Using Satellite-Based Evapotranspiration Estimates to Improve the Structure of a Simple Conceptual Rainfall-Runoff Model

Tirthankar Roy[1], Hoshin V. Gupta[1], Aleix Serrat-Capdevila[1,2] and Juan B. Valdes[1]

[1] Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ, USA
[2] Water Global Practice, The World Bank, Washington DC, USA

Correspondence to: Tirthankar Roy (royt@email.arizona.edu)

Abstract. Daily, quasi-global (50°N-S and 180°W-E), satellite-based estimates of actual evapotranspiration at 0.25° spatial resolution have recently become available, generated by the Global Land Evaporation Amsterdam Model (GLEAM). We investigate the use of these data to improve the performance of a simple lumped catchment-scale hydrologic model driven by satellite-based precipitation estimates to generate streamflow simulations for a poorly gauged basin in Africa. In one approach, we use GLEAM to constrain the evapotranspiration estimates generated by the model, thereby modifying daily water balance and improving model performance. In an alternative approach, we instead change the structure of the model to improve its ability to simulate actual evapotranspiration (as estimated by GLEAM). Finally, we test whether the GLEAM product is able to further improve the performance of the structurally modified model. Results indicate that while both approaches can provide improved simulations of streamflow, the second approach significantly improves the simulation of actual evapotranspiration, which substantiates the importance of making '*diagnostic structural improvements*' to hydrologic models whenever possible.

1 Introduction

1.1 Statement of the Problem

As a primary mechanism in the surface-to-atmosphere portion of the water cycle, evapotranspiration (ET) plays a crucial role in the water and energy budgets of a hydrologic system. In practice, ET can be estimated either from model simulations or from remotely sensed observations. For example, ET can be estimated as a residual of water balance computations, or via a land-surface energy budget (e.g. Monteith, 1965; Priestley and Taylor, 1972), and simple empirical physically based schemes (Hargreaves and Samani, 1985) can be applied in data-scarce regions. Ultimately, the quality of a model-derived estimate of ET depends on the various sources of uncertainty (inputs, parameters, process representation, structure, etc.) inherent to the model-based scheme used, and common problems include both over- and under-estimation of evaporative fluxes (Trambauer et al., 2014). Recently, methods that use satellite-based remotely sensed climatic and environmental observations provide an alternate approach to the estimation of ET and its different components (e.g. Bastiaanssen et al., 1998; Arboleda et al., 2005).

Several studies have advocated and/or implemented the idea of using physically consistent estimates for the parameters of hydrologic models (Pokhrel et al., 2008, 2012; Savenije, 2010; Schaefli et al., 2011; Kumar et al., 2013; Troch et al., 2015 and references therein). However, in catchment-scale modeling, it is common practice to use parameter estimates that are calibrated by adjusting the simulated streamflows to try and match observed data. If due care is not implemented during the calibration strategy, this approach can result in conceptually unrealistic estimates for the parameters. Such a result defeats an important purpose of using conceptual/physically based models (as opposed to empirical data-based models), which is to help us better understand the dynamical behavior of the system.

In principle, the potential of such models can be better realized by incorporating more information about the physical system during model development. Such information can take various forms and be incorporated in different ways. Evapotranspiration (ET) can be used to constrain model parameters that are sensitive to the ET process (Winsemius et al., 2008; van Emmerik et al., 2015). Alternatively, ET can be used as a calibration target along with streamflow within a multi-objective setting (Zhang et al., 2009). There has also been a recent drive towards structurally flexible models that are able to both better characterize the uncertainty associated with model structure and use additional information to help reduce such uncertainty (Wagener et al., 2001; Marshall et al., 2006; Clark et al., 2008; Savenije, 2010; Schaefli et al., 2011; Fenicia et al., 2008a, 2008b, 2011; Bulygina and Gupta, 2009, 2010, 2011; Martinez and Gupta, 2011; Nearing, 2013; Nearing and Gupta, 2015; Clark et al., 2015).

A variety of satellite-based remotely sensed estimates of daily precipitation have been available for some time (e.g. Hsu et al., 1997; Joyce et al., 2004; Huffman et al., 2007; Funk et al., 2014), making it possible to consider the model-based generation of streamflow simulations for ungaged locations. Recently, satellite-based remotely sensed estimates of daily ET have become available, based on a variety of different retrieval algorithms of varying complexity (e.g. Bastiaanssen et al., 1998; Arboleda et al., 2005; Miralles et al., 2011a). Worldwide evaluations suggest that satellite-based ET estimates are strongly correlated (~0.83) with ground-based observations made at flux towers (Demaria and Serrat-Capdevila, 2016).

For this study, we use the Global Land Evaporation Amsterdam Model (GLEAM) as the source of the satellite-based ET (SET) data. In the GLEAM algorithm, ET is computed using only a small number of satellite-based inputs, which makes it particularly beneficial for application to sparsely gauged basins. Miralles et al. (2011) have shown that GLEAM estimates of evaporation are strongly correlated (0.80) with annual cumulative evaporation estimated via eddy covariance at 43 stations, and have very low (-5%) average bias. The correlations at individual stations are strong (0.83) for all vegetation and climate conditions, and improve to 0.9 for monthly time series (Miralles et al., 2011). McCabe et al. (2016) reported satisfactory statistical performance ($R^2 = 0.68$; Root Mean Square Difference = 64 Wm^{-2}; Nash-Sutcliffe Efficiency = 0.62) of GLEAM when compared against data from 45 globally-distributed eddy-covariance stations. Michel et al. (2016) compared Priestley-Taylor Jet Propulsion Laboratory model (PT-JPL), Moderate Resolution Imaging Spectroradiometer evaporation product (PM-MOD), Surface Energy Balance System (SEBS), and GLEAM simulations against 22 FLUXNET tower-based flux observations and found GLEAM and PT-JPL to more closely match in-situ observations for the selected towers and reference period (2005-2007). Their extended analysis over 85 towers also had a similar overall outcome. Miralles et al. (2016)

compared three process-based ET methods (PM-MOD, GLEAM and PT-JPL) against surface water balance from 837 globally distributed catchments, and reported that GLEAM and PT-JPL provide more realistic estimates of ET. They found these two products to provide superior overall performance for most ecosystem and climate regimes, whereas PM-MOD tends to underestimate the flux in tropics and subtropics.

5   While previous studies have used SET estimates to *constrain the parameters* of hydrologic models (Winsemius et al., 2008; van Emmerik et al., 2015), the recent interest in diagnostic improvements to model structure (Gupta et al., 2008, 2012; Gupta and Nearing, 2014) suggests that it would be potentially more valuable to use the ET data to actually *improve the model structure* when possible. This study attempts to explore this possibility in the context of using satellite-based data to drive a streamflow simulation model for a poorly gauged basin in Africa.

10   **1.2 Objectives and Scope**

In this study, we explore the use of the GLEAM daily SET product (Miralles et al., 2011; Martens et al., 2016) to improve the performance of a simple lumped catchment scale hydrologic model driven by satellite-based precipitation estimates to generate streamflow simulations for a poorly gauged basin in Africa. We first use the GLEAM product to constrain the evapotranspiration estimates generated by the model, thereby improving the daily water balance. Next, we instead
15   change the structure of the model to make it more physically consistent and improve its ability to simulate actual evapotranspiration (as estimated by GLEAM). Finally, we test whether the use of GLEAM SET can further improve the performance of the structurally modified model, and whether there is any decline in model performance if GLEAM SET data become unavailable.

2 Study Area, Data and Methodology

20   **2.1 Study Area**

This study is carried out for the Nyangores River basin, which is a sub-basin of the Mara River basin flowing through Kenya (Fig. 1). The Nyangores River basin has an aerial coverage of 697 km^2 and is located at the northeastern side of the Mara River Basin (Location: 33°88'E 35°90'E 0°28'S 1°97'S). The perennial Nyangores River originates from the Mau Escarpment (3000 m ASL) fault scarp passing through the western side of the Great Rift Valley in Kenya. It then merges with
25   the Amala River at the Napuiyapi swamp (2932 m ASL) to form the Mara River, which flows all the way to Lake Victoria at Musoma Bay, Tanzania (1130 m ASL). Mara River basin (or Nyangores River basin) has two wet seasons consequent to the yearly oscillations of the inter-tropical convergence zone (ITCZ), the primary wet season occurring during March to May and the secondary during October to December. The long-term mean rainfall in the Mau Escarpment is around 1500 mm. The rainfall in the basin is influenced by factors like topography, elevation gradient, regional influence of Lake Victoria, sea-
30   surface temperature (SST) of the Indian Ocean, etc.
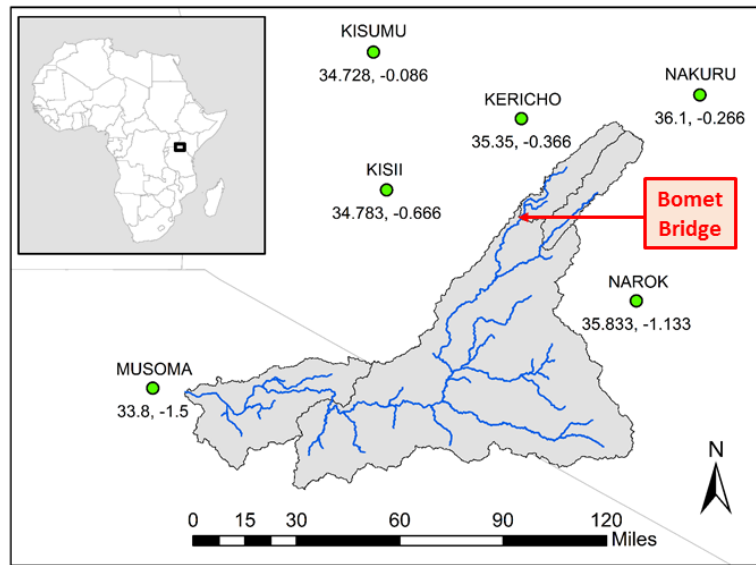
Fig. 1. The Mara River basin and the Nyangores River sub-basin (Roy et al., 2017). The discharge station is located at Bomet Bridge (red dot). Meteorological stations (green dots) are located in the surrounding regions.

2.2 Data

2.2.1 Estimates of Actual Evapotranspiration

The source of the SET data used in this study is the Global Land Evaporation Amsterdam Model (GLEAM) Version 3.0. GLEAM comprises a set of algorithms that use remotely sensed climatic and environmental observations to estimate various components of ET. Satellite-based observations of surface net radiation and near-surface air temperature are processed via the Priestley-Taylor Equation (Priestley and Taylor, 1972) to calculate Potential Evapotranspiration (PET), which is then converted to Actual Evapotranspiration (AET) by incorporating an evaporative stress factor obtained from microwave observations of vegetation optical depth (as a proxy for vegetation water content) and root-zone soil moisture (simulations). Interception loss is calculated using the Gash analytical model (Gash, 1979).

Three different versions of the GLEAM datasets are currently available, depending on the satellite observations used. The version used in this study (GLEAM_v3.0b) is based solely on satellite observations, is quasi-global (50°N-S, 180°W-E), has a spatial resolution of 0.25°, and has a daily temporal coverage of 13 years (2003 to 2015).

Fig. 2 shows the annual mean of GLEAM AET (GAET) over the entire Mara River basin. As can be seen, GAET increases towards the western side of the basin. The annual average GAET varies between 900 mm/year to 1200 mm/year. We computed corresponding estimates of PET via the Hargreaves Equation (HPET) using temperature data collected from the six met stations surrounding the Mara River basin (Fig. 1); maximum and minimum temperatures were averaged to obtain time series of maximum and minimum temperatures and processed through the Hargreaves Equation to calculate PET. For a small number (~0.6%) of days, the lumped GAET values were found to be larger than the lumped HPET values; for these few anomalous values, HPET was replaced by GAET. Fig. 3 shows the time series of HPET and GAET for the Nyangores River basin.

4
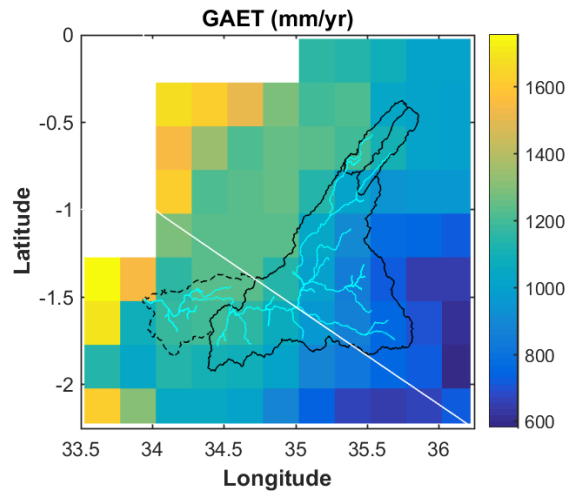
Fig. 2. Annual mean of GAET over the entire Mara River basin.
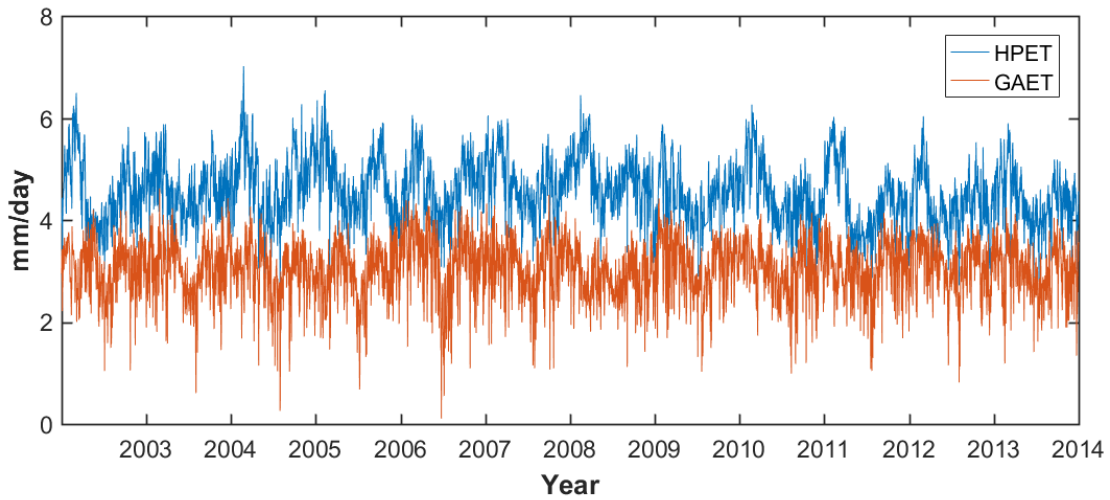
Fig. 3. Time series of HPET and GAET for the Nyangores River basin.

2.2.2 Estimates of Precipitation

The Real Time Multi-satellite Precipitation Analysis of the NASA Tropical Rainfall Measuring Mission (TMPA-RT) combines information from multiple satellites to produce a quasi-global (50°N-S, 180°W-E), near-real-time (March 1, 2000 to near-present) precipitation product at $0.25° \times 0.25°$ spatial and 3-hourly temporal resolution (this product is the real-time version of TMPA (Huffman et al., 2007)). Until it was shut down on 8 April 2015 due to fuel deficiency and battery issues in the satellite, TMPA used to include the TRMM Microwave Imager (TMI) products. TRMM was the first satellite dedicated for precipitation studies. The after-real-time TMPA product also incorporates rain gauge information wherever feasible. In this study, we aggregated the 3-hourly TMPA-RT data to daily level, resampled from the coarse resolution ($0.25° \times 0.25°$) to a resolution of $0.05° \times 0.05°$, and implemented a bias correction using the "Climate Hazards Group InfraRed Precipitation with Station" product (CHIRPS; Funk et al., 2014, 2015) and rain gauge measurements (details not included).

5

2.2.3 Estimates of Streamflow

Streamflow data were computed using the calibrated stage-discharge relationship for the Bomet Bridge discharge station (Station ID: 1LA03; Location: 0°47′23.50″S 35°20′47.45″E) on the Nyangores River (drainage area approximately 697 km^2), which is one of the two main tributaries of the Mara River. Data is available for the period Jan 1, 1996 to Jun 30, 2010, during which time only about ~8% of the records are missing.

2.2.4 Estimates of Temperature

We computed PET using the Hargreaves Equation (Hargreaves and Samani, 1985), the annual mean of which closely matched the reported PET value for the study area (WREM, 2008). The temperature data used in the Hargreaves Equation were extracted from the Global Surface Summary of the Day (GSOD) product produced by the National Climatic Data Center (NCDC) in Asheville, NC. The daily temperature data includes multiple observations and are available in three forms: maximum, minimum, and average.

2.3 The Hydrologic Model HyMod-V2

The spatially lumped HyMod Version 1 (HyMod-V1) conceptual rainfall-runoff model with five/six parameters has previously been used in several studies (Boyle et al., 2000; Vrugt et al., 2003, 2009; Moradkhani et al., 2005; Duan et al., 2007; Wang et al., 2009; Razavi and Gupta, 2016). The model is driven using daily precipitation and PET data to generate daily estimates of AET (HAET: HyMod-generated AET) and streamflow. Nonlinear vertical flow processes are controlled by a two-parameter soil moisture accounting module based on the rainfall excess model proposed by Moore (1985). Horizontal flows are simulated in a linear routing module that includes a Nash Cascade for quick-flow routing (overland flow) and a linear reservoir for slow-flow routing (baseflow).

In this study, we improved the ET process parameterization within the model by relying on satellite-based AET estimates provided by GLEAM. In this regard, we were careful to ensure that the structural modifications (1) do not over-complicate the model since that defeats the whole purpose of having simpler models such as HyMod, (2) do not require large number of additional model parameters, (3) are more physically consistent, (4) consistently produce improved ET simulations, and finally (5) do not deteriorate the streamflow simulations. We refer to the structurally modified version of the model as HyMod Version 2 (HyMod-V2), as shown in Fig. 4. As can be seen, a new ET resistance module is added to the soil moisture accounting module of the model. A detailed description of the ET resistance module is provided in Section 2.4. Details of the overall model structure and process equations are presented in Appendix A.
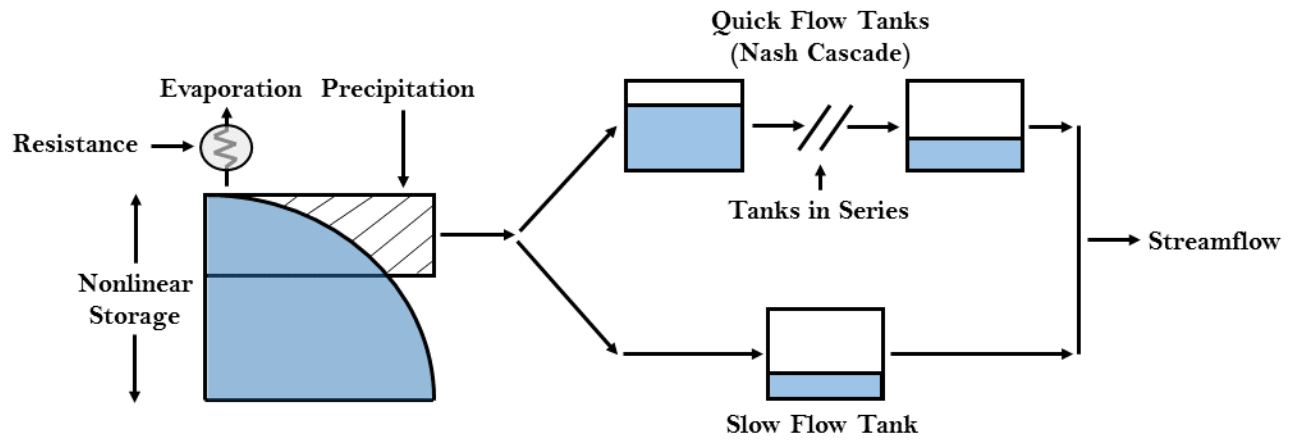
Fig. 4. Schematic diagram of HyMod-V2.

2.4 Study Approach

We conducted the investigation in two stages. The first stage consists of five steps designed to improve model performance with respect to both streamflow and evapotranspiration (as assessed against data), but *without* making structural modifications to the model. The strategy includes using GAET to *constrain* simulated evapotranspiration, recalibration of model parameters, and constraint adjustments. In doing so, we specifically do *not* directly assimilate GAET into the model (either by a Bayesian data assimilation "nudging" procedure such as the Ensemble Kalman Filter, or by direct insertion), so that the model's representation of overall water balance is not compromised. Accordingly, while we *are* extracting information from the GLEAM product, we do so via a process of "*constraining*" rather than "*assimilation*". In the second stage, we modify the structure of the model to directly improve its ability to simulate ET (using GAET as the target) by trying to capture the physics of the underlying processes more accurately. The steps followed in stage one are repeated so that results of the different strategies can be compared.

Conceptually, the main difference between Stage-I and Stage-II is that, in the former, the information provided by GAET is used only to constrain the evapotranspiration fluxes of the (recalibrated) model, whereas in the latter the information contained in GAET is used to alter the model structure. While the former provides a temporary improvement to model performance, achieved as long as GLEAM data are available, the latter is expected to provide a lasting improvement to model performance that should persist even when GLEAM data are not available. In the final step of Stage II, we check to see whether the GAET product contains residual information that, not having yet been used to improve the model structure, remains useful for improving model performance via the constraining operation.

The entire study approach is summarized in Fig. 5. As can be seen, only Step-1 is different for both the stages (Stage-I and Stage-II), while the remaining four steps (Step 2-5) are similar. Thus, in each stage, there are five steps altogether. Stage-I Step-1 is for generating benchmark simulations using the calibrated model but without any ET constraint or structural modifications. On the other hand, Stage-II Step-I has four different cases (A-D) corresponding to different structural modifications in the ET process parameterization. Both the benchmark model from Stage-I Step-I and the best performing

model from Stage-II Step-I are used in the following steps. Step-2 is based on imposing the ET constraint but without recalibration, meaning that the same set of calibrated parameters as in the benchmark step is further used in this step. Step-3 is based on recalibrating the model while imposing the ET constraint. Step-4 is conceptually similar to Step-3, however, additionally, some constraint adjustments (Eq. 1 and 2) are applied and the adjustment parameters are calibrated together with model parameters (to match the simulated and observed streamflows). Finally, in Step-5, we remove the ET constraint to see whether the performance of the new model will decline when satellite ET data becomes unavailable (note this is no longer the benchmark model since we recalibrated the parameters in Step 4).

$$y = f(x, \cdots) \text{ and } x = \boldsymbol{a} \cdot \text{GAET} \tag{1}$$

$$y = f(x, \cdots) \text{ and } x = \boldsymbol{a} \cdot \text{GAET}^{\boldsymbol{b}} \tag{2}$$

In Eq. 1 and 2, $\boldsymbol{y}$ represents the streamflows after the adjustment of GAET constraint, and $\boldsymbol{a}$ and $\boldsymbol{b}$ are the parameters of the adjustment formulations ($\boldsymbol{a}$ controls the variance and $\boldsymbol{b}$ controls the degree of non-linearity).
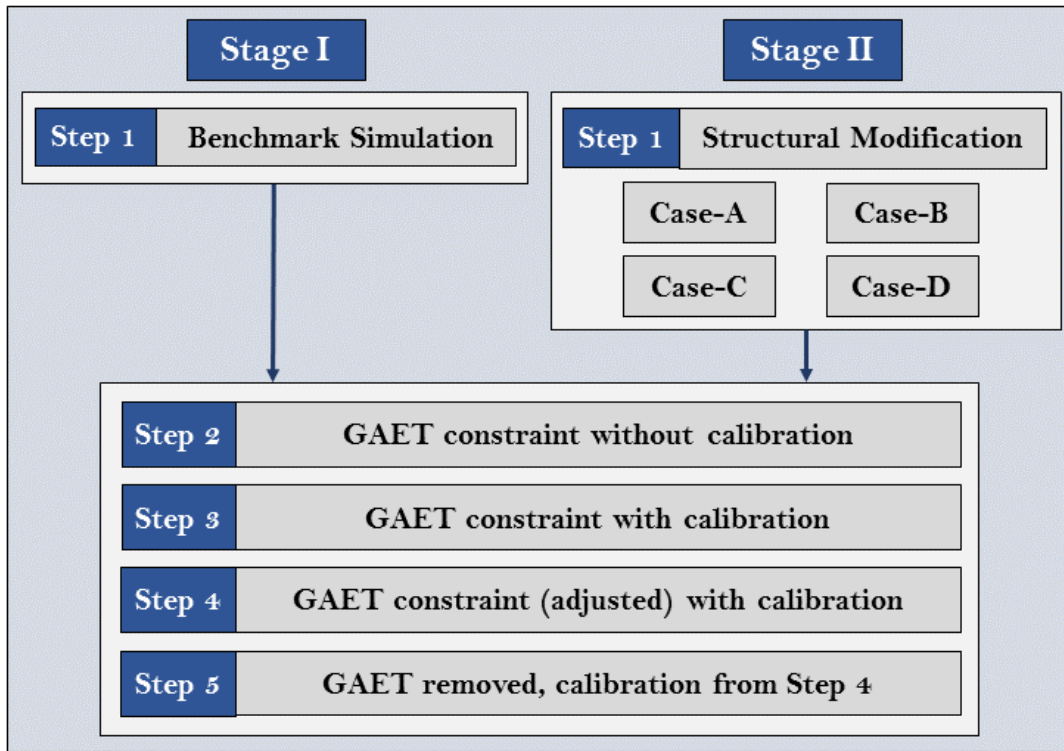
Fig. 5. The approach followed in this study.

ET Constraining

The ET constraint was imposed by modifying the original ET equation of the model (Eq. A5) from $\text{HAET} = \min\{\text{PET}, \text{C}\}$ to the new form $\text{HAET} = \min\{\text{PET}, \text{GAET}, \text{C}\}$. Note that this is <u>not</u> a structural modification to the form of the process equation, rather GAET sets the upper limit of HAET in this case, which is more realistic than using PET directly as the upper limit.

Structural Modifications

The ET process parameterization within the original model (HyMod-V1) is modified in Stage-II Step-1 to improve its ability to reproduce GAET more accurately without deteriorating the streamflow simulations. Four ET equations of progressive complexity and physical basis are tested. In each case, the model parameters are recalibrated to match the simulated streamflows to the observed data. The final result is a structurally modified model called HyMod-V2.

More specifically, the ET equation of HyMod-V1 is multiplied by a function $K(\cdot)$ such that $0 \leq K(\cdot) \leq 1$. This function acts as a resistance to the ET flux of the model. Four different forms for $K(\cdot)$ that represent incremental increases in complexity and physical basis (Table 1) are tested. Writing the main ET equation in the general form:

$$Y_t = K_t \cdot X_t \cdot EDR_t \tag{3}$$

where Y_t is the AET generated by HyMod ($HAET_t$), X_t is the soil moisture storage (C_{SMA_t}), and EDR_t is the *evaporation demand ratio* computed as $\min\left\{1, \frac{PET_t}{X_t}\right\}$. The most general form for K_t is given by:

$$K_t = K_{min} + [K_{max} - K_{min}] \cdot f(\psi_t) \tag{4}$$

where, K_{min} and K_{max} are lower and the upper limits for K, and ψ_t is the ratio of actual to maximum storage capacity ($\psi_t = X_t/X_{max}$).

Table 1: K-function in different cases.

Cases	K_{max}	K_{min}	$f(\psi_t)$	Additional Parameters
A	1	0	1	None
B	K_0	0	1	K_0
C	K_0	0	X_t/X_{max}	K_0
D	K_{max}	$\gamma \cdot K_{max}$	$(X_t/X_{max})^{BE}$	K_{max}, γ, BE

2.5 Calibration Methodology and Benchmark Model Calibration

Calibration of the model (and adjustment) parameters was performed using the SCE-UA algorithm (Duan et al., 1992). In all cases, the calibration runs were carried out using 10 complexes and 25 loops. Model simulated streamflows were matched against the observed streamflows in the λ-transformed space to minimize the effects of skewness and reduce heteroscedasticity. The λ-transformation was applied after modifying the original equation as proposed by Box and Cox, (1964) (see Appendix B), and the value of the λ parameter was calculated from the observed streamflow records. The performance criterion used was the Mean Squared Error (MSE) of transformed flows. The model was run continuously for the 7.5-year period Jan 2003 to June 2010, with the first 4 years (2003 to 2006) used for calibration and the remaining 3.5 years

(2007 to mid-2010) used to provide an additional assessment of model performance. Results are shown for the "*calibration (4-years)*", "*evaluation (3.5 years)*" and "*total (7.5 years)*" simulation periods.

2.6 Metrics Used for Performance Evaluation

Four metrics are used in this study to assess the model performance (Table 2). These metrics measure performance in regards to overall mean squared error, bias, variability, and correlation (see Gupta et al., 2009), are computed in the transformed space where applicable (e.g. for streamflows), and are normalized to be comparable.

Table 2: Performance evaluation metrics used in this study.

Metrics	Equations
Normalized Mean Square Error (NMSE)	$MSE = mean((O_i - S_i)^2); \ NMSE = \frac{MSE}{var(O)}$
Normalized Bias in Mean (NBμ)	$NB\mu = \dfrac{mean(S) - mean(O)}{mean(O)}$
Normalized Bias in Standard Deviation (NBσ)	$NB\sigma = \dfrac{std(S) - std(O)}{std(O)}$
Correlation Coefficient (ρ)	$\rho = \dfrac{\sum_{i=1}^{N}(O_i - mean(O))(S_i - mean(S))}{\sqrt{\sum_{i=1}^{N}(O_i - mean(O))^2 \sum_{i=1}^{N}(S_i - mean(S))^2}}$
O: Observed flows; S: Simulated flows; N: Number of data points.	

3 Results

3.1 Results from Stage-I (ET Constraints)

The performance of the benchmark model HyMod-V1, driven using TMPA-RT satellite-based precipitation and with parameters calibrated to match simulated streamflow to observed data, is reported in Table 5. The NMSE varies between 0.56 (calibration period) and 0.84 (evaluation period), where NMSE = 0.56 means that on average only about 44% (1.0 - 0.56 = 0.44) of the variability in the flows has been explained. This is not surprising given the use of a simple lumped conceptual model driven by satellite-based estimates of precipitation for a poorly gauged basin. The flow biases are small (NBμ < 15%) indicating that long-term water balance is approximately preserved. The calibrated values of the model parameters are reported in Appendix C.

Table 3 presents a comparison of the model generated HAET with the GAET data (for the total 7.5 year simulation period). HAET tends to be larger on average, varies over a wider range, is considerably more skewed, and is less kurtotic. Some of the reasons for this can be understood from the time-series plot and scatterplot shown in Fig. 6. The behavior of HAET tends to be more erratic and, although both HAET and GAET show seasonal patterns, the former regularly drops to zero or near zero (explained by the very simple, threshold-like, ET process representation in the model, which does not contain a resistance term). The result is that HAET and GAET are not well correlated (Fig. 6b) and have different shapes for their

10

empirical probability distributions (Fig. 7). Even if we were to ignore the time-steps when HAET drops closer to zero, HAET is strongly positively biased (too large), which results from trying to satisfy the potential evapotranspiration (PET).

Table 4 reports a water balance estimate WBAET of the mean annual AET for the basin, obtained by subtracting mean annual streamflow (at the discharge station) from mean annual precipitation (estimated from TMPA-RT). WBAET is similar in magnitude to GAET, and we have GAET < WBAET < HAET < HPET, indicating that the AET computed by the model tends to be a little high.

Table 3: Descriptive statistics of GAET and HAET.

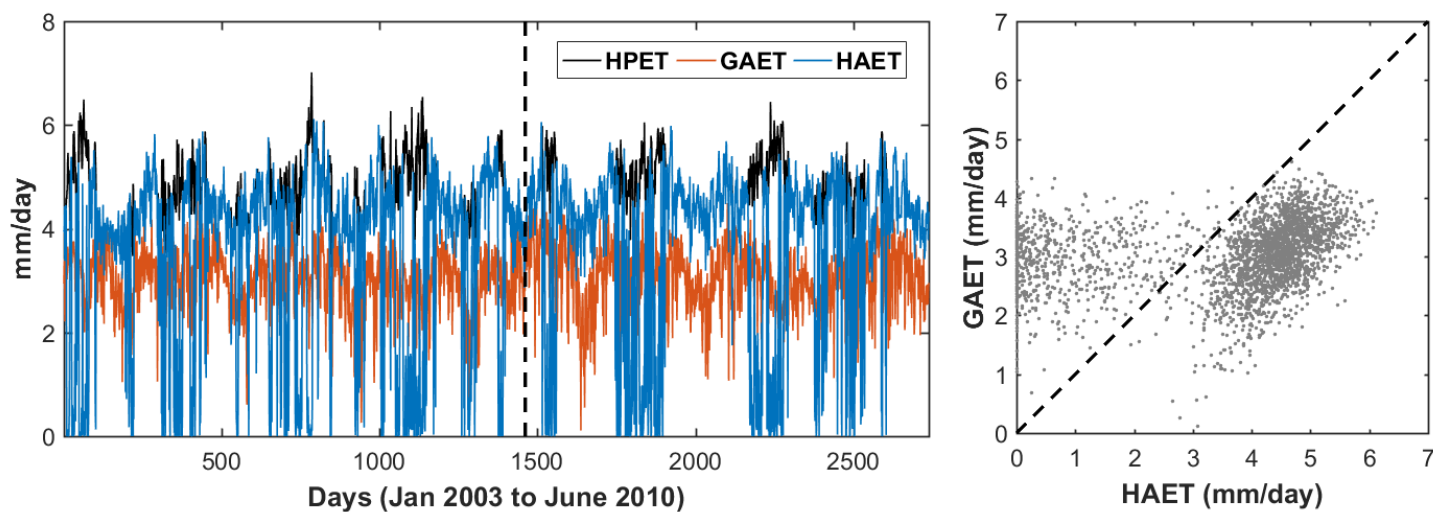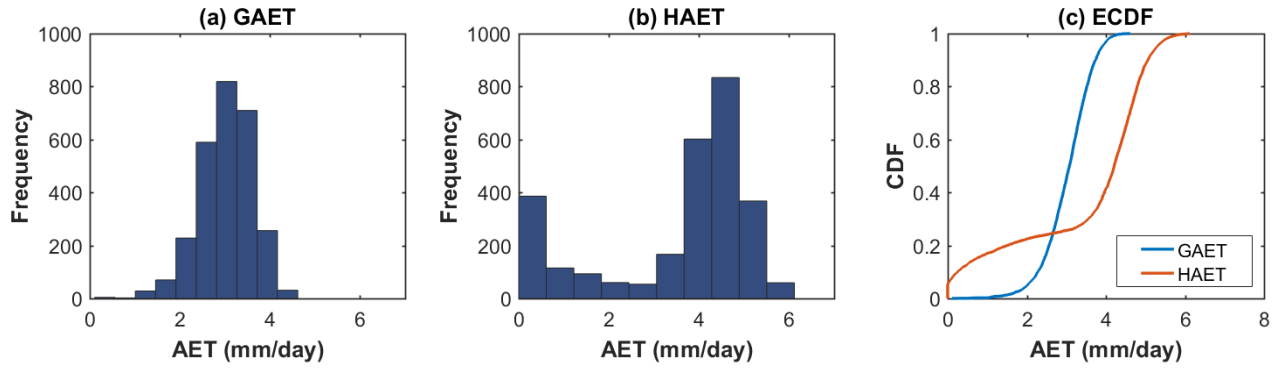| Statistics	GAET	HAET
Maximum	4.62	6.12
Minimum	0.119	0.00
Mean	3.03	3.52
Median	3.08	4.21
Mode	0.11	0.00
Std. Dev.	0.59	1.72
Skewness	-0.58	-1.03
Kurtosis	3.84	2.62

Fig. 6. Time series and scatter plots of HPET, GAET, and HAET.

11

Fig. 7. Histogram and ECDF plots of GAET and HAET.

Table 4: Annual mean of AETs and HPET.

Source	Annual Mean (mm)
GAET	1100
HAET	1263
WBAET	1146
HPET	1704

The benchmark model was constrained using the GAET estimates in Step-2 but the model parameters were not recalibrated. As can be seen in Table 5, the model performance has become significantly worse due to streamflow becoming positively biased. Given that GAET < HAET, on average in the previous step, this makes sense because imposing GAET as a constraint alters the water balance of the model. To try and fix this water-balance problem, we recalibrated the parameters of the model to improve the match to observed streamflows while continuing to use GAET to constrain HAET in the model (Step-3). Although the large positive bias was reduced (Table 5) and the NMSE statistic was improved as compared to Step-2, most of the error statistics deteriorated for all three periods (calibration, evaluation, and total simulation) as compared to Step-1. Importantly, this calibration step resulted in an unrealistically large value of 17.36 meters for the size of the soil moisture storage (previously a more realistic 0.76 meters). This value is clearly conceptually and physically inconsistent (the realistic range is about zero to 2 meters), and while it improved calibration period performance, the lack of consistency is reflected in the sharp deterioration in performance during evaluation. Unrealistic parameter values, such as this, are indications of either severe errors in the data, or structural errors in the model. Since GAET agrees well with WBAET on average, it is likely that the major cause here is model structural inadequacy (Gupta et al., 2012). For completeness, we next checked (Step-4) to see whether this problem could instead be resolved by implementing an empirical adjustment to GAET.

We tested two empirical constraint adjustment schemes (Eq. 1 and 2) applied to the GAET data, and calibrated the additional parameters (from these equations) along with HyMod-V1 parameters. Results for both schemes were similar, but Eq. 2 provided slightly better performance for the evaluation and total simulation periods and so we selected Eq. 2. Compared to the benchmark (Step-1), NMSE and NBµ calibration period statistics reduced from 0.56 to 0.43 and 12% to 6%, respectively (Table 5) while ρ increased from (0.76/0.66/0.72; Cal/Eval/Tot) to (0.83/0.74/0.78). Perhaps more important, the calibrated

12

value of parameter H is now 0.65 meters, which is within the conceptually acceptable range. Finally, the model obtained in Step-4 was run without the use of GAET to see how well the model would perform if GAET data were to become unavailable. The results (Table 5) results indicate that model performance does not deteriorate significantly when GAET data become unavailable and, in some cases, the performance is better than the benchmark.

5

Table 5: Streamflow error statistics for calibration, evaluation, and total simulation (in parenthesis) for all five different cases in **Stage-I** analysis.

Metrics	Step-1	Step-2	Step-3	Step-4	Step-5
Calibration					
NMSE	0.56	1.68	0.64	0.43	0.60
NBμ	0.09	0.32	0.12	0.09	-0.09
NBσ	-0.12	-0.15	-0.19	-0.06	-0.04
ρ	0.76	0.81	0.74	0.83	0.75
Evaluation (Total Simulation)					
NMSE	0.84 (0.77)	2.13 (2.17)	0.92 (1.19)	0.88 (0.75)	0.64 (0.56)
NBμ	0.14 (0.14)	0.38 (0.38)	0.09 (0.22)	0.15 (0.16)	-0.01 (-0.02)
NBσ	-0.04 (-0.08)	-0.03 (-0.10)	0.04 (-0.01)	0.17 (0.02)	0.14 (0.05)
ρ	0.66 (0.72)	0.71 (0.76)	0.61 (0.69)	0.74 (0.78)	0.73 (0.74)

3.2 Results from Stage-II (Structural Modifications)

10   Results from Stage-I confirm that GAET constraining can improve the overall performance of HyMod. However, for operational implementation, the method requires real-time estimates of SET, which could sometimes pose a challenge for practical applications. To overcome the need for real-time data availability, a simple approach could be to establish a functional relationship between HAET and GAET from the historical records and use that relationship to adjust HAET. In our case, however, HAET and GAET did not show a sufficiently strong relationship (Fig. 6). Therefore, we instead investigated

15   whether we could use the historical GAET data to improve the structure of the model itself.

Our previous results (Fig. 6) showed that HAET generated by HyMod-V1 did not match well with GAET. This is likely because the entire soil moisture storage was exposed to the ET process. Consequently, it is common for all of the soil moisture to evaporate away during a single time step, leaving no water available for evaporation at the next time step (provided no precipitation is added), so that HAET drops to zero. This tendency can be reduced by modifying the ET process

20   representation so that HAET more closely follows GAET.

Step-1 in the Stage-II analysis has four different cases as shown in Table 1. The first case (Case-A) is identical to the benchmark step (Step-1) in the Stage-I analysis, where the calibrated HyMod-V1 is run without GAET estimates. In Case-B, $K_t = K_0$ is applied as a constant multiplier to the ET equation (see Table 1), thereby acting as a constant surface resistance to ET. Calibration (of all of the model parameters) resulted in improved error statistics (Table 6). The estimate obtained for the

13

surface resistance was $K_0 = 0.73$. However, we again obtained a conceptually unrealistic value (H = 9.5 meters) for the soil moisture storage parameter. In Case-C, the more complex form $K_t = K_0 \cdot f(\psi_t)$ was used (see Table 1). This produced a model performance (Table 6) comparable to that of the previous, but with a more realistic calibrated value of the soil moisture storage (H = 0.90 meters). Interestingly, the calibrated value for K_0 was 1, implying that K_0 becomes irrelevant once $f(\psi_t)$ is introduced to the ET equation.

Finally, the most complex form $K_t = K_{min} + [K_{max} - K_{min}] \cdot f(\psi_t)$ was introduced in Case-D. The calibration parameter K_0 was used to represent K_{max}, and K_{min} was defined as $K_{min} = \gamma \cdot K_{max}$ via a second calibration parameter γ (ranging from 0 to 1) (see Table 1). Results indicate that although the calibration error statistics (Table 6) are similar to that of Case-C, the evaluation and total simulation statistics are better. The calibrated value of parameter 'BE' (derived by transforming the parameter 'be'; see Eq. B1) was 0.86, indicating only a mildly non-linear relationship between ψ_t and K (or HAET). The minimum and the maximum limits of K_0 were close to zero and one, respectively, confirming the findings of Case-C that once $f(\psi_t)$ is introduced to the ET equation, the need for K_{max}, and K_{min} parameters largely disappears.

Table 6: Streamflow error statistics for calibration period, evaluation period, and total simulation period (in parenthesis) for all four different cases in **<u>Stage-II Step-1</u>** analysis.

| Metrics	Case-A	Case-B	Case-C	Case-D
Calibration				
NMSE	0.56	0.52	0.51	0.51
NBμ	0.09	0.10	0.10	0.10
NBσ	-0.12	-0.19	-0.04	-0.04
ρ	0.76	0.77	0.80	0.80
Evaluation (Total Simulation)				
NMSE	0.84 (0.77)	0.65 (0.77)	0.88 (0.72)	0.84 (0.70)
NBμ	0.14 (0.14)	0.07 (0.16)	0.14 (0.14)	0.14 (0.13)
NBσ	-0.04 (-0.08)	0.00 (-0.06)	0.16 (0.10)	0.16 (0.10)
ρ	0.66 (0.72)	0.70 (0.75)	0.73 (0.78)	0.74 (0.78)

Final Model Selection from Stage-II Step-1

In this section, we address two main questions: (a) Does the structural modification of the model (to the representation of the ET process) improve ET estimation? If so, then (b) what level of complexity is adequate? Table 7 presents the streamflow and AET performance statistics for the total simulation period for the four cases. Since Case-A provides very poor error statistics for AET (e.g. NMSE = 8.93 and NBσ = 1.89), we disregard this case. Although Case-B shows the best NBσ (-0.06) statistics for streamflow, and the best NMSE (1.28) and NBμ (0.12) statistics for AET, the value obtained for the soil moisture storage capacity (H) was unrealistic; we therefore also disregard this case. Comparison of Case-C and Case-D shows that while their streamflow and AET simulations are similar (Fig. 8), Case-D provides slightly better NMSE (0.70) and NBμ (0.13) statistics for streamflow and slightly better ρ (0.49) statistics for AET (Table 7 and Fig. 9). We therefore selected the

14

most complex form $K_t = K_{min} + [K_{max} - K_{min}] \cdot f(\psi_t)$ for the ET function (Case-D). The corresponding model is hereafter referred to as 'HyMod-V2'. Note that for practical applications, any simpler structural modification (Case-B or Case-C) can be adapted, if proves convenient.

Comparing the streamflow error statistics of Stage-I Step-4 (Table 5) and Stage-II Step-1 Case-D (Table 6), we see that they are quite similar, indicating that the ET constraining (first approach) and diagnostic structural improvement (second approach) strategies produce dynamical behaviors that are similar (as measured by the four performance metrics used).

Table 7: Streamflow and AET error statistics in total simulation for all four cases in **Stage-II Step-1** analysis.

Metrics	Case-A	Case-B	Case-C	Case-D
Streamflow				
NMSE	0.77	0.77	0.72	0.70
NBμ	0.14	0.16	0.14	0.13
NBσ	-0.08	-0.06	0.10	0.10
ρ	0.72	0.75	0.78	0.78
AET				
NMSE	8.93	1.28	1.70	1.71
NBμ	0.18	0.12	0.17	0.17
NBσ	1.89	-0.26	-0.10	-0.13
ρ	0.22	0.43	0.48	0.49
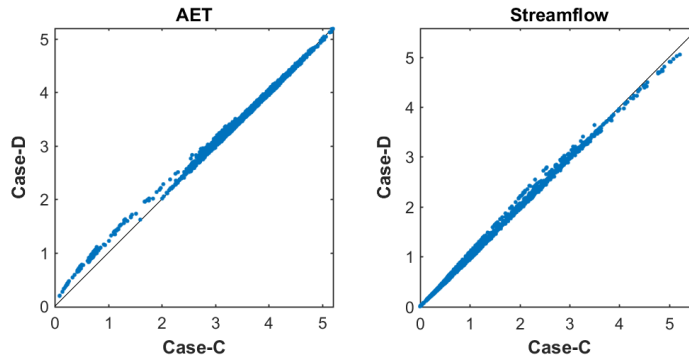
Fig. 8. Scatter plots of streamflow and AET from Case-C and Case-D.
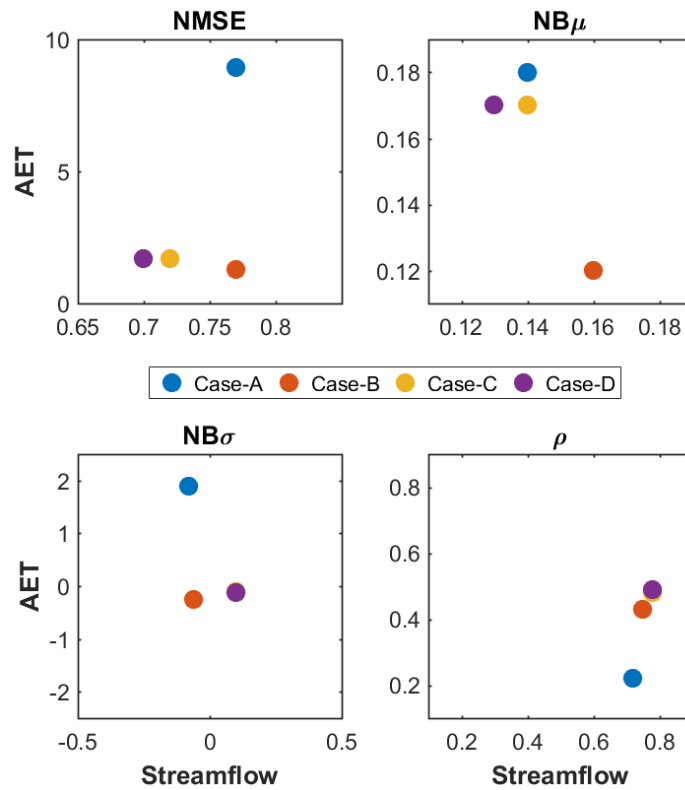
Fig. 9. Streamflow and AET error statistics in total simulation for all four cases in Stage-II Step-1 analysis.

Is Further Improvement Possible?

　　The modified model (HyMod-V2) selected in the previous section (Stage-II Step-1 Case-D) was next used with GAET in Step-2 to Step-5 (see Fig. 5) to address two questions: (1) Could more information from GAET be incorporated (via constraining) into the model, or is the improved model structure (without GAET) already good enough? (2) Is the constraint adjustment on GAET (Step-4) still relevant once the model structure has been improved?

　　When GAET was used to constrain the ET process (Step-2) in HyMod-V2 without model recalibration (parameters used were from Case-D), there was significant overestimation bias evident in the simulations of streamflow (Table 8). Clearly, recalibration of the modified model was necessary to see if the model performance could be improved any further. The recalibration of HyMod-V2 improved the error statistics (Table 8); compare these results with Stage-I Step-3 results in Table 4 derived the same way for HyMod-V1. While a small improvement was obtained for the soil moisture storage capacity parameter H (reduced from 17.4 meters to 12.8 meters), this value remained conceptually inconsistent (too large). Overall, the error statistics deteriorated compared to the best results from Case-D. The HyMod-V2 parameters were then calibrated along with the parameters of the GAET adjustment equation (using Eq. 2), as in Step I-4. Although the results improved (Table 8), and the value of H parameter became conceptually realistic (0.88 meters), the results were not significantly different from Case-D in Step-1. Finally, when the GAET data were made unavailable, the model performance remained stable as also seen in Stage-I Step-5 results.

16

Therefore, in regards to the two questions that motivated this section, the results indicate that; (1) once information from GAET was incorporated into the model as a modification to the structure, there was no further need for the use of GAET to constrain the simulation of ET (use of GAET even caused some of the results to deteriorate), and (2) implementation of a constraint adjustment to GAET (Step-4) in the structurally modified model (Stage-II) did not improve the overall results.

Table 8: Streamflow error statistics for calibration, evaluation, and total simulation (in parenthesis) for all five different cases in **Stage-II** analysis.

Metrics	Step-1	Step-2	Step-3	Step-4	Step-5
Calibration					
NMSE	0.51	1.82	0.64	0.51	0.50
NBμ	0.10	0.32	0.12	0.10	0.10
NBσ	-0.04	0.16	-0.19	-0.04	-0.04
ρ	0.80	0.81	0.74	0.80	0.80
Evaluation (Total Simulation)					
NMSE	0.84 (0.70)	2.46 (2.29)	0.92 (1.19)	0.84 (0.70)	0.84 (0.70)
NBμ	0.14 (0.13)	0.37 (0.36)	0.10 (0.22)	0.14 (0.14)	0.14 (0.13)
NBσ	0.16 (0.10)	0.38 (0.31)	0.05 (-0.01)	0.16 (0.10)	0.16 (0.09)
ρ	0.74 (0.78)	0.70 (0.77)	0.61 (0.69)	0.74 (0.78)	0.74 (0.78)

### 3.3 Overall Comparison and Analysis of Uncertainty

Fig. 10 compares the streamflow time series obtained from Stage-I Step-4 (constraining ET) and Stage-II Step-1 Case-D (selected model after structural modification) against the benchmark (Stage-I Step-1) in both actual and λ-transformed space. Simulations from the structurally modified model HyMod-V2 follow the observations most closely, followed by the simulations from Stage-I Step-4 (ET constraining) and benchmark. Clearly, while the streamflow simulations are improved by both ET constraining and model structural modification, the latter performs the best.
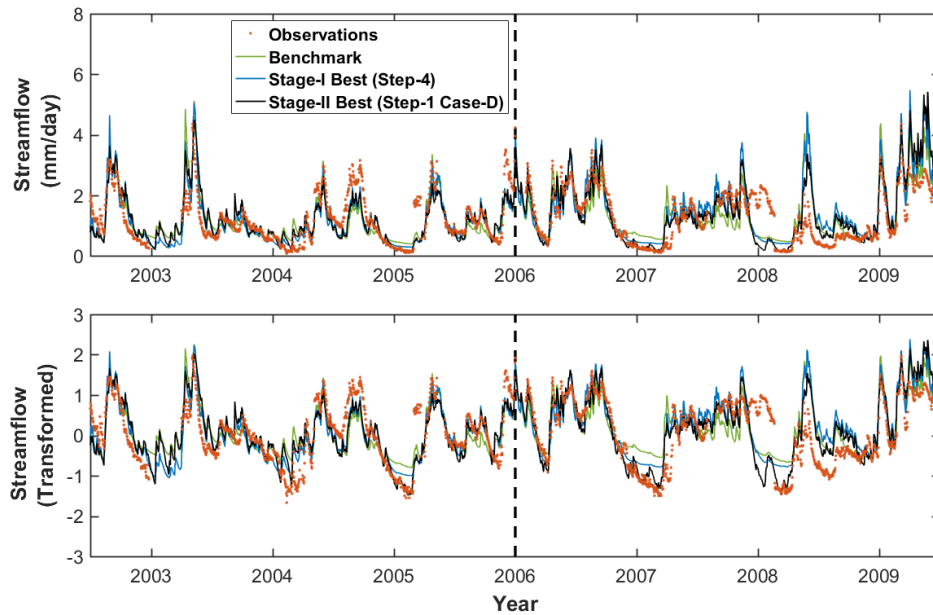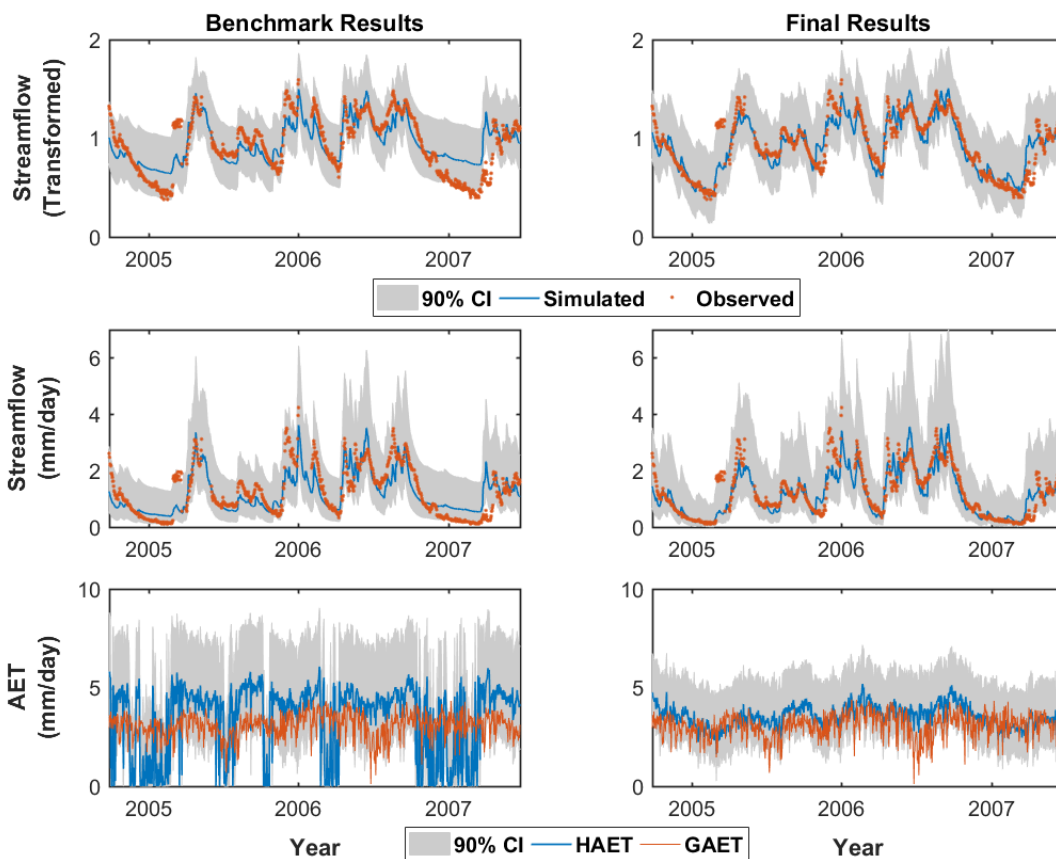
Fig. 10. Time series plots of streamflow for the best simulations in Step I and Step II, the benchmark simulation, and the observations.

Using the best model from Stage-II (HyMod-V2), we next investigate the change in simulation uncertainties for streamflow and AET due to the model structural improvement. The calibration period residual distributions (assumed stationary) were calculated in the λ-transformed space and superimposed on the daily estimates of the corresponding variables for the total simulation period. Fig. 11 shows the histograms of calibration period residuals for the benchmark and final (Stage-II Step-1 Case-D) steps. In both cases (AET and streamflow), the residuals become more normally distributed, with the improvement being more prominent for AET. This result is expected, since HyMod-V1 in Step-I showed poor performance in regards to AET. Overall, the structural modification is clearly beneficial.
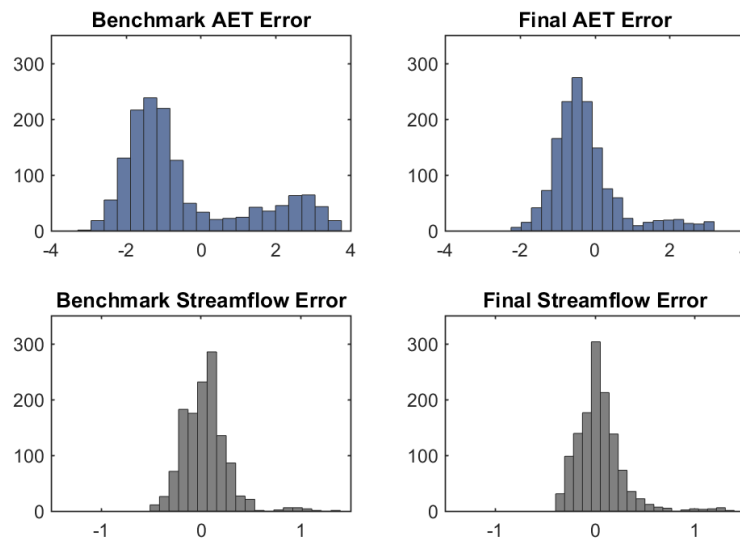
Fig. 11. AET and streamflow error distributions for the benchmark and the final steps.

Fig. 12 shows the streamflow and AET time series along with their corresponding 90% confidence intervals for the benchmark and the final steps. Both the streamflow and AET simulations improve as a result of the model structural modification. Although the streamflow uncertainty bounds have not narrowed significantly, the flow series is clearly less biased and tracks the recessions better. Meanwhile the AET simulations have improved significantly: (a) the bias has been reduced, (b) the uncertainty bounds are narrower, and (c) the erratic behavior originally seen in the AET simulations (frequent drops to zero) has disappeared. Further, although the improvement in streamflow performance is evident from the statistics in Table 5 and 6, the improved behavior is even more apparent in Fig. 12 where the model can be seen to now track the recessions quite well.

Fig. 12. Time series plots of streamflow (λ-transformed and actual) and AET for the benchmark and the final steps. For clarity, we only show a window of 1000 days.

19

4 Discussion

In this study, we have explored two different approaches for using the recently available satellite-based AET dataset GLEAM to improve the realism and performance of the conceptual catchment-scale hydrologic model HyMod. In the first approach, GAET is used as a constraint to the ET process equation in the model, while in the second approach, the model structure has been modified so that the ET process parameterizations become more physically consistent and realistic. We avoided making the model overly complicated in terms of its structural representations and/or have large number of parameters, since both of these would defeat its main purpose of being a simple model. Our goal was to increase the realism within the model and improve its performance in simple manners. Furthermore, we also made sure that the improvements in some particular process simulation (e.g. AET) do not deteriorate model's performance for simulating some other process (e.g. streamflow). Our results show that both the approaches (process constraining and structural modification) can improve the simulations of streamflow, while the later also significantly improves the AET simulations. Clearly, the satellite-based ET datasets (GLEAM in this case) can significantly benefit the process of hydrologic modeling in poorly gauged basins.

The use of ET data as a constraint can improve streamflow forecasts, provided some additional processing steps are implemented. If the GAET data are used directly as a constraint to the ET equation, the model tends to show bias in streamflow simulations. This behavior can be attributed to the fact that once GAET is incorporated, the water balance within the model is altered, the effects of which are reflected in terms of bias in the simulated streamflows. The type of this bias, of course, is subject to change depending on the dataset. While recalibration of the model with the ET constraint improves the performance, it can result in conceptually unrealistic estimates of certain parameters (H in this case). However, the model produces conceptually realistic values of the H parameter if some adjustments are made in the GAET constraints, instead of using them directly. Note that constraint adjustment is not similar to bias correction; for the latter we need the 'ground truth' of actual ET. Therefore, the adjustment process is <u>not</u> necessarily indicative of the presence of any actual bias in GAET estimates. The adjustment factors are model parameters that correspond to the structural deficiencies within the model. They may or may not be necessary as the structure changes. As we have seen in Stage-II Step-4, the constraint adjustment became irrelevant once the structure of the model itself was improved. We also found that a simple adjustment (using a multiplicative factor) can perform equally well as a more complex alternative (power function).

Improving the model structure provides several other benefits. For example, a model that simulates ET more accurately can be a suitable candidate for real-time forecasting applications. This type of a model can also prove useful for projecting future water availability. Although ET plays a significant role in the hydrological cycle, traditionally, for conceptual models, the main focus has mainly been drawn towards improving their streamflow simulation performance, while making the ET process overly simplified (e.g. simple water budget). In this study, we show that by incorporating simple but physically consistent structural changes, the ET simulation performance can be improved significantly. In poorly gauged basins, the satellite-based estimates of AET provides useful information to carry out this improvement.

In this study, we tested several conceptually reasonable structural modifications to the model of varying levels of complexity and physical basis (Case-A to Case-D), and selected the one that provided the best simulations of both GAET and

observed streamflow. We found that relatively simple changes to the model's ET equation significantly improved the ET simulations as assessed by GAET. While our goal was to improve the AET and streamflow simulations, we were also careful to ensure that the model parameters remain conceptually realistic. We saw that using a simple multiplicative factor (parameter K) as a resistance to AET produced excellent streamflow and AET forecasts (Case-B), but resulted in an unrealistic estimate of basin storage capacity (parameter H). In contrast, inclusion of a soil moisture dependent function $f(\psi_t)$, resulted in a more realistic estimate of basin storage capacity without compromising the streamflow and AET simulations. Once the model structure was appropriately modified to provide good simulations, the model simulations were robust/stable, and there was no need to impose the ET constraint (with/without constraint adjustment). The modified model structure provided significantly improved AET forecasts with much narrower uncertainty intervals (see Fig. 12), along with reduced bias in streamflow and improved tracking of the streamflow recessions.

Overall, by incorporating an additional source of external information in a sensible manner (here by structural modification), the need for calibration can be reduced (note that the model was not calibrated against GAET); see the extensive discussion by Gharari et al., (2014) and Bahremand (2016) on this topic. Nevertheless, given the simplistic nature of the hydrologic models and the large uncertainties that exist therein, some degree of calibration will generally remain important and relevant. We do not mean to imply, therefore, that calibration is not essential, because we will rarely (correctly) know everything we need to know about the system we are modeling. Instead, we should be aware of the strengths and weaknesses involved in the use of calibration and apply it carefully in such a way that useful information is gained about the underlying nature of the actual physical system. In this study, we demonstrate the need for both approaches. On the one hand, improving the model structure resulted in improved AET simulations without any need for calibration (to AET). On the other hand, the best streamflow performance was achieved when the structurally modified model was tuned via a calibration procedure.

Note that this study is based on testing the model on a single basin using a single satellite-based AET product. While not demonstrating universal applicability, the results are clearly indicative and the methodology illustrates how such data can be used to investigate potential improvements to the structures of simple catchment scale models used for hydrologic studies in data scarce regions. A rigorous analysis of the methodology over multiple basins is a future research need. For more detailed process-based models, the ET process parameters can be calibrated against some reliable satellite-based AET estimates (e.g. GLEAM), or the process representation itself can be improved by adapting some similar strategies that these AET products follow.

4.4 Conclusions

In conclusion, SET data can be used to improve model performance in different ways. However, strategies that result in model structural modifications can generally be expected to provide longer lasting benefits than ones that simply update or constrain the state trajectories of the model. This is because structural modifications can both improve the initial estimates of the state at each time step, and sustain these improvements into future time steps (Bulygina and Gupta, 2009, 2010, 2011; Nearing and Gupta, 2015). In contrast, even though data assimilation to directly adjust state estimates can improve model performance, inadequacies in model structure will tend to cause the state estimates to drift away from their more appropriate

values over time, so that performance deteriorates markedly when the constraining data are not available. Of course, we have only tested a *"constraining"* strategy to assimilating ET information, which is a relatively simple form of data assimilation (DA) (Houser et al., 1998), and more sophisticated approaches such as the Ensemble Kalman Filter (EnKF) could instead be implemented. However, the efficiency of the EnKF for soil moisture retrieval has been shown to be as low as 30% (Nearing et al., 2013a, 2013b) and so it is not clear that more sophisticated forms of DA are justified, especially given the large uncertainties associated with both the data and the model structure for this poorly gauged catchment. We leave such investigation for future work.

Code and Data Availability

Data and codes (HyMod-V2 in Matlab) used in this study are available on request from the corresponding author, Tirthankar Roy (royt@email.arizona.edu).

Conflicts of Interests

The authors express no conflicts of interests.

Acknowledgements

APPENDIX

Appendix A: Original HyMod Equations

The benchmark version of the spatially lumped conceptual rainfall-runoff model HyMod has six parameters. The model is driven by mean daily precipitation and PET data to generate daily estimates of AET and streamflow. It has two main components, a two-parameter soil moisture accounting (SMA) module based on the Moore (1985) rainfall excess concept, and a linear routing (ROUT) module with parallel quick-flow (fast overland flow) and slow-flow (baseflow) pathways. In the SMA module, the *state variable* (soil moisture storage, C) and the *indicator variable* (storage height, H) are non-linearly related via the following equation (Moore, 1985):

$$C(t) = C_{max}\left(1 - \left(1 - \frac{H(t)}{H_{max}}\right)^{1+b}\right) \tag{A1}$$

where the *maximum storage capacity* (C_{max}) and the maximum indicator height (H_{max}) are related as:

$$C_{max} = \frac{H_{max}}{1 + b} \tag{A2}$$

First, the *initial storage* (C_{beg}) is calculated from the initial indicator height (H_{beg}) using Eq. A1. Next, H_{max} is subtracted from the sum of precipitation (P) and H_{beg} to calculate *overland flow* (OV) as:

$$OV = P + H_{beg} - H_{max} \tag{A3}$$

Infiltration (I) is then calculated by subtracting OV from P:

$$I = P - OV \tag{A4}$$

and an intermediate indicator height (H_{int}) is computed by adding I to H_{beg}, and used to calculate the *intermediate storage* (C_{int}) via Eq. A2. By subtracting C_{int} from the sum of I and C_{beg} we obtain the *interflow* (IF). Finally, the *total runoff* is obtained by adding together OV and IF.

Finally, the HyMod AET (called HAET) is taken to be the smaller of available water C_{int} and potential demand PET (which is provided as input to the model):

$$HAET = \min\{PET, C_{int}\} \tag{A5}$$

and the storage at the end of the time step is computed by subtracting AET from C_{int}:

$$C_{end} = C_{int} - HAET \tag{A6}$$

The power coefficients in HyMod ('BE' in Table 1 and 'b' in Eq. A1 & A2) can have values ranging from 0 to infinity. For calibration it is useful to be able to impose finite values to the feasible ranges of the parameters; therefore we applied the following transformation (Eq. A7) which converts the [0,inf) range of parameter BE to the [0,2) range of transformed parameter 'be' so that the search can be conducted on finite range of parameter 'be' instead (similarly for parameter 'b' in Eq. A1 and Eq. A2):

$$BE = \ln(1 - be/2)/\ln(0.5); \quad be = [0, 2) \tag{A7}$$

23

Appendix B: The λ-Transformation Used

The λ-transformation on streamflows used in this work is given by the equation:

$$TQ_t = \left(\frac{Q_t}{\mu_{Qobs}}\right)^\lambda \tag{B1}$$

where Q_t and TQ_t represent streamflows in the actual space and the transformed space, μ_{Qobs} is the mean of the observations in the actual space, and λ is the transformation parameter that reduces the skewness. This expression differs slightly from the

5   form $TQ_t = \frac{(Q_t)^\lambda - 1}{\lambda}$ recommended by (Box and Cox, 1964), in that the flows are normalized by the mean μ_{Qobs} instead of by 1.0 before transformation, and the transformed flows all remain positive. This form works as long as the transformation parameter $\lambda \neq 0$ which is true in our case; if $\lambda = 0$, then one should use $TQ_t = \ln(Q_t)$ as discussed by Box and Cox (1964).

Appendix C: Calibrated HyMod (Actual and Modified) Parameters

The following table provides calibrated parameters of the actual and the modified HyMod models.

Para	Stage-I Step-1	Stage-I Step-3	Stage-I Step-4	Stage-II Step-1 Case-B	Stage-II Step-1 Case-C	Stage-II Step-1 Case-D	Stage-II Step-3	Stage-II Step-4
H	761.0	17364.0	646.4	9494.0	903.7	866.0	12763.8	878.7
B	1.93	1.95	1.24	1.87	0.29	0.34	1.86	0.32
α	0.48	0.37	0.67	0.38	0.31	0.27	0.45	0.31
Nq	1.44	4.54	1.25	4.22	4.80	4.71	3.51	4.50
Ks	0.00	0.09	0.00	0.10	0.10	0.09	0.10	0.10
Kq	0.10	0.22	0.10	0.19	0.24	0.24	0.18	0.20
Kmax	-	-	-	0.73	1.00	1.00	1.00	1.00
γ	-	-	-	-	-	0.00	1.00	0.00
BE	-	-	-	-	-	0.86	0.06	0.90
a	-	-	1.33	-	-	-	-	1.82
b	-	-	0.93	-	-	-	-	2.00

10

References

Arboleda, A., Ghilain, N. and Gellens-Meulenberghs, F.: The LSA-SAF evapotranspiration product - first results with MSG, in Proceedings of the 2005 EUMETSAT meteorological satellite data user's conference, Dubrovnik, Croatia., 2005.

Bahremand, A.: HESS Opinions: Advocating process modeling and de-emphasizing parameter estimation, Hydrol. Earth Syst. Sci., 20(4), 1433–1445, doi:10.5194/hess-20-1433-2016, 2016.

Bastiaanssen, W. G. M., Pelgrum, H., Wang, J., Ma, Y., Moreno, J. F., Roerink, G. J. and van der Wal, T.: A remote sensing surface energy balance algorithm for land (SEBAL)., J. Hydrol., 212–213, 213–229, doi:10.1016/S0022-1694(98)00254-6, 1998.

Box, G. E. P. and Cox, D. R.: An Analysis of Transformations, J. R. Stat. Soc. Ser. B, 26(2), 211–252, 1964.

Boyle, D. P., Gupta, H. V. and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, Water Resour. Res., 36(12), 3663–3674, doi:10.1029/2000WR900207, 2000.

Bulygina, N. and Gupta, H.: Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, Water Resour. Res., 45(12), W00B13, doi:10.1029/2007WR006749, 2009.

Bulygina, N. and Gupta, H.: How Bayesian data assimilation can be used to estimate the mathematical structure of a model, Stoch. Environ. Res. Risk Assess., 24(6), 925–937, doi:10.1007/s00477-010-0387-y, 2010.

Bulygina, N. and Gupta, H.: Correcting the mathematical structure of a hydrological model via Bayesian data assimilation, Water Resour. Res., 47(5), W05514, doi:10.1029/2010WR009614, 2011.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. a., Vrugt, J. a., Gupta, H. V., Wagener, T. and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44(12), W00B02, doi:10.1029/2007WR006735, 2008.

Demaria, E. and Serrat-Capdevila, A.: Validation of remote sensing-estimated hydrometeorological variables, in Earth Observation for Water Resources Management: Current Use and Future Opportunities for the Water Sector, edited by L. E. García, D. J. Rodríguez, M. Wijnen, and I. Pakulski, pp. 167–194, The World Bank., 2016.

Duan, Q., Sorooshian, S. and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28(4), 1015–1031, doi:10.1029/91WR02985, 1992.

Duan, Q., Ajami, N. K., Gao, X. and Sorooshian, S.: Multi-model ensemble hydrologic prediction using Bayesian model averaging, Adv. Water Resour., 30(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014, 2007.

van Emmerik, T., Mulder, G., Eilander, D., Piet, M. and Savenije, H.: Predicting the ungauged basin: model validation and realism assessment, Front. Earth Sci., 3(October), 1–11, doi:10.3389/feart.2015.00062, 2015.

Fenicia, F., McDonnell, J. J. and Savenije, H. H. G.: Learning from model improvement: On the contribution of complementary data to process understanding, Water Resour. Res., 44(6), W06419, doi:10.1029/2007WR006386, 2008a.

Fenicia, F., Savenije, H. H. G., Matgen, P. and Pfister, L.: Understanding catchment behavior through stepwise model concept

improvement, Water Resour. Res., 44(1), W01402, doi:10.1029/2006WR005563, 2008b.

Fenicia, F., Kavetski, D. and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47(11), W11510, doi:10.1029/2010WR010174, 2011.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Sci. Data, 2, 150066, doi:10.1038/sdata.2015.66, 2015.

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C. and Verdin, A. P.: A Quasi-Global Precipitation Time Series for Drought Monitoring., 2014.

Gash, J. H. C.: An analytical model of rainfall interception by forests, Q. J. R. Meteorol. Soc., 105(443), 43–55, doi:10.1002/qj.49710544304, 1979.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H. and Savenije, H. H. G.: 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, doi:10.5194/hess-18-4839-2014, 2014.

Gupta, H. V and Nearing, G. S.: Debates-the future of hydrological sciences: A (common) path forward? Using models and data to learn: A systems theoretic perspective on the future of hydrological science, Water Resour. Res., 50(6), 5351–5359, doi:10.1002/2013WR015096, 2014.

Gupta, H. V., Wagener, T. and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22(18), 3802–3813, doi:10.1002/hyp.6989, 2008.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G. and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resour. Res., 48(8), W08301, doi:10.1029/2011WR011044, 2012.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Appl. Eng. Agric., 1(2), 96–99, 1985.

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H. and Goodrich, D. C.: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, Water Resour. Res., 34(12), 3405–3420, doi:10.1029/1998WR900001, 1998.

Hsu, K., Gao, X., Sorooshian, S. and Gupta, H. V.: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks, J. Appl. Meteorol., 36(9), 1176–1190, doi:10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2, 1997.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P. and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates

at Fine Scales, J. Hydrometeorol., 8(1), 38–55, doi:10.1175/JHM560.1, 2007.

Joyce, R. J., Janowiak, J. E., Arkin, P. a. and Xie, P.: CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution, J. Hydrometeorol., 5(3), 487–503, doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2, 2004.

5   Kumar, R., Samaniego, L. and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resour. Res., 49(1), 360–379, doi:10.1029/2012WR012195, 2013.

Marshall, L., Sharma, A. and Nott, D.: Modeling the catchment via mixtures: Issues of model specification and validation, Water Resour. Res., 42(11), W11409, doi:10.1029/2005WR004613, 2006.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W.

10   A. and Verhoest, N. E. C.: GLEAM v3.0: satellite-based land evaporation and root-zone soil moisture, 2016.

Martinez, G. F. and Gupta, H. V.: Hydrologic consistency as a basis for assessing complexity of monthly water balance models for the continental United States, Water Resour. Res., 47(12), W12540, doi:10.1029/2011WR011229, 2011.

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D. and Wood, E. F.: The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data, Geosci. Model Dev., 9(1), 283–305,

15   doi:10.5194/gmd-9-283-2016, 2016.

Michel, D., Jiménez, C., Miralles, D. G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., McCabe, M. F., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F. and Fernández-Prieto, D.: The WACMOS-ET project &ndash; Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms, Hydrol. Earth Syst. Sci., 20(2), 803–822, doi:10.5194/hess-20-803-2016, 2016.

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

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F. and Fernández-Prieto, D.: The WACMOS-ET project &ndash; Part

25   2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci., 20(2), 823–842, doi:10.5194/hess-20-823-2016, 2016.

Monteith, J. L.: Evaporation and environment, in Proceedings of a symposium of the Society for Experimental Biology, pp. 205–234., 1965.

Moore, R. J.: The probability-distributed principle and runoff production at point and basin scales, Hydrol. Sci. J., 30(2), 273–

30   297, doi:10.1080/02626668509490989, 1985.

Moradkhani, H., Hsu, K.-L., Gupta, H. and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resour. Res., 41(5), n/a-n/a, doi:10.1029/2004WR003604, 2005.

Nearing, G. S.: Diagnostics And Generalizations Of Parametric State Estimation, The University of Arizona., 2013.

Nearing, G. S. and Gupta, H. V.: The quantity and quality of information in hydrologic models, Water Resour. Res., 51(1), 524–538, doi:10.1002/2014WR015895, 2015.

Nearing, G. S., Gupta, H. V., Crow, W. T. and Gong, W.: An approach to quantifying the efficiency of a Bayesian filter, Water Resour. Res., 49(4), 2164–2173, doi:10.1002/wrcr.20177, 2013a.

5 Nearing, G. S., Gupta, H. V. and Crow, W. T.: Information loss in approximately Bayesian estimation techniques: A comparison of generative and discriminative approaches to estimating agricultural productivity, J. Hydrol., 507, 163–173, doi:10.1016/j.jhydrol.2013.10.029, 2013b.

Pokhrel, P., Gupta, H. V. and Wagener, T.: A spatial regularization approach to parameter estimation for a distributed watershed model, Water Resour. Res., 44(12), n/a-n/a, doi:10.1029/2007WR006615, 2008.

10 Pokhrel, P., Yilmaz, