at: https://bnpb.go.id/infografis.

Boyle, D. P. (2001) *Multicriteria Calibration of Hydrologic Models*. *PhD thesis*, University of Arizona, Tucson, Arizona.

Chiew, F. H. S. (1990) *Estimating Groundwater Recharge Using an Integrated Surface and Groundwater Model*. *PhD thesis*, University of Melbourne, Melbourne, Australia.

Chiew, F. & McMahon, T. (1994) Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments, *Journal of Hydrology*, **153**, 383–416. https://doi.org/10.1016/0022-1694(94)90200-3.

Chiew, F., Peel, M., Western, A., (2002) Application and testing of the simple rainfall–runoff model SIMHYD. In: Singh, V. & Frevert, D. (eds.) *Mathematical Models of Small Watershed Hydrology*. Chelsea, Michigan: Water Resources Publications LLC, pp. 335–367.

Chiew, F. H. S., Zheng, H. & Potter, N. J. (2018) Rainfall–runoff modelling considerations to predict streamflow characteristics in ungauged catchments and under climate change, *Water*, **10** (10), 1319. https://doi.org/10.3390/w10101319.

Desclaux, T., Lemonnier, H., Genthon, P., Soulard, B. & Le Gendre, R. (2018) Suitability of a lumped rainfall-runoff model for flashy tropical watersheds in New Caledonia, *Hydrological Sciences Journal*, **63** (11), 1689–1706. https://doi.org/10.1080/02626667.2018.1523613.

dos Santos Silva, F. D., da Costa, C. P. W., dos Santos Franco, V., Gomes, H. B., da Silva, M. C. L., dos Santos Vanderlei, M. H. G., Costa, R. L., da Rocha Júnior, R. L., Cabral Júnior, J. B., dos Reis, J. S., Cavalcante, R. B. L., Tedeschi, R. G., de Jesus da Costa Barreto, N., Nogueira Neto, A. V., dos Santos Jesus, E. & da Silva Ferreira, D. B. (2023) Intercomparison of different sources of precipitation data in the Brazilian legal Amazon, *Climate*, **11** (12), 241. https://doi.org/10.3390/cli11120241.

Duan, Q., Sorooshian, S. & Gupta, V. (1992) Effective and efficient global optimization for conceptual rainfall–runoff models, *Water Resources Research*, **28** (4), 1015–1031. https://doi.org/10.1029/91WR02985.

Duan, Q., Sorooshian, S. & Gupta, V. (1994) Optimal use of the SCE-UA global optimization method for calibrating watershed models, *Journal of Hydrology*, **158**, 265–284. https://doi.org/10.1016/0022-1694(94)90057-4.

Fenicia, F., Savenije, H. H. G., Matgen, P. & Pfister, L. (2008) Understanding catchment behaviour through stepwise model concept improvement, *Water Resources Research*, **44** (1), W01402. https://doi.org/10.1029/2006WR005563.

Fukushima, Y. (1988) A model of river flow forecasting for a small forested mountain catchment, *Hydrological Processes*, **2**, 167–185. https://doi.org/10.1002/hyp.3360020207.

Fulazzaky, M. A. (2010) Water quality evaluation system to assess the status and the suitability of the Citarum river water to different uses, *Environmental Monitoring and Assessment*, **168**, 669–684. https://doi.org/10.1007/s10661-009-1142-z.

Funk, C., Peterson, P., Landsfield, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. & Michaelsen, J. (2015) The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes, *Scientific Data*, **2**, 150066. https://doi.org/10.1038/sdata.2015.66.

Greenfield, B. J. (1984) *The Thames Water Catchment Model (Internal Report)*. Reading, UK: Technology and Development Division, Thames Water.

Guillaume, J. (2013) *Hydromad R Package*. Available at: http://hydromad.catchment.org/.

Gunawan, G. (2021) Hydrological modelling of Air Bengkulu river watershed in Indonesia using SUH and HEC-HMS models, *IOP Conference Series: Materials Science and Engineering*, **1173** (1), 012021. https://doi.org/10.1088/1757-899X/1173/1/012021.

Hall, J. W., Lempert, R. J., Keller, K., Hackbarth, A., Mijere, C. & McInerney, D. J. (2012) Robust climate policies under uncertainty: a comparison of robust decision making and info-gap methods, *Risk Analysis*, **32** (10), 1657–1672. https://doi.org/10.1111/j.1539-6924.2012.01802.x.

