Accounting for the influence of vegetation and landscape improves model transferability in a tropical savannah region

Gao, Hongkai; Hrachowitz, Markus; Sriwongsitanon, Nutchanart; Fenicia, Fabrizio; Gharari, Shervan; Savenije, Huub

DOI
10.1002/2016WR019574

Publication date
2016

Document Version
Final published version

Published in
Water Resources Research

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2016WR019574

Key Points:

- Landscape and vegetation heterogeneity control hydrological function
- Accounting for these heterogeneities improves model transferability
- Method improves prediction in ungauged basins

Supporting Information:

- Supporting Information S1

Correspondence to:
H. Gao,
h.gao-1@tudelft.nl or
hongkai.gao@asu.edu

Citation:
Gao, H., M. Hrachowitz, N. Sriwongsitanon, F. Fenicia, S. Gharari, and H. H. G. Savenije (2016), Accounting for the influence of vegetation and landscape improves model transferability in a tropical savannah region, *Water Resour. Res.*, *52*, doi:10.1002/2016WR019574.

Received 27 JUL 2016
Accepted 23 SEP 2016
Accepted article online 27 SEP 2016

Accounting for the influence of vegetation and landscape improves model transferability in a tropical savannah region

Hongkai Gao[1], Markus Hrachowitz[1], Nutchanart Sriwongsitanon[2], Fabrizio Fenicia[3], Shervan Gharari[1,4,5], and Hubert H. G. Savenije[1]

[1]Water Resources Section, Delft University of Technology, Delft, Netherlands, [2]Department of Water Resources Engineering, Kasetsart University, Bangkok, Thailand, [3]Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, [4]Global Institute for Water Security, School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Canada, [5]Luxembourg Institute of Science and Technology, Belvaux, Luxembourg

Abstract Understanding which catchment characteristics dominate hydrologic response and how to take them into account remains a challenge in hydrological modeling, particularly in ungauged basins. This is even more so in nontemperate and nonhumid catchments, where—due to the combination of seasonality and the occurrence of dry spells—threshold processes are more prominent in rainfall runoff behavior. An example is the tropical savannah, the second largest climatic zone, characterized by pronounced dry and wet seasons and high evaporative demand. In this study, we investigated the importance of landscape variability on the spatial variability of stream flow in tropical savannah basins. We applied a stepwise modeling approach to 23 subcatchments of the Upper Ping River in Thailand, where gradually more information on landscape was incorporated. The benchmark is represented by a classical lumped model (FLEX^{L}), which does not account for spatial variability. We then tested the effect of accounting for vegetation information within the lumped model (FLEX^{LM}), and subsequently two semidistributed models: one accounting for the spatial variability of topography-based landscape features alone (FLEX^{T}), and another accounting for both topographic features and vegetation (FLEX^{TM}). In cross validation, each model was calibrated on one catchment, and then transferred with its fitted parameters to the remaining catchments. We found that when transferring model parameters in space, the semidistributed models accounting for vegetation and topographic heterogeneity clearly outperformed the lumped model. This suggests that landscape controls a considerable part of the hydrological function and explicit consideration of its heterogeneity can be highly beneficial for prediction in ungauged basins in tropical savannah.

1. Introduction

Tropical savannah is the second most common climate type by land area [*Peel et al*., 2007]. Whereas in temperate, humid, or tropical rain forest regions, as illustrated by a large body of literature [*Parajka et al*., 2013; *Farrick and Branfireun*, 2014], the hydrology of tropical savannah regions is understood to lesser extent [*Hrachowitz et al*., 2011; *Wohl et al*., 2012]. Partly, this is the result of data scarcity, but also of special characteristics of savannahs. Savannah regions are characterized by considerable intra-annual hydrological variability with pronounced dry and wet seasons [*Peel et al*., 2007; *Sriwongsitanon and Taesombat*, 2011] as well as by comparably elevated aridity indices ($I_A = E_P/P$) of $1 < I_A < 2$, due to high evaporative demand. Being water-limited, evaporative processes, and the temporal dynamics in the partitioning between evaporative fluxes and runoff throughout, the year are generally more controlled by vegetation characteristics than in energy-limited, temperate regions [*Miyazawa et al*., 2014]. This is highlighted by the typically large water storage capacities that are accessible to roots, which buffer the seasonal variations in water availability, provide plants with continuous access to water and generally increase the nonlinearity in storage-discharge relationships [*Montanari et al*., 2006; *Gao et al*., 2014a]. Due to the large temporal variability of wetness states in the different compartments of its hydrological system, runoff processes in savannah catchments are strongly dominated by threshold processes; more so than in temperate and wet climates that experience less moisture variability.

With some exceptions [e.g., *Pitman*, 1973; *Petheram et al*., 2012; *Caballero et al*., 2013], the vast majority of hydrological models have been developed for use in temperate and humid regions, which frequently struggle to meaningfully accommodate the strong threshold behavior [e.g., *Perrin et al*., 2007]. In spite of efforts toward improved model regionalization techniques [e.g., *Hughes*, 2006; *Kapangaziwiri et al*., 2012], model performance in arid regions is generally worse than in other regions [*Parajka et al*., 2013].

Spatial transferability of model structures and parameters (hereafter referred to as model transferability) is an important validation test for hydrological models. The ability to regionalize or transfer models was an important objective and challenge of the IAHS decade on Predictions in Ungauged Basins (PUB) [*Sivapalan et al*., 2003; *Blöschl et al*., 2013; *Hrachowitz et al*., 2013] as it is tightly linked to the issues of model scaling [*Sivapalan and Kalma*, 1995; *Blöschl*, 2001] and consistency [*Martinez and Gupta*, 2011; *Euser et al*., 2013, 2015; *Hrachowitz et al*., 2014*; Fovet et al*., 2015].

For model applications, it is in general common that parameters have to be recalibrated or models even have to be redesigned to describe hydrological processes in different catchments even if they are spatially close to each other. This practice is symptomatic of a limited understanding of how catchment characteristics relate to model parameters and structure. Moreover, it is not applicable in ungauged basins, where time series for model calibration are not available.

In order to provide confidence of a generalizable understanding of hydrological processes, and of the ability to predict beyond the range of observed data, various model performance tests have been proposed [e.g., *Andréassian et al*., 2009]. Most often, models are evaluated with respect to temporal transferability, using split-sample or differential split-sample tests [*Donnelly-Makowecki and Moore* 1999; *Hartmann and Bárdossy*, 2005; *Refsgaard et al*., 2014]. Such tests, however, do not guarantee that a model is transferable to other regions. A more stringent test is spatial model transferability between proxy catchments with similar hydrological function [*Klemeš*, 1986]. This test increases the confidence that the hydrological processes at play are correctly represented, and that the model can be used outside the range of calibration [e.g., *Blöschl et al*., 2013; *Gupta et al*., 2014, *Fenicia et al*., 2016].

To achieve model spatial transferability, one among a wide range of strategies is to assume that catchments with similar climatic (e.g., rainfall, potential evaporation) and physical properties (e.g., landscape, vegetation, soils, geology, area, etc.) have similar response behavior, and to classify entire catchments based on these characteristics [*Bárdossy*, 2007]. The main limitation of this approach is that, given the high variability of climatic and physical properties between catchments, it is difficult to identify groups of catchments with similar characteristics [*Andréassian et al*., 2009]. Given the scarcity of data in many regions of the world this, in turn, puts serious constraints on the application of classification schemes based on this approach.

Another common approach for model transferability is parameter regionalization [e.g., *Hundecha and Bárdossy*, 2004; *McIntyre et al*., 2005; *Parajka et al*., 2005*; Laaha and Blöschl*, 2006; *Bárdossy*, 2007*; Blöschl et al*., 2013; *Viglione et al*., 2013]. For example, *Merz and Blöschl* [2004] looked for spatially regionalized patterns of parameters in the HBV model for over 300 catchments in Austria. When applied using lumped models at the scale of entire catchments, this approach did not result into clear relationships between model parameters and catchment properties [e.g., *Merz and Blöschl*, 2004].

In order to improve the link between model parameters at the large-scale and observable characteristics at the small-scale, *Samaniego et al*. [2010] proposed a multiscale parameterization regionalization (MPR) method, where model parameters are linked to land surface characteristics at the finest spatial resolution available by transfer functions, and then upscaled to the model grid size. This approach was successfully applied in many regions [e.g., *Kumar et al*., 2013].

An alternative approach for model transferability is based on the concept of hydrological response units (HRUs). This approach, instead of defining hydrological similarity at the scale of entire catchments, considers smaller areas within a catchment which are considered hydrologically similar [*Flügel*, 1996]. Previous studies applied this concept to develop models such as the Soil and Water Assessment Tool (SWAT), where HRUs are mainly based on soil type and landuse [*Arnold et al*., 1995], or Dynamic TOPMODEL [*Beven and Freer*, 2001], which is mainly based on topographic information.

The lack of suitable data for detailed HRU definition, together with equifinality problems due to increased model complexity, frequently hinders the application of such modeling strategies in practice. It is therefore critical to systematically analyze the individual and combined effects of different landscape characteristics such as of topography, topology, soils, vegetation, and geology, to develop an understanding on how much and which information is necessary to meaningfully define HRUs and which of these factors are first-order controls on model transferability.

The coevolution of topography, vegetation, soil texture, climate, and geology, suggests that these landscape factors are correlated [*Sivapalan and Blöschl*, 2015]. We may therefore not need to consider all these aspects in our models [*Savenije*, 2010; *Troch et al*., 2013], as information of one of these aspects could in principle be derived from the others. While this is in detail problematic for individual processes such as soil formation, that act on very long-time scales and that may never reach equilibrium, there is growing evidence that vegetation and its influence on the hydrological system can adjust relatively quickly, i.e., at time scales of a few years, to changes and disturbances [e.g., *Troch et al*., 2009]. It is hypothesized that this information can be used, at the time scale of interest for many hydrological applications (i.e., from subdaily up to decadal) to distinguish between functionally different landscape classes characterized by different dominant hydrological processes to increase the representation of hydrological process heterogeneity in a semi-distributed yet parsimonious way. This allows us to keep models as simple as possible, data requirements, and parameters equifinality low. Based on this assumption, *Savenije* [2010] proposed a topography-driven modelling approach (FLEX-Topo), in which only topography and land cover (vegetation) characteristics are considered as landscape features for deriving HRUs.

Gharari et al. [2014] found that, if adequately constrained by expert knowledge, the relatively complex FLEX-Topo can be quite robust compared to a standard lumped model even without calibration. Similarly, a recent study by *Hrachowitz et al*. [2014] highlighted the value of increased process complexity introduced by HRUs to adequately reproduce a wide range of hydrological signatures and thus the system integrated response characteristics of a catchment. *Gao et al*. [2014a] showed for mountainous catchments in cold, arid regions that FLEX-Topo, which explicitly accounts for topography and land cover information, can perform substantially better than lumped models without this information.

Hillslopes and riparian zones have different runoff generation mechanisms, which are revealed by field hydrometric, soil moisture, isotopic, and solute measurements [*McGlynn and McDonnell*, 2003; *Detty and McGuire*, 2010]. Vegetation dynamics in space and time influence interception, infiltration, transpiration, percolation, and even groundwater dynamic [*Rodriguez-Iturbe*, 2000; *Cleverly et al*., 2006; *Yu, et al*., 2010]. Following the evidence that topography and vegetation are, besides other factors, such as geology [e.g., *Fenicia et al*., 2014, 2016], first-order controls on the hydrological behavior of catchments, the objective of this study is to evaluate the importance of topographic and vegetation-induced heterogeneity for the hydrological function of catchments and to test to which degree the explicit utilization of readily available topographic and vegetation information can improve model transferability in a region characterized by a tropical savannah climate.

2. Study Site and Data

2.1. Study Site Introduction

The Ping River is one of the main tributaries of the Chao Phraya, which drains more than one-third of Thailand and is the country's largest river basin [*Sriwongsitanon and Taesombat*, 2011; *Visessri*, 2014; *Visessri and McIntyre*, 2015]. The study sites are 23 catchments of the Upper Ping River basin (UPRB; Figure 1), with areas ranging from 128 to 14814 km^2 and complex nested relations (Figure 1b). Most of these catchments are dominated by forest (80% in 2005, Figure 1h) [*Sriwongsitanon and Taesombat*, 2011; *Visessri*, 2014; *Visessri and McIntyre*, 2015], and the landscape is characterized by steep hillslopes intersected by wetlands. The average annual rainfall between 1988 and 2005 was around 1200 mm/a, and runoff was around 270 mm/a [*Taesombat and Sriwongsitanon*, 2009]. The climate of this region is tropical savannah (Aw in Köppen–Geiger climate classification), characterized by South Asian Monsoon, with hot wet summers and hot dry winters. Red-yellow podzolic soils, is the dominant soil type (Figure 1i), which overlays a complex geology, dominated by quartzite, phyllite, schist, sandstone, shale, tuff, and alluvial deposits.

Figure 1g shows a landscape classification in the three classes wetland, terrace, and hillslope (the classification approach is detailed in section 3.3.1). Of the 23 study catchments, in particular the catchments P.5, P.73, P.76, and P.85 are relatively flat and have higher proportions of wetland and terrace landscapes than other catchments (Figures 1c, 1d, and 1g). The dominant vegetation of these catchments is deciduous forest, shrubs, and agriculture, with the lowest long-term catchment average dry season NDVI of the study catchments (Table 1 and Figure 1e). The remaining catchments are characterized by steep hillslopes, denser

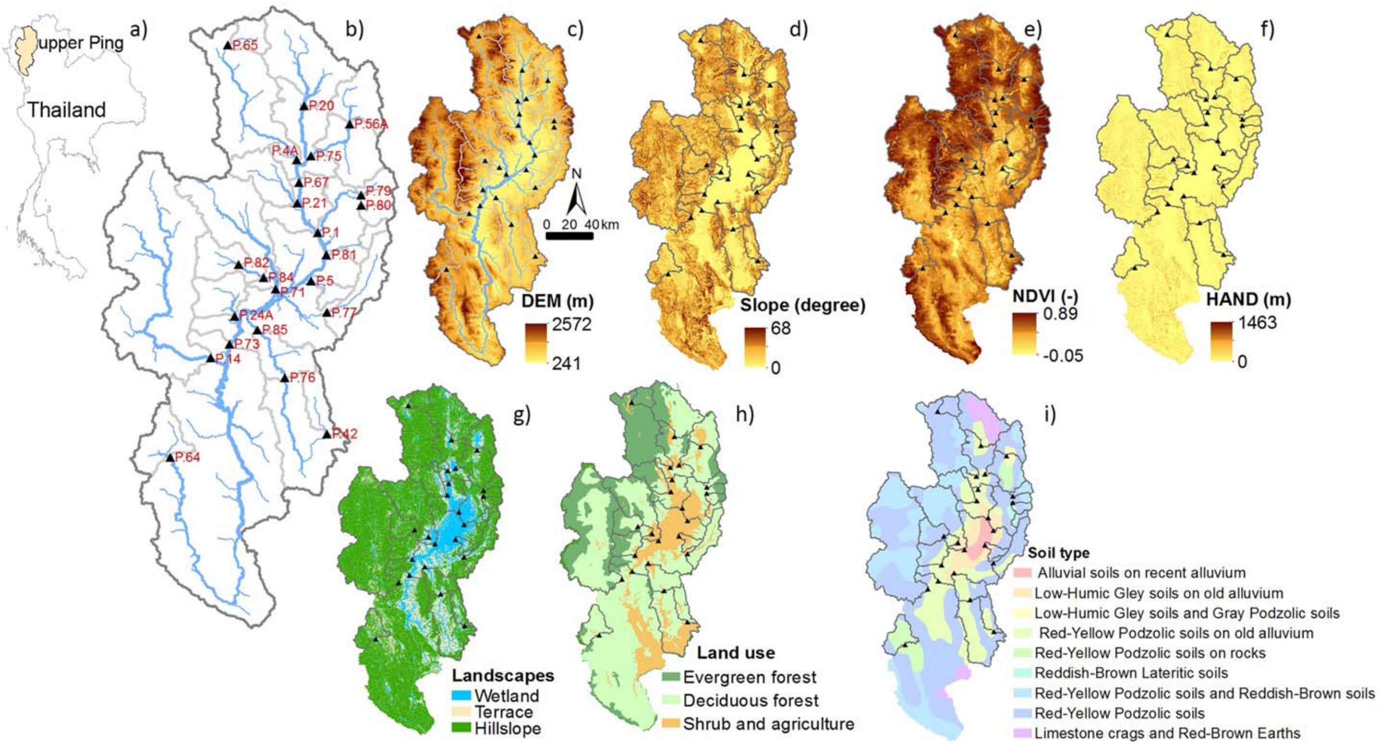

Figure 1. (a) Location of the Upper Ping River basin (UPRB) in Thailand; (b) 23 study catchments and the location of their runoff gauging stations; (c) DEM of the UPRB; (d) slope map of the UPRB; (e) dry seasonal NDVI (Normalized Difference Vegetation Index) map of the UPRB; (f) HAND (Height Above the Nearest Drainage) map of the UPRB; (g) landscapes classification map of the UPRB based on HAND and slope; (h) land use map of the UPRB; and (i) the soil types of the UPRB.

vegetation, and higher dry season NDVI. A detailed summary of catchment characteristics is given in Figure 1 and Table 1 where the marked differences in landscape characteristics between the study catchments can be clearly seen. For example, consider the lowland catchments P.5 and P.73. More than 40% of their areas are covered by wetlands (p_W) and only 13 and 35% of the respective catchment areas are classified as hillslopes (p_H). In contrast, P.14 and P.79 are more upland type of areas with wetland proportions $p_W < 5\%$ and hillslope proportions $p_H > 80\%$. In addition, considerable vegetation variability between the study catchments is indicated by the catchment average NDVI values that range between $\sim$0.55 (P.5) and 0.80 (P.79).

2.2. Data Set

Daily rainfall, runoff, and temperature data were available from the Thailand Meteorological Department and Royal Irrigation Department. The daily areal rainfall distribution across the UPRB was generated using 68 stations in and around the UPRB by thin plate spline extrapolation [*Taesombat and Sriwongsitanon*, 2009]. The daily potential evaporation (Table 1) was calculated using the Hargreaves equation [*Hargreaves*, 1975], with daily maximum and minimum temperature as forcing data from three close-by stations: Chiangmai (18°47′N, 98°59′E), Lamphun (18°34′N, 99°02′E), Maejo (18°55′N, 99°00′E). Note, that a range of studies suggests that at the catchment scale simple temperature and radiation-based evaporation models, such as the Hargreaves model, can be considered sufficiently accurate [e.g., *Oudin et al.*, 2005; *Kleidon et al.*, 2014]. This is because both typically exhibit strong correlations with other components of the energy balance and thus serve as an adequate integrated indicator for the catchment-scale energy budget [*Ambach*, 1988; *Allen et al.*, 1998; *Hock*, 2003].

The Digital Elevation Model (DEM) used in this study was the Shuttle Radar Topography Mission (SRTM) product with a spatial resolution of 90 m. Values of the Normalized Difference Vegetation Index (NDVI) were obtained from the MOD13Q1 product with 250 m spatial and 16 days temporal resolution. Both the DEM and NDVI (2000-2011) data were downloaded from https://lpdaac.usgs.gov/data_access/usgs_earthexplorer.

Since there is only a limited number of meteorological stations available and due to the marked elevation difference within the UPRB, temperatures for the estimation of the potential evaporation were elevation adjusted using the environmental lapse rate of 0.006°Cm^{-1}.

Table 1. Summary of Catchment Characteristics[a]

Code	Name	Area (km^2)	Average Elevation (m)	Calibration Period	Validation Period	Data Quality Score	p_W (%)	p_T (%)	p_H (%)	Runoff Coefficient	P (mm/a)	E_p (mm/a)	NDVI
P.1	Nawarat Bridge	6307	799	1999–2005	2006–2012	32	12.5	17.5	70	0.174	1247	1561	0.716
P.4A	Ban Mae Taeng	1902	1026	1985–1995	1996–2005	39	5.6	11.5	83.0	0.153	1274	1547	0.773
P.5	Sing Phithak Bridge	151	540	2005–2008	2009–2012	30	40.7	24.3	35.0	0.292	914	1672	0.547
P.14	Ban Kaeng Ob Luang	3853	990	1985–1995	1996–2005	41	4.2	12.2	83.6	0.271	1253	1556	0.733
P.20	Chiang Dao	1355	777	1985–1995	1996–2005	42	9.7	17.4	72.9	0.226	1223	1606	0.732
P.21	Ban Rim Tai	515	724	1985–1995	1996–2005	41	10.5	23.0	66.5	0.195	1273	1619	0.759
P.24	Ban Sop Tia	452	937	1985-1995	1996-2005	39	6.5	15.9	77.6	0.358	978	1568	0.748
P.42	Ban Mae Bon Mai	315	669	1985–1993	1994–2001	41	12.15	27.3	60.6	0.143	950	1644	0.682
P.56	Ban SahaKhon Rom Klao	529	448	1999–2005	2006–2011	29	15.5	17.8	66.8	0.263	1282	1698	0.707
P.64	Ban Luang	495	979	1999–2005	2006–2011	29	7.3	26.0	66.7	0.432	1124	1559	0.693
P.65	Ban Muang Pog	240	1121	1993–1997	1998–2001	28	5.8	11.4	82.9	0.354	1284	1525	0.729
P.67	Ban Mae Tae	5236	1057	1999–2005	2006–2011	29	10.1	16.4	73.5	0.180	1580	1533	0.721
P.71	Ban Klang	1798	837	1996–2000	2001–2005	36	10.1	15.5	74.4	0.156	997	1582	0.744
P.73	Ban Sop Soi	14814	767	1999–2005	2006–2012	31	41.3	45.7	13.0	0.208	1181	1610	0.678
P.75	Ban Cho Lae	3090	1097	1999–2005	2006–2011	31	11.6	17.6	70.8	0.177	1345	1528	0.716
P.76	Ban Mae E-Hai	1541	582	2000–2002	2003–2005	29	16.0	41.6	42.4	0.132	980	1655	0.671
P.77	Ban Sop Mae Sapuad	550	637	1999–2002	2003–2005	32	9.9	19.6	70.6	0.137	1110	1648	0.732
P.79	Ban Mae Wan	134	820	2001–2005	2006–2011	30	2.6	7.0	90.4	0.464	1300	1598	0.802
P.80	Ban Pong Din	128	1146	2001–2005	2006–2011	31	7.3	20.1	72.6	0.329	1212	1493	0.765
P.81	Ban Pong	1134	456	2002–2006	2007–2011	29	14.9	17.4	67.7	0.218	1136	1690	0.711
P.82	Ban Mae Win	388	930	2003–2007	2008–2011	28	5.6	14.2	80.2	0.411	1205	1551	0.756
P.84	Ban Mae Chaem	482	630	2003–2007	2008–2011	28	7.2	16.6	76.2	0.202	1197	1649	0.736
P.85	Ban Lai Khaeo	1996	308	2003–2007	2008–2011	29	14.8	35.4	49.8	0.105	1132	1703	0.607

[a]The higher the data quality score, the better the data quality (46 represents the best possible data quality in this study). p_W, p_T, and p_H represent areal proportions of wetlands, terraces and hillslopes, respectively. Runoff coefficients were computed as the long-term runoff coefficients for the observation period. P and E_p represent mean annual precipitation and mean annual potential evaporation. NDVI was calculated as the long-term average NDVI in dry seasons.

Based on the approach proposed by *Visessri and McIntyre* [2015], we calculated the data quality scores of the 23 study catchments (Table 1), depending on the length of the flow record (years), the flow records overlapping with rainfall record (years), the number of possible outliers in the flow records, the frequency of revision of rating curve, missing rainfall data from rain gauges located within the subcatchment, the size of catchments, and the elevations of catchments. The maximum possible score of 46 according to *Visessri and McIntyre* [2015] indicates the best possible data quality score given the available data.

3. Methodology and Model Setups

To assess the importance of a more complete representation of the topography and vegetation-induced process heterogeneity for the spatial transferability of models, four model setups with increasing complexity were tested in this study. The models are a combination of interconnected reservoirs and transfer functions, as conceptualized in the SUPERFLEX modeling framework [*Fenicia et al.*, 2011]. These four models are detailed below.

To evaluate the models' potential for spatial transferability and to assess the relative importance of accounting for topography and/or vegetation-induced process heterogeneity for improving model transferability, a "leave-p-out-cross-validation strategy" was chosen [*Shao*, 1993]: from the study catchments, one was chosen in turn as donor catchment to calibrate the four models, which were then transferred and tested in the other receiver catchments. This procedure was repeated so that each catchment served as donor catchment once. In the end, a rank sum test was conducted to investigate whether FLEX^{LM}, FLEX^{T}, and FLEX^{TM} significantly improved transferability compared with FLEX^{L}.

3.1. FLEX^{L}

FLEX^{L} is a lumped conceptual hydrological model (Figure 2), which consists of four reservoirs: the interception reservoir S_i (mm), the unsaturated reservoir S_u (mm), the fast response reservoir S_f (mm), the slow response reservoir S_s (mm), and two lag functions representing the lag time from storm to peak flow (T_{lagF}), and the lag time of recharge from the root zone to the groundwater (T_{lagS}). In total, there are 11 free calibration parameters, including S_{uMax}, in FLEX^{L}. The relevant model equations are given in supporting information Table S1 and the prior parameter distributions in Table 2.

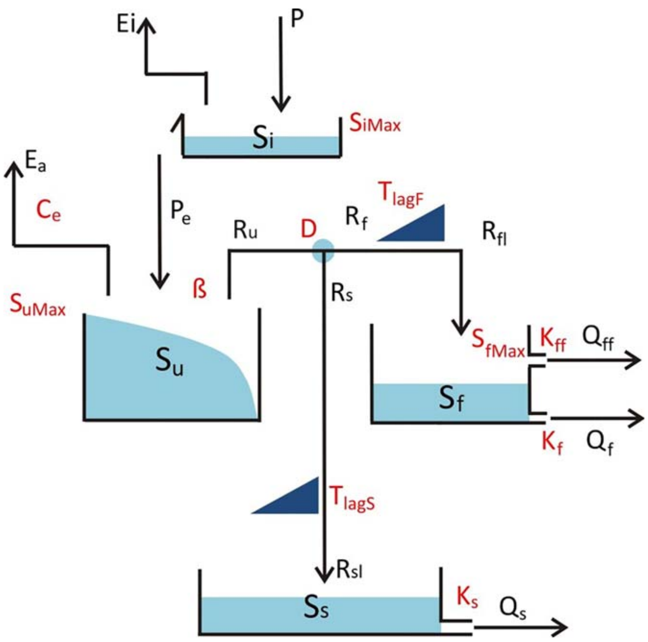

Figure 2. Model structure of FLEX^{L}.

3.2. FLEX^{LM}

FLEX^{LM} has the same structure as FLEX^{L}, (it is therefore also lumped) and differs from it in the number of calibration parameters. In FLEX^{LM} S_{uMax} is not estimated by calibration, but it is directly determined from observed climate and vegetation data using the Mass Curve Technique (MCT) [*Gao et al.*, 2014b]. It therefore has 10 free calibration parameters.

3.2.1. The MCT Technique

The rationale behind the MCT technique is that the size of the root zone reservoir (i.e., S_{uMax}) is "designed" by the ecosystem so that it allows for sufficient water storage to overcome dry spells with a certain return period. In analogy to engineers designing drinking water reservoirs, the MCT allows to estimate the required root zone storage capacity S_{uMax} of a system based on cumulative input and cumulative demand. The highest deficit between input and demand over a specified time period will then be an estimate of the required storage capacity. The system input is the effective precipitation (the difference between precipitation and interception evaporation), i.e., $P_E = P - E_I$, where P is precipitation, assuming an average interception rate of $E_I = 2$ mm d^{-1} during rainy days. The long-term average system water demand, i.e., transpiration, has been estimated as $E_{Ta} = P_E$-Q, where Q is the observed runoff from the system. These estimates were subsequently translated into long-term average dry season transpiration E_{Td} based on the long-term catchment-average NDVI ratio between wet and dry seasons. The storage requirement was then estimated as the maximum deficit between cumulative P_E and cumulative E_{Td} over dry periods. These values were ranked so as to obtain storage requirements for droughts with different return periods. In line with *Gao et al*. [2014b], the root zone storage capacity S_{uMax} was selected based on a drought of once in 20 years. For further details of the MCT approach, the reader is referred to *Gao et al*. [2014b].

3.3. FLEX^{T}

The model structure of FLEX^{T} is shown in Figure 3 and consists of three parallel model components, which represent functionally distinct landscape units, classified according to their topographic characteristics. The main difference between these components is the architecture and parameterization of the unsaturated root zone reservoirs S_u, which is distributed in FLEX^{T} (Figure 3 and supporting information Table S2). There are 13 free calibration parameters in FLEX^{T}. All equations are listed in supporting information Table S2, and prior parameter distributions are shown in Table 2.

Table 2. Uniform Prior Parameter Distributions of the FLEX^{L} and FLEX^{T}

FLEX^{L}				FLEX^{T}			
Parameter	Range	Parameters	Range	Parameter	Range	Parameters	Range
S_{iMax} (mm)	(0.1, 6)	K_{ff} (d)	(1, 9)	S_{iMax} (mm)	(0.1, 6)	D	(0, 1)
S_{uMax} (mm)	(10, 1000)	T_{lagF} (d)	(0, 5)	S_{uMaxH} (mm)	(10, 1100)	K_{fH} (d)	(1, 20)
β (-)	(0, 3)	T_{lagS} (d)	(0, 100)	β (-)	(0, 2)	T_{lagF} (d)	(0, 5)
C_e (-)	(0.1, 1)	K_f (d)	(1, 40)	C_e (-)	(0.1, 1)	T_{lagS} (d)	(1, 90)
D (-)	(0, 1)	K_s (d)	(10, 500)	P_{Max} (mm/d)	(0, 3)	K_f (d)	(1, 40)
S_{fMax} (mm)	(10, 500)			S_{uMaxT} (mm)	(10, 1000)	K_s (d)	(10, 400)
				S_{uMaxW} (mm)	(10, 1000)		

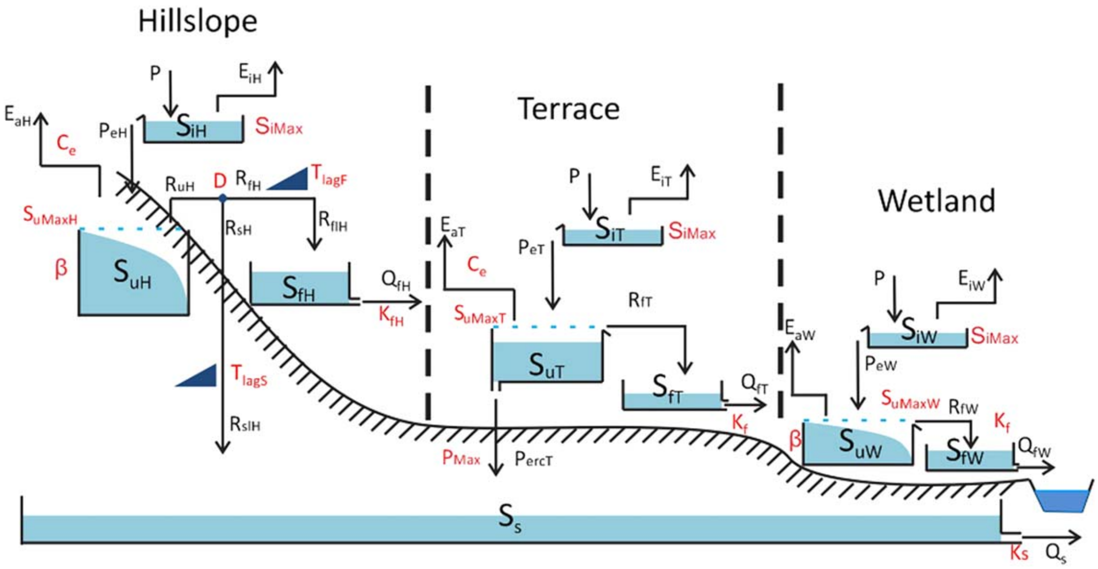

Figure 3. Model structure of FLEX^T.

3.3.1. Landscape Classification

Based on the recently formalized metric Height Above the Nearest Drainage (HAND) [*Rennó et al*., 2008], topographic data were used to subdivide the study catchments into landscape classes with different hydrological function [*Savenije*, 2010; *Nobre et al*., 2011]. These classes were then associated with individual models characterized by different architectures and different parameter values, operating in parallel and linked by a groundwater reservoir. The combined inflow generated from the individual landscape classes into the common groundwater reservoir was then computed as the area-weighted average of the outflows from the individual landscape classes. This strategy was found to be valuable for providing more robust representations of the observed system dynamics in a range of previous studies in contrasting environments [*Gao et al*., 2014a; *Gharari et al*., 2014; *Hrachowitz et al*., 2014]. From preliminary on-site terrain analysis, three dominant landscape classes or hydrological response units could be identified for the study catchments: wetlands, terraces, and hillslopes. Following the suggestions of *Rennó et al*. [2008] and *Gharari et al*. [2011], the combined use of HAND and local slope allowed the definition of the three HRUs. Locations with HAND $<$ 5m were treated as wetlands, locations with HAND $>$ 5 m, and slope $<$ 0.1 were classified as terraces and locations with HAND $>$ 5m and slope $>$ 0.1 were regarded as hillslopes. The threshold values were selected based on local expert knowledge and experience in other catchments [*Rennó et al*., 2008; *Gharari et al*., 2011]. The different landscapes proportions of the 23 catchments are shown in Figure 1g and Table 1.

3.3.2. Model Structure Rationale

To fulfil the contrasting functions of water retention and drainage, forested hillslopes are often characterized by larger root zone storage capacities (S_{uMaxH}) than wetlands or grass hillslopes, due to the need for forests to buffers for dry periods [cf., *Savenije*, 2010]. To provide drainage for hillslope vegetation, water in excess of the storage capacity is split into one part that is routed through a fast reservoir to the channel by subsurface storm flow (R_{fH}) and another part that recharges the groundwater reservoir (R_{sH}).

In contrast to the importance of lateral drainage of water on hillslopes, the main direction of water movement on terraces can be expected to be more vertical, due to its flatter topography. Thus, most infiltrating water in terraces is either stored or recharges (P_{ercT}) the groundwater reservoir. Depending on soil and bedrock characteristics, generation of lateral flow from terraces (R_{fT}) is likely to require larger storm events or more extended wet periods than generation from hillslopes.

On wetlands, the root zone storage capacity (S_{uMaxW}) is relatively low due to the shallow groundwater table. As a consequence, wetlands are characterized by higher runoff coefficients than hillslopes, as excess water will be directly and rapidly routed to the stream as soon as the relatively small storage capacity is exceeded, which is in strong agreement with findings from many experimental studies [e.g., *Freer et al*., 2004; *Detty and McGuire*, 2010]. In contrast to the other landscape units, wetlands are frequently close to saturation, and transpiration is therefore rather energy than water limited. The main functions of wetlands are thus lateral drainage and transpiration.

The three landscape units described are connected by a common groundwater reservoir, recharged by hillslopes (R_{slH}) and terraces (P_{ercT}), and with an upward flux which sustains the evaporative demand of the wetland vegetation.

Transpiration rates are estimated through different methods. On hillslopes and terraces, actual transpiration was computed based on soil moisture and potential evaporation, as in the lumped model FLEX^{L}. On wetlands, due to the sufficient water supply, actual transpiration is assumed to occur at potential rates after canopy interception [*Mohamed et al*., 2012].

3.4. FLEX^{TM}

To test the integrated influence of topographic and vegetation information on model performance and model transferability, we linked some parameters of FLEX^{T} to vegetation information. FLEX^{TM} is the model that incorporates these linkages. The number of calibration parameters reduces from 13 (FLEX^{T}) to 11 (FLEX^{TM}).

In order to account to some degree for differential transpiration dynamics, and thus potentially distinct root zone storage capacities, the catchment integrated estimates of S_{uMax} were adjusted for the individual landscape classes (e.g., wetlands, terraces, and hillslopes) in FLEX^{TM}. There is evidence that suggests a functional relationship between NDVI and transpiration rates [e.g., *Boegh et al*., 1999]. Postulating a direct, and in the absence of any further information, linear dependence between transpiration rates and S_{uMax} (i.e., plant available water) in water limited environments and under stable climatic conditions, which is plausible given the results of recent studies [e.g., *Gentine et al*., 2012; *Gao et al*., 2014b], then allowed the respective root zone storage capacities on hillslopes (S_{uMaxH}) and on terraces (S_{uMaxT}) to be estimated from estimates of catchment integrated S_{uMax} as obtained from the MCT (see FLEX^{LM}) and the long-term average dry season NDVI in the two different landscapes ($I_{NDVI,H}$ on hillslopes and $I_{NDVI,T}$ on terraces). In contrast, the shallow groundwater levels in wetlands imply very low storage capacities S_{uMaxW}, which are a consequence of the energy rather than water limitation in that landscape unit and which are unlikely to be reflected by NDVI. Thus, under the assumption of $S_{uMax,W}$ being lower than the storage capacities in hillslopes and terraces, it was estimated as a free calibration parameter between the limits $0 < S_{uMaxW} < \min(S_{uMaxH}, S_{uMaxT})$ so that the following holds:

$$S_{uMax} = c(p_H S_{uMaxH} + p_T S_{uMaxT}) + p_W S_{umaxW} \tag{1}$$

with

$$S_{uMaxH} = S_{uMax} \frac{I_{NDVI,H}}{I_{NDVI}} \tag{2}$$

$$S_{uMaxT} = S_{uMax} \frac{I_{NDVI,T}}{I_{NDVI}} \tag{3}$$

where p_H, p_T, and p_W are the areal proportions of hillslope, terrace, and wetland classes, respectively, in the individual catchments and c is a rescaling factor to maintain the catchment integrated S_{uMax} and, which, as all other terms are known, varies with S_{uMaxW}.

4. Model Evaluation

4.1. Objective Functions

In order to select parameter sets that reproduce modeled simulations in high agreement with different aspects of the hydrological response, frequently not catered for by calibration to single objective functions, a multiobjective calibration strategy [*Gupta et al*., 1998] was applied in this study. The two objective functions used are Kling-Gupta efficiencies of flows (I_{KGE}), and of the logarithm of flows (I_{KGL}). These objective functions were chosen as they emphasize different parts of the hydrograph. While I_{KGE} can help to identify parameter sets that can best reproduce high flow dynamics, I_{KGL} identifies those that can better reproduce the hydrograph during low flows. The use of these two objective functions therefore helps to identify parameter sets that enforce a balanced system response which is more likely to be hydrologically consistent than parameter sets selected based on only one single objective function [*Gupta et al*., 1998]. The Pareto-optimal parameters sets were obtained with the MOSCEM-UA algorithm [*Vrugt et al*., 2003] with the number of complexes reflecting the number of parameters n and the number of initial samples set to $10n^2$ and a total number of 50,000 model iterations for all the models structures.

4.2. Experimental Design of Transferability Test

Model performance of the four tested models (FLEX^{L}, FLEX^{LM}, FLEX^{T}, FLEX^{TM}) in the 23 study catchments for calibration, validation, and transferability was assessed by the two objective functions (I_{KGE}, I_{KGL}). The detailed procedure of the experiment is listed as follows:

1. Calibrate one model based on one catchment with half the time series of the rainfall-runoff data. The Pareto-optimal parameter sets are retained.
2. Use the calibrated Pareto-optimal parameter sets to perform a temporal split-sample to assess the model performance during the validation period.
3. Identify catchments with poor data quality and exclude them from the subsequent transferability tests. Catchments with poor data quality were excluded because it is illogical to transfer the model parameters that failed to reproduce the hydrological processes of a donor catchment to a receiver catchment. Poor data quality was assessed based on the quality index proposed by *Visessri and McIntyre* [2015] and by the models' performance in temporal-validation. In particular, catchments with a quality index lower than 29 and with I_{KGE} and I_{KGL} lower than 0.5 in temporal-validation were excluded from the analysis.
4. Use all retained catchments in turn as donor catchments and transfer both, model and Pareto-optimal parameter sets of each individual model, to the remaining (receiver) catchments and model the respective flows without further calibration; note that for FLEX^{T} and FLEX^{TM}, the local landscape proportions and for FLEX^{LM} and FLEX^{TM} the local climate-derived S_{uMax} estimates of the respective receiver catchments are used rather than the values from the donor catchment.
5. Follow the same procedure for all four model setups.
6. Analyze model transferability in different landscapes by evaluating the performance of the four model set-ups with respect to the calibration objective functions. Compared with FLEX^{L} model, "substantial improvement" is defined as the objective function increased over 0.5; the increase of (0.1–0.5) means "moderate improvement"; reversely, < -0.5 indicates "substantial deterioration," and (-0.5 to -0.1) means "moderate deterioration." In addition, the transferred models are evaluated according to their ability to reproduce flow duration curves, individual hydrograph components, and the water balance of different landscapes in a way that is consistent with our understanding of the system functioning.

5. Results

5.1. Calibration Validation and Catchment Selection

Figure 4 illustrates the performance of the four tested models with respect to their objective functions, both, in calibration and validation periods for the sets of Pareto-optimal solutions. The corresponding parameter values in the 23 study catchments are shown in supporting information Figures S1 and S2. For most models and catchments, the mean model performance in the calibration period well exceeds I_{KGE} and $I_{KGL} = 0.5$. This indicates that all four models exhibit comparable skill to fit the observed hydrograph during the calibration period and that a well-justified choice of the most suitable, i.e., hydrologically consistent, model among the tested ones is not warranted by this information alone. In contrast, the performances in the validation period are significantly lower for a range of catchments (e.g., P.56, P.64, P.65, P.73, P.76, P.79, P.82, P.84, and P.85), which mainly coincide with those with lower data quality scores (Table 1). It can be argued that the lower performances in the validation period point, besides to data quality issues, toward model structural deficiencies or unsuitable parameters. However, due to the overall similarity of the performances of the four models, even in the validation period, little can be inferred on the importance of topographic heterogeneity and vegetation from this temporal split-sample test.

Applying the approach mentioned in point (3) of section 4.2, we found that 14 out of 23 catchments (Table 1 and Figure 4), which are in general those with the highest data quality scores, are suitable for use in the following model transferability test. The elevated proportion of catchments that exhibited poor data quality strongly underlines the challenges of acquiring reliable observations in many tropical environments.

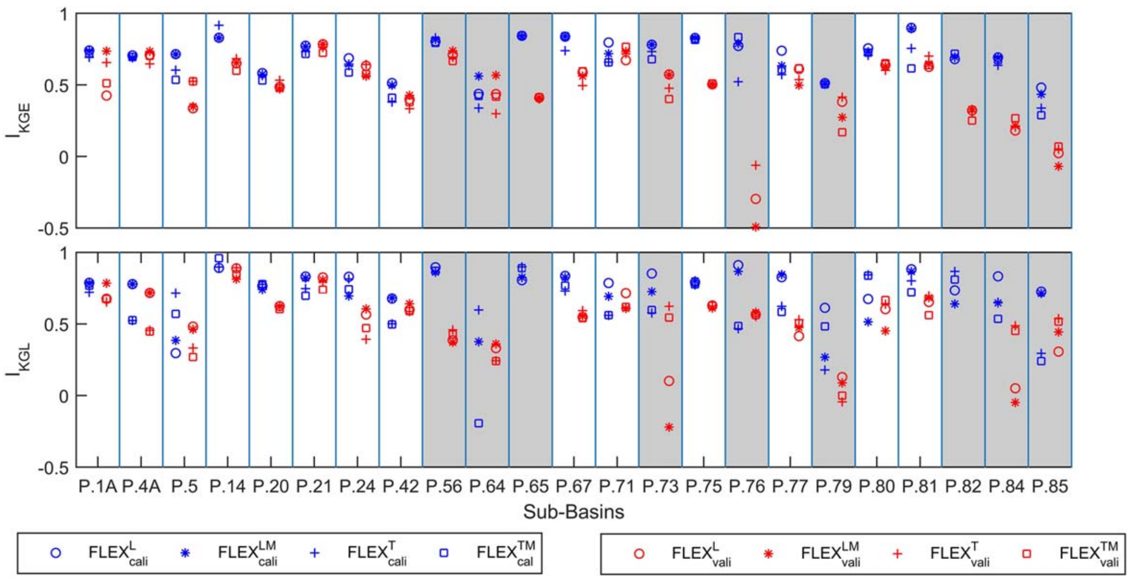

Figure 4. Calibration (blue) and time-validation (red) results of the four tested models in the 23 study catchments. The symbols indicate the mean (a) I_{KGE} and (b) I_{KGL} of all Pareto-optimal solutions. Grey boxes indicate the catchments excluded for the spatial transferability test due to insufficient performance with respect to at least on objective function, i.e., I_{KGE} or $I_{KGL} < 0.5$. Note that based on this test, no model clearly outperforms the others.

5.2. Model Transfer

5.2.1. Transferability of FLEX^L

In the transferability tests (Figure 5), i.e., when transferring a calibrated model from the calibration (donor) catchment to receiver catchments without further calibration, the model performance of FLEX^L during the calibration period deteriorated dramatically in most cases: the mean I_{KGE} across all catchments and Pareto-optimal solutions dropped from 0.74 to 0.18 whereas the mean I_{KGL} decreased from 0.75 to 0.19. Supporting information Table S3 shows the results with respect to two objective functions of all individual transfers tested in this study. It clearly shows the difficulty to transfer calibrated lumped model to adjacent catchments. Interestingly, it can be seen that FLEX^L has a reasonable transferability potential when transferring from P.20 or P.21 to other catchments. This indicates that these two catchments have a hydrological function that comes closest to that of the remaining catchments and that the selection of donor catchments is essential if the model does not account for a sufficiently high level of process heterogeneity.

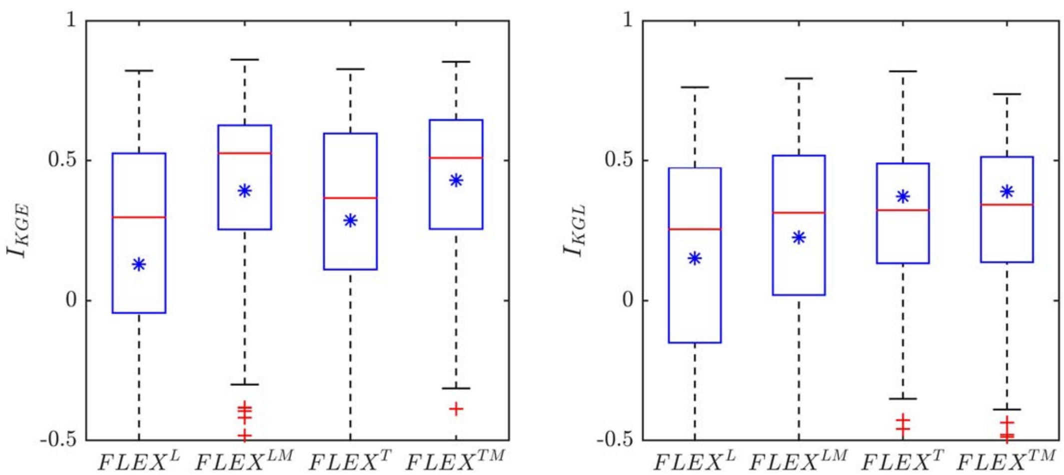

Figure 5. Overall model transferability results of the four models for all 14 study catchments used in the transferability test, as indicated by the distributions of I_{KGE} and I_{KGL} of all Pareto-optimal solutions. The red lines in boxes indicate median values, and blue stars indicate average values.

5.2.2. Transferability of FLEX^{LM}

The values of S_{uMax} for the 14 study catchments, independently obtained from observed hydroclimatological data by the MCT, are between 189mm (P.5) and 697mm (P.1). Figure 5 summarizes the mean performances across all Pareto-optimal solutions applied in all receiver catchments. Using the S_{uMax} values of the individual catchments in FLEX^{LM}, thus accounting for intercatchment differences in vegetation when treating them as receiver catchments and without any further calibration of the remaining parameters transferred from the donor catchments, increased the mean I_{KGE} from 0.18 to 0.44 and thus also improved model transferability compared to FLEX^{L}. A rank sum test supports these observations, indicating that the distributions of model performances of FLEX^{L} and FLEX^{LM} in the receiver catchments are significantly different ($p < 0.001$). Figure 6 provides a detailed illustration of the performance changes of FLEX^{LM} in the individual receiver catchment with respect to FLEX^{L}. It can be seen that 16% and 32% of the transfer cases, respectively, experience a substantial or moderate improvement, while in only 10% of the cases a (moderate) deterioration was observed. The improvement of the mean I_{KGL} compared to FLEX^{L} was on average 0.08, with less significance ($p = 0.061$). However, in 6% and 23% of the cases substantial or moderate improvement, respectively, was observed, with only moderate deterioration in 10% of the cases. The results suggest that when using donor catchments with comparatively low NDVI compared to the receiver catchments (e.g., P.5, P.80) the effects of accounting for differences in vegetation by adjusting S_{uMax} are beneficial. In contrast, donor catchment with NDVI values closer to the average of the receiver catchments (e.g., P.20, P.21, and P.81) exhibit less performance improvements.

5.2.3. Transferability of FLEX^{T}

The explicit consideration of several landscape units with distinct hydrological function in the semidistributed FLEX^{T} setup led, when transferred to receiver catchments, to an average improvement of the mean I_{KGE} from 0.18 to 0.36 compared to FLEX^{L} (Figure 5), with the distribution of solutions being significantly different ($p = 0.021$) according to a rank sum test, with substantial and moderate improvement in 13% and 24% of the transfer cases, respectively, and moderate deterioration in only 11% of the cases (Figure 6). The average improvement of I_{KGL} is from 0.19 (FLEX^{L}) to 0.41 (FLEX^{T}), with 21% and 17% of the cases exhibiting substantial or moderate improvement, respectively, and 17% moderate deterioration. No clear pattern could however be distinguished which types of catchments benefit most from explicitly accounting for topographic heterogeneity. It was however observed that the transfer from one of the donor catchments that results in the strongest performance improvements in the receiver catchments is characterized by rather strong landscape heterogeneity, i.e., no clear dominance of one class (P.5). In contrast, donor catchments characterized by more homogeneous landscapes, i.e., the dominance of one specific landscape class, exhibit the lowest overall performance improvements when accounting for landscape heterogeneity (e.g., P.4A, P.14, P.20, and P.75). This is an indication that the incorporation of landscape information is in particular beneficial for model transfer cases in which donor catchments are characterized by significant landscape heterogeneity while the receiver catchments are rather homogenous. In contrast, when receiver catchments are characterized by more landscape heterogeneity than a donor catchment, the results suggest that such donor catchments carry insufficient information on the composition of runoff processes that can be extracted by calibration to act as donor catchment [e.g., *Wooldridge et al*., 2002; *Nijzink et al*., 2016].

5.2.4. Transferability of FLEX^{TM}

The combined utilization of vegetation and topographic information in FLEX^{TM} resulted in considerable improvements of model transferability with respect to both objective functions, in general outperforming the other models (Figure 5). On average an improvement of mean I_{KGE} for the receiver catchments to 0.46 compared to FLEX^{L} ($I_{KGE} = 0.18$) can be observed. Similarly, an improvement from 0.19 (FLEX^{L}) to 0.42 was found for mean I_{KGL}. For both objective functions, the improvements are statistically significant as suggested by a rank sum test ($p < 0.001$ and $p < 0.001$). While the results indicate moderate to substantial performance improvements of mean I_{KGE} for 32% and 20% of the transfer cases, respectively, moderate or substantial improvements of mean I_{KGL} were observed in 26% and 24% of the transfers (Figure 6). In contrast, only 5% and 18% of the cases, respectively, indicate a moderate performance deterioration. On balance, it can be said that FLEX^{TM} provides the overall strongest improvements for model transferability in the study region. Its improvements for reproducing the high flow responses (i.e., I_{KGE}) in the receiver catchments are more pronounced than those for the low flows (i.e., I_{KGL}). Accounting for intercatchment vegetation differences (FLEX^{LM}), and thus differences of how catchments partition, store and release water, is somewhat more important for the improvement of high flows in receiver catchments than accounting for intracatchment landscape heterogeneity (FLEX^{T}). The latter, however, was in this study found to be slightly more relevant for improving the representation of low flows in

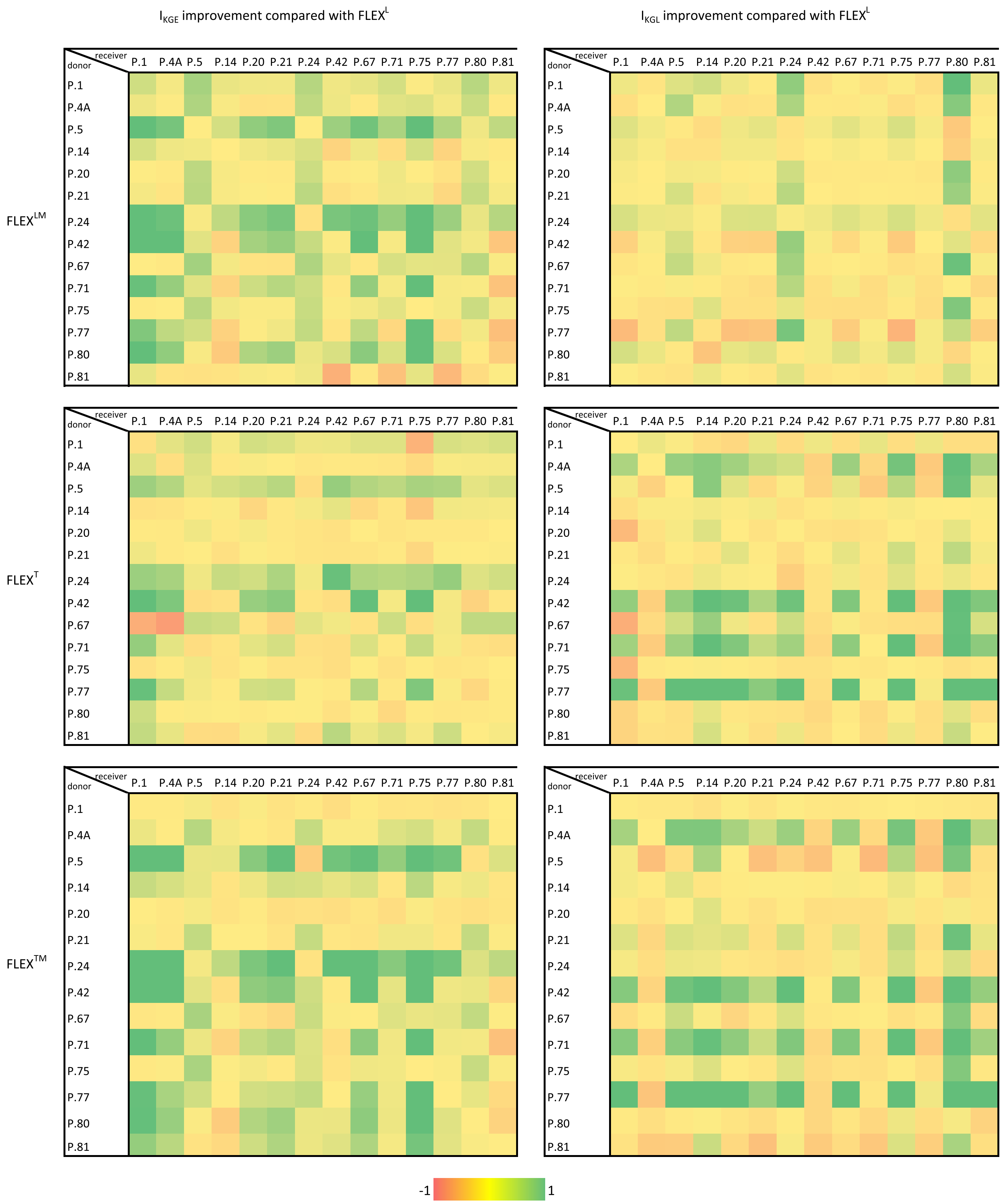

Figure 6. The improvement of three hydrological models (FLEX^{LM}, FLEX^{T}, and FLEX^{TM}) compared with FLEX^{L} for the two calibration objective functions (I_{KGE}, I_{KGL}) after model transfer from donor to receiver catchments in 14 catchments. The values of objective functions are based on the mean value of each objective function from the set of Pareto-optimal solutions. Shades of green indicate slight (light green) to strong (darker green) improvement in terms of mean I_{KGE} and I_{KGL}, respectively; shades of red indicate slight (orange) to strong (red) deterioration, while yellow indicates no significant change. FLEX^{TM} shows the best improvement for both I_{KGE} and I_{KGL}.

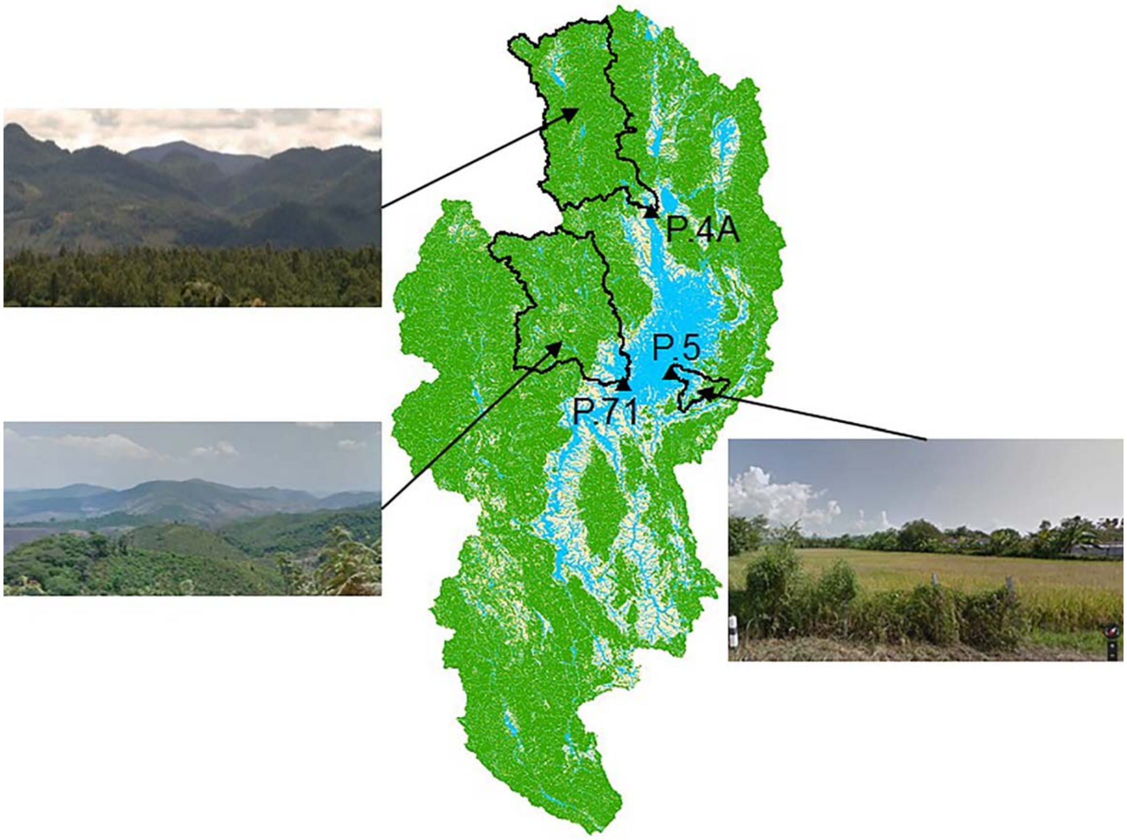

Figure 7. Comparison of landscapes for two specific transfer cases: (a) transfer between catchments with similar landscape (donor: P.71, receiver: P.4A), (b) transfer between distinct landscapes (donor: P.5, receiver: P.4A).

the receiver catchments. The overall results provide some evidence for the value of explicitly considering both landscape-driven intracatchment process heterogeneity and the influence of intercatchment vegetation differences using independent estimates of root zone storage capacity to increase model transferability. Note, however, that in a few specific cases, FLEX^{TM} cannot outperform the other models tested in this study. This is for example the case for transferring the model from P.80 to P.14 (Figure 6) and highlights the potential influence of additional factors not considered in this study, such as geology or soil types. In most cases in this study, however, the transferability of the four models follows the sequence FLEX^{TM} $>$ FLEX^{LM} $>$ FLEX^{T} $>$ FLEX^{L}.

5.3. Two Transfer Cases

To better illustrate the considerable influence of heterogeneity in topography and vegetation cover (using S_{uMax} and NDVI as proxies) on the composition of dominant processes in a catchment and thus for spatial model transferability, we will in the following discuss two specific transfer cases in detail. One transfer case concerns two catchments with similar landscapes and size (Figure 7), i.e., from P.71 to P.4A, both of which are dominated by hillslopes ($p_W < 11\%$, $p_T < 16\%$, $p_H > 70\%$) and dense vegetation cover (NDVI > 0.70). The second transfer case concerns catchments with very different landscape features (Figure 7), i.e., from P.5 to P.4A. Catchment P.5 has a larger proportion of wetlands and terraces ($p_W \sim 40\%$, $p_T \sim 24\%$, $p_H \sim 35\%$) and a less dense vegetation cover than P.4A (NDVI ~ 0.55). The hydrograph and FDC of the receiver catchment P.4A are shown in Figures 8 and 9, and which are generated by all the Pareto-optimal parameter sets obtained from the donor catchments P.71 and P. 5.

We found that when transferring between similar catchments, with comparable topography and vegetation cover, i.e., from P.71 to P.4A, not surprisingly, FLEX^{L} performs much better than from P.5 to P.4A with different landscapes.

While the FLEX^{L} does largely pick up the timing of peaks (Figure 8) especially when transferring from P.5 to P.4A, it significantly overestimates the flow (Figures 8 and 9), indicating that the flux partitioning in P.4A is

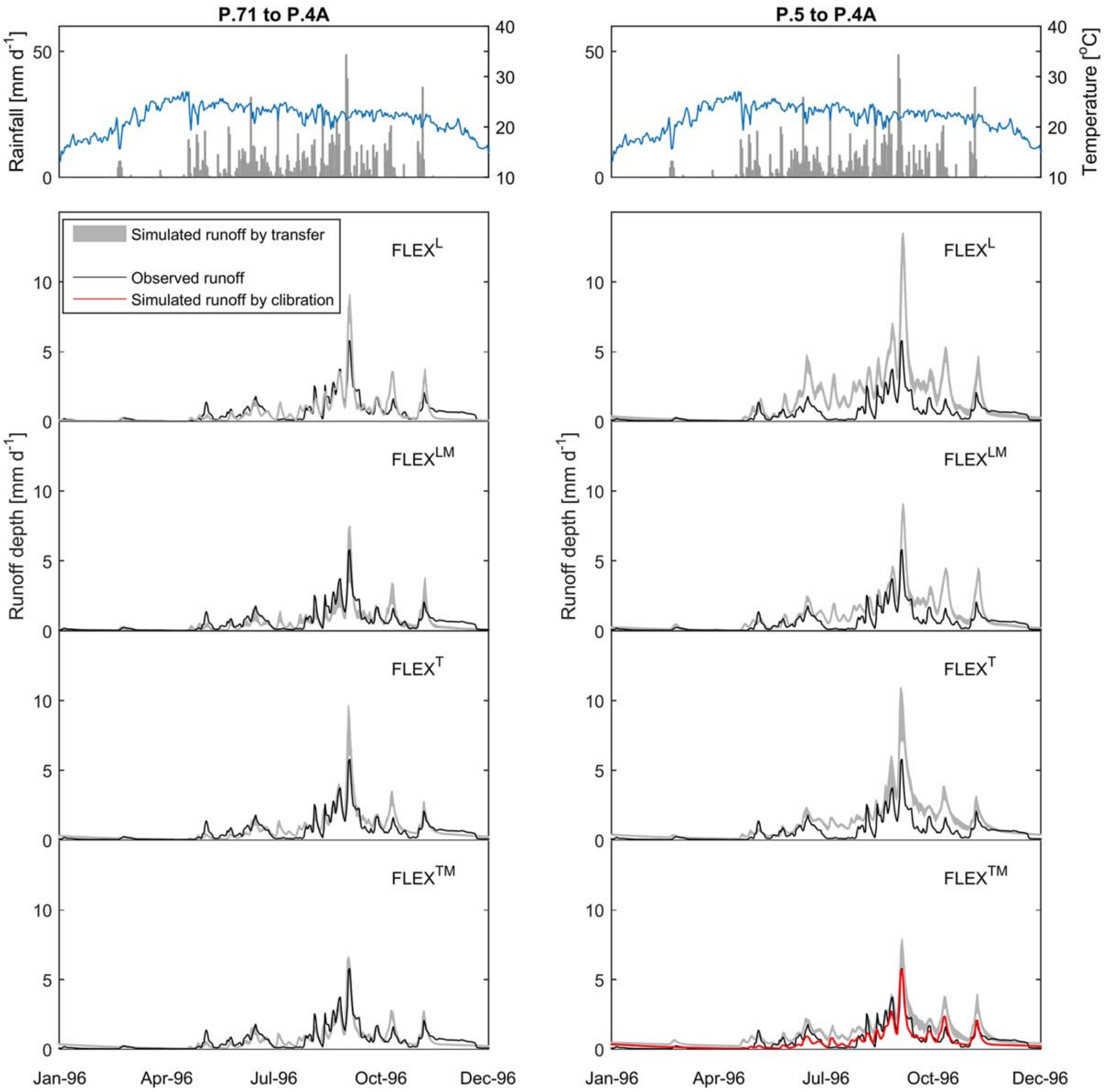

Figure 8. Observed (black line), modeled runoff by transfer (grey-shaded area), and calibrated hydrographs (red line) of the four models in the receiver catchment P.4A for (a) transfer of both model structure and parameters from donor catchment P.71 characterized by similar landscapes and (b) from donor catchment P.5 characterized by different landscape composition. This illustrates that accounting topographic and vegetation information allows more robust model transferability among catchments with different landscapes.

not adequately reproduced. The results further illustrate that directly transferring parameters obtained from FLEX^{L} without consideration of the differences in hydrological function may result in serious misrepresentations of the system response, especially peak flows, in the receiver catchments, as demonstrated by many previous studies [e.g., *Heuvelmans et al*., 2004; *Uhlenbrook et al*., 2010; *Gao et al*., 2014a], strongly indicating an insufficient representation of processes heterogeneity in the FLEX^{L} model.

When considering either vegetation (FLEX^{LM}) or topography (FLEX^{T}), however, the simulations in the receiver catchments experience significant performance improvements. The model transfer is most successful in this example catchment when involving both vegetation information, in terms of spatially distributed, a priori estimated values of S_{uMax}, and topography, in terms of spatial process heterogeneity, in FLEX^{TM}.

5.4. The Modeled Water Balance of FLEX^{TM}

The modeled components of the flow generated in the individual landscape units for the two selected illustrative cases (transfer from P.71 and P.5 to P.4A) are given in Figure 10. The results clearly suggest that in the donor catchments (P.71 and P.5), most runoff toward the end of dry and the beginning of wet season is

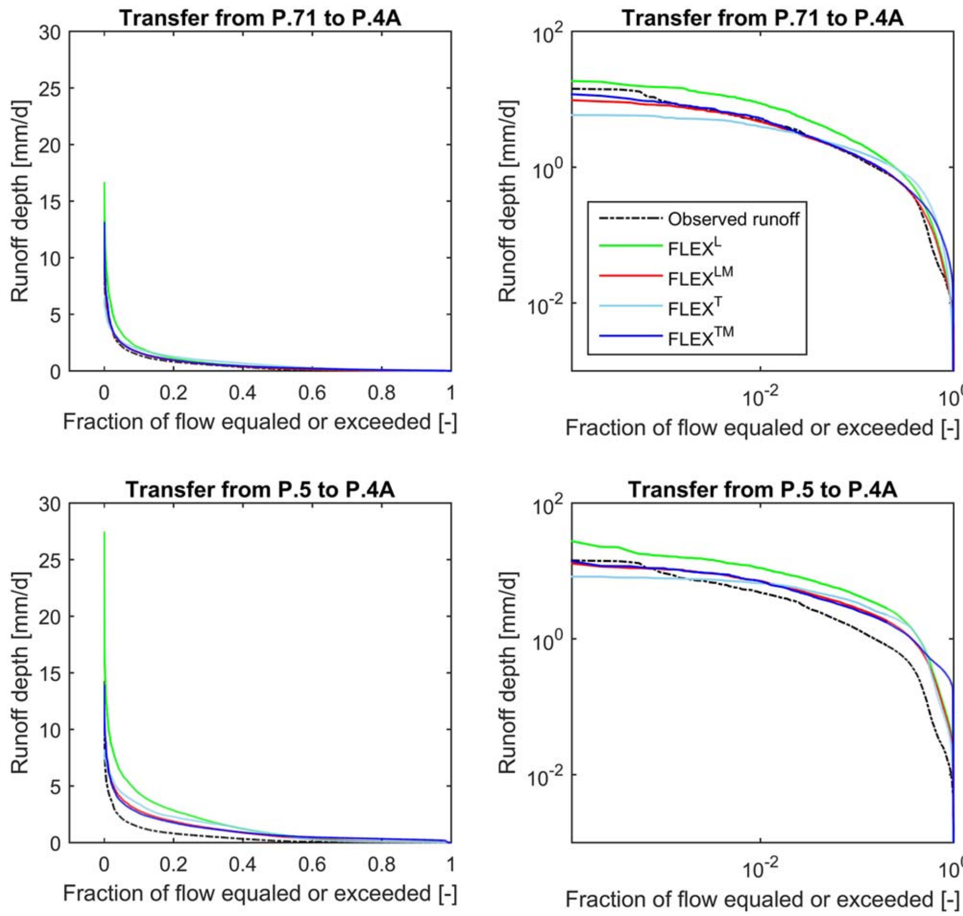

Figure 9. Observed and modeled flow duration curves (in normal scale, left and in semi-log scale, right) of receiver catchment P.4A, when using the model parameters obtained from calibration in the donor catchments (two top figures) P.71 and (two bottom figures) P.5. Lumped model can also reproduce FDC while transferring among similar catchments, but fails to reproduce FDC of catchments with different landscape features.

generated from wetlands, in spite of their limited area (Table 1), thus clearly reflecting the intended process conceptualization. Further, runoff generation from hillslopes, in particular during wetting-up conditions, does not reflect its dominant areal proportion (Table 1), due to the considerable higher storage deficits, sustained by deeper groundwater tables and higher interception evaporation and plant transpiration rates than in the agriculturally dominated wetlands. However, with the increase of soil moisture, the proportion of runoff from hillslopes becomes more important during the wet seasons. It was also observed that in the model, terraces do not generate direct runoff, except for the end of the wet season, although they cover a considerable area in the catchments, which mirrors the a priori defined hydrological function of the terrace landscape unit. P.4A has a similar hydrological characteristic, but due to the dominance of hillslopes, a larger amount of runoff is generated from there. The similarity of simulated hydrograph components, no matter if transferred from P.71 or P.5, indicates a certain level of robustness of FLEX^{TM} to reproduce model-internal dynamics that are in line with our understanding of catchment functioning. In general, these results suggest that a model allowing for some degree of landscape heterogeneity, such as FLEX^{TM}, provides means for a more meaningful representation of the distinct flow dynamics in wet and dry seasons, caused by runoff contributions from different parts of the landscape, which is also beneficial for model transfer.

In Table 3, the modeled fluxes for the P.4A catchment and its different individual landscapes elements, obtained by Pareto-optimal parameters from P.5, are given. The results suggest that wetlands, due to their limited storage capacity, generate considerably more runoff per unit area (554–746mm/a) than hillslopes (141–318mm/a). Terraces contribute least to storm flow (34–71mm/a) due to their flat slopes and, related to that, the elevated amounts of percolation (345–573mm/a). By comparison, we found that the fast runoff

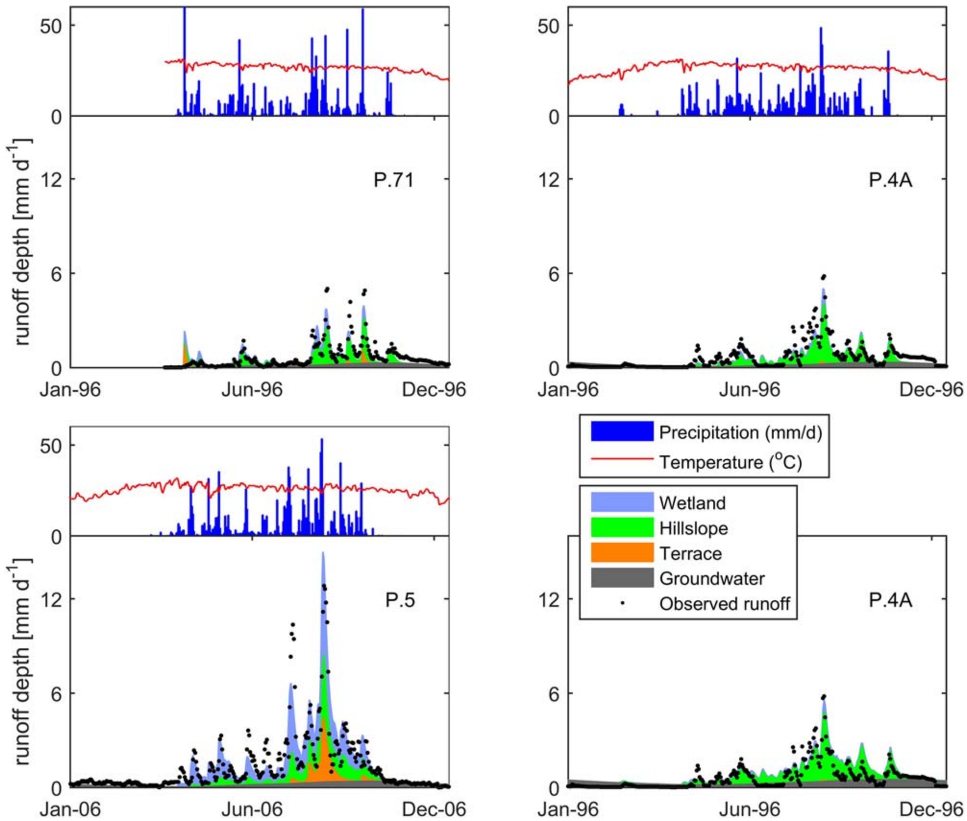

Figure 10. Modeled hydrograph components obtained by FLEX^{TM}, (here: average values of all modeled hydrographs obtained from the set of feasible solutions), using the model parameters calibrated in donor catchments (top left) P.71 and (bottom left) P.5, respectively, in the receiver catchment P.4A (right). Different proportion of landscapes with distinctive water balance allows better model transferability.

generated from wetlands of each unit is over 2–5 times larger than the runoff generated from each unit of hillslopes, not to mention terraces.

It is worthwhile to check the simulated evaporation from different landscapes as well. The modeled evaporation and transpiration from wetlands (996–1253mm/a) is higher than from hillslopes (520–759 mm/a) and terraces (445–749 mm/a). Being less moisture constrained, wetlands not only generate larger proportion of runoff but also contribute larger proportions of evaporation and transpiration than other landscapes. The overall water balance clearly shows, for that example, that FLEX^{L} considerably underestimates evaporative fluxes and overestimates runoff when transferred from P.5 (Figure 11). The other model setups (FLEX^{T}, FLEX^{LM}, FLEX^{TM}) that allow some flexibility for the transfer of the parameters controlling water availability for transpiration, i.e., the root zone storage capacity (S_{uMax}), produce considerably more adequate estimates

Table 3. Modeled Water Balance of Individual Fluxes in Catchment P.4A Using the Set of Pareto-Optimal Parameters Obtained From Calibration in P.5[a]

Entire Catchment		Hillslope		Terrace		Wetland	
Fluxes		Fluxes		Fluxes		Fluxes	
P (mm/a)	1274	P (mm/a)	1274	P (mm/a)	1274	P (mm/a)	1274
Q_m (mm/a)	(282, 427)	Q_{fH} (mm/a)	(141, 318)	Q_{fT} (mm/a)	(34, 71)	Q_{fW} (mm/a)	(554, 746)
E_i (mm/a)	(163, 382)	E_{aH} (mm/a)	(520, 759)	E_{aT} (mm/a)	(445, 749)	E_{aW} (mm/a)	(996, 1253)
Q_s (mm/a)	(108, 134)	R_{slH} (mm/a)	(110, 141)	P_{ercT} (mm/a)	(345, 573)		

[a]Rainfall (P), interception (E_i), modeled runoff (Q_m), and groundwater flow (Q_s) are for the entire catchments. The remaining fluxes are given per unit area of each landscape unit. E_{aH}, E_{aT}, E_{aW} indicate transpiration from hillslope, terrace, and wetland. Q_{fH}, Q_{fT}, and Q_{fW} indicate subsurface flow from hillslope, and saturated overland flow from terrace and wetland separately. R_{slH} indicates preferential recharge of groundwater on hillslope; P_{ercT} indicates the percolation on terrace. The values in brackets indicate the ranges of the modeled fluxes as obtained by the Pareto-optimal parameter sets.

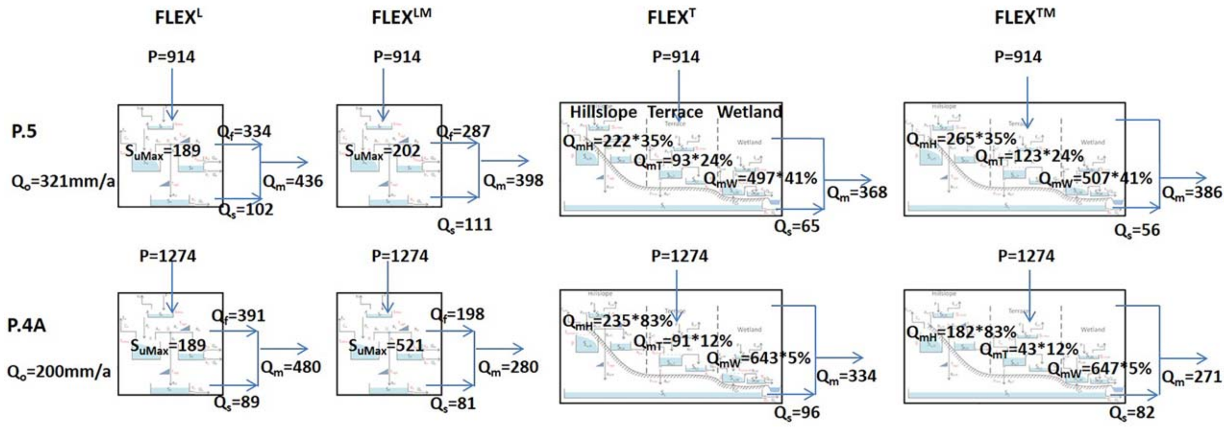

Figure 11. The modeled water balance of the four models for the transfer from donor catchment P.5 to receiver catchment P.4A. The fluxes (mm/a) are the mean modeled values obtained by all Pareto-optimal solutions calibrated based on P.5. P is the long-term mean precipitation; E are the long-term mean evaporative fluxes from the entire catchment; E_H, E_T, and E_W are the evaporative fluxes per unit area from the individual landscape classes; Q_o is the long-term observed catchment runoff; and Q_m is the long-term modeled catchment runoff. Three models ($FLEX^{LM}$, $FLEX^T$, and $FLEX^{TM}$) applying different approaches to consider landscapes heterogeneity allow more realistic water balance simulation while being transferred.

of annual flow volumes in P.5. This underlines the importance of a plausible inter and intracatchment representation of root zone storage capacities, and thus the role of vegetation and its intracatchment distribution as controlling factor for flux partitioning in such water.

6. Discussion

This study is primarily meant to highlight the considerable explanatory power of landscape and ecosystem heterogeneity to characterize the heterogeneity of the dominant hydrological processes. The study is not meant to provide a full regionalization framework or to present an operational model for the study area, and therefore no recalibration or additional regionalization efforts of model parameters were attempted. In the following sections, we discuss the implications and potential of using landscape and ecosystem properties for model transferability, model consistency, and prediction in ungauged basins.

6.1. The Influence of Vegetation and Topography on Model Transferability

As emphasized in many previous studies, spatial proximity does not necessarily entail hydrological similarity [e.g., *Bárdossy et al*., 2005; *Ali et al*., 2012]. It is therefore not surprising to find here that a direct transfer of lumped models and their parameters without further calibration to other study catchments does in general fail. However, it can be observed that in the validation period the lumped $FLEX^L$ model is, for most catchments, outperformed by one or more of the other three models, but in particular by $FLEX^{TM}$ that allows for both topographic heterogeneity and a priori estimates of the root zone storage capacity without further recalibration. This suggests that, in the study region, both topography and vegetation contain relevant information on runoff generation processes.

More specifically, the inclusion of vegetation information by adapting S_{uMax} (i.e., the storage capacity required by vegetation to overcome dry periods with a specific return period) to the observed environmental conditions improved model transferability in most study catchments, even if all other model parameters were left unchanged. The results provide supporting evidence that—first-order accurate—estimates of the root zone storage capacity S_{uMax}, based on climate data and simple vegetation indicators as suggested in recent studies [*Gao et al*., 2014b; *de Boer-Euser et al*., 2016], can be efficiently used for hydrological modeling, to replace detailed information on soils and rooting depth. For example, due to less dense vegetation, P.5 requires a smaller root zone storage capacity than the more densely vegetated P.4A catchment. Thus, when using P.5 (S_{uMax}=189mm) as donor catchment and directly transferring its root zone storage capacity to P.4A using $FLEX^L$, the buffering capacity is too low to supply sufficient water for the evaporative water demand, resulting in significant overestimation of both the annual runoff volume (Figure 11) and peak flow events (Figures 9 and 10). In contrast, adapting S_{uMax} to the denser vegetation cover and, implicitly, the higher canopy water demand of P.4A in $FLEX^{LM}$ (S_{uMax} = 521 mm, Figure 11) significantly increases the model's ability to moderate high flows and to better reproduce mean annual flows (and thus also transpiration amounts) that are closer to the observations. The results therefore clearly illustrate the role of vegetation in partitioning water fluxes and the value of using climate-based estimates of root zone storage capacities in hydrological models.

In addition, few studies explicitly investigated the influence of topography on model transferability, although numerous hydrological models [e.g., *Beven and Kirkby*, 1979; *Reggiani et al*., 2000; *Gharari et al*., 2014] have been developed based on topographic information. Here, the strategy to use landscape heterogeneity as a proxy for process heterogeneity has proven effective for improving both model consistency and transferability. For example, P.5 has a large spatial extent of wetlands and terraces (Table 1, Figures 1g and 7), which are characterized by a relatively shallow root zone storage capacity. In contrast, P.4A is dominated by hillslopes (Table 1, Figures 1g and 7), with a larger S_{uMax} which allows the development of higher moisture deficits and which, in turn after a dry spell requires more water to establish hydrological connectivity. FLEX^{T} takes the proportions of three different landscapes into account, considering the differences in root storage capacity between them (Figure 11). This significantly improves model transferability as compared to the lumped FLEX^{L} model, in particular when transferring to catchments with more topographic heterogeneity.

In fact, the major difference between the lumped model and the other models is the description of the unsaturated root zone reservoir (S_u, Figure 11). While FLEX^{LM} accounts for vegetation differences between catchments by direct estimation of S_{uMax} from climatic and vegetation data, FLEX^{T}, in contrast, determines S_{uMax} in different landscapes based on topographic information and assumptions on the position of the groundwater table, with FLEX^{TM} combining both approaches. The significantly improved transfer capacity of these models compared to FLEX^{L}, indicates that a more detailed description of the root zone storage capacity is essential to improve model transferability in the study region and ultimately to adequately represent the dominant hydrological processes.

6.2. Hydrological Consistency of Model Results

The semidistributed FLEX^{T} and FLEX^{TM} model setups not only exhibit a generally better performance than the lumped FLEX^{L} model, but also, and maybe more importantly, produce internal process dynamics that are broadly consistent with the modelers' understanding and expectation of how different parts of the system should respond differently. The results suggest that already a limited level of additional process heterogeneity, as encapsulated in a semidistributed model formulation, has considerable value for reproducing observed response dynamics. Note, that variability in hydrological response dynamics between catchments is clearly influenced by factors (e.g., geology, soils, etc.) other than landscape and vegetation as well. However, the general results of this study provide supporting evidence that both, landscape and vegetation, cannot only exert considerable influence but that these influences are relatively unproblematic to meaningfully implement and parameterize at an adequate scale of interest in a model independent of further calibration efforts. The chosen strategy therefore has a crucial advantage over standard applications of semidistributed models: in spite of a relatively high number of parameters, the number of free calibration parameters is comparably low as landscape proportions and root zone storage capacities are directly estimated from data, thereby efficiently limiting the adverse effects of equifinality. Nevertheless, with the FLEX^{TM} model with 11 calibrated parameters, significant equifinality remains, which may lead to some of the parameter values having ambiguous significance and uncertainty in simulated results (Table 3). Work is needed to further reduce the need for calibration.

The use of a parallel structure in these models is supported by field experiments [*Zhao*, 1984; *Pfister*, 2006]. For example, the results of many tracer and piezometer experiments highlight the difference in runoff generation mechanisms between riparian areas and hillslopes [*McGlynn and McDonnell*, 2003; *Molenat et al*., 2008; *Detty and McGuire*, 2010]. Here, the individually modeled hydrograph components (Figure 10 and section 5.5) and evaporation (Table 3) meet the expected system internal dynamics and do not contradict with existing experimental knowledge in hillslope and catchment hydrology [*Yu, et al*., 2001; *McGlynn and McDonnell*, 2003]. Specifically, in the beginning of a wet season, most peak flows are generated from wetlands. Gradually, more water is discharged from hillslopes as the catchment wets up and the soil moisture deficits on the hillslopes are eventually reduced. During large storm events, terraces become saturated, therefore contributing to runoff. Lumped representations of a catchment (i.e., FLEX^{L}) cannot reproduce these characteristic features of the hydrological response. However due to the relatively complex model structure and the elevated numbers of parameters, it is difficult to meaningfully test models such as FLEX^{TM} by temporal split-sample validation at one specific study site. With the help of a transferability test, the designed model setups could be tested more robustly by transferring from a calibrated donor catchment to receiver catchments, indicating an increasingly robust transfer performance for the suite of suggested models. Although this clearly indicates increasingly adequate process representations, the suggested model concepts still require quantitative evaluation against more detailed experimental data.

6.3. Implications for Regionalization and PUB

Previous research efforts on prediction in ungauged basins (PUB) focused mostly on the development of suitable parameter regionalization techniques for relatively simple, lumped models [e.g., *Merz and Blöschl*, 2004; *Parajka et al.*, 2005] applying the same model structure in donor and receiver catchments. Although shown to be valuable for use in catchments with similar geomorphic features, the use of lumped models, even if suitable for calibration in individual catchment, disregards the fact that field experiments underline the importance of distinguishing between dominant hydrological processes in different landscapes [*McGlynn and McDonnell*, 2003; *Molenat et al.*, 2008; *Detty and McGuire*, 2010] within and between individual catchments. In such cases, the mere regionalization of parameters in lumped models may, depending on the model, not be sufficient to account for this heterogeneity, thus requiring more flexible model structures with parallel components. These components can introduce heterogeneity by different architectures, different parameter values, or both. This is clearly highlighted by the results of the present study, which demonstrate the value of such a landscape driven, semidistributed setup of parallel model structures with distinct processes and parameter values. It was shown here, that calibrating FLEX^{T} in a donor catchment and transferring it by merely adjusting the areal proportions of the individual landscape units without further calibration or regionalized parameter estimates to receiver catchments can significantly improve the results in comparison to the transfer of FLEX^{L}. This further supports earlier studies [e.g., *Hrachowitz et al.*, 2014] that already the distinction in different landscape units, based on readily available information, and contains considerable information on the hydrological function of the system within a hydroclimatically homogeneous region.

Similarly, adjustments of S_{uMax} in FLEX^{LM} according to the local precipitation and transpiration characteristics in receiver catchments proved, without recalibration of other parameters, highly beneficial for model transferability and underpins the importance of this parameter for meaningful flux partitioning. The use of direct S_{uMax} estimates is, however, at this point limited to catchments where sufficient data, i.e., at least either estimates of average annual flow or actual evaporation (e.g., from remote sensing products), are available. If these are available, the combination of topographic heterogeneity and vegetation heterogeneity (FLEX^{TM}) is a potentially powerful tool for predicting the hydrological response in ungauged catchments, provided some level of calibration is possible within the same hydroclimatic region. Future studies may want to test to which level the suggested approach is complementary to traditional regionalization techniques to improve our ability to predict flows in ungauged catchments. In general, it may be noted that the suggested techniques for spatial model transfer are expected to be particularly useful in relatively dry environments, such as in the tropical savannah study region, where the pronounced differences between wet and dry seasons and the associated changes in hydrologically active areas within catchments, together with the dominance of evaporative fluxes over stream flow, and through transpiration, the importance of vegetation control the hydrological response. Efforts to predict runoff in ungauged basins in such semiarid climate zones benefit from the suggested model transfer method, as no longer local or regional empirical transfer functions for parameter regionalization are required.

Notwithstanding these findings, some limitations and open questions remain. On the one hand, the direct estimation of S_{uMax} is still dependent on reliable estimates of catchment-averaged precipitation, potential evaporation, and runoff. Although precipitation data are, through remote sensing products, globally available, the uncertainty in these data may cause a considerable bias in the S_{uMax} estimates. Likewise, runoff observations are typically subject to uncertainties, in particular for high flows [e.g., *Coxon et al.*, 2015; *McMillan and Westerberg*, 2015] and, similarly, the choice of the method to potential evaporation may also somewhat affect the results, although several previous studies suggest that these effects are minor at the catchment-scale [e.g., *Oudin et al.*, 2005; *Kleidon et al.*, 2014]. An alternative would be to use remotely sensed evaporation estimates, but these still contain considerable uncertainty as well. In addition, direct estimation of S_{uMax} should be made subject to the different survival strategies of ecosystems, such as deciduous forests, grasses, evergreen forests, or Eucalyptus species that develop very deep root systems to tap ground water. On the other hand, it should be noted that even within hydroclimatically homogeneous regions, catchments can be characterized by distinct landscape units. In other words, even if a donor catchment accounts for several different landscape units, a receiver catchment may require additional or other (e.g., glacier, bare rock, etc.) landscape units. Finally, the suggested approach does not yet account for model-structural distinctions due to geological factors which may be dominant [e.g., *Uhlenbrook et al*., 2004; *Fenicia et al*., 2014, 2016].

7. Conclusions

Landscapes are essential in determining the dominant runoff generation mechanisms, model structure, and parameterization. In this study, we developed four conceptual models with increasing complexity: a classical lumped model serving as benchmark (FLEX^L), a lumped model with independent estimates of root zone storage capacity derived from hydroclimatological data (FLEX^{LM}), a topography based semidistributed model (FLEX^T), and a topography-based semidistributed model with independent estimates of root zone storage capacity derived from climate data and NDVI (FLEX^{TM}). Fourteen study catchments, located in a tropical savannah region, that is hydrologically less well understood than temperate regions, were used to test the performance of the four model structures, by a calibration validation and transferability test. The calibration and time-validation results indicate that all four models can, to a certain extent, mimic the hydrological behavior of the study catchments. But model performance in space-validation differs significantly between models. We found that:

1. Both, intercatchment heterogeneity in vegetation and intracatchment heterogeneity in topography can explain a considerable part of differences in hydrological function between catchments in this tropical savannah region.
2. Accounting for both vegetation and topographic heterogeneity allowed to distinguish catchment characteristics and to considerably increase model transferability without the need for empirical regionalization relationships. Accounting for topography and vegetation simultaneously gave best transferability results. Individually, accounting for vegetation heterogeneity by local adaptation of parameters describing the root zone storage capacity through independent, climate data-based estimation may be more beneficial than purely topographic information in this specific climate.
3. The semidistributed modelling approach (FLEX^T, FLEX^{TM}) allows a more plausible representation of the system internal dynamics and the dominant runoff generation mechanisms than the lumped model (FLEX^L). Explicitly allowing for topographic heterogeneity in a model can be highly valuable for predicting flows in ungauged tropical savannah catchments. If at least estimates of annual flow averages are available in an otherwise ungauged catchment, vegetation heterogeneity does also bear considerable potential to improve flow predictions under data-limited conditions in this tropical savannah basin.
4. The results in general underline, for this climate region, the importance of (a) vegetation and its within-catchment distribution for the partitioning of evaporative fluxes and runoff, due to high evaporative demand and the seasonally markedly changing wetness conditions and (b) seasonally changing contributing areas due to pronounced dry and wet periods and the distinct hydrological function of different parts of the landscape.

Notation

P	precipitation (mm/d)
E_p	means the potential evaporation, (mm/d)
p_H	area proportion of hillslopes
p_T	area proportion of terraces
p_W	area proportion of wetlands
I_{NDVI}	dry seasonal averaged NDVI of entire catchment
$I_{NDVI,H}$	dry seasonal averaged NDVI of the hillslopes
$I_{NDVI,T}$	dry seasonal averaged NDVI of the terraces
I_{KGE}	Kling-Gupta Efficiency of hydrograph
I_{KGL}	Kling-Gupta Efficiency of the logarithmic hydrograph

FLEX^L Model

E_i	interception (mm/d)
E_a	transpiration (mm/d)
P_e	means the effective rainfall after interception (mm/d)

Q_{ff} overland flow (mm/d)
Q_f subsurface storm flow (mm/d)
Q_s groundwater flow (mm/d)
Q_m simulated runoff (mm/d)
R_u generated runoff of certain rainfall event (mm/d)
R_f generated fast runoff from the runoff generation routine (mm/d)
R_s generated slow runoff from the runoff generation routine (mm/d)
R_{fl} discharge into the fast response reservoir after the convolution (mm/d)
R_{sl} discharge into the slow response reservoir after the convolution (mm/d)
S_{iMax} maximum storage capacity of interception reservoir (mm)
S_{uMax} averaged storage capacity of unsaturated reservoir (mm)
β shape parameter of the tension water storage capacity curve
C_e threshold controls actual evaporation and transpiration
D splitter between surface runoff and groundwater recharge
T_{lagF} parameter represents the time lag between storm and fast runoff generation (d)
T_{lagS} lag time from preferential flow and percolation flow to the groundwater (d)
S_{fMax} threshold of generating overland flow (mm)
K_{ff} recession parameter of overland runoff (d)
K_f recession parameter of subsurface storm flow (d)
K_s recession parameter of groundwater flow (d)

FLEX^T Model

E_{iH} interception from hillslope (mm/d)
E_{iT} interception from terrace (mm/d)
E_{iW} interception from wetland (mm/d)
E_{aH} transpiration from hillslope (mm/d)
E_{aT} transpiration from terrace (mm/d)
E_{aW} transpiration from hillslope (mm/d)
P_e effective rainfall after interception (mm/d)
Q_{fH} subsurface storm flow on hillslope (mm/d)
Q_{fT} saturated overland flow on terrace (mm/d)
Q_{fW} saturated overland flow on wetland (mm/d)
R_{slH} recharge to the groundwater reservoir by preferential flow on hillslopes (mm/d)
P_{ercT} recharge to the groundwater reservoir by percolation on terraces (mm/d)
Q_s groundwater flow (mm/d)
C_R capillary rise from groundwater into unsaturated reservoir on wetland (mm/d)
S_{uMaxH} root zone storage capacity on hillslope (mm)
S_{uMaxT} root zone storage capacity on terrace (mm)
S_{uMaxW} root zone storage capacity on wetland (mm)
P_{Max} maximum percolation capacity from terrace to groundwater (mm/d)
K_{fH} recession parameter of subsurface storm flow on hillslope (d)
K_{fTW} recession parameter of saturated overland flow on wetland and terrace (d)

Acknowledgments
We are grateful to Royal Irrigation Department (RID) and Thai Meteorological Department (TMD) for proving the hydrological and meterological data in the upper Ping River basin in Thailand. Both land use map and soil type map are provided by Land Development Department in Thailand. The DEM and NDVI data were downloaded from https://lpdaac.usgs.gov/data_access/usgs_earthexplorer. We thank Suraphan Santiyanon for data processing. Constructive discussions and suggestions from Wei Shao, Luis Samaniego, and Rohini Kumar are highly appreciated. The authors also thank Alberto Viglione and two anonymous reviewers for their critical and highly constructive comments.

References

Ali, G., D. Tetzlaff, C. Soulsby, J. J. McDonnell, R. Capell (2012), A comparison of similarity indices for catchment classification using a cross-regional dataset, *Adv. Water Resour.*, *40*, 11–22, doi:10.1016/j.advwatres.2012.01.008.
Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration-Guidelines for computing crop water requirements, *FAO Irrig. Drain. Pap., 56*, vol. 300, D05109 p., FAO, Rome.
Ambach, W. (1988), Interpretation of the positive-degree-days factor by heat balance characteristics—West Greenland, *Nord. Hydrol.*, *19*, 217–224.
Andréassian, V., C. Perrin, L. Berthet, N. Le Moine, J. Lerat, C. Loumagne, L. Oudin, T. Mathevet, M. H. Ramos, and A. Valéry (2009), HESS Opinions "Crash tests for a standardized evaluation of hydrological models", *Hydrol. Earth Syst. Sci. Discuss.*, *6*(3), 3669–3685.
Arnold, J. G., R. Srinavasan, R. S. Muttiah, and J. R. Williams (1995), Large-area hydrologic modeling and assessment: Part I. Model development, *J. Am. Water Resour. Assoc.*, *34*, 73–89.
Bárdossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, *11*(2), 703–710.
Bárdossy, A., G. G. S. Pegram, and L. Samaniego (2005), Modeling data relationships with a local variance reducing technique: Applications in hydrology, *Water Resour. Res. 41*, W08404, doi:10.1029/2004WR003851.
Beven, K., and J. Freer (2001), A dynamic TOPMODEL, *Hydrol. Processes*, *15*(10), 1993–2011.

Beven, K., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, *24*(1), 43–69.
Blöschl, G. (2001), Scaling in hydrology, *Hydrol. Processes*, *15*(4), 709–711.
Blöschl, G., M. Sivapalan, T. Wagener, A. Viglione, and H. Savenije (2013), *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales*, Cambridge Univ. Press, Cambridge.
Boegh, E., H. Soegaard, N. Hanan, P. Kabat, and L. Lesch (1999), A remote sensing study of the NDVI–Ts relationship and the transpiration from sparse vegetation in the Sahel based on high-resolution satellite data. *Remote Sens. Environ.*, *69*(3), 224–240, doi:10.1016/S0034-4257(99)00025-5.
Caballero, L. A., Z. M. Easton, B. K. Richards, and T. S. Steenhuis (2013), Evaluating the bio-hydrological impact of a cloud forest in Central America using a semi-distributed water balance model. *J. Hydrol. Hydromech.*, *61*, 9, doi:10.2478/jhh-2013-0003.
Cleverly, J. R., C. N. Dahm, J. R. Thibault, D. E. McDonnell, and J. E. Allred Coonrod (2006), Riparian ecohydrology: Regulation of water flux from the ground to the atmosphere in the Middle Rio Grande, New Mexico, *Hydrol. Processes*, *20*(15), 3207–3225, doi:10.1002/hyp.6328.
Coxon, G., J. Freer, I. K. Westerberg, T. Wagener, R. Woods, and P. J. Smith (2015), A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations, *Water Resour. Res.*, *51*, 5531–5546, doi:10.1002/2014WR016532.
De Boer-Euser, T., H. K. McMillan, M. Hrachowitz, H. C. Winsemius, and H. H. G. Savenije (2016), Influence of soil and climate on root zone stotrage capacity, *Water Resour. Res.*, *52*, 2009–2024, doi:10.1002/2015WR018115.
Detty, J., and K. McGuire (2010), Topographic controls on shallow groundwater dynamics: Implications of hydrologic connectivity between hillslopes and riparian zones in a till mantled catchment, *Hydrol. Processes*, *24*(16), 2222–2236.
Donnelly-Makowecki, L. M., and R. D. Moore (1999), Hierarchical testing of three rainfall–runoff models in small forested catchments, *J. Hydrol.*, *219*(3), 136–152.
Euser, T., H. C. Winsemius, M. Hrachowitz, F. Fenicia, S. Uhlenbrook, and H. H. G. Savenije (2013), A framework to assess the realism of model structures using hydrological signatures, *Hydrol. Earth Syst. Sci.*, *17*(5), 1893–1912.
Euser, T., M. Hrachowitz, H. C. Winsemius, and H. H. G. Savenije (2015), The effect of forcing and landscape distribution on performance and consistency of model structures, *Hydrol. Processes.*, *29*(17), 3727–3743.
Farrick K. K., and B. A. Branfireun (2014), Infiltration and soil water dynamics in a tropical dry forest: It may be dry but definitely not arid, *Hydrol. Processes*, *28*(14), 4377–4387, doi:10.1002/hyp.10177.
Fenicia, F., D. Kavetski, and H. H. G. Savenije (2011), Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, *Water Resour. Res.*, *47*, W11510, doi:10.1029/2010WR010174.
Fenicia, F., D. Kavetski, H. H. G. Savenije, M. P. Clark, G. Schoups, L. Pfister, and J. Freer (2014), Catchment properties, function, and conceptual model representation: Is there a correspondence?, *Hydrol. Processes*, *28*(4), 2451–2467.
Fenicia, F., D. Kavetski, H. H. G. Savenije, and L. Pfister (2016), From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, *Water Resour. Res.*, *52*, 954–989, doi:10.1002/2015WR017398.
Flügel, W.-A. (1996), Hydrological Response Units (HRUs) as modeling entities for hydrological river basin simulation and their methodological potential for modeling complex environmental process systems, *Erde*, *127*, 42–62.
Fovet, O., L. Ruiz, M. Hrachowitz, M. Faucheux, and C. Gascuel-Odoux (2015), Hydrological hysteresis and its value for assessing process consistency in catchment conceptual models, *Hydrol. Earth Syst. Sci.*, *19*(1), 105–123.
Freer, J. E., H. McMillan, J. J. McDonnell, and K. J. Beven (2004), Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, *J. Hydrol.*, *291*(3–4), 254–277.
Gao, H., M. Hrachowitz, F. Fenicia, S. Gharari, and H. H. G. Savenije (2014a), Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, *Hydrol. Earth Syst. Sci.*, *18*(5), 1895–1915.
Gao, H., M. Hrachowitz, S. J. Schymanski, F. Fenicia, N. Sriwongsitanon, and H. H. G. Savenije (2014b), Climate controls how ecosystems size the root zone storage capacity at catchment scale, *Geophys. Res. Lett.*, *41*, 7916–7923, doi:10.1002/2014GL061668.
Gentine, P., P. D'Odorico, B. R. Lintner, G. Sivandran, and G. Salvucci (2012), Interdependence of climate, soil, and vegetation as constrained by the Budyko curve, *Geophys. Res. Lett.*, *39*, L19404, doi:10.1029/2012GL053492.
Gharari, S., M. Hrachowitz, F. Fenicia, and H. H. G. Savenije (2011), Hydrological landscape classification: Investigating the performance of HAND based landscape classifications in a central European meso-scale catchment, *Hydrol. Earth Syst. Sci.*, *15*(11), 3275–3291.
Gharari, S., M. Hrachowitz, F. Fenicia, H. Gao, and H. H. G. Savenije (2014), Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, *Hydrol. Earth Syst. Sci.*, *18*(12), 4839–4859.
Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*(4), 751–763.
Gupta, H. V., C. Perrin, G. Blöschl, A. Montanari, R. Kumar, M. Clark, and V. Andréassian (2014), Large-sample hydrology: A need to balance depth with breadth, *Hydrol. Earth Syst. Sci.*, *18*(2), 463–477.
Hargreaves, G. H. (1975), Moisture availability and crop production, *Trans. Am. Soc. Agric. Eng.*, *18*(5), 980–984.
Hartmann, G., and A. Bárdossy (2005), Investigation of the transferability of hydrological models and a method to improve model calibration, *Adv. Geosci.*, *5*, 83–87.
Heuvelmans, G., B. Muys, and J. Feyen (2004), Evaluation of hydrological model parameter transferability for simulating the impact of land use on catchment hydrology, *Phys. Chem. Earth*, *29*(11–12), 739–747.
Hock, R. (2003), Temperature index melt modelling in mountain areas, *J. Hydrol.*, *282*, 104–115.
Hrachowitz, M., R. Bohte, M. L. Mul, T. A. Bogaard, H. H. G. Savenije, and S. Uhlenbrook (2011), On the value of combined event runoff and tracer analysis to improve understanding of catchment functioning in a data-scarce semi-arid area, *Hydrol. Earth Syst. Sci.*, *15*, 2007–2024.
Hrachowitz, M., O. Fovet, L. Ruiz, T. Euser, S. Gharari, R. Nijzink, J. Freer, H. H. G. Savenije, and C. Gascuel-Odoux (2014), Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, *Water Resour. Res.*, *50*, 7445–7469, doi:10.1002/2014WR015484.
Hrachowitz, M., et al. (2013), A decade of predictions in ungauged basins (PUB): A review, *Hydrol. Sci. J.*, *58*(6), 1198–1255.
Hughes, D. A. (2006), Water resources estimation in less developed regions: Issues of uncertainty associated with a lack of data, in *Prediction in Ungauged Basins: Promise Progress*, pp. 72–79, IAHS Publ., CEH Wallingford, Oxfordshire, U. K.
Hundecha, Y., and A. Bárdossy (2004), Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *J. Hydrol.*, *292*(1–4), 281–295.
Kapangaziwiri, E., D. A. Hughes, and T. Wagener (2012), Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa, *Hydrol. Sci. J.*, *57*, 1000–1019, doi:10.1080/02626667.2012.690881.
Kleidon, A., M. Renner, and P. Porada (2014), Estimates of the climatological land surface energy and water balance derived from maximum convective power, *Hydrol. Earth Syst. Sci.*, *18*, 2201–2218.
Klemeš, V. (1986), Dilettantism in hydrology: Transition or destiny?, *Water Resour. Res.*, *22*(9), 177S–188S, doi:10.1029/WR022i09Sp0177S.

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

Laaha, G., and G. Blöschl (2006), A comparison of low flow regionalisation methods: Catchment grouping, *J. Hydrol.*, *323*(1–4), 193–214.

Martinez, G. F., and H. V. Gupta (2011), Hydrologic consistency as a basis for assessing complexity of monthly water balance models for the continental United States, *Water Resour. Res.*, *47*, W12540, doi:10.1029/2011WR011229.

McGlynn, B. L., and J. J. McDonnell (2003), Quantifying the relative contributions of riparian and hillslope zones to catchment runoff, *Water Resour. Res.*, *39*(11), 1310, doi:10.1029/2003WR002091.

McIntyre, N., H. Lee, H. Wheater, A. Young, and T. Wagener (2005), Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, *41*, W12434, doi:10.1029/2003WR002091.

McMillan, H. K., and I. K. Westerberg, I. K. (2015). Rating curve estimation under epistemic uncertainty, *Hydrol. Processes*, *29*(7), 1873–1882.

Merz, R., and G. Blöschl (2004), Regionalisation of catchment model parameters, *J. Hydrol.*, *287*(1–4), 95–123.

Miyazawa, Y., M. Tateishi, H. Komatsu, V. Ma, T. Kajisa, H. Sokh, N. Mizoue, and T. Kumagai (2014), Tropical tree water use under seasonal waterlogging and drought in central Cambodia, *J. Hydrol.*, *515*, 81–89, doi:10.1016/j.jhydrol.2014.04.049.

Mohamed, Y. A., W. G. M. Bastiaanssen, H. H. G. Savenije, B. J. J. M. van den Hurk, and C. M. Finlayson (2012), Wetland versus open water evaporation: An analysis and literature review, *Phys. Chem. Earth*, *47–48*, 114–121.

Molenat, J., C. Gascuel-Odoux, L. Ruiz, and G. Gruau (2008), Role of water table dynamics on stream nitrate export and concentration in agricultural headwater catchment (France), *J. Hydrol.*, *348*(3–4), 363–378.

Montanari, L., M. Sivapalan, and A. Montanari (2006), Analysis of process controls on streamflow response in an Australian tropical catchment, in *Predictions in Ungauged Basins: Promise and Progress*, pp. 38–46, IAHS Publ.

Nijzink, R. C., L. Samaniego, J. Mai, R. Kumar, S. Thober, M. Zink, D. Schäfer, H. H. G. Savenije, and M. Hrachowitz (2016), The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models, *Hydrol. Earth Syst. Sci.*, *20*(3), 1151–1176, doi:10.5194/hess-20-1151-2016.

Nobre, A., L. Cuartas, M. Hodnett, C. Rennó, G. Rodrigues, A. Silveira, M. Waterloo, and S. Saleska (2011), Height above the nearest drainage: A hydrologically relevant new terrain model, *J. Hydrol.*, *404*(1), 13–29.

Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, and C. Loumagne (2005), Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2: Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, *J. Hydrol.*, *303*(1), 290–306.

Parajka, J., R. Merz, and G. Blöschl (2005), A comparison of regionalisation methods for catchment model parameters, *Hydrol. Earth Syst. Sci.*, *9*(3), 157–171.

Parajka, J., A. Viglione, M. Rogger, J. L. Salinas, M. Sivapalan, and G. Blöschl (2013), Comparative assessment of predictions in ungauged basins-Part 1: Runoff-hydrograph studies, *Hydrol. Earth Syst. Sci.*, *17*(5), 1783–1795, doi:10.5194/hess-17-1783-2013.

Peel, M. C., B. L. Finlayson, and T. A. McMahon (2007), Updated world map of the Köppen-Geiger climate classification, *Hydrol. Earth Syst. Sci.*, *11*(5), 1633–1644, doi:10.5194/hess-11-1633-2007.

Perrin, C., L. Oudin, V. Andreassian, C. Rojas-Serna, C. Michel, and T. Mathevet (2007), Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models, *Hydrol. Sci. J.*, *52*, 131–151, doi:10.1623/hysj.52.1.131.

Petheram, C., P. Rustomji, F. H. S. Chiew, and J. Vleeshouwer (2012), Rainfall–runoff modelling in northern Australia: A guide to modelling strategies in the tropics, *J. Hydrol.*, *462–463*, 28–41, doi:10.1016/j.jhydrol.2011.12.046.

Pfister, L. (2006), *Study of the water cycle components in the Attert River Basin (CYCLEAU), report*, Fonds Natl. de la Rech., Luxembourg. [Available at: http://onlinelibrary.wiley.com/doi/10.1029/2010WR009525/epdf.]

Pitman, W. V. (1973), *A Mathematical Model for Generating Monthly River Flows From Meteorological Data in Southern Africa*, Univ. of Witwatersrand, Johannesburg, South Africa.

Refsgaard, J. C., et al. (2014), A framework for testing the ability of models to project climate change and its impacts, *Clim. Change*, *122*(1–2), 271–282.

Reggiani, P., M. Sivapalan, and S. M. Hassanizadeh (2000), Conservation equations governing hillslope responses: Exploring the physical basis of water balance, *Water Resour. Res.*, *36*(7), 1845–1863.

Rennó, C., A. Nobre, L. Cuartas, J. Soares, M. Hodnett, J. Tomasella, and M. Waterloo (2008), HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia, *Remote Sens. Environ.*, *112*(9), 3469–3481.

Rodriguez-Iturbe, I. (2000), Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamics, *Water Resour. Res.*, *36*(1), 3–9, doi:10.1029/1999WR900210.

Samaniego, L., R. Kumar, and S. Attinger (2010), Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resour. Res.*, *46*, W05523, doi:10.1029/2008WR007327.

Savenije, H. H. G. (2010), HESS opinions ''Topography driven conceptual modelling (FLEX-Topo)'', *Hydrol. Earth Syst. Sci.*, *14*(12), 2681–2692.

Shao, J. (1993), Linear model selection by cross-validation, *J. Am. Stat. Assoc.*, *88*(422), 486–494.

Sivapalan, M., and G. Blöschl (2015), Time scale interactions and the coevolution of humans and water, *Water Resour. Res.*, *51*, 6988–7022, doi:10.1002/2015WR017896.

Sivapalan, M., and J. D. Kalma (1995), Scale problems in hydrology: Contributions of the robertson workshop, *Hydrol. Processes*, *9*(3-4), 243–250.

Sivapalan, M., et al. (2003), IAHS decade on predictions in ungauged basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci.*, *48*(6), 857–880.

Sriwongsitanon, N., and W. Taesombat (2011), Effects of land cover on runoff coefficient, *J. Hydrol.*, *410*(3–4), 226–238.

Taesombat, W., and N. Sriwongsitanon (2009), Areal rainfall estimation using spatial interpolation techniques, *ScienceAsia*, *35*, 268–275.

Troch, P. A., G. F. Martinez, V. R. N. Pauwels, M. Durcik, M. Sivapalan, C. Harman, P. D. Brooks, H. Gupta, and T. Huxman (2009), Climate and vegetation water use efficiency at catchment scales, *Hydrol. Processes*, *23* (16), 2409–2414, doi:10.1002/hyp.7358.

Troch, P. A., G. Carrillo, M. Sivapalan, T. Wagener, and K. Sawicz (2013), Climate-vegetation-soil interactions and long-term hydrologic partitioning: Signatures of catchment co-evolution, *Hydrol. Earth Syst. Sci.*, *17*(6), 2209–2217.

Uhlenbrook, S., S. Roser, and N. Tilch (2004), Hydrological process representation at the meso-scale: The potential of a distributed, conceptual catchment model, *J. Hydrol.*, *291*(3), 278–296.

Uhlenbrook, S., Y. Mohamed, and A. S. Gragne (2010), Analyzing catchment behavior through catchment modeling in the Gilgel Abay, Upper Blue Nile River Basin, Ethiopia, *Hydrol. Earth Syst. Sci.*, *14*(10), 2153–2165.

Viglione, A., J. Parajka, M. Rogger, J. L. Salinas, G. Laaha, M. Sivapalan, and G. Blöschl (2013), Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria, *Hydrol. Earth Syst. Sci.*, *17*(6), 2263–2279.

Visessri, S. (2014), Flow prediction in data scarce catchments: A case study of Northern Thailand, PhD dissertation, Dept. of Civil and Environmental Engineering, Imp. Coll. London, London.

Visessri, S., and N. McIntyre (2015), Regionalisation of hydrological responses under land use change and variable data quality, *Hydrol. Sci. J.*, *61*(2), 302–320, doi:10.1080/02626667.2015.1006226.
Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.
Wohl, E., et al. (2012), The hydrology of the humid tropics, *Nat. Clim. Change*, *2*(9), 655–662, doi:10.1038/nclimate1556.
Wooldridge, S. A., J. D. Kalma, S. W. Franks, and G. Kuczera (2002), Model identification by space–time disaggregation: A case study from eastern Australia, *Hydrol. Processes*, *16*(2), 459–477, doi:10.1002/hyp.329.
Yu, Z., T. N. Carlson, E. J. Barron, and F. W. Schwartz (2001), On evaluating the spatial-temporal variation of soil moisture in the Susquehanna River Basin, *Water Resour. Res.*, *37*(5), 1313–1326.
Yu, Z., H. Lü, Y. Zhu, S. Drake, and C. Liang (2010), Long-term effects of revegetation on soil hydrological processes in vegetation-stabilized desert ecosystems, *Hydrol. Processes*, *24*(1), 87–95.
Zhao, R.-J. (1984), *Watershed Hydrological Modeling: Xinanjiang Model and Shanbei Model,1984*.[in Chinese], Waterpower Press, Beijing.