Remote sensing-aided rainfall-runoff modelling in the tropics of Costa Rica

Saúl Arciniega-Esparza[1], Christian Birkel[2,3], Andrés Chavarría-Palma[2], Berit Arheimer[4], J. Agustín Breña-Naranjo[5,6]

5 [1] Hydrogeology Group, Faculty of Engineering, Universidad Nacional Autónoma de México, Mexico City, 04510, Mexico.
[2] Department of Geography and Water and Global Change Observatory, University of Costa Rica, San José, Costa Rica
[3] Northern Rivers Institute, University of Aberdeen, Aberdeen, Scotland.
[4] Swedish Meteorological and Hydrological Institute, Norrköping, Sweden.
[5] Institute of Engineering, Universidad Nacional Autónoma de México, Mexico City, Mexico.
10 [6] Instituto Mexicano de Tecnología del Agua, Jiutepec, Morelos, Mexico.

Correspondence to: Saúl Arciniega-Esparza (sarciniegae@comunidad.unam.mx)

Abstract. Streamflow simulation across the tropics is limited by the lack of data to calibrate and validate large-scale
15 hydrological models. Here, we applied the process-based, conceptual HYPE (Hydrological Predictions for the Environment)
model to quantitively assess Costa Rica's water resources at a national scale. Data scarcity was compensated using adjusted
global topography and remotely-sensed climate products to force, calibrate, and independently evaluate the model. We used a
global temperature product and bias-corrected precipitation from CHIRPS (Climate Hazards Group InfraRed Precipitation
with Station data) as model forcings. Daily streamflow from 13 gauges for the period 1990-2003 and monthly MODIS
20 (Moderate Resolution Imaging Spectroradiometer) potential evapotranspiration (PET) and actual evapotranspiration (AET)
for the period 2000-2014 were used to calibrate and evaluate the model applying four different model configurations (M1, M2,
M3, M4). The calibration consisted of step-wise parameter constraints preserving the best parameter sets from previous
simulations in an attempt to balance the variable data availability and time periods. The model configurations were
independently evaluated using hydrological signatures such as the baseflow index, runoff coefficient, and aridity index, among
25 others. Results suggested that a two-step calibration using monthly and daily streamflow (M2) was a better option instead of
calibrating only with daily streamflow (M1), with a similar mean Kling-Gupta Efficiency (KGE~0.53) for daily streamflow
time-series, but with improvements to reproduce the flow-duration-curves, with a median Root Squared Median Error (RMSE)
of 0.42 for M2 and a median RMSE of 1.15 for M1. Additionally, including AET (M3 and M4) in the calibration statistically
improved the simulated water balance and better matched hydrological signatures, with a mean KGE of 0.49 for KGE in M3-
30 M4, in comparison to M1-M2 with mean KGE<0.3. Furthermore, statistical tests of Kruskal-Wallis and Mann-Whitney support
a similar model performance for M3 and M4, suggesting that monthly PET-AET and daily streamflow (M3) represents an
appropriate calibration sequence for regional modelling. Such a large-scale hydrological model has the potential to be used
operationally across the humid tropics informing decision making at relatively high spatial and temporal resolution.

1 Introduction

35 Tropical regions differ from temperate regions by larger energy inputs, more intense atmospheric dynamics, higher precipitation rates, larger streamflow, and sediment yields (Dehaspe et al., 2018; Esquivel-Hernández et al., 2017; Wohl et al., 2012). Moreover, tropical regions are among the fastest-changing environments, with a hydrological cycle pressurized by population growth (Wohl et al., 2012; Ziegler et al., 2007), land use/cover modifications (Gibbs et al., 2010), and altered precipitation and runoff patterns (Esquivel-Hernández et al., 2017) due to climate change. Central America, the northern

40 boundary of the humid tropics, was identified by Giorgi (2006) as the most sensitive tropical region to climate change due to the location between two major water bodies, the Pacific Ocean and the Caribbean Sea.

Increasing concerns about the effects of human activities and climate change on tropical catchments demand an accurate quantification of the water balance components in space and time to guarantee the future water resources availability for ecosystems and socio-economic activities (Esquivel-Hernández et al., 2017; Wohl et al., 2012). Hydrological models have

45 been widely used to assess the spatio-temporal variability of water resources and to provide insights into potential future climate and management decisions (Andersson et al., 2015; Xiong and Zeng, 2019).

However, models also implicitly include many uncertainties (Beven, 2012). For example, Birkel et al. (2020) and Dehaspe et al. (2018) highlighted those hydrological models that are useful to predict streamflow but showed limitations to assess water partitioning and storage changes required for water management in the humid tropics. Modelling in the tropics is further

50 hampered by the lack of good quality hydrometric data used to drive models and for calibration (Westerberg and Birkel, 2015; Westerberg et al., 2014). Moreover, a decrease in hydrological measurements and monitoring networks in many tropical regions occurred during the last three decades (Wohl et al., 2012), limiting the applicability of hydrological models or reducing their performance to simulate streamflow in Central America (Westerberg et al., 2014) and South America (Guimberteau et al., 2012). Model calibration mostly leads to several combinations of parameters with similar streamflow response, i.e.,

55 equifinality (Beven, 2012; Xiong and Zeng, 2019), and it is therefore desirable to reduce or constrain model parameters uncertainty. Moreover, some case studies around the world have found that soil model parameters can be relatively insensitive to streamflow simulations (Massari et al., 2015; Rajib et al., 2018b; Silvestro et al., 2015).

Opportunities to provide more realistic internal hydrological partitioning exist in the form of including additional variables to streamflow, such as, e.g., tracers and remotely sensed variables of evapotranspiration and soil moisture (Dal Molin et al., 2020;

60 Rakovec et al., 2016; Xiong and Zeng, 2019; Massari et al., 2015). The latter, however, may come at the expense of increased complexity for model calibration and evaluation (Arheimer et al., 2020; Her and Seong, 2018; Massari et al., 2015; Xiong and Zeng, 2019; Zhang et al., 2018) since the non-linearity increases the complexity for data assimilation (Massari et al., 2015; Rajib et al., 2018a, 2018b). In addition, hydrological signatures can improve model realism through the synthesis of many simultaneous catchment processes at different scales (Arheimer et al., 2020; Sawicz et al., 2011). Hydrological signatures can

65 be used to increase our understanding of water balance partitioning and hydrological similarity across different scales (e.g., Arciniega-Esparza et al., 2016; Beck et al., 2015; Kirchner, 2009; Troch et al., 2009) and have been applied to improve model

evaluation (e.g., Andersson et al., 2015; Arheimer et al., 2020; Dal Molin et al., 2020; Raphael-Tshimanga and Hughes, 2014; Westerberg et al., 2014). Despite uncertainties in observed hydrological signatures (Westerberg and McMillan, 2015), there is potential to identify model weaknesses and to ultimately produce a more well-balanced catchment representation.

70   Most hydrological models have been developed since the 1970s to solve different needs at catchment scales (Pechlivanidis and Arheimer, 2015; Todini, 2007). Nevertheless, water management increasingly requires detailed hydrological information over larger, aggregated spatial domains instead of a single catchment (Arheimer et al., 2020; Rojas-Serna et al., 2016). Global hydrological models can serve this purpose but suffer from rather coarse spatial resolution and increased computational cost (Kumar et al., 2013; Sood and Smakhtin, 2015). Distributed landscape characteristics at large scales such as soil, topography,

75   and land cover can result in complex hydrological models with many calibrated model parameters (Gurtz et al., 1999) and resulting in greater uncertainty. However, distributed model parameterization based on landscape characteristics also promises the advantage of predicting the hydrological response of ungauged basins (Hrachowitz et al., 2013; Pechlivanidis and Arheimer, 2015). Therefore, the question as to how complex or simple a hydrological model should be remains an open science debate considering that simpler models can lead to similar results in comparison with more complex and more highly

80   parameterized models (Archfield et al., 2015; Rojas-Serna et al., 2016).

An alternative to simulate the hydrology at large spatial scales by means of semi-distributed, conceptual hydrological models together with global data of precipitation, evapotranspiration, and soil moisture (Andersson et al., 2015; Brocca et al., 2020). Conceptual models fall in the category between very simple bucket models and physically-based, distributed models maintaining the numbers of parameters more limited with the possibility to still gain insights into the hydrological processes

85   governing a set of neighboring catchments (e.g., Beven (2012) for a model classification). Moreover, recent hydrological studies implemented data assimilation from remote sensing and global products of soil moisture (Kwon et al., 2020; Massari et al., 2015; Silvestro et al., 2015), snow depth (Infante-Corona et al., 2014), evapotranspiration (Lin et al., 2018; Rajib et al., 2018a, 2018b) and terrestrial water storage (Getirana et al., 2020; Reager et al., 2015) often in combination with conceptual models to reduce or constrain the model parameter uncertainty and to help with model evaluation (e.g., Sheffield et al., 2018).

90   Such an approach needs testing in tropical regions such as Central America, located on the narrow continental bridge (<40 km in places) that connects North with South America. The relatively smaller landmass also results in relatively smaller-sized catchments that quickly convert coarse-scale global products unsuitable for modelling. Additionally, remotely-sensed climatological model input data is a source of error over complex topographical landscapes such as Central America (Maggioni et al., 2016); for example, Zambrano-Bigiarini et al. (2017) estimated KGE<-2 in high elevation areas (>2000 m.a.s.l.) of Chile

95   using seven precipitation products, in comparison with low and mid-elevation areas (<1000 m.a.s.l.) that showed KGE>0.7. Therefore, this paper aims to test the use of the large-scale conceptual but process-based semi-distributed HYPE model (Lindström et al., 2010) exploring strategies to improve regional modelling of tropical data-scarce regions, incorporating different time steps and global gridded products for the complex topographical regions of Costa Rica. We, therefore, used the potential and actual evapotranspiration, PET, and AET products, respectively, from MODIS (Moderate Resolution Imaging

100   Spectroradiometer) additionally to streamflow time series to calibrate the model followed by a posteriori independent

evaluation of hydrological signatures calculated from these global data sets. The model was calibrated using a step-wise procedure tracking the most effective strategy to constrain the parameter space and to reduce the model uncertainty.

Our specific objectives are the following:

1. Adjust the open-source, conceptual rainfall-runoff model HYPE to simulate Costa Rica's catchment hydrology at the national scale using remotely-sensed global climate data and landscape products to drive and evaluate the model under four different step-wise calibration strategies.

2. Analyze the effect of remotely-sensed PET and AET data on model calibration and its capability to improve the simulated water balance and matching hydrological signatures.

2 Study area and data

The study area corresponds to Costa Rica, located on the Central American Isthmus, between 8 and 11°N latitude and 82 and 86°W longitude. Costa Rica covers ~51,000 km^2 between the neighboring countries, Nicaragua to the north and Panama to the south. Costa Rica is characterized by an elevation range from 0 to ~3,840 meters above sea level (m.a.s.l.) due to a mountain range of volcanic origin dividing the country from northwest to south-east into the Pacific and Caribbean drainage basins. Notably, the proximity to the two large water bodies (the Pacific Ocean and the Caribbean Sea) differentiates the atmospheric water dynamics resulting in a marked gradient of tropical rainfall patterns east and west of the continental divide (Maldonado et al., 2013).

Fig. 1.a shows the study area boundaries, the precipitation gauges (blue dots), the monitored catchments (red polygons), and their respective streamflow gauges (black squares), as well as the catchments used within the HYPE model (gray polygons). In situ data consisted of 75 precipitation stations obtained from the National Meteorological Service (IMN in Spanish) containing a minimum length of 10 years of data overlapping the period from 1981 to 2017. This period was selected to compare ground precipitation records with precipitation from global remote sensing products. Moreover, 13 streamflow gauges with daily records from 1990 to 2003 were obtained from the Costa Rican Electricity Institute (ICE). The attributes and climate properties of monitored catchments are shown in Table 1, with catchment areas ranging from 74 to 4,772 km^2 that cover a total extension of ~10,508 km^2 (~21 % of Costa Rican territory), with mean catchment elevations ranging from 330 to 2,600 m.a.s.l. Regarding model simulations, more than 600 nested catchments covering the whole country were delineated using the 30 m Shuttle Radar Topography Mission elevation model (SRTM) (Bamler, 1999) and the terrain analysis toolset from SAGA GIS v.6.4 (Conrad et al., 2015), where the fill sinks algorithm by Wang and Liu (2006) was applied with a minimum slope parameter of 0.0001 degrees and the flow accumulation top-down algorithm together with the single flow direction (D8) configuration. These parameters were defined following previous experience using SAGA with SRTM. Several issues were found during the delineation of catchments on flat terrain, where computed water courses differed from the actual river network. We corrected the computed river network using the vector layers of the main rivers from OpenStreetMaps (OSM), forcing the water courses

following Monteiro et al. (2018). The final catchments ranged from 3 to 500 km^2 with a median value of 65 km^2 and a main river length from 2.5 to 75 km, and a median value of 15.2 km.

135   Figure 1.b and Fig. 1.c show the soil types and land uses across Costa Rica, respectively. Soil types were derived from SoilGrids (Hengl et al., 2017) (see dataset description in Table 2) and compared to national scale soil maps. Sand content and clay content at 1 m depth were used to classify the soil types from the USDA classification criteria in SAGA GIS tools. Furthermore, in order to reduce the number of model parameters, only the four most frequent soil types were considered (Fig. 1.b). The predominant soil texture is clay loam covering an extension of ~35,360 km^2 (69 % of Costa Rica), mainly across low

140   elevation areas. Clay soils cover an extension of ~9,740 km^2 (~19 % of Costa Rica) and are located mainly along the Pacific basin. Moreover, in high elevations loamy soils predominate, covering an extension of ~3,800 km^2 (7 % of total area). The land use classes were obtained from the Climate Change Initiative Land Cover (CCI LC), where 19 unique land covers were found for Costa Rica. Similarly, the land use was reclassified to the four most common categories (Fig. 1.c), where the predominant land uses were tree cover (~65 %) and mosaic cover (~34 %, that includes shrubs, grassland, sparse vegetation,

145   croplands). Urban areas represent less than 0.5 % of Costa Rica. Different CCI LC tree categories were grouped into a single tree class, with ~87% corresponding to broadleaved evergreen trees. Greater 98% of the reclassified urban areas correspond to the original urban land use from CCI LC. The mosaic reclassification was mainly composed of mosaic natural vegetation and croplands (54 and 13%, respectively). Finally, the water reclassification consisted in 93% water bodies and flooded shrub areas.

150   The climatological space-time series were obtained from remote sensing and global products, described in Table 2. The precipitation grid was obtained from the Climate Hazards Group InfraRed Precipitation with Satellite data (CHIRPS) version 2 (Funk et al., 2015), and the mean daily temperature was obtained from the CPC Global Daily Temperature product provided by the NOAA/OAR/ESRL PSL (https://psl.noaa.gov/). Temperature exhibited low seasonality, with mean values ranging from 27 °C in coastal regions and 20 °C in the central region at around 1000 m.a.s.l. (Esquivel-Hernández et al., 2017). Figure 1.d

155   shows the seasonality of monthly precipitation from CHIRPS using the index proposed by Walsh and Lawler (1981), where lower values (<0.3) correspond to a more uniform monthly precipitation, and higher values (>0.8) indicate that annual precipitation is concentrated over a few months. Such seasonality is widely controlled by air masses that reach Costa Rica at the Caribbean littoral, accumulating more humidity on the Caribbean slope (Sáenz and Durán-Quesada, 2015), shown as dark blue areas in Fig. 1.d. Meanwhile, the humidity along the Pacific basin is highly influenced by the migration of continental air

160   masses of the Intertropical Convergence Zone (ITCZ), which establishes the rainy season in May-June and in September-November (Esquivel-Hernández et al., 2017; Muñoz et al., 2008).

The yearly cycle of wet and dry deviations in the ocean-atmosphere is linked to changes in the sea surface temperature of both the Pacific Ocean and the Caribbean Sea, where the El Niño Southern Oscillation (ENSO) is associated to a decrease of the mean annual precipitation across the Pacific basin, and an increase of precipitation in the Caribbean basin (Muñoz et al., 2008).

165

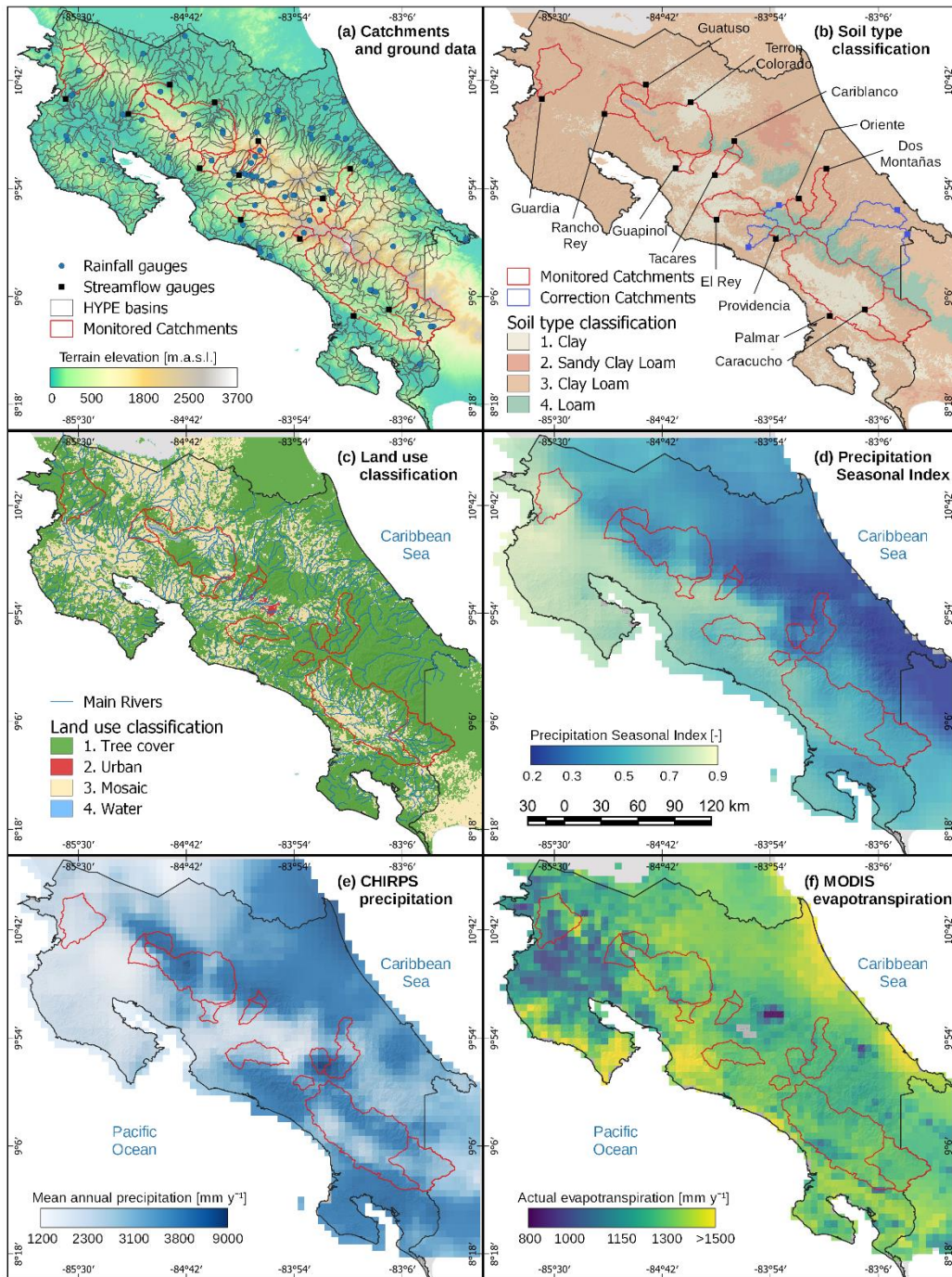



Figure 1. Study area (a) rainfall gauges (blue dots), monitored catchments (red polygons) and sub-basins used in the HYPE model (gray polygons), (b) soil type at 1 m depth from SoilGrids where blue polygons correspond to catchments used for rain correction but not for calibration, (c) major land use categories from CCI Land Cover, (d) Precipitation Seasonal Index

with dark blue colors corresponding to uniform monthly precipitation and yellow colors to a more seasonal precipitation regime, (e) mean annual precipitation for the period 1981-2017 from CHIRPS and (f) mean annual actual evapotranspiration for the period 2000-2014 from MODIS.

Moreover, the cold phase La Niña is the cause of an increase in precipitation in the Pacific basin and a decrease in the Caribbean (Waylen et al., 1996). Overall, the mean annual precipitation averaged ~3,000 mm for Costa Rica with maxima as high as 9,000 mm, observed in the headwaters of the Reventazón catchment at the northwest of the Talamanca Mountain range and the Caribbean basin (Fig. 1.e). The minimum annual precipitation of 1,200 mm y^{-1} is observed on the northern Pacific basin in the Bebedero and Tempisque catchments.

The rainfall patterns across Costa Rica are reflected in the streamflow responses of catchments on the Pacific and Caribbean sides. The daily streamflow tends to be higher in the Caribbean basin (9.2 mm d^{-1} in comparison to 4.2 mm d^{-1} in the Pacific side, computed from observed streamflow records, Table 1), mainly due to the seasonal climate across the Pacific slope with reduced water availability during the dry season from December to April. Furthermore, the stream length of rivers on the Caribbean slope tends to be longer in comparison with rivers from the Pacific basin (see the river network in Fig. 1.c).

Table 1. Physical and climatological properties of the monitored catchments. Streamflow gauges were grouped according to their location on the Caribbean and the Pacific basins.

Zone	Station	Area [km²]	Elevation [m.a.s.l.]			Slope [%]	Prec [mm/year]	EI	AI	Qt mean daily [m³ s⁻¹]
			Min	Mean	Max					
Caribbean	Cariblanco	75	761.7	1850.8	2829.5	26.5	3079	0.44	0.54	9.02
	Oriente	229	586.5	1413.7	2740.1	40.1	4202	0.32	0.42	29.90
	Dos Montanas	660	108.5	1316.7	3191.0	37.2	3551	0.36	0.45	53.79
	Terron Colorado	2061	18.6	734.7	2312.5	21.3	3175	0.44	0.56	149.66
	Guatuso	242	7.8	520.6	1881.7	18.9	3968	0.35	0.45	29.38
Pacific	Providencia	122	1365.8	2573.0	3479.7	44.3	2750	0.50	0.68	6.68
	Tacares	201	584.3	1404.8	2723.8	17.6	2714	0.49	0.61	11.27
	Guapinol	210	178.2	1070.4	2173.3	23.7	2879	0.49	0.66	10.13
	Caracucho	1133	60.9	1251.0	3252.7	24.7	3091	0.43	0.56	72.92
	El Rey	656	48.6	1142.1	2501.5	36.5	2648	0.51	0.67	31.96
	Rancho Rey	320	0.0	490.7	1902.4	15.3	2234	0.57	0.86	9.35
	Guardia	961	0.0	336.5	1898.2	11.0	1787	0.68	1.07	23.66
	Palmar	4771	0.0	1009.2	3791.0	30.3	3176	0.41	0.55	305.45

Potential evapotranspiration and actual evapotranspiration were obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) product (Mu et al., 2011) distributed by the Numerical Terradynamic Simulation Group at the University of Montana, USA, which compared well to few available ground stations in Costa Rica, with errors from -0.3 to 0.7 mm/year (Esquivel-Hernández et al., 2017). Even though several products of AET and PET are available at higher temporal

resolution (such as GLEAM at a daily time step, Miralles et al., 2011), the spatial resolution of these products is at least five times lower than MODIS (~5x5 km, ~25 km^2). Since ~70% of our delimited catchments are smaller than 100 km^2, the spatial resolution of the global products plays an important role to capture the spatial variability of the water balance for modelling.

Figure 1.f shows the mean annual AET from MODIS, which spatially ranges from 547 to 1612 mm. The highest AET values were observed at the coast (Caribbean and Pacific). Moreover, the lower AET values overlap with low humidity zones and sparse vegetation areas (northwestern Costa Rica), as well as higher elevation cloud cover that decreases soil evaporation (Caribbean slope mountain region).

Table 2. Remote sensing and global products used in this study.

Dataset	Variable	Coverage and Resolution	Period	Scale	Data type	Reference
CHIRPSv2.0	Precipitation (P)	50°S- 50°N, ~5 km	1981-present	daily	Merged remote sensing interpolated and calibrated using more than 14,000 rain gauges	Funk et al. (2015)
MOD16	Evapotranspiration (AET) and Potential Evap. (PET)	Global, ~5 km	2000-2014	monthly	AET and latent heat flux based on the Penman-Monteith equation incorporated remote sensed MODIS products	Mu et al. (2011)
CPC Global Temperature	Temperature (Tmin, Tmax, Tmean)	89.75S-89.75N, ~50 km	1979-present	daily	Gridded temperature from 6000 ~ 7000 global stations	https://psl.noaa.gov/
CCI Land Cover	Vegetation cover (Land Use)	Global, 0.3 km	1993-2015	annual	Land Cover maps derived from MERIS remote sensing products and classification models	Bontemps et al. (2013)
SoilGrids	Silt, sand and clay content	Global, 0.25 km	-	-	Soil properties derived from soil profiles and machine learning	Hengl et al. (2017)
SRTM	Land elevation	30 m	-	-	SAR interferometry	Bamler (1999)

3 Materials and Methods

3.1 The HYPE model structure and set-up

We used the Hydrological Predictions for the Environment (HYPE) version 5.9, a semi-distributed hydrological model for the assessment of water resources and water quality at small and large scales (Lindström et al., 2010) to simulate the hydrological response of Costa Rican catchments. The HYPE model could be considered as the evolution of the distributed Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Lindström et al., 1997). HYPE was developed by the Swedish Meteorological

and Hydrological Institute (SMHI) as the operative model for drought and flood forecasting across Sweden (Pechlivanidis et al., 2014). Moreover, HYPE was recently applied to other climatic regions (Andersson et al., 2017; Arheimer et al., 2018; Berg et al., 2018; Lindström, 2016; Pugliese et al., 2018; Tanouchi et al., 2019), including a global scale application (Arheimer et al., 2020).

The HYPE model allows simulating the water balance and nutrient fluxes at a daily or sub-daily scale using precipitation and temperature as forcings (SMHI, 2018). The model structure (Fig. 2.a) describes the major water pathways and fluxes, ensuring mass conservation at the catchment and sub-catchment scale. Furthermore, each sub-catchment is divided into the most fundamental spatial soil and land use classes (SLCs) depending on the classification of soil types, land cover, climate, and elevation, as shown in Fig. 2.b. The SLCs in HYPE provide the capability to predict streamflows in ungauged basins since the parameters that regulate the fluxes and storages are linked to each SLC, with a maximum of three layers of different soil thickness, as shown in (Fig. 2.b). Water bodies such as lakes and rivers may be considered as a SLC, where lakes can be defined as natural lakes or regulated dams with multiple water outputs. For full details of the HYPE model, see the description by Lindström et al. (2010) and the open-access code references at https://hypeweb.smhi.se/model-water/.

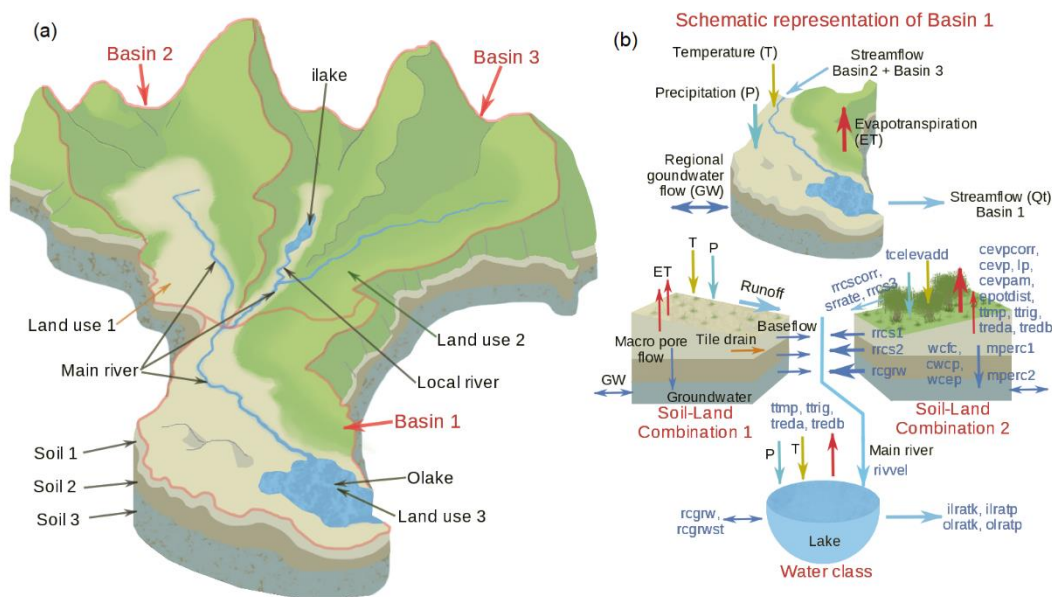



Figure 2. Schematic representation of the HYPE model (a) division into sub-basins and local classes according to topography, land use, and soil classes and (b) the model structure of Basin 1 considering two main soil-land combinations and lake properties. In (b), the simulated hydrological processes and variables are shown in black, while parameter names are given in blue. A full description of parameters (in blue) is available in https://hypeweb.smhi.se/model-water/. The *ilake* parameter corresponds to an internal lake, the *olake* parameter to an outlet lake, T is temperature, P is precipitation, ET is evapotranspiration, Qt is streamflow, GW is groundwater.

For this study, Costa Rica was divided into 605 catchments (Fig. 1.a) with 12 SLCs obtained from the spatial combination of soil types and land cover maps shown in Fig. 1.b and Fig. 1.c, respectively. The outlet lakes (lakes that discharge to downstream catchments) and internal lakes (lakes that discharge into the main river or tributaries) were set up as different SLCs to consider the water bodies that regulate the streamflow. The largest water body in Costa Rica is the Arenal reservoir, located in the San Carlos River catchment (Fig. 1.b). The Arenal reservoir is an artificial lake for hydropower purposes with an average surface area of 87.8 km^2 and a depth that ranges from 30 to 60 m. The Arenal reservoir was implemented as a natural lake since operational rules are confidential.

Soil thickness varied for different SLCs with a maximum soil thickness of 3 m under forest and a minimum of 2 m for bare soil cover, following Arheimer et al. (2020). Furthermore, delimited catchments were classified according to their elevation and location (Pacific basin and Caribbean basin), applying regional factors to correct the hydrological behavior of lowland and mountainous catchments with similar SLCs, resulting in six defined regions.

Daily time series of precipitation from CHIRPS and temperature from NOAA for the period 2000-2014 were extracted for each catchment using GRASS GIS (Neteler et al., 2012), where datasets were resampled to 1 km using the nearest neighbor criteria and spatially averaged for each catchment. The climatological forcings were resampled due to the small size of some catchments (area of ~1 km^2). Arheimer et al. (2020) recommended the computation of the average of the nearest grids to obtain the forcings instead of deriving the data from the nearest pixel.

3.2 Precipitation correction

Rainfall estimations from satellites are subject to systematic errors that may produce uncertainty in hydrological simulations (Goshime et al., 2019; Grillakis et al., 2018; Infante-Corona et al., 2014; Wörner et al., 2019). The CHIRPS product already incorporates a bias correction procedure but uses few concentrated ground stations in Costa Rica. The performance of CHIRPS estimating annual water balances is shown in Fig. S1.a and Fig. S1.c, with frequent underestimation in monitored catchments. Therefore, we applied a linear scaling to further correct the bias between the product and ground precipitation from 75 available stations across Costa Rica (Fig. 1.a). The corrected precipitation was estimated as:

$$CHIRPSc(t) = CHIRPS(t) * BF, \tag{1}$$

Where CHIRPSc is the bias-corrected precipitation at time t, CHIRPS is the original precipitation at time t, and BF is the bias factor. The bias factor was estimated as:

260 $$BF = \frac{\mu(P)}{\mu(CHIRPS)},$$ (2)

Where $\mu(P)$ is the mean of the historical precipitation from ground stations, and $\mu(CHIRPS)$ is the mean of the historical precipitation from CHIRPS. Note that $\mu(P)$ and $\mu(CHIRPS)$ were computed using the common study period. The simple linear bias correction was preferred over more complex methods due to the lack of a long common period for all stations. Therefore,

265 we used the individual records of more than 60 available stations covering a period from 1980 to 2010 to better capture the complex topography and resulting rainfall patterns.

Some monitored catchments exhibiting higher annual streamflow than annual precipitation could not be corrected due to groundwater contributions from neighboring catchments (Genereux and Jordan, 2005; Genereux et al., 2002), under-catch at rainfall gauges (Frumau et al., 2011), and the insufficient number of precipitation stations to correct the CHIRPS database at

270 a national scale. Nevertheless, errors in climatological data have been found the most common issue for water balance modelling in Central America (Westerberg et al., 2014; Birkel et al., 2012). In that sense, an additional approach was implemented to reduce the unrealistic relationship between streamflow and precipitation, which consisted in the creation of virtual stations at the catchment centroid where the new bias factor was computed as:

275 $$BF2 = \frac{\mu(Qty) + \mu(AETy)}{\mu(CHIRPSy)},$$ (3)

Where μ refers to the mean value, Qty is the annual streamflow from 1990 to 2003, AETy is the MODIS's annual actual evapotranspiration from 2001 to 2014, CHIRPSy is the CHIRPS's annual precipitation from 1990 to 2014. BF2 adjusts the long-term precipitation volume to ensure that the water balance is preserved, avoiding underestimation of streamflow and

280 evapotranspiration. Four streamflow gauges in addition to those shown in Table 1 were used to cover more spatial area in high elevations for the correction of satellite-based precipitation. These streamflow gauges were omitted from the model calibration-validation procedure due to their shorter records (lower than 7 years). The location of the four streamflow gauges and their catchments are shown in Fig. 1.b.

Finally, BF points from precipitation stations and BF2 from virtual points were interpolated using the Inverse Distance

285 Weighted (IDW) (Shepard, 1968) method with an exponent value of 2 at the original CHIRPS resolution. The interpolated map of the bias factor was used to spatially correct the time series of CHIRPS across Costa Rica.

3.3 Evapotranspiration and temperature correction

HYPE incorporates four methods for PET estimation (SMHI, 2018). After initial tests, we found that the monthly PET signal

290 from MODIS in Costa Rica can be reproduced by only using temperature as forcing, where PET is computed as:

$$PET = (cevp * cseason) * (temp - ttmp) * (1 + cevpcorr),\tag{4}$$

Where PET is the daily potential evapotranspiration (in mm), temp is the daily mean air temperature (°C), cevp is an
evapotranspiration parameter that depends on the land use (mm °C-1 d-1), ttmp is a threshold temperature for
evapotranspiration (°C), cevpcorr is a correction factor for evapotranspiration, and cseason is a factor computed as:

$$cseason = 1 + cevpam * sin\left(\frac{2*\pi*(dayno-cevpph)}{365}\right),\tag{5}$$

Where cevpam is a correction factor, dayno is the day of the year and cevpph is a factor to correct the phase of the sine function
to correct the potential evapotranspiration (set as zero in this study). To deal with the coarse spatial resolution of the
temperature database (0.5°), a correction factor that depends on catchment elevation was computed (SMHI, 2018):

$$tempc = temp - \frac{tcelevadd*elev}{100},\tag{6}$$

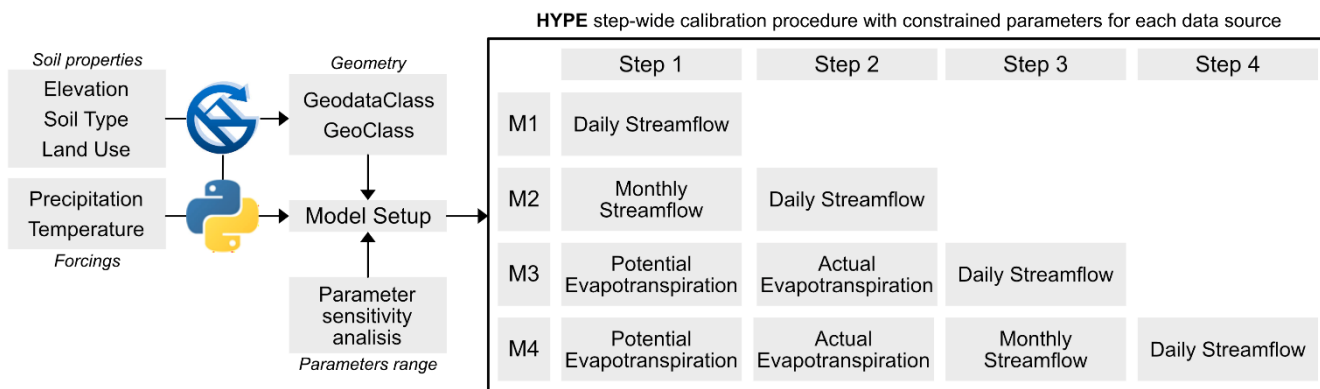
Where tempc is the corrected air temperature (in °C), temp is the original air temperature (°C), tcelevadd is a calibrated
parameter that corrects temperature (°C $100^{-1}m^{-1}$), and elev is the mean catchment elevation (m). Since only few (< 10)
temperature station records were available, a bias correction procedure was not possible, but measured temperature closely
followed the environmental lapse rate (Esquivel-Hernandez et al., 2017). The parameters cevp, cevcorr, cevpam, and tcelevadd
are part of the Monte Carlo simulation and their ranges are shown in Table 3.

3.4 Model calibration procedure

Figure 3 shows the workflow adopted for model calibration, which involves a qualitative parameter sensitivity analysis to find
the most suitable range of values for the automatic calibration. The initial parameter ranges were obtained from manual
iterations of one parameter at a time to facilitate automatic calibration (Infante-Corona et al., 2014).
We considered four model configurations to analyze the effect of including PET and AET into model calibration:

- Model configuration 1 (M1), calibrated using only daily streamflow (Qt).
- Model configuration 2 (M2), calibrated using monthly streamflow followed by daily streamflow.
- Model configuration 3 (M3), incorporates a calibration using monthly PET and AET, followed by daily streamflow.

320 • Model configuration 4 (M4), similar to M3 additionally used monthly streamflow before the daily streamflow calibration.

For comparison purposes, the M1 was chosen as the baseline model configuration, which usually is standard in hydrological practice. The steps are described in Fig. 3. The common period between Qt and PET-AET is relatively short (3 years), resulting in Qt and PET-AET calibration using different steps. The automatic calibration consisted of a step-wise procedure, where each
325 model configuration was calibrated for different fluxes (daily Qt, monthly Qt, monthly AET, monthly PET). The parameter names and initial ranges used for the calibration steps and their configuration are shown in Table 3 and Fig. 2. A final step (not shown in Fig. 3) consisted in calibrating the curve discharge parameters of the Arenal reservoir using observed water levels. However, the Arenal infrastructure does not contribute to the downstream basins and has a poor impact on the regional model calibration. Moreover, as previously stated, the reservoir was simulated as a lake since operational rules are unknown.
330 The streamflow records were divided into the period from 1991 to 1999 for calibration and from 2000 to 2003 for validation. The PET and AET calibration period was established from 2002 to 2010 and the validation period from 2011 to 2014. In both cases, we ran two years prior to calibration for model warm-up since our modeling tests showed that using two years was enough to stabilize the effects of initial conditions of water content in soil layers, rivers and reservoirs. The 13 monitored catchments were used for streamflow calibration. For PET and AET calibration steps, only the 130 catchments within the 13
335 monitored catchments were used since our tests showed that using the 605 catchments did not significantly increase the model performance but increased the calibration time by five. The simulations of the 605 catchments were used to compute the metrics for the calibration and validations periods.

340 **Figure 3**. Schematic representation of the HYPE model calibration strategy considering a step-wise procedure to constrain parameters. Four model configurations (M1, M2, M3, M4) were established using different data sets or different time scales. From each calibration step, the 10th and 90th values of the best-fit parameters were used to constrain the parameters of the next step.

A total of 86 parameters were used to build the HYPE model structure consisting of 36 parameters linked to 4 soil types, 24 parameters linked to 4 land covers, 6 for the general structure, 12 for the regional correction of PET and temperature, and 8 for lake discharge. The Monte Carlo (MC) routine for parameter sampling and sensitivity analysis included in HYPE was used for calibration, and the model configurations were run 10,000 times for each step, except for M1, which used 20,000 runs to cover more parameter combinations since this configuration only used daily streamflow. Despite the lower computational efficiency of the MC with respect to other optimization schemes (such as gradient-based methods), the MC routines are more flexible in accounting for multiple parameters sets in complex models (Beven, 2006). The 10th and 90th percentiles of the resulting parameters from the best 100 runs were used to constrain the parameters for the next calibration step.

Table 3. Parameter names and initial ranges of the step-wise parameter estimation for each model configuration. Columns M1, M2, M3, and M4 correspond to parameters optimized during each step of the configuration, and Qt, PET, and AET correspond to the observed time series used to calibrate the model configurations. Qtd means daily streamflow, Qtm monthly streamflow, PETm monthly potential evapotranspiration, and AETm monthly actual evapotranspiration.

				M1	M2		M3 and M4		M3	M3	M4
Model configuration				Step1	Step1	Step2	Step1	Step2	Step3	Step3	Step4
Random samples				20000	10000	10000	10000	10000	10000	10000	10000
Calibration and validation time step				Day	month	day	month	month	day	month	day
Variable for objective function				Qt	Qt	Qt	PET	AET	Qt	Qt	Qt
Calibration period (YY-YY)				91-99	91-99	91-99	02-10	02-10	91-99	91-99	91-99
Validation period (YY-YY)				00-03	00-03	00-03	11-14	11-14	00-03	00-03	00-03
Type	**Process**	**Parameter**	**Initial Rage**	**Step1**	**Step1**	**Step2**	**Step1**	**Step2**	**Step3**	**Step3**	**Step4**
Regional	PET	cevpcorr	-0.35-0	X	X		X				
	Runoff	rrcscorr	-0.1-0.1	X	X				X	X	
Soil	Runoff	srrate	0.01-0.3	X	X	X			X	X	X
	Runoff	rrcs1	0.01-0.3	X	X	X			X	X	X
	Baseflow	rrcs2	0.001-0.2	X	X				X	X	
	Percolation	mperc1	1-100	X	X			X			
	Percolation	mperc2	1-100	X	X			X			
	Water content	wcfc	0.2-1.0	X	X			X			
	Water content	wcwp	0.01-1.0	X	X			X			
	Water content	wcep	0.1-0.8	X	X			X	X	X	
Land	PET	ttmp	1.0-14	X	X		X	X			
	PET	cevp	0.05-0.4	X	X		X				
	AET	ttrig	0.5-1.5	X	X			X			
	AET	treda	0.01-1.5	X	X			X			

			C1	C2	C3	C4	C5	C6	C7	C8
AET	tredb	0.5-15	X	X			X			
Runoff	srrcs	0.01-0.3	X	X	X			X	X	X
General	AET	lp	0.4-1.4	X	X		X	X			
	PET	cevpam	0.15-0.4	X	X		X				
	Temperature	tcelevadd	-0.35 to -0.1	X	X		X				
	PET	epotdist	0.5-2.0	X	X		X				
	Streamflow	rivvel	0.1-0.5	X	X	X			X	X	X
	Runoff	rrcs3	0.0001-0.01	X	X	X			X	X	X

3.5 Model calibration and validation using hydrological signatures

360   The CHIRPS product was evaluated with ground records using the False Alarm Rate (FAR, computed with Eq. (7)), Probability of Detection (PD, computed with equation 8), and Threat Score (TS, computed with Eq. (9)):

$$FAR = \frac{falsealarms}{hits+falsealarms}, \tag{7}$$

$$PD = \frac{hits}{hits+misses}, \tag{8}$$

365   $$TS = \frac{hits}{hits+falsealarms+misses}, \tag{9}$$

Where hits are days with precipitation detected by CHIRPS and ground rain gauges, false alarms are days where precipitation was detected only by CHIRPS, and misses are days where precipitation was detected only by rain gauges.

The model performance was evaluated using the Kling-Gupta Efficiency (KGE) computed as (Kling and Gupta, 2009):

370

$$KGE = 1 - \sqrt{(r-1)^2 + (\propto -1)^2 + (\beta-1)^2}, \tag{10}$$

$$r = CC = \frac{cov(xo,xs)}{\sigma_o \sigma_s}, \tag{11}$$

$$\propto = \frac{\sigma_s}{\sigma_o}, \tag{12}$$

$$\beta = \frac{\mu_s}{\mu_o}, \tag{13}$$

375

Where suffixes o and s correspond to observations and simulations, respectively; μ is the mean, x is the time series (streamflow, actual evapotranspiration or potential evapotranspiration), σ is the standard deviation, r or CC is the correlation coefficient, α is the agreement between amplitude and β is the bias. KGE was chosen as the objective function for calibration since it equally captures maximum and minimum flows (e.g., Arheimer et al., 2020; Pechlivanidis and Arheimer, 2015; Rajib et al., 2018a;

380 Rakovec et al., 2016; Xiong and Zeng, 2019), and has been described as a relatively balanced metric with slightly more focus on high flows (Garcia et al., 2017). Non-transformed data was used since Santos et al. (2018) advice against the use of log-transformed discharge with the KGE for low flow evaluation. Furthermore, other statistical criteria were computed to facilitate assessing the performance of the model configurations, such as the Pearson correlation coefficient (computed with Eq. (11)), Mean Absolute Error (MAE, computed with Eq. (14)), Nash-Sutcliffe efficiency (NSE, computed with Eq. (15)), Root Mean

385 Square Logarithmic Error (RMSLE, computed with Eq. (16)); and relative bias (Eq. (17)):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_s(i) - x_o(i)|, \tag{14}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(x_s(i) - x_o(i))^2}{\sum_{i=1}^{n}(x_o(i) - \mu_o)^2}, \tag{15}$$

$$RMSLE = \sqrt{\frac{\sum_{i=1}^{n}(\log(x_s(i)) - \log(x_o(i)))^2}{n}}, \tag{16}$$

390 $$Bias = \frac{\mu_s - \mu_o}{\mu_o}, \tag{17}$$

Furthermore, hydrological signatures were calculated to independently assess how well the calibrated model configurations reproduce different hydrological criteria. The hydrological signatures used in this study are shown in Table 4 and the Budyko curve was constructed from the Aridity and Evaporative Index (AI, EI).

Finally, the non-parametric tests of Kruskal-Wallis (Kruskal and Wallis, 1952) and Mann-Whitney (Mann and Whitney, 1947)

395 were used to detect statistically different performances.

Table 4. Hydrological signatures used as independent performance evaluation criteria

Signature	Equation	Description
Mean.Qtd	$\mu = \frac{1}{n}\sum_{i=1}^{n} Qd(i)$	Mean flow of daily streamflow series
Median.Qtd	$m = \frac{1}{2}\left(Qd\left(\frac{n}{2}\right) + Qd\left(\frac{n+1}{2}\right)\right)$	Median value of daily streamflow series
Slope.Qtd	$slope = \frac{Qd_{0.33} - Qd_{0.66}}{0.66 - 0.33}$	Slope of the flow duration curve
CV.Qtd	$CV = \frac{\mu(Qd)}{\sigma(Qd)}$	Variation coefficient, ratio between mean and standard deviation

SC	$SC = \dfrac{1}{N}\sum\limits_{y=1}^{N}\dfrac{\sum_{i=1}^{365} Qd(y,i)}{\sum_{i=1}^{365} P(y,i)}$	Streamflow Coefficient, mean of annual streamflow divided by annual precipitation
BFI	$BFI = \dfrac{1}{N}\sum\limits_{y=1}^{N}\dfrac{\sum_{i=1}^{365} Qb(y,i)}{\sum_{i=1}^{365} Qd(y,i)}$	Base Flow Index, mean of annual baseflow divided by annual streamflow
AI	$AI = \dfrac{1}{N}\sum\limits_{y=1}^{N}\dfrac{\sum_{i=1}^{365} PET(y,i)}{\sum_{i=1}^{365} P(y,i)}$	Aridity Index, mean of annual potential evapotranspiration divided by annual precipitation
EI	$EI = \dfrac{1}{N}\sum\limits_{y=1}^{N}\dfrac{\sum_{i=1}^{365} AET(y,i)}{\sum_{i=1}^{365} P(y,i)}$	Evaporative Index, mean of annual actual evapotranspiration divided by annual precipitation
FDC	$sort(Qd)$	Flow Duration Curve, a plot displaying the statistical distribution of daily streamflow in a decreasing order

4 Results

4.1 Remote sensing input data bias correction and evaluation

Comparing precipitation from CHIRPS with annual streamflow and streamflow plus evapotranspiration (assuming long-term balance P-Qt-AET=0) showed underestimated annual precipitation (as shown in Fig. S1.a and Fig. S1.c from the Supplementary Material), leading to unrealistic water balance values. The interpolated bias correction factor (BF, Fig. 4.a) showed overestimated CHIRPS rainfall in blue and underestimations in red. The BF ranged from 0.65 to 1.57 with an average of 1.06±0.14, where the higher disagreements between the ground precipitation and satellite-merged precipitation were observed along the Pacific basin. The underestimation reached 30 to 35 % in the north of the Gulf of Nicoya and in the southwest of the Providencia catchment. Underestimation of CHIRPS across the Caribbean slope was mainly observed in the Terron Colorado and Cariblanco catchments, with a BF between 1.2 and 1.4. Moreover, the largest overestimation of CHIRPS was observed for the Guanacaste region (BF=0.65-0.8), downstream of the Tacares catchment (BF=~0.8), and to the south-east of Costa Rica (BF=0.8-0.85).

For modelling purposes, we evaluated the temporal synchronicity of rainfall versus streamflow (Fig. 4.b) using cross-correlation between daily streamflow and catchment-scale daily precipitation from CHIRPS, where the x-axis corresponds to the lag time in days. Most of the monitored catchments exhibited the highest correlation within lag time zero, indicating that the hydrological response of catchments tends to occur within the same day. Nevertheless, the Cariblanco and Rancho Rey catchments exhibited a poor correlation (ρ<0.3), which means a lack of synchrony between daily satellite-merged precipitation and streamflow.

The bias correction improved the annual precipitation where CHIRPSc was consistent with annual streamflow, and the long-term water balance was mostly preserved, as observed in Fig. S1.b and Fig. S1.d. Figure S2 shows the mean absolute error (MAE) normalized by mean precipitation for CHIRPS and bias-corrected CHIRPS (CHIRPSc), both with respect to the 75 precipitation stations, where boxplots correspond to the variability of normalized MAE estimated by each point. The average

normalized MAE at a daily scale was estimated at 1.2±0.12 mm/mm for CHIRPS and 1.17±0.08 mm/mm for CHIRPSc, 0.30±0.09 and 0.27±0.07 mm/mm at a monthly scale, and 0.16±0.09 and 0.11±0.04 mm/mm at an annual scale, respectively. Figure S2.d shows the probability of success and failure of CHIRPSc to detect rainy or dry days with respect to ground stations, where the probability was computed from a single time series merged from the 75 station records. Furthermore, Figure S2.e to Fig. S2.g shows the False Alarm Rate (FAR), Probability of Detection (PD), and Threat Score (TS), respectively. Results indicated that CHIRPSc detected true rainy and dry days with a similar probability (0.31 to 0.34) compared to the in situ observed rainfall. Whereas the FAR ranged from 0.15 to 0.38 with the larger values (i.e., incorrect detection of dry days as rainy days by CHIRPS) to the southeast, and the PD showed larger values (i.e., better performance of CHIRPS to detect rainy days) on the Pacific basin (median of ~0.69) in comparison with the rain gauges on the Caribbean basin (median PD of ~0.54). The TS showed similar results to PD (Fig. S2.g), with a better performance of CHIRPS on the Pacific basin (median of ~0.53) than for the Caribbean basin (median of ~0.46). Such low capacity of CHIRPS to detect rainy days in the Caribbean basin could affect the performance of the hydrological model during peak flows.

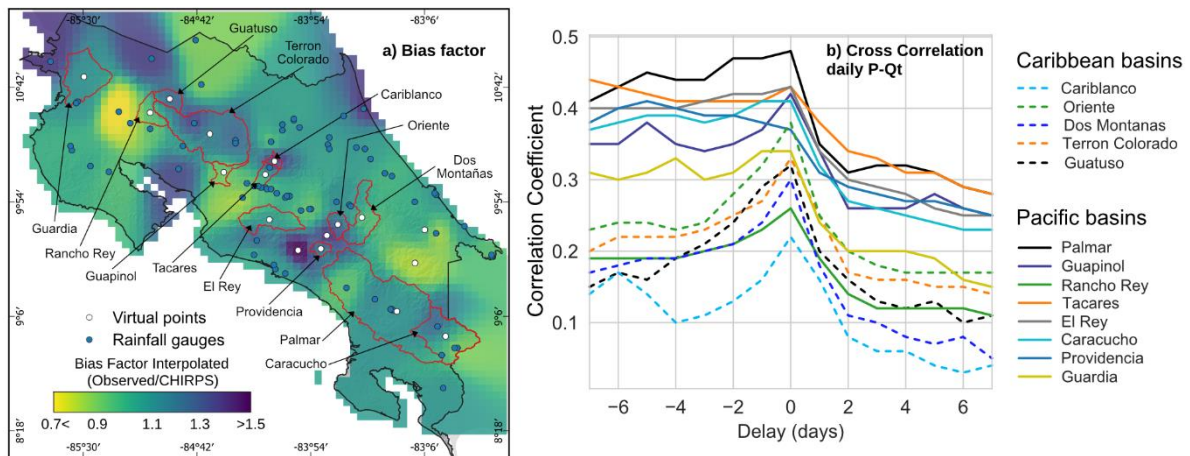



Figure 4. Performance of the CHIRPS precipitation product in representing observed rainfall in Costa Rica. (a) interpolated bias factor where red areas indicate underestimation and dark blue areas overestimation of CHIRPS with respect to ground stations, and (b) cross correlations between daily precipitation from CHIRPS and observed daily streamflow, where high correlation values without delay (at zero) indicate that precipitation and high flows tend to occur at the same day. A high correlation with negative delays indicates that precipitation occurred on average before the streamflow response.

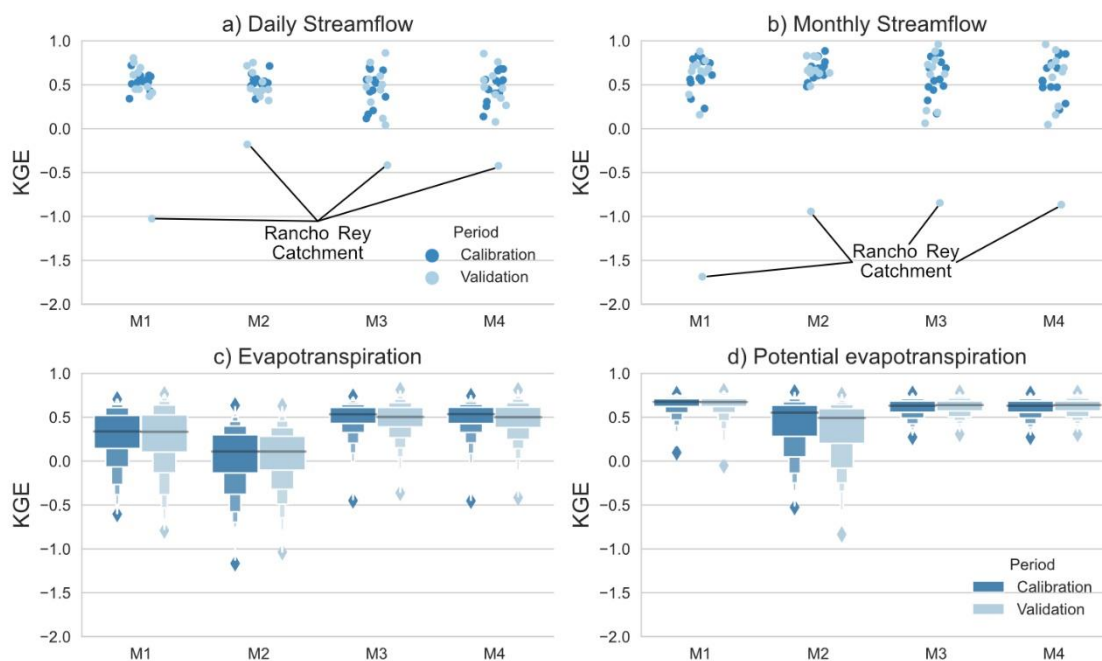
4.2 Model performance and parameter uncertainty

Figure 5 shows the comparison of the model configurations' performance for the calibration (dark blue color) and validation (light blue color) periods. Simulated daily streamflow for the 13 gauged catchments was similar for baseline configuration

(M1) and M2 during the calibration period (1991-1999) (Fig. 5.a) with a mean KGE of 0.54 ±0.09 and 0.53±0.08, respectively.

445 Nevertheless, M1 showed a median NSE of 0.33, in comparison to M2 that showed a median NSE of 0.25. The comparison of the correlation coefficient, median absolute error, and Nash-Sutcliffe Efficiency metrics is shown in Figure S3. Moreover, metrics for M3 and M4 were slightly poorer due to a larger dispersion across the sample, with a mean KGE of 0.45±0.2 and 0.47±0.17, respectively, and a mean NSE of 0.23±0.2 and 0.21±0.21. For the validation period (2000-2003), the mean KGE decreased by ~0.08, but with similar performance for NSE.

450 The configuration M2 best reproduced monthly streamflow for the calibration period (Fig. 5.b), with a mean KGE of 0.67±0.11, whereas the configurations M1, M3, and M4 showed a mean KGE of ~0.60, also driven by larger dispersion along the KGE scale. The NSE also supports M2 as the best monthly streamflow predictor for the calibration period (Fig. S3), with a median value of 0.54, in contrast to M1 that showed a median NSE of 0.43. The four model configurations preserved performance for the validation period, and in some cases, the KGE even increased, as was the case for the Palmar, Caracucho,

455 and El Rey catchments (not shown). Nevertheless, the Rancho Rey catchment exhibited poor performance during the validation period (KGE<0 and NSE<-2) for daily and monthly scales since the four configurations overestimated streamflow. We present more details for the Rancho Rey that could explain the catchment behavior and its performance in the following sections.

460 **Figure 5**. The range of KGE values for the calibration (dark blue color) and validation periods (light blue color). (a) the KGE statistical dispersion for daily streamflow, (b) the KGE statistical dispersion for monthly streamflow, (c) boxplots of KGE values for AET, and (d) boxplot of KGE for PET. Streamflow calibration period from 1991 to 1999 and validation period from 2000 to 2003. PET and AET calibration period from 2001 to 2010 and validation period from 2011 to 2014. Since

configurations M1 and M2 were calibrated only with streamflow, panels (c) and (d) are for comparison purposes only showing the effect of including PET and AET in the calibration procedure.

Figure 5.c and Fig. 5.d show the effect of including AET and PET in the calibration steps, and the KGE was computed by aggregating the complete domain (605 nested catchments). The calibration consisted of 130 nested catchments within the monitored catchments. Furthermore, M1 and M2 were only plotted for comparison purposes since these configurations were calibrated with streamflow. From Fig. 5.c, we observed that simulated monthly AET for the calibration period (2002-2010) improved for M3 and M4 with a mean KGE of ~ 0.49±0.17 with respect to M1 with a mean KGE of 0.29±0.29 and M2 (0.04±0.33). The higher performance of AET was also observed for M3 and M4 according to the correlation coefficient and MAE (Fig. S3). Surprisingly, the baseline configuration (M1) showed a slightly better performance of simulated monthly PET, with a mean KGE of 0.64±0.09, whereas M3 and M4 showed a mean KGE of ~0.61±0.10, and M2 a mean KGE of 0.43±0.28 (Fig. 5.d). The monthly AET and monthly PET performance were similar for the validation period (2011-2014) (Fig. S3).

The results from Fig. 6 suggested that the best performances of daily and monthly streamflow for the calibration period (2001-2009) were obtained for catchments in the south Pacific, such as the Palmar, Caracucho, El Rey, and Guapinol catchments with KGE's higher than 0.55 (NSE>0.4, as shown in Fig. S4) for daily streamflow and KGE higher than 0.8 (NSE>0.63) for monthly streamflow. Nevertheless, the mid-Pacific basin also resulted in the Tacares and Providencia catchments exhibiting the worst performances for monthly streamflow and the configurations M3 and M4 with KGE<0.3 and NSE<0 (Fig. 6.g and Fig. 6.h).

The spatially distributed KGE on the last two panels of Fig. 6 shows the improvement by including AET and PET in the calibration steps (panels (k), (l), (o), (p)), unlike the case of daily and monthly Qt where no significant improvements were observed using the four calibration procedures. The calibrated monthly AET simulated with M1 showed low efficiency (KGE<0.2 and NSE<0) for ~182 catchments of the Pacific basin but an acceptable performance (KGE>0.6 and NSE>0.2) for monthly PET. The M2 exhibited poor performance across the simulation domain for AET and low efficiency of PET in the southeast Caribbean. Additionally, M3 and M4 showed similar results with acceptable performance (KGE>0.6 and NSE>0.2) for ~179 catchments, most of them located in the northeast. The median MAE for M3 and M4 was ~11% lower (MAE=~15 mm) than for M1, and the median correlation coefficient was ~39% higher (CC>0.63) for M3 and M4. Surprisingly, the simulated PET with M3 and M4 was similar to PET from M1. The performance of the calibrated water level on the Arenal reservoir was relatively low for all configurations (KGE~0.35, NSE<-0.1, and CC~0.36), affected by the unknown withdrawals from the reservoir during the driest months (Apr-Jul).

Fig. 7 shows the model parameter ranges from the 100 best-fit simulations resulting from the last calibration step for each configuration. The red dots from Fig. 7 correspond to the optimal parameters used for modelling, where multiple red dots and boxes for each model are shown by soil type and land use.

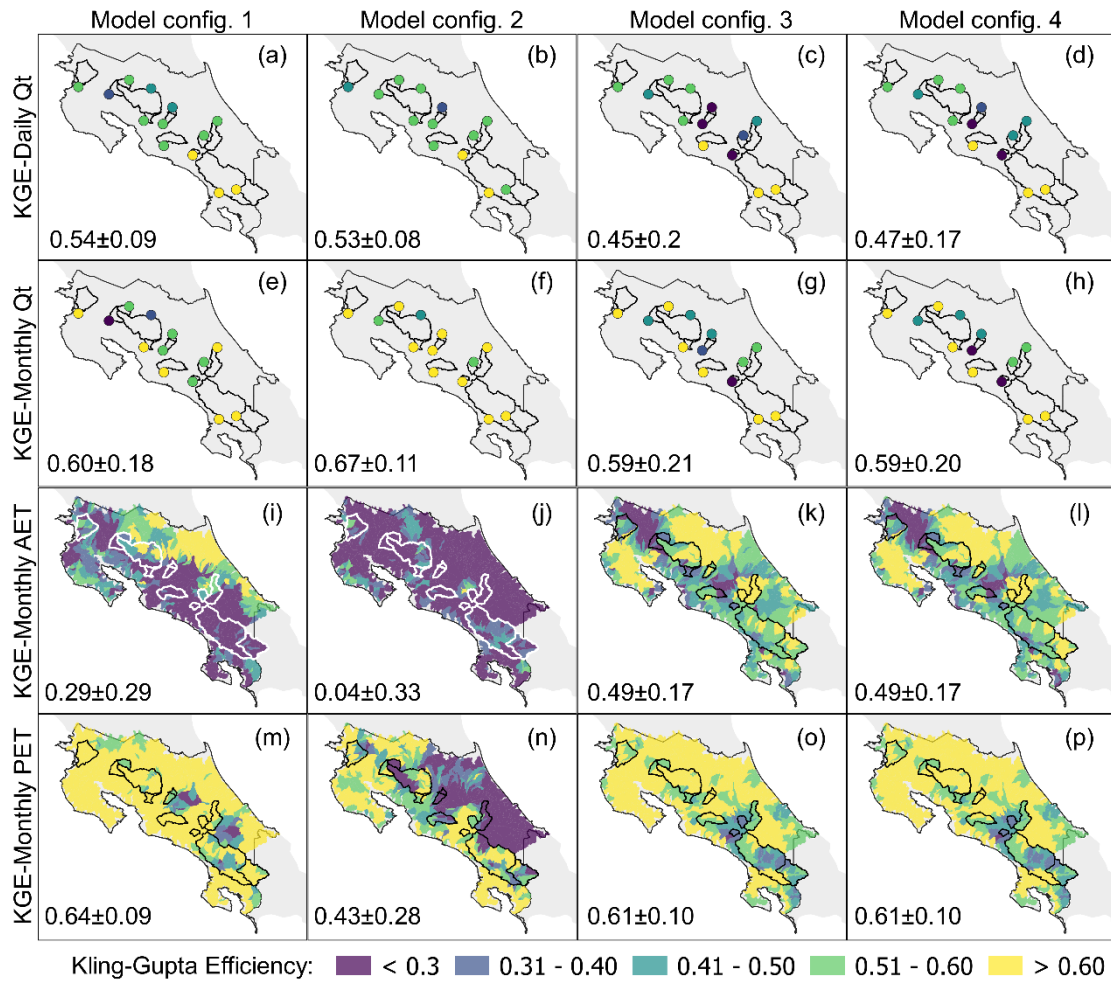



Figure 6. Matrix of spatially distributed KGE results for the calibration period (streamflow from 1991 to 1999, PET and AET from 2001 to 2010), where green and blue colors reflect better performance. Configurations M1 and M2 were calibrated only with streamflow. Nevertheless, PET and AET panels are comparative to show the effect of including such variables in the calibration procedure. The mean±std of KGE is shown on the lower left for each panel.

A large dispersion with a coefficient of variation (CV=std/mean) greater than 0.35 was observed for runoff response parameters (srrate, fraction for surface runoff; srrcs, recession coefficient for surface runoff) and baseflow parameters (rrcs1, recession coefficient for uppermost soil layer; rrcs2, recession coefficient for lowest soil layer). The impacts of monthly streamflow on calibration were observed for the general model parameters of rivvel (river velocity) and rrcs3 (deep layer recession coefficient) with constrained posterior parameter distributions for configurations M2 and M4 and larger velocities and baseflow discharge for M2 with respect to M4.

510   The soil type and land use coverage influence the calibrations' parametrization. M2 and M4 showed constrained distributions of parameters srrate and rrcs1 for clay-loam soil (third class), the most frequent soil type in the monitored catchments (Fig. 1.b). The bottom panel at Fig. 7 shows the spatial distribution of the srrate parameter, with similar values for M2, M3, and M4 and the most frequent soil classes (clay and clay-loam).

The soil parameters that regulate the soil water content (wcwp, wcep) showed similar distributions with the median value of
515   the fraction of soil water available for evapotranspiration (wcfc). The effective porosity (wcep) was slightly higher for configurations M1 and M2, but the final parameters (red dots) differed between models. Furthermore, for M3 and M4, the parameters lp and cevpam exhibited constrained distributions with a CV of 0.12 and 0.11, respectively. In comparison, M1 and M2 showed CV values of ~0.25 and ~0.28 for lp and cevpam.

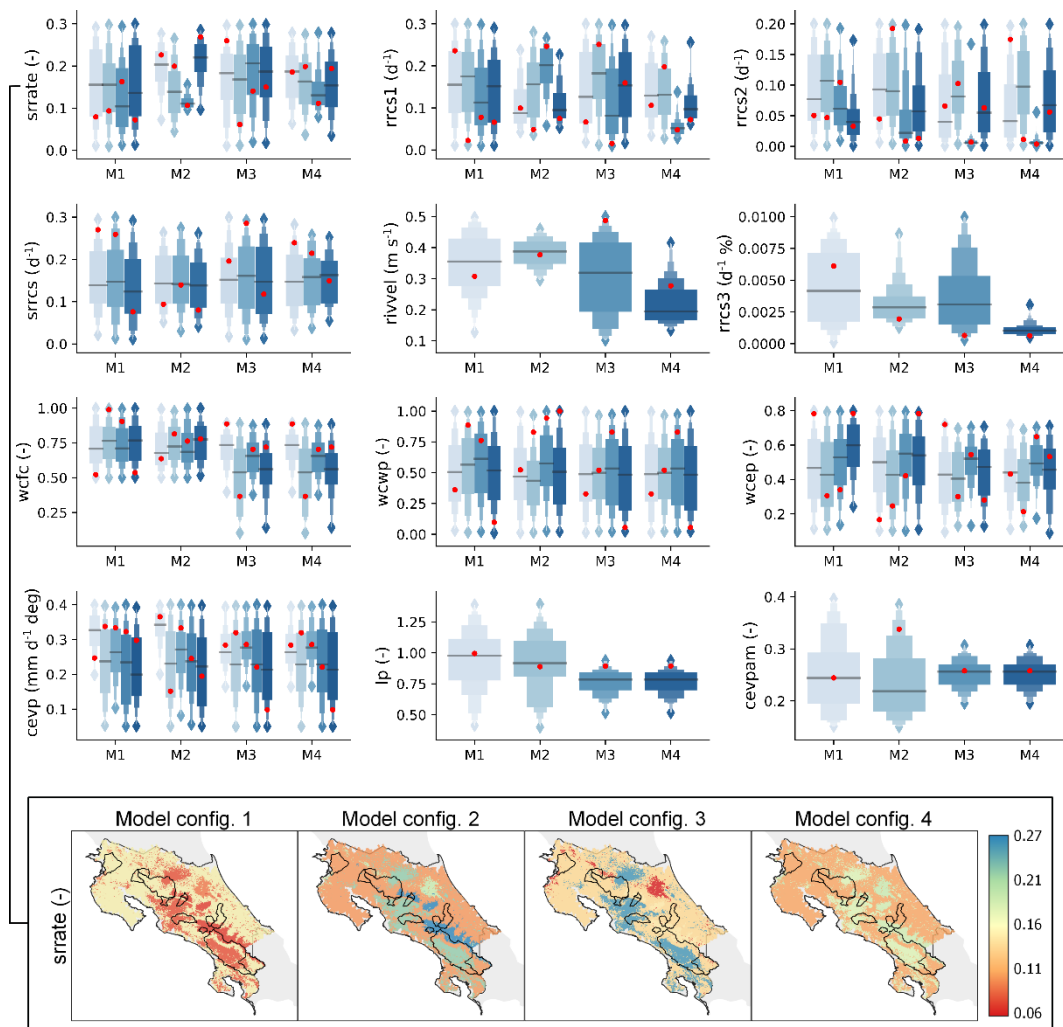



Figure 7. A *posteriori* parameter distribution for the 100 best-fit simulations from the last calibration step for each configuration, where red dots correspond to the optimal parameters. Multiple boxes by model configuration correspond to parameters for different soil and land classes. The first two parameter panels correspond to streamflow components, the third panel to water content parameters, and the last panel to PET and AET processes (see Table 3 for reference). For comparison purposes, the bottom panel shows the spatial variability of the best-fit calibrated srrate (-) parameter for each configuration.

4.3 Evaluating streamflow simulations and hydrological signatures

The step-wise calibration improved the model performance in different aspects. Figure 8 shows the comparison of the hydrological simulations for two monitored catchments contrasting the best simulation with the highest KGE performance (Palmar catchment) and the worst simulation with the lowest KGE performance (Rancho Rey catchment).

The Palmar catchment exhibited an acceptable performance (KGE>0.5 and NSE>0.45) for daily streamflow, but all configurations underestimated the highest peak flows during the calibration and validation periods. For the Rancho Rey catchment, the observed highest peak flows during 1997 were three times larger than simulated peak flows. Underestimation of simulated peak flows was related to the poor capabilities of CHIRPSc to detect heavy storms since observed peak flows were not associated with large precipitation amounts (Fig. 8.a). In the Palmar and Rancho Rey catchments, the baseline (M1) underestimates low flows by one and two orders of magnitude during the dry season, respectively (Fig. 8.d and Fig. 8.j). Moreover, daily streamflow for M2 exhibited a mean relative bias (Eq. (17)) of -0.046 with respect to M1 due to underestimation of peak flows in all monitored catchments, and M3-M4 showed a mean relative bias of 0.156. The mean relative bias with respect to baseline (M1) using the logarithm of daily streamflow ranged from 0.24 to 0.29; hence, configuration M1 tends to generate lower flows during the dry period.

At a monthly scale, streamflow was preserved by the model configurations in several catchments, except for Rancho Rey where simulated streamflow was on average two times larger than the observed streamflow during rainy season (Fig 8.l). Such overestimation indicated that the bias factor was insufficient to correct the global precipitation product or large discharge measurement errors. Furthermore, all configurations reproduced the seasonality of AET and PET from MODIS (not shown), but M3 and M4 underestimated the AET and PET in Palmar with a good performance for AET in Rancho Rey. Moreover, simulated monthly soil moisture (SM) content was independently compared with the catchment average soil moisture content from the LPRM product for the period 2012 to 2016. The simulated SM for M1 followed the seasonal behavior of the LPRM product in the Palmar catchment, matching the absolute LPRM % SM content. The LPRM product uses SM from the upper 5 cm against the 50 cm of the upper layer defined for all the model configurations. However, all model configurations have shown a high correlation (CC>0.7) in both catchments (Palmar and Rancho Rey), matching the seasonality.

The observed and simulated flow duration curves for all monitored catchments are shown for the period 1991-2003 in Fig. 9. The baseline (M1, red dashed line) underestimated the median and low flows in several catchments (Guardia, Rancho Rey, Guatuso, Terron Colorado, Caracucho, El Rey, Guapinol), with a median RSME of 1.15 considering all monitored catchments,

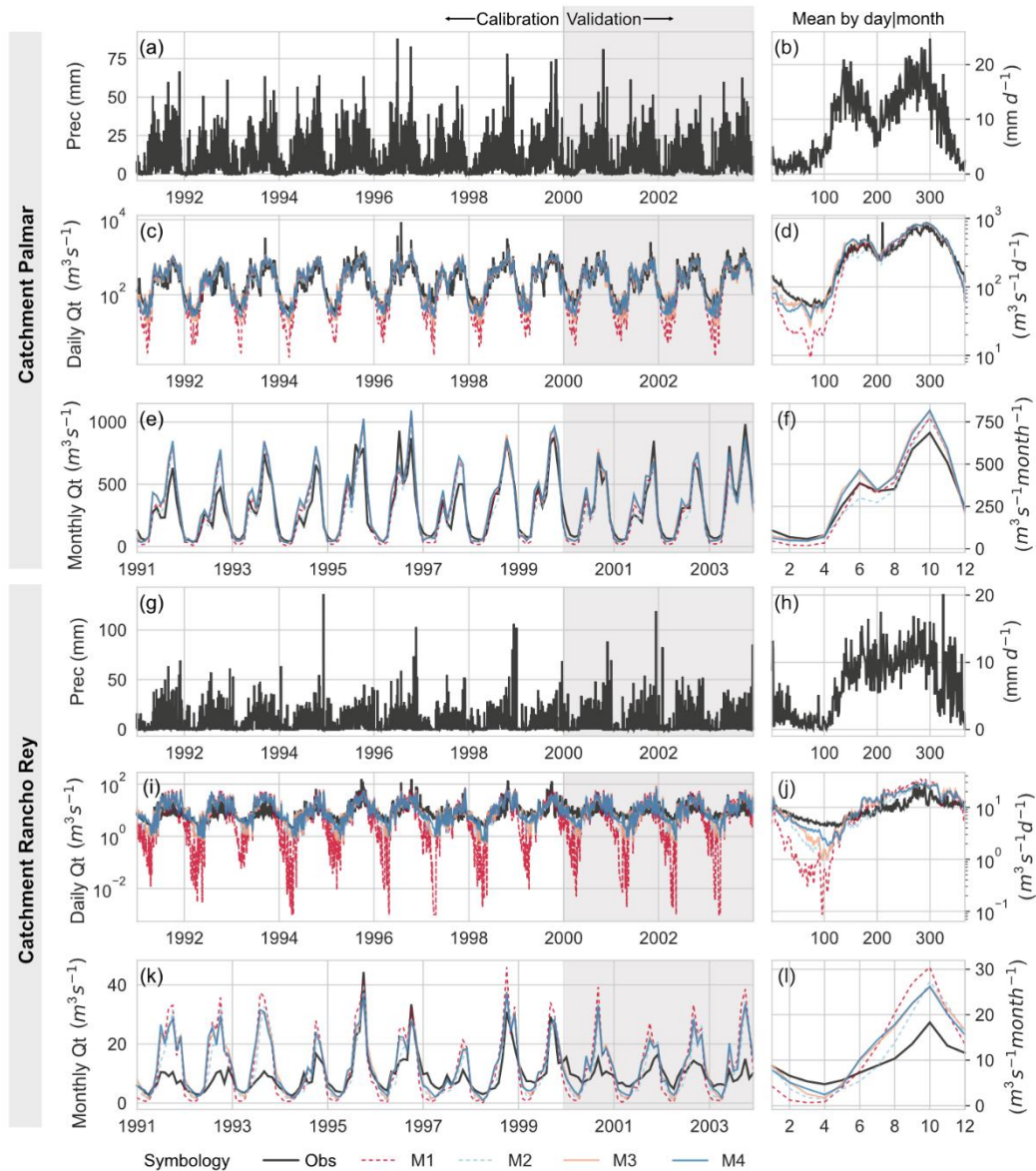
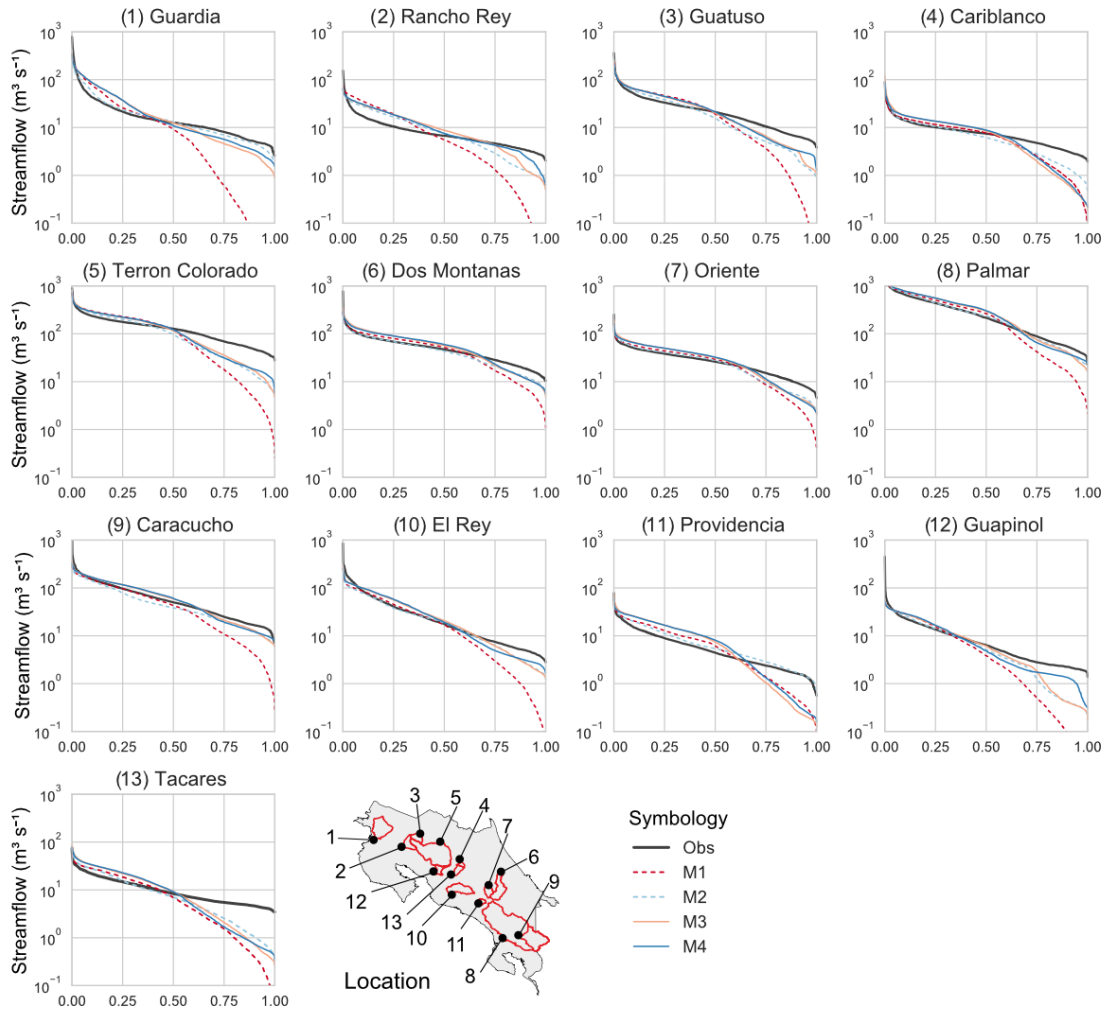
two times larger than other model configurations. The M2 (blue dashed line) exhibited the best performance for median and
555   low flows (with a median RSME of 0.42), whereas M3 (orange line) and M4 (blue line) showed similar results to M2. Higher
efficiencies for median and low flows were obtained for catchments that exhibited higher cross-correlation with precipitation
(Fig. 4.b), as was the case for Palmar, Caracucho, and El Rey, among others.

560   **Figure 8**. Simulated versus observed time series for catchments with the best streamflow KGE performance (Palmar
watershed) and the worst streamflow simulation (Rancho Rey watershed). Black lines correspond to observations where
streamflow is shown from 1990 to 2003. Right axes (b, d, f, h, j, l) show the mean daily and monthly time series.

565

Figure 9. A comparison of observed flow duration curves (FDC) as a hydrological signature and the matching simulated FDC for each model configuration. The simulated period was from 1991 to 2003.

The simulated and observed hydrological signatures are shown in Fig. 10, where simulations covered the period 1991-2014, and observations covered different periods depending on available records. The comparison of hydrological signatures by monitored watershed is shown in Table S1 and Table S2. The simulated long-term mean annual water balance (Prec-Qt-AET) was mostly closed in all catchments (~0 mm), with average values of -1.80 mm (M1), -0.62 mm (M2), -0.79 mm (M3), and -1.17 mm (M4), hence the water balance was significantly improved when compared to the observed water balance. Indeed, observed (Prec-Qt-AET) yielded values going from -800 to 600 mm. Such variability in observed water balances may be

575 related to the short common period of data but also due to the discrepancies between data sources (in situ, interpolated and merged, remotely sensed), as can be observed in Fig. S4 from the Supplementary Material. Figure S4 shows how the long-term water balance using the observed data differs from the Budyko curve in all monitored catchments (Westerberg et al., 2014), while the simulations fitted the theoretical curve.

The spatial distribution of baseflow indices (BFI) derived from M2, M3 and M4 exhibited similarities with respect to the
580 observations. Simulated BFI showed an overwhelming groundwater contribution to streamflow with relatively similar average values of 0.70, 0.69, 0.68, and 0.74 for the model configurations M1, M2, M3, and M4, respectively.

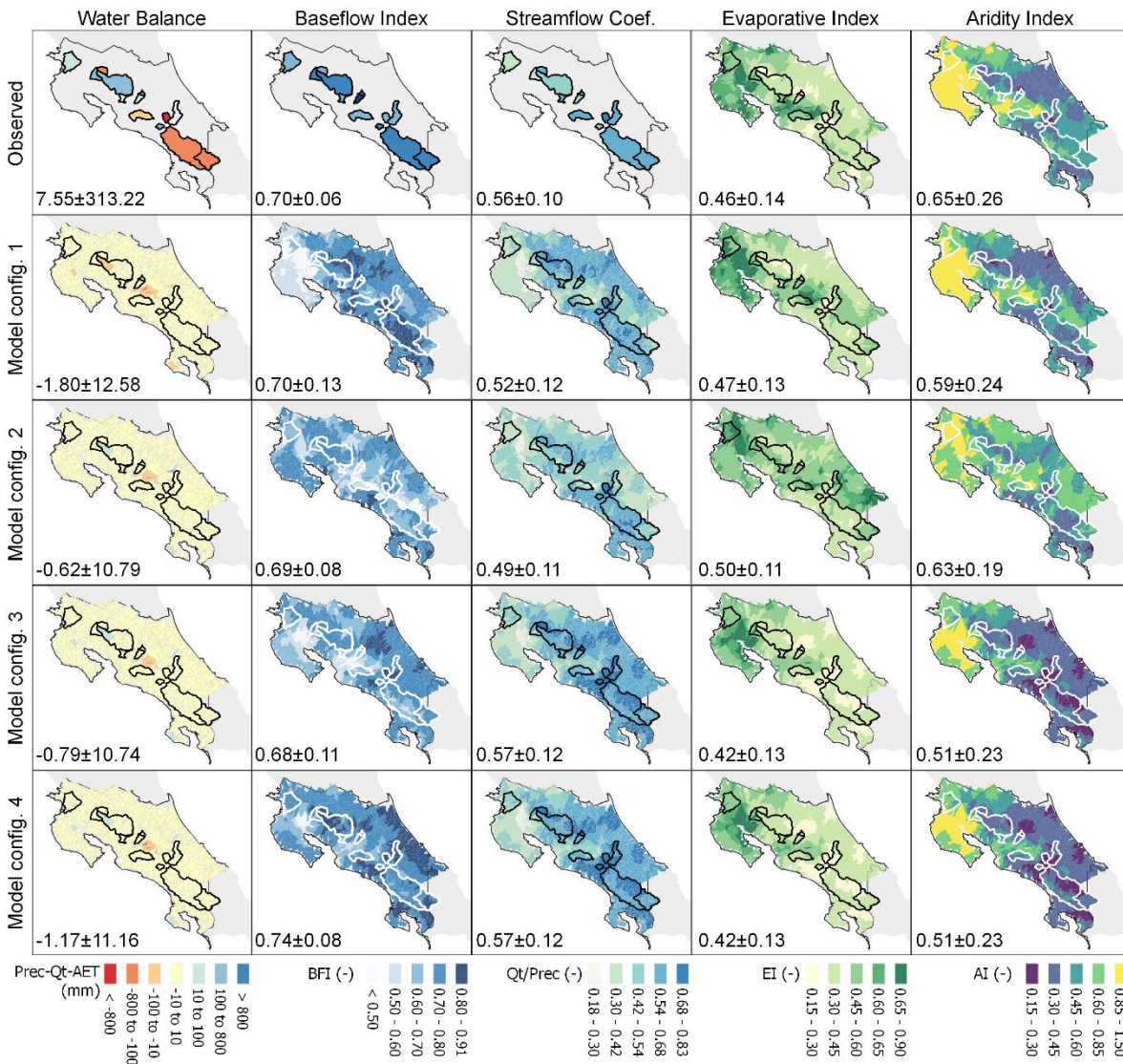



Figure 10. Observed and simulated spatial distribution of hydrological signatures. BFI: Baseflow Index (Baseflow / Streamflow), EI: Evaporative Index (AET / P), AI: Aridity Index (PET / P). Signatures from observations were obtained for the periods 2000-2003 (Balance), 1991-2003 (BFI, SC), and 2002-2014 (EI, AI). Signatures from simulations were obtained for the period 1991-2014, and the mean±std are shown on the lower left for each panel.

Larger differences were observed in the northwest and southwest when comparing the BFI of M1 with respect to other configurations, whereas M4 resulted in larger contributions of baseflow to streamflow in coastal areas of the Caribbean. Similar spatial patterns were obtained for the streamflow coefficient (SC=Qt/Prec), with low values in the drier northwest and higher values for catchments that receive more rainfall (Fig. 1.e). In contrast, the M2 indicated that a lower amount of precipitation became streamflow, with an average value of 0.49, in comparison with M1, M3, and M4 that showed medians of 0.52, 0.57, and 0.57, respectively. Moreover, M1 and M2 followed the spatial patterns of observed SC due to their higher streamflow performance.

The Evaporative index (EI) and Aridity index (AI) were similar for M3 and M4 due to similar model parameters. The spatial distribution of observed EI from MODIS was reproduced by M3 and M4, whereas AI spatial patterns were preserved by M1. Besides, M1 and M3-M4 showed similar spatial patterns for EI and AI across the north, but differences were observed in the south, where M1 indicated lower water availability attributed to higher evaporative ratios (higher EI). M2 simulated the driest catchments with an average value of EI and AI of 0.50 and 0.63, respectively, whereas M1 showed median values of 0.47 and 0.59, and M3-M4 values of 0.42 and 0.51.

4 Discussion

4.1 Remote sensing and global products as model forcings and calibration series

Daily precipitation from CHIRPS was preferred over other global precipitation products because of a relatively higher spatial resolution and good performance across different climates and biomes (e.g., Bayissa et al., 2017; Ullah et al., 2019; Zambrano-Bigiarini et al., 2017). Nevertheless, CHIRPS showed a large bias and more rainy days with respect to ground precipitation across Costa Rica (Fig. 4.a). The results suggested that at large scales, the precipitation bias was compensated since the mean bias factor (BF) was ~1, but underestimation of precipitation was observed in mountainous regions as wells as large overestimations in the drier northwest (Fig. 4). Similarly, results from Chile by Zambrano-Bigiarini et al. (2017) indicated that an evaluation and even correction of global products is necessary and may be related to the poor density of ground gauges used to generate CHIRPS in those regions, especially across regions with steep topography (Funk et al., 2015). The latter authors also reported an underestimation during extreme rainfall events and an overestimation during rainy days.

Our simple, linear bias correction of CHIRPS showed better performance at monthly and annual scales and solved the water balance inconsistencies of most catchments (Fig. 10). However, the cross-correlation between daily precipitation and

streamflow remains unchanged by the bias correction. Not surprisingly, our results showed that catchments with highly correlated streamflow and daily precipitation exhibited better performance than catchments with low correlations. Several studies highlighted those meteorological forcings are the largest source of uncertainty in hydrological modelling (e.g., Arheimer et al., 2020; Dal Molin et al., 2020; Lin et al., 2018; Wörner et al., 2019), whereas more complex bias correction techniques (e.g., quantile mapping) may improve the results (e.g., Goshime et al., 2019; Wörner et al., 2019). However, the lack of matching daily streamflows with precipitation inputs and intense rain events might persist. Nonetheless, Infante-Corona et al. (2014) suggested that global products can achieve better streamflow simulation results than sparse ground precipitation data, whereas Westerberg and Birkel (2015) found that in situ precipitation in Costa Rica may require corrections to achieve better model results.

The global CPC temperature dataset used was not bias-corrected due to the lack of sufficient in situ measurements. Temperature can introduce large errors in hydrological simulations if used for the estimation of potential evapotranspiration and actual evapotranspiration (Andersson et al., 2015). We corrected the temperature data set using elevation and a lapse rate parameter (Eq. (6)) during the calibration steps. The corrected temperature closely followed the environmental lapse rate of 6ºC temperature decrease per 1000 m elevation gain and improved model performance (KGE) by on average ~10%. Global evapotranspiration estimates showed differences compared to ground estimations in different regions due to the influence of, e.g., irrigation, vegetation dynamics, and uncertainty in climatological forcings (Pan et al., 2019; Velpuri et al., 2013). We used the PET and AET from MODIS16 for calibration and evaluation due to reported good performance for different applications around the world (e.g., Lin et al., 2018; Mu et al., 2013; Pan et al., 2019; Rajib et al., 2018b; Tang et al., 2011; Velpuri et al., 2013). Nevertheless, MODIS AET has shown poor performance at point-scale in different regions (e.g., Liu et al., 2015; Weerasinghe et al., 2019), but better performance when aggregated at the catchment scale (Velpuri et al., 2013). Additionally, Wohl et al. (2012) recognized that dense vegetation and frequent cloudiness in the humid tropics are challenges for satellite monitoring of AET. Unfortunately, the low density of eddy covariance towers and lack of comparative studies for tropical climates are limiting factors to validate MODIS16 in Central America. Nonetheless, among the few existing studies available, Esquivel-Hernandez et al. (2017) compared the MODIS16 PET product against 10 Priestley-Taylor station data derived PET estimates in Costa Rica and found relatively small errors from -0.33 to +0.36 mm/year.

4.2 HYPE performance in data-scarce tropical catchments

Simulated daily streamflow showed reasonable performance using the four model configurations, where M2 (calibrated Qt monthly + Qt daily) improved low flows simulations in comparison with M1 (Fig. 9). Moreover, our configuration using a step-wise calibration for Costa Rica resulted in improved streamflow performance compared to the global model by Arheimer et al. (2020). The main shortcomings of the four configurations were underestimated larger peak flows (Fig. 8), but such errors were associated with the precipitation input rather than model capabilities. The spatial comparison of streamflow simulations indicated that catchments in the southwest performed best (KGE>0.5 and NSE>0.45, Fig. 6) compared to other areas. The

southwestern Pacific is characterized by a moderate precipitation seasonality (Fig. 1.d) with a low bias of the precipitation product (Fig. 4.a) and better performance to detect rainy and dry days (Fig. S2), compared with the tropical climate gradient of the dry to humid tropics in Costa Rica. Furthermore, we found overestimated monthly streamflows in the drier northwestern region of Costa Rica (Rancho Ray and Guardia catchments with higher mean daily and monthly simulated streamflows, as is shown in Table S2). Previous studies have noted that HYPE overestimates streamflow in dry environments (Arheimer et al., 2020). Finally, the streamflow overestimation could also be related to the precipitation bias of CHIRPSc and possibly to the nonuniform spatial distribution of our streamflow observational sample, with more "wetter" catchments used for calibration. Indeed, precipitation overestimation persisted in drier environments despite the bias correction (Fig. 4), associated with the lack of ground precipitation records to correct the CHIRPS product in headwater catchments, such as Rancho Rey. Furthermore, the low correlation of daily precipitation of CHIRPS with rain gauges resulted in low NSE values (<0.2) for daily streamflow due to the unsynchronized peak flows, mainly in the Caribbean slope.

Results from the Kruskal-Wallis test suggested median KGE in AET and PET from M3-M4 were statistically different compared to the baseline configuration M1 (the statistic H~183 and p-value<0.05 for AET, and H~72 and p-value<0.05 for PET). Nevertheless, the Kruskal-Wallis test and Mann-Whitney test indicated that median and the distribution of KGE for daily and monthly streamflow are similar among all model configurations (p-values>0.05). Hence, statistical significance using AET and PET for calibration is related to improvements in water partitioning in soil layers (Fig. 6 and Fig. 7) rather than runoff generation.

Such multi-objective calibration trade-offs were previously observed by, e.g., Zhang et al. (2018). Larger improvements were obtained for AET simulation of M3 and M4, whereas M1 (calibration only daily Qt) showed similar performance of PET in both periods (calibration and validation). The worst PET and AET simulations were observed for M2 since the monthly aggregation ignores an accurate representation of spatial water partitioning to match the monthly hydrograph (Rajib et al., 2018b). Our results also suggested that low flows were improved using PET and AET for calibration (Figure 9), where FDC exhibited an average RMSLE (Eq. (16)) value of ~0.5±0.22 compared to 1.1±0.53 from M1, constraining vertical fluxes and regulating discharge from soil layers (Massari et al., 2015; Rakovec et al., 2016). The constrained posterior parameter distributions using MODIS PET, and AET in calibration steps decreased the variability of parameters related to evapotranspiration processes (lp, cevpam). However, the soil water content parameters (wcfc, wcwp, wcep) ranges were similar among model configurations (Fig. 7). Additionally, including monthly Qt into the calibration routine also constrained parameters related to soil layer discharge (srrate, rrcs1, rrcs3) for the most frequent soil types on the monitored catchments. Such results highlight that remotely sensed PET and AET are useful to constrain some parameters and that the combination of data sources representing different modeled hydrological processes helps constrain model uncertainties, particularly for large-scale domains (Rajib et al., 2018b). Despite the partly different calibration periods used due to limited data availability, similar record lengths (8 years calibration and three years in validation) resulted in consistent results from M1 to M4.

Despite the generally reasonable performance of our model configurations, we found some issues when comparing our results with previous efforts, such as Birkel et al. (2012), who modeled the streamflow in the Sarapiqui River basin (data not used in

this study for calibration) with the HBVlight model (Seibert, 2005) for the period from 1983 to 1991 and obtained a NSE of 0.74 after a rainfall correction of the underestimated observed precipitation. Our configuration M4 resulted in a NSE=-5.6 due to rainfall underestimation (Frumau et al., 2011). In contrast, our model configuration M4 reflected improvements with respect to the global product of Arheimer et al. (2020). For example, we obtained a KGE=0.73 and NSE=0.46 for the streamflow simulation of the Palmar catchment, in comparison to the global product showing a KGE lower than 0.3. Such results reflect that more data is required to improve the streamflow response at local scales.

4.3 Independent model evaluation using hydrological signatures

A hydrological model useful for water management should be able to mimic streamflow seasonality and to realistically represent the large-scale physical processes of the water partitioned by vegetation interception and the soil matrix into evapotranspiration and discharge (Arheimer et al., 2020; Kwon et al., 2020; Pechlivanidis and Arheimer, 2015; Rajib et al., 2018b; Rakovec et al., 2016; Xiong and Zeng, 2019). We, therefore, independently evaluated the four configurations using a range of hydrological signatures (Table 4) following Westerberg and McMillan (2015) in an attempt to single out the sought-after well-balanced model for use in decision making. However, using multiple signatures also complicated the interpretation of simulations since daily streamflow (Qtd), and monthly streamflow (Qtm) indicated improvements in different configurations (see Table S1 and Table S2).

Significant spatial variations in hydrological signatures were observed between M1-M2 and M3-M4 since implementing a spatial calibration of AET improved the representativeness of the more complex large-scale climate gradient. Similar results were found in catchments from the United States (Lin et al., 2018; Rajib et al., 2018a) and worldwide (Arheimer et al., 2020). The model configurations M3 and M4 better reproduced the spatial variability between the Pacific and Caribbean basins and the north-south gradient of the AI and EI (Esquivel-Hernandez et al., 2017). Furthermore, the resulting hydrological signatures of M3 and M4 were consistent with previous small catchment scale studies that showed that runoff coefficients tend to be larger than the evaporative index (Dehaspe et al., 2018; Gómez-Delgado et al., 2011). Results also suggested that the event streamflow response is dominated by quick near-surface soil water discharge (Dehaspe et al., 2018), with streamflow being fed by groundwater during dry periods resulting in BFI values exceeding 0.7 (Birkel et al., 2012). In contrast to Westerberg et al. (2014), who calibrated Central American catchments using FDC information, we used the observed FDCs as an independent hydrological signature (Fig. 9). The configurations M2 to M4 outperformed the baseline (M1), supporting the notion that only streamflow used for calibration is not enough to produce a well-balanced model.

5 Conclusions

This study is the first attempt to apply the process-based, conceptual rainfall-runoff HYPE model at the national scale of Costa Rica (~600 simulated catchments). Due to the lack of detailed observational data available in Costa Rica, as in most parts of

715   the world's tropics, we used different global topography, soil, land use products, daily streamflow from 13 gauges, the bias-corrected global precipitation product CHIRPS (with 75 ground stations) and remotely sensed MODIS16 PET and AET products to improve the performance of HYPE in a step-wise calibration procedure towards a well-balanced model useful for water resources management. The calibrated model configurations were independently evaluated using a suite of hydrological signatures. We summarize our main findings here:

720   - Bias was observed in precipitation from CHIRPS, with underestimation in mountainous regions and overestimation in the driest region with around 1,000 mm of annual rainfall in Costa Rica.

- CHIRPS showed ~10% more days with rainfall in comparison with ground precipitation but could not capture extreme rainfall events, which ultimately impacts streamflow simulation.

- Our bias correction procedure using the linear-scaling technique reduced the annual water balance inconsistencies
725   (Prec<Q+ET) by 25%. Still, a more complex methodology is required to improve the daily precipitation depth and timing.

- The temperature could efficiently be used with an elevation correction; nevertheless, a higher resolution temperature product or downscaling approach would improve the many micro-climates across the complex topography in Costa Rica.

730   - HYPE successfully reproduced major processes (evapotranspiration, runoff, baseflow discharge) of tropical catchments in Costa Rica, where we obtained acceptable performance for daily streamflow (median KGE from 0.4 to 0.6) and good performance for monthly streamflow (median KGE from 0.6 to 0.9, and median NSE from 0.4 to 0.55) with best-fit results for PET and AET of KGE=0.6 and KGE=0.5, respectively.

- Model calibration using monthly and daily streamflow (M2) improved the performance of the low flows in
735   comparison to only daily streamflow (M1) calibration, where the average RMSLE of FDC was computed as 0.42±0.22 for M2, compared to 1.14±0.53 from M1.

- Remotely-sensed PET and AET constrained the soil type and land cover parameters associated with the evapotranspiration process.

- Statistical difference of AET and PET performance was observed for M3-M4 with respect to M1 and M2, but not for
740   daily and monthly (KGE) streamflow simulations.

- Including PET and AET in calibration (M3 and M4) slightly decreased the overall streamflow performance (average KGE of 0.47±0.17 compared to 0.54±0.09 from baseline M1) at the expense of an improved and more well-balanced median and low flow (average RMSLE for FDC of ~0.5±0.22 compared to 1.1±0.53 from baseline M1) simulation and evapotranspiration water partitioning.

745   - Simulated hydrological signatures (aridity index, evaporative index, baseflow index, streamflow coefficient, flow duration curve) differed for each calibrated model, but configurations M3 and M4 more realistically mimicked the spatial distribution of all tested hydrological signatures.

We conclude that M3 and M4 were promising model configurations to quantitively assess water resources in Costa Rica and that PET-AET and daily streamflow (M3) and PET-AET, daily and monthly streamflow (M4) represent an appropriate calibration sequence for regional modelling. Improvements to these models could be achieved by incorporating more independent data into the calibration process, such as soil moisture and groundwater level and storage data. However, all global products crucially depend on an evaluation and even correction, which needs observational in situ data. Nonetheless, we hope to have provided a way forward towards a large-sale operational hydrological model for the humid tropics of Costa Rica and potentially in other humid regions of the world.

Data availability

The hydrological simulations of model configuration 4 (M4) represent the HYPE for Costa Rica version 1.0 dataset (HYPE CR 1.0) freely available online (https://doi.org/10.5281/zenodo.4029572) and through an interactive web app visualization tool (https://zaul-ae.gitbook.io/oacg-hidrologia/).

Author contribution

All the authors contributed to the study's conception and design. The data collection was made by Saul Arciniega-Esparza and Christian Birkel. The data processing and the model construction were made by Saul and Andrés Chavarría-Palma. Results were analyzed by Saul and Christian. Berit Arheimer and Agustín Breña-Naranjo contributed to the methodology review and discussion of the model results. Christian and Agustín critically revised the work, and all the authors approved the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

Saúl Arciniega-Esparza was supported by the CONACYT graduate scholarship program and the *Programa de Maestría y Doctorado en Ingeniería* at UNAM. Christian Birkel would like to acknowledge funding from the Observatorio del Agua y Cambio Global (OACG) under UCR grant ED-3319. The authors also thank the Center for Geophysical Research (CIGEFI) and Dr. Ana Maria Duran at UCR for sharing meteorological station data under the IMN contract: IMN-DIM-CM-117-0917. We thank Alejandra González Hernández for her support in editing and translating the manuscript.

The authors thank research centers that provide free access to their databases and models, such as the Climate Hazards Center at the University of California, Santa Barbara, the Numerical Terradynamic Simulation Group at the University of Montana,

780 the National Oceanic and Atmospheric Administration, the Swedish Meteorological and Hydrological Institute, among others. The authors also thank the two anonymous reviewers that improved the content of this paper.

References

Andersson, J. C. M., Ali, A., Arheimer, B., Gustafsson, D., and Minoungou, B.: Providing peak river flow statistics and

785 forecasting in the Niger River basin. Physics and Chemistry of the Earth, 100, 3–12. https://doi.org/10.1016/j.pce.2017.02.010, 2017.

Andersson, J. C. M., Pechlivanidis, I. G., Gustafsson, D., Donnelly, C., and Arheimer, B.: Key factors for improving large-scale hydrological model performance, Eur. Water, 49(December), 77–88, 2015.

Arciniega-Esparza, S., Breña-Naranjo, J. A., and Troch, P. A.: On the connection between terrestrial and riparian vegetation:

790 The role of storage partitioning in water-limited catchments, (December), doi:10.1002/hyp.11071, 2016.

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W. H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., and Over, T.: Accelerating advances in continental domain hydrologic modeling, Water Resour. Res., 51, 10078–10091, doi:https://doi.org/10.1002/2015WR017498, 2015.

795 Arheimer, B., Hjerdt, N., and Lindström, G.: Artificially Induced Floods to Manage Forest Habitats Under Climate Change. Frontiers in Environmental Science, 6(September), 1–8. https://doi.org/10.3389/fenvs.2018.00102, 2018.

Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., and Pineda, L.: Global catchment modelling using World-Wide HYPE (WWH), open data and stepwise parameter estimation. Hydrology and Earth System Sciences Discussions, 1–34. https://doi.org/10.5194/hess-24-535-2020, 2020.

800 Bamler, R.: The SRTM mission: A world-wide 30m resolution DEM from SAR interferometry in 11 days. Photogrammetric Week, 1999.

Bayissa, Y., Tadesse, T., Demisse, G., and Shiferaw, A.: Evaluation of satellite-based rainfall estimates and application to monitor meteorological drought for the Upper Blue Nile Basin, Ethiopia, Remote Sens., doi:10.3390/rs9070669, 2017.

Beck, H. E., de Roo, A., and van Dijk, A. I. J. M.: Global Maps of Streamflow Characteristics Based on Observations from

805 Several Thousand Catchments*, J. Hydrometeorol., 16(4), 1478–1501, doi:10.1175/JHM-D-14-0155.1, 2015.

Berg, P., Donnelly, C., and Gustafsson, D.: Near-real-time adjusted reanalysis forcing data for hydrology. Hydrology and Earth System Sciences, 22(2), 989–1000. https://doi.org/10.5194/hess-22-989-2018, 2018.

Beven, K.: A manifesto for the equifinality thesis. Journal of Hydrology, 320(1–2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007, 2006.

810   Beven, K.: Rainfall-runoff modelling: The primer. 2nd edition, Wiley, Chichester, UK, 2012.

Birkel, C., Duvert, C., Correa, A., Munksgaard, N. C., Maher, D. T., and Hutley, L. B.: Tracer-Aided Modeling in the Low-Relief, Wet-Dry Tropics Suggests Water Ages and DOC Export Are Driven by Seasonal Wetlands and Deep Groundwater, Water Resour. Res., doi:10.1029/2019WR026175, 2020.

Birkel, C., Soulsby, C., and Tetzlaff, D.: Modelling the impacts of land-cover change on streamflow dynamics of a tropical

815   rainforest headwater catchment, Hydrol. Sci. J., 57(8), 1543–1561, doi:10.1080/02626667.2012.728707, 2012.

Bontemps, S., Defourny, P., Radoux, J., Van Bogaert, E., Lamarche, C., Achard, F., et al.: Consistent Global Land Cover Maps for Climate Modeling Communities: Current Achievements of the ESA's Land Cover CCI. In ESA Living Planet Symposium, 2013.

Brocca, L., Massari, C., Pellarin, T., Filippucci, P., Ciabatta, L., Camici, S., Kerr, Y. H., and Fernández-Prieto, D.: River flow

820   prediction in data scarce regions: soil moisture integrated satellite rainfall products outperform rain gauge observations in West Africa, Sci. Rep., 10(1), doi:10.1038/s41598-020-69343-x, 2020.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., et al.: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geoscientific Model Development. https://doi.org/10.5194/gmd-8-1991-2015, 2015.

Dal Molin, M., Schirmer, M., Zappa, M., and Fenicia, F.: Understanding dominant controls on streamflow spatial variability

825   to set up a semi-distributed hydrological model: The case study of the Thur catchment, Hydrol. Earth Syst. Sci., 24(3), 1319–1345, doi:10.5194/hess-24-1319-2020, 2020.

Dehaspe, J., Birkel, C., Tetzlaff, D., Sánchez-Murillo, R., Durán-Quesada, A. M., and Soulsby, C.: Spatially distributed tracer-aided modelling to explore water and isotope transport, storage and mixing in a pristine, humid tropical catchment, Hydrol. Process., 32(21), 3206–3224, doi:10.1002/hyp.13258, 2018.

830   Esquivel-Hernández, G., Sánchez-Murillo, R., Birkel, C., Good, S. P., and Boll, J.: Hydroclimatic and ecohydrological resistance/resilience conditions across tropical biomes of Costa Rica. Ecohydrology, 10(6), 1–12. https://doi.org/10.1002/eco.1860, 2017.

Frumau, K. F. A., Bruijnzeel, L. A. S., and Tobón, C.: Precipitation measurement and derivation of precipitation inclination in a windy mountainous area in northern Costa Rica. Hydrological Processes, 25(3), 499–509.

835   https://doi.org/10.1002/hyp.7860, 2011.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., et al.: The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. Scientific Data, 2, 1–21. https://doi.org/10.1038/sdata.2015.66, 2015.

Garcia, F., Folton, N. and Oudin, L.: Which objective function to calibrate rainfall–runoff models for low-flow index

840   simulations?, Hydrol. Sci. J., 62(7), 1149–1166, doi:10.1080/02626667.2017.1308511, 2017.

Genereux, D. P., and Jordan, M.: Interbasin groundwater flow and groundwater interaction with surface water in a lowland rainforest, Costa Rica: A review. Journal of Hydrology, 320(3–4), 385–399. https://doi.org/10.1016/j.jhydrol.2005.07.023, 2006.

Genereux, D. P., Wood, S. J., and Pringle, C. M.: Chemical tracing of interbasin groundwater transfer in the lowland rainforest

845 of Costa Rica. Journal of Hydrology. https://doi.org/10.1016/S0022-1694(01)00568-6, 2002.

Getirana, A., Jung, H. C., Arsenault, K., Shukla, S., Kumar, S., Peters-Lidard, C., Maigari, I., and Mamane, B.: Satellite Gravimetry Improves Seasonal Streamflow Forecast Initialization in Africa, Water Resour. Res., 56(2), doi:10.1029/2019WR026259, 2020.

Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N. and Foley, J. A.: Tropical forests were

850 the primary sources of new agricultural land in the 1980s and 1990s, Proc. Natl. Acad. Sci. U. S. A., doi:10.1073/pnas.0910275107, 2010.

Giorgi, F.: Climate change hot-spots, Geophys. Res. Lett., doi:10.1029/2006GL025734, 2006.

Gómez-Delgado, F., Roupsard, O., Le Maire, G., Taugourdeau, S., Pérez, A., Van Oijen, M., Vaast, P., Rapidel, B., Harmand, J. M., Voltz, M., Bonnefond, J. M., Imbach, P., and Moussa, R.: Modelling the hydrological behaviour of a coffee agroforestry

855 basin in Costa Rica, Hydrol. Earth Syst. Sci., 15(1), 369–392, doi:10.5194/hess-15-369-2011, 2011.

Goshime, D. W., Absi, R., and Ledésert, B.: Evaluation and Bias Correction of CHIRP Rainfall Estimate for Rainfall-Runoff Simulation over Lake Ziway Watershed, Ethiopia, Hydrology, 6(68), 1–22 [online] Available from: doi:10.3390/hydrology6030068, 2019.

Grillakis, M., Koutroulis, A., Tsanis, I., Grillakis, M., Koutroulis, A., and Tsanis, I.: Improving Seasonal Forecasts for Basin

860 Scale Hydrological Applications, Water 2018, Vol. 10, Page 1593, 10(11), 1593, doi:10.3390/W10111593, 2018.

Guimberteau, M., Drapeau, G., Ronchail, J., Sultan, B., Polcher, J., Martinez, J.-M., Prigent, C., Guyot, J.-L., Cochonneau, G., Espinoza, J. C., Filizola, N., Fraizy, P., Lavado, W., De Oliveira, E., Pombosa, R., Noriega, L., and Vauchel, P.: Discharge simulation in the sub-basins of the Amazon using ORCHIDEE forced by new datasets, Hydrol. Earth Syst. Sci., 16, 911–935, https://doi.org/10.5194/hess-16-911-2012, 2012.

865 Gurtz, J., Baltensweiler, A., and Lang, H.: Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins, Hydrol. Process., 13(17), 1999.

Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al.: SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE. https://doi.org/10.1371/journal.pone.0169748, 2017.

Her, Y., and Seong, C.: Responses of hydrological model equifinality, uncertainty, and performance to multi-objective

870 parameter calibration, J. Hydroinformatics, 20(4), 864–885, doi:10.2166/hydro.2018.108, 2018.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)-a review, Hydrol. Sci. J., 58(6), 1198–1255,

875 doi:10.1080/02626667.2013.803183, 2013.

Infante-Corona, J. A., Lakhankar, T., Pradhanang, S., and Khanbilvardi, R.: Remote sensing and ground-based weather forcing data analysis for streamflow simulation, Hydrology, 1(1), 89–111, doi:10.3390/hydrology1010089, 2014.

Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resour. Res., 45, doi:10.1029/2008WR006912, 2009.

880 Kling, H. and Gupta, H.: On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability, J. Hydrol., doi:10.1016/j.jhydrol.2009.04.031, 2009.

Kruskal, W. H. and Wallis, W. W.: Use of Ranks in One-Criterion Variance Analysis, J. Am. Stat. Assoc., 47(260), 283–621, 1952.

Kumar, R., Livneh, B., and Samaniego, L.: Toward computationally efficient large-scale hydrologic predictions with a 885 multiscale regionalization scheme, Water Resour. Res., 49(9), 5700–5714, doi:10.1002/wrcr.20431, 2013.

Kwon, M., Kwon, H. H., and Han, D.: A hybrid approach combining conceptual hydrological models, support vector machines and remote sensing data for rainfall-runoff modeling, Remote Sens., 12(11), doi:10.3390/rs12111801, 2020.

Lin, P., Rajib, M. A., Yang, Z. L., Somos-Valenzuela, M., Merwade, V., Maidment, D. R., Wang, Y., and Chen, L.: Spatiotemporal Evaluation of Simulated Evapotranspiration and Streamflow over Texas Using the WRF-Hydro-RAPID 890 Modeling Framework, J. Am. Water Resour. Assoc., 54(1), 40–54, doi:10.1111/1752-1688.12585, 2018.

Lindström, G.: Lake water levels for calibration of the S-HYPE model. Hydrology Research, 47(4), 672–682. https://doi.org/10.2166/nh.2016.019, 2016.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 model. J. Hydrol. 201: 272–288, 1997.

895 Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., and Berit, A.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrology Research, 4(41.3), 295–319. https://doi.org/10.2166/nh.2010.007, 2010.

Liu, Z., Shao, Q., and Liu, J.: The performances of MODIS-GPP and -ET products in China and their sensitivity to input data (FPAR/LAI), Remote Sens., 7(1), 135–152, doi:10.3390/rs70100135, 2015.

900 Maggioni, V., Meyers, P. C., and Robinson, M. D.: A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) era, J. Hydrometeorol., 17(4), 1101–1117, doi:10.1175/JHM-D-15-0190.1, 2016.

Maldonado, T., Alfaro, E., Fallas-López, B., and Alvarado, L.: Seasonal prediction of extreme precipitation events and frequency of rainy days over Costa Rica, Central America, using Canonical Correlation Analysis. Advances in Geosciences, 905 33, 41–52. https://doi.org/10.5194/adgeo-33-41-2013, 2013.

Mann, H. B. and Whitney, D. R.: On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat., 18, 50–60, 1947.

Massari, C., Brocca, L., Tarpanelli, A., and Moramarco, T.: Data assimilation of satellite soil moisture into rainfall-runoff modelling: A complex recipe?, Remote Sens., 7(9), 11403–11433, doi:10.3390/rs70911403, 2015.

910   Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations. Hydrology and Earth System Sciences, 15(2), 453–469. https://doi.org/10.5194/hess-15-453-2011, 2011.

Monteiro, E. S. V., Fonte, C. C., and de Lima, J. L. M. P.: Analyzing the potential of OpenStreetMap data to improve the accuracy of SRTM 30 DEM on derived basin delineation, slope, and drainage networks. Hydrology, 5(3), 1–27.

915   https://doi.org/10.3390/hydrology5030034, 2018.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sensing of Environment, 115(8), 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019, 2011.

Mu, Q., Zhao, M. and Running, S. W.: MODIS Global Terrestrial Evapotranspiration (ET) Product (MOD16A2/A3), Algorithm Theor. Basis Doc, 2013.

920   Muñoz, E., Busalacchi, A. J., Nigam, S., and Ruiz-Barradas, A.: Winter and summer structure of the Caribbean low-level jet. Journal of Climate, 21(6), 1260–1276. https://doi.org/10.1175/2007JCLI1855.1, 2008.

Neteler, M., Bowman, M. H., Landa, M., and Metz, M.: GRASS GIS: A multi-purpose open source GIS, Environ. Model. Softw., doi:10.1016/j.envsoft.2011.11.014, 2012.

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S.,

925   Lombardozzi, D., Ottle, C., Poulter, B., and Zaehle, S.: Evaluation of global terrestrial evapotranspiration by state-of-the-art approaches in remote sensing, machine learning, and land surface models, Hydrol. Earth Syst. Sci. Discuss., 1–51, doi:10.5194/hess-2019-409, 2019.

Pechlivanidis, G. I., and Arheimer, B.: Large-scale hydrological modelling by using modified PUB recommendations: The India-HYPE case, Hydrol. Earth Syst. Sci., 19(11), 4559–4579, doi:10.5194/hess-19-4559-2015, 2015.

930   Pechlivanidis, G.I., Bosshard, T., Spångmyr, H., Lindström, G., Gustafsson, D., and Arheimer, B.: Uncertainty in the Swedish Operational Hydrological Forecasting Systems. ASCE proceedings: Vulnerability, Uncertainty, and Risk: 253–262, 2014.

Pugliese, A., Persiano, S., Bagli, S., Mazzoli, P., Parajka, J., Arheimer, B., et al.: A geostatistical data-assimilation technique for enhancing macro-scale rainfall-runoff simulations. Hydrology and Earth System Sciences, 22(9), 4633–4648. https://doi.org/10.5194/hess-22-4633-2018, 2018

935   Rajib, A., Evenson, G. R., Golden, H. E., and Lane, C. R.: Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters, J. Hydrol., 567(November), 668–683, doi:10.1016/j.jhydrol.2018.10.024, 2018a.

Rajib, A., Merwade, V., and Yu, Z.: Rationale and Efficacy of Assimilating Remotely Sensed Potential Evapotranspiration for Reduced Uncertainty of Hydrologic Models, Water Resour. Res., 54(7), 4615–4637, doi:10.1029/2017WR021147, 2018b.

940   Rakovec, O., Kumar, R., Attinger, S. and Samaniego, L.: Improving the realism of hydrologic model functioning through multivariate parameter estimation, Water Resour. Res., 52, 7779–7792, doi:10.1002/ 2016WR019430, 2016.

Raphael Tshimanga, M., and Hughes, D. A.: Basin-scale performance of a semidistributed rainfall-runoff model for hydrological predictions and water resources assessment of large rivers: The Congo River, Water Resour. Res., 50, 1174–188, doi:10.1111/j.1752-1688.1969.tb04897.x, 2014.

945 Reager, J. T., Thomas, A. C., Sproles, E. A., Rodell, M., Beaudoing, H. K., Li, B., and Famiglietti, J. S.: Assimilation of GRACE terrestrial water storage observations into a land surface model for the assessment of regional flood potential, Remote Sens., 7(11), 14663–14679, doi:10.3390/rs71114663, 2015.

Rojas-Serna, C., Lebecherel, L., Perrin, C., Andréassian, V., and Oudin, L.: How should a rainfall-runoff model be parameterized in an almost ungauged catchment? A methodology tested on 609 catchments, Water Resour. Res., 52, 4765–

950 4784, doi:https://doi.org/10.1002/2015WR018549, 2016.

Sáenz, F., and Durán-Quesada, A. M.: A climatology of low level wind regimes over Central America using a weather type classification approach. Frontiers in Earth Science, 3(April), 1–18. https://doi.org/10.3389/feart.2015.00015, 2015.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15(9), 2895–2911,

955 doi:10.5194/hess-15-2895-2011, 2011.

Sheffield, J., Wood, E. F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., and Verbist, K.: Satellite Remote Sensing for Water Resources Management: Potential for Supporting Sustainable Development in Data-Poor Regions, Water Resour. Res., 54(12), 9724–9758, doi:10.1029/2017WR022437, 2018.

Seibert, J.: HBV light version 2 User's Manual. Sweden: Department of Earth Sciences, Hydrology, 2005.

960 Silvestro, F., Gabellani, S., Rudari, R., Delogu, F., Laiolo, P., and Boni, G.: Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data, Hydrol. Earth Syst. Sci., 19(4), 1727–1751.

SMHI (2018). HYPE model description, 1–113, 2015.

Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data, in Proceedings of the 1968 23rd ACM National Conference, ACM 1968., 1968.

965 Sood, A. and Smakhtin, V.: Revue des modèles hydrologiques globaux, Hydrol. Sci. J., 60(4), 549–565, doi:10.1080/02626667.2014.950580, 2015.

Tang, R., Li, Z. L., and Chen, K. S.: Validating MODIS-derived land surface evapotranspiration with in situ measurements at two AmeriFlux sites in a semiarid region, J. Geophys. Res. Atmos., 116(4), 1–14, doi:10.1029/2010JD014543, 2011.

Tanouchi, H., Olsson, J., Lindström, G., Kawamura, A., and Amaguchi, H.: Improving Urban Runoff in Multi-Basin

970 Hydrological Simulation by the HYPE Model Using EEA Urban Atlas: A Case Study in the Sege River Basin, Sweden. Hydrology, 6(1), 28. https://doi.org/10.3390/hydrology6010028, 2019.

Todini, E.: Hydrological catchment modelling: Past, present and future, Hydrol. Earth Syst. Sci., 11(1), 468–482, doi:10.5194/hess-11-468-2007, 2007.

Troch, P. A., Martinez, G. F., Pauwels, V. R. N., Durcik, M., Sivapalan, M., Harman, C., Brooks, P. D., Gupta, H. and Huxman,

975 T.: Climate and vegetation water use efficiency at catchment scales, Hydrol. Process., 23, 2409–2414, doi:10.1002/hyp, 2009.

Ullah, W., Wang, G., Ali, G., Fiifi, D., Hagan, T., Bhatti, A. S., and Lou, D.: Comparing Multiple Precipitation Products against In-Situ Observations over Different Climate Regions of Pakistan, Remote Sens., 11(6), 628, doi:https://doi.org/10.3390/rs11060628, 2019.

980 Velpuri, N. M., Senay, G. B., Singh, R. K., Bohms, S., and Verdin, J. P.: A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET, Remote Sens. Environ., 139, 35–49, doi:10.1016/j.rse.2013.07.013, 2013.

Walsh, R. P. D., and Lawler, D. M.: Rainfall seasonality spatial patterns and change through time. Weather, 36, 201–208, 1981.

Wang, L. and Liu, H.: An efficient method for identifying and filling surface depressions in digital elevation models for 985 hydrologic analysis and modelling, Int. J. Geogr. Inf. Sci., 20(2), 193–213, doi:10.1080/13658810500433453, 2006.

Waylen, P. R., Caviedes, C. N., and Quesada, M. E.: Interannual variability of monthly precipitation in Costa Rica. Journal of Climate, 1996.

Weerasinghe, I., Griensven, A. Van, Bastiaanssen, W., Mul, M., and Jia, L.: Can we trust remote sensing ET products over Africa?, Hydrol. Earth Syst. Sci. Discuss., (June), 1–27, doi:10.5194/hess-2019-233, 2019.

990 Westerberg, I. K. and Birkel, C.: Observational uncertainties in hypothesis testing: Investigating the hydrological functioning of a tropical catchment, Hydrol. Process., 29(23), 4863–4879, doi:10.1002/hyp.10533, 2015.

Westerberg, I. K., and McMillan, H. K.: Uncertainty in hydrological signatures, Hydrol. Earth Syst. Sci., 19(9), 3951–3968, doi:10.5194/hess-19-3951-2015, 2015.

Westerberg, I. K., Gong, L., Beven, K. J., Seibert, J., Semedo, A., Xu, C. Y., and Halldin, S.: Regional water balance modelling 995 using flow-duration curves with observational uncertainties, Hydrol. Earth Syst. Sci., 18(8), 2993–3013, doi:10.5194/hess-18-2993-2014, 2014.

Wohl, E., Barros, A., Brunsell, N., Chappell, N. A., Coe, M., Giambelluca, T., Goldsmith, S., Harmon, R., Hendrickx, J. M. H., Juvik, J., McDonnell, J. and Ogden, F.: The hydrology of the humid tropics, Nat. Clim. Chang., 2(9), 655–662, doi:10.1038/nclimate1556, 2012.

1000 Wörner, V., Kreye, P. and Meon, G.: Effects of bias-correcting climate model data on the projection of future changes in high flows, Hydrology, 6(2), doi:10.3390/hydrology6020046, 2019.

Xiong, L. and Zeng, L.: Impacts of introducing remote sensing soil moisture in calibrating a distributed hydrological model for streamflow simulation, Water (Switzerland), 11(4), doi:10.3390/w11040666, 2019.

Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., and Ribbe, L.: Temporal and spatial evaluation of satellite-based 1005 rainfall estimates across the complex topographical and climatic gradients of Chile, Hydrol. Earth Syst. Sci., 21(2), 1295–1320, doi:10.5194/hess-21-1295-2017, 2017.

Ziegler, A. D., Giambelluca, T. W., Plondke, D., Leisz, S., Tran, L. T., Fox, J., Nullet, M. A., Vogler, J. B., Minh Troung, D., and Tran Duc Vien.: Hydrological consequences of landscape fragmentation in mountainous northern Vietnam: Buffering of Hortonian overland flow, J. Hydrol., doi:10.1016/j.jhydrol.2007.01.031, 2007.

1010 Zhang, R., Liu, J., Gao, H., and Mao, G.: Can multi-objective calibration of streamflow guarantee better hydrological model accuracy?, J. Hydroinformatics, 20(3), 687–698, doi:10.2166/hydro.2018.131, 2018.