Harlan, D., Wangsadikpura, M. & Munajat, C. M. (2010) Rainfall–runoff modeling of Citarum Hulu river basin by using GR4J, *Proceedings of the World Congress on Engineering*, **2**, 1607–1611.

Hatmoko, W., Levina & Diaz, B. (2020) Comparison of rainfall–runoff models for climate change projection – case study of Citarum river basin, Indonesia, *IOP Conference Series: Earth and Environmental Science*, **423**, 012045. https://doi.org/10.1088/1755-1315/423/1/012045.

Haylock, M. & McBride, J. (2001) Spatial coherence and predictability of Indonesian wet season rainfall, *Journal of Climate*, **14** (18), 3882–3887. https://doi.org/10.1175/1520-0442(2001)014 < 3882:SCAPOI > 2.0.CO;2.

Jakeman, A. J., Littlewood, I. G. & Whitehead, P. G. (1990) Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments, *Journal of Hydrology*, **117**, 275–300. https://doi.org/10.1016/0022-1694(90)90097-H.

Jansen, K. F., Teuling, K. F., Craig, J. R., Dal Molin, M., Knoben, W. J. M., Parajka, J., Vis, M. & Melsen, L. A. (2021) Mimicry of a conceptual hydrological model (HBV): what's in a name?, *Water Resources Research*, **57** (5), e2020WR029143. https://doi.org/10.1029/2020WR029143.

Jones, A. E., Bell, V. A., Folwell, S. S., Moore, R. J., Farquharson, F. A. & Jones, D. A. (2002) *PDM and KW Model Calibration for the Citanduy Basin. South Java Flood Control Sector Project*. Contract Report to GISTEC. Wallingford, UK: Centre for Ecology & Hydrology.

Jourdain, N. C., Gupta, A. S., Taschetto, A. S., Ummenhofer, C. C., Moise, A. F. & Ashok, K. (2013) The Indo-Australian monsoon and its relationship to ENSO and IOD in reanalysis data and the CMIP3/CMIP5 simulations, *Climate Dynamics*, **41**, 3073–3102. https://doi.org/10.1007/s00382-013-1676-1.

Julian, M. M., Brenning, A., Kralisch, S. & Fink, M. (2019) Modelling of hydrological responses in the upper Citarum basin based on the spatial plan of West Java province 2029 and climate change, *International Journal of Technology*, **10** (5), 866–875. https://doi.org/10.14716/ijtech.v10i5.2376.

Kachroo, R. K. (1992) River flow forecasting. part 5. application of a conceptual model, *Journal of Hydrology*, **133** (1–2), 141–178. https://doi.org/10.1016/0022-1694(92)90150-T.

Kling, H., Fuchs, M. & Paulin, M. (2012) Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, *Journal of Hydrology*, **424**, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011.

Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C. & Woods, R. A. (2019) Supplement of Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v1.2: an open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations, *Geoscientific Model Development*, **12**, 2463–2480. https://doi.org/10.5194/gmd-12-2463-2019.

Lehner, B., Verdin, K. & Jarvis, A. (2008) New global hydrography derived from spaceborne elevation data, *Eos Transactions AGU*, **89** (10), 93–94. https://doi.org/10.1029/2008EO100001.

Li, W., Cheng, X. & Zhu, D. (2024) Towards the hydrological effects of land use change in karst area, a case study in Lijiang river basin, China, *Journal of Hydrology*, **630**, 130629. https://doi.org/10.1016/j.jhydrol.2024.130629.

Lindström, G., Johansson, B., Persson, M., Gardelin, M. & Bergström, S. (1997) Development and test of the distributed HBV-96 hydrological model, *Journal of Hydrology*, **201**, 272–288. https://doi.org/10.1016/S0022-1694(97)00041-3.

Luo, X., Wu, W., He, D., Li, Y. & Ji, X. (2019) Hydrological simulation using TRMM and CHIRPS precipitation estimates in the lower Lancang-Mekong River Basin, *Chinese Geographical Science*, **29** (1), 13–25. https://doi.org/10.1007/s11769-019-1014-6.

Massey, F. J. (1951) The Kolmogorov-Smirnov test for goodness of fit, *Journal of the American Statistical Association*, **46** (253), 68–78. https://doi.org/10.2307/2280095.

Mathias, S. A. (2023) *Hydraulics, Hydrology and Environmental Engineering*. Berlin: Springer. https://doi.org/10.1007/978-3-031-41973-7.

Mathias, S. A., McIntyre, N. & Oughton, R. H. (2016) A study of non-linearity in rainfall-runoff response using 120 UK catchments, *Journal of Hydrology*, **540**, 423–436. https://doi.org/10.1016/j.jhydrol.2016.06.039.

Mishra, B. K., Rafiei Emem, A., Masago, Y., Kumar, P., Regmi, R. K. & Fukushi, K. (2017) Assessment of future flood inundations under climate and land use change scenarios in the Ciliwung River Basin, Jakarta, *Journal of Flood Risk Management*, **11** (S2), S1105–S1115. https://doi.org/10.1111/jfr3.12311.

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V. & Kumar, R. (2019) On the choice of calibration metrics for 'high-flow' estimation using hydrologic models, *Hydrology and Earth System Sciences*, **23**, 2601–2614. https://doi.org/10.5194/hess-23-2601-2019.

Moore, R. J. (2007) The PDM rainfall-runoff model, *Hydrology & Earth System Sciences*, **11**, 483–499. https://doi.org/10.5194/hess-11-483-2007.

Moore, R. J. & Bell, V. A. (2001) *Comparison of Rainfall-Runoff Methods for Flood Forecasting Part 1: Literature Review of Models*. Bristol, UK: Environment Agency.

Narsey, S. Y., Brown, J. R., Colman, R. A., Delage, F., Power, S. B., Moise, A. F. & Zhang, H. (2020) Climate change projections for the Australian monsoon from CMIP6 models, *Geophysical Research Letters*, **47** (13), e2019GL086816. https://doi.org/10.1029/2019GL086816.

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

Nastiti, K. D., Kim, Y., Jung, K. & An, H. (2015) The application of rainfall-runoff-inundation (RRI) model for inundation case in Upper Citarum watershed, West Java-Indonesia, *Procedia Engineering*, **125**, 166–172. https://doi.org/10.1016/j.proeng.2015.11.024.

Nielsen, S. A. & Hansen, E. (1973) Numerical simulation of the rainfall-runoff process on a daily basis, *Hydrology Research*, **4** (3), 171–190. https://doi.org/10.2166/nh.1973.0013.

Paul, P. K., Zhang, Y., Ma, N., Mishra, A., Panigrahy, N. & Singh, R. (2021) Selecting hydrological models for developing countries: perspective of global, continental, and country scale models over catchment scale models, *Journal of Hydrology*, **600**, 126561. https://doi.org/10.1016/j.jhydrol.2021.126561.

Perrin, C., Michel, C. & Andréassian, V. (2003) Improvement of a parsimonious model for streamflow simulation, *Journal of Hydrology*, **279**, 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7.

Petheram, C., Rustomji, P., Chiew, F. H. S. & Vleeshouwer, J. (2012) Rainfall-runoff modelling in northern Australia: a guide to modelling strategies in the tropics, *Journal of Hydrology*, **462–463**, 28–41. https://doi.org/10.1016/j.jhydrol.2011.12.046.

Rahayu, R., Mathias, S. A., Reaney, S., Vesuviano, G., Suwarman, R. & Ramdhan, A. M. (2023) Impact of land cover, rainfall and topography on flood risk in West Java, *Natural Hazards*, **116**, 1735–1758. https://doi.org/10.1007/s11069-022-05737-6.

Rahmawati, N., Rahayu, K. & Yuliasari, S. T. (2021) Performance of daily satellite-based rainfall in groundwater basin of Merapi Aquifer System, Yogyakarta, *Theoretical and Applied Climatology*, **146** (1), 173–190. https://doi.org/10.1007/s00704-021-03731-9.

R Core Team (2019) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://r-project.org.

Rickards, N., Thomas, T., Kaelin, A., Houghton-Carr, H., Jain, S. K., Mishra, P. K., Nema, M. K., Dixon, H., Rahman, M. M., Horan, R., Jenkins, A. & Rees, G. (2020) Understanding future water challenges in a highly regulated Indian river basin – modelling the impact of climate change on the hydrology of the upper Narmada, *Water*, **12** (6), 1762. https://doi.org/10.3390/w12061762.

Ruane, A. C., Goldberg, R. & Chryssanthacopoulos, J. (2015) Climate forcing datasets for agricultural modeling: merged products for gap-filling and historical climate series estimation, *Agricultural and Forest Meteorology*, **200**, 233–248. https://doi.org/10.1016/j.agrformet.2014.09.016.

Rusli, S. R., Weerts, A. H., Taufiq, A. & Bense, V. F. (2021) Estimating water balance components and their uncertainty bounds in highly groundwater-dependent and data-scarce area: an example for the upper citarum basin, *Journal of Hydrology: Regional Studies*, **37**, 100911. https://doi.org/10.1016/j.ejrh.2021.100911.

Santos, L., Thirel, G. & Perrin, C. (2018) Continuous state-space representation of a bucket-type rainfall-runoff model: a case study with the GR4 model using state-space GR4 (version 1.0), *Geoscientific Model Development*, **11**, 1591–1605. https://doi.org/10.5194/gmd-11-1591-2018.

Savenije, H. H. G. (2010) Topography driven conceptual modelling (FLEX-Topo), *Hydrology & Earth System Sciences*, **14**, 2681–2692. https://doi.org/10.5194/hess-14-2681-2010.

Sekaranom, A. B., Nurjani, E., Hadi, M. P. & Marfai, M. A. (2018) Comparsion of TRMM precipitation satellite data over central java region – Indonesia, *Quaestiones Geographicae*, **37** (3), 97–114. https://doi.org/10.2478/quageo-2018-0028.

Senjaya, T., Yudianto, D., Yuebo, X. & Adidarma, W. K. (2020) Application of TRMM in the hydrological analysis of Upper Bengawan Solo River Basin, *Journal of the Civil Engineering Forum*, **6** (3), 309–318. https://doi.org/10.22146/jcef.57125.

Son, K. & Sivapalan, M. (2007) Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, *Water Resources Research*, **43** (1), W01415. https://doi.org/10.1029/2006WR005032.

State of the Tropics (2014) *State of the Tropics 2014 Report*. Townsville, Australia: James Cook University.

Suroso, S., Santoso, P. B., Birkinshaw, S., Kilsby, C., Bardossy, A. & Aldrian, E. (2023) Assessment of TRMM rainfall data for flood modelling in three contrasting catchments in Java, Indonesia, *Journal of Hydroinformatics*, **25** (3), 797–814. https://doi.org/10.2166/hydro.2023.132.

Tan, B. Q. & O'Connor, K. M. (1996) Application of an empirical infiltration equation in the SMAR conceptual model, *Journal of Hydrology*, **185**, 275–295. https://doi.org/10.1016/0022-1694(95)02993-1.

Tanaka, Y., Minggat, E. & Roseli, W. (2021) The impact of tropical land-use change on downstream riverine and estuarine water properties and biogeochemical cycles: a review, *Ecological Processes*, **10**, 40. https://doi.org/10.1186/s13717-021-00315-3.

Thober, S., Kumar, R., Wanders, N., Marx, A., Pan, M., Rakovec, O., Samaniego, L., Sheffield, J., Wood, E. F. & Zink, M. (2018) Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming, *Environmental Research Letters*, **13** (1), 014003. https://doi.org/10.1088/1748-9326/aa9e35.

van den Brink, F. (2009) *Modelling the Discharge of the Cidanau River in West Java with the HBV Model*. *Bachelor thesis*, LabMath-Indonesia/Universiteit Twente, Bandung, Indonesia/Twente, The Netherlands.

van der Besselaar, E., van der Schrier, G., Cornes, R. C., Iqbal, A. S. & Klein Tank, A. M. G. (2017) SA-OBS: a daily gridded surface temperature and precipitation dataset for Southeast Asia, *Journal of Climate*, **30** (14), 5151–5165. https://doi.org/10.1175/JCLI-D-16-0575.1.

van Kempen, G., van der Wiel, K. & Melsen, L. A. (2021) The impact of hydrological model structure on the simulation of extreme runoff events, *Natural Hazards and Earth System Sciences*, **21** (3), 961–976. https://doi.org/10.5194/nhess-21-961-2021.

Vernimmen, R. R. E., Hooijer, A., Mamenun, Aldrian, E. & van Dijk, A. I. J. M. (2012) Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia, *Hydrology and Earth System Sciences*, **16**, 133–146. https://doi.org/10.5194/hess-16-133-2012.

Vesuviano, G., Griffin, A. & Stewart, E. (2022) Flood frequency estimation in data-sparse Wainganga Basin, India, using continuous simulation, *Water*, **14** (18), 2887. https://doi.org/10.3390/w14182887.

Vu, D. T., Dang, T. D., Pianosi, F. & Galelli, S. (2023) Calibrating macroscale hydrological models in poorly gauged and heavily regulated basins, *Hydrology and Earth System Sciences*, **27** (19), 3485–3504. https://doi.org/10.5194/hess-27-3485-2023.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V. & Sorooshian, S. (2001) A framework for development and application of hydrological models, *Hydrology & Earth System Sciences*, **5**, 13–26. https://doi.org/10.5194/hess-5-13-2001.

Wang, J., Zhuo, L., Han, D., Liu, Y. & Rico-Ramirez, M. A. (2023) Hydrological model adaptability to rainfall inputs of varied quality, *Water Resources Research*, **59**, e2022WR032484. https://doi.org/10.1029/2022WR032484.

Wati, T., Hadi, T. W., Sopaheluwakan, A. & Hutasoit, L. M. (2022) Statistics of the performance of gridded precipitation datasets in Indonesia, *Advances in Meteorology*, **2022**, 7995761. https://doi.org/10.1155/2022/7995761.

Wiwoho, B. S., Astuti, I. S., Alfarizi, I. A. G. & Sucahyo, H. R. (2021) Validation of three daily satellite rainfall products in a Humid Tropic Watershed, Brantas, *Indonesia: Implications to Land Characteristics and Hydrological Modelling. Hydrology*, **8** (4), 154. https://doi.org/10.3390/hydrology8040154.

Xie, P., Chen, M. & Shi, W. (2010). 'CPC unified gauge-based analysis of global daily precipitation', *24th Conference on Hydrology*, Vol 2. Atlanta, GA. American Meteorological Society.

Yanto, Livneh, B., Rajagopalan, B. & Kasprzyk, J. (2017a) Hydrological model application under data scarcity for multiple watersheds, Java Island, Indonesia, *Journal of Hydrology: Regional Studies*, **9**, 127–139. https://doi.org/10.1016/j.ejrh.2016.09.007.

Yanto, Livneh, B. & Rajagopalan, B. (2017b) Development of a gridded meteorological dataset over Java island, Indonesia 1985–2014, *Scientific Data*, **4**, 170072. https://doi.org/10.1038/sdata.2017.72.

Ye, S., Yaeger, M., Coopersmith, E., Cheng, L. & Sivapalan, M. (2012) Exploring the physical controls of regional patterns of flow duration curves – Part 2: role of seasonality, the regime curve, and associated process controls, *Hydrology & Earth System Sciences*, **16**, 4447–4465. https://doi.org/10.5194/hess-16-4447-2012.

Yoneyama, K. & Zhang, C. (2020) Years of the maritime continent, *Geophysical Research Letters*, **47** (12), e2020GL087182. https://doi.org/10.1029/2020GL087182.

Yoshida, K., Tanaka, K., Noda, K., Homma, K., Maki, M., Hongo, C., Shirakawa, H. & Oki, K. (2017) Quantitative evaluation of spatial distribution of nitrogen loading in the citarum river basin, Indonesia, *Journal of Agricultural Meteorology*, **73** (1), 31–44. https://doi.org/10.2480/agrmet.D-15-00020.

Zhang, S., Chen, J. & Gu, L. (2022) Overall uncertainty of climate change impacts on watershed hydrology in China, *International Journal of Climatology*, **42** (1), 507–520. https://doi.org/10.1002/joc.7257.

Zhang, Y., He, Y. & Song, J. (2023) Effects of climate change and land use on runoff in the Huangfuchuan Basin, China, *Journal of Hydrology*, **626**, 130195. https://doi.org/10.1016/j.jhydrol.2023.130195.

Zhao, R.-J. (1992) The Xinanjiang model applied in China, *Journal of Hydrology*, **135** (1–4), 371–381. https://doi.org/10.1016/0022-1694(92)90096-E.

First received 11 February 2025; accepted in revised form 13 May 2025. Available online 20 May 2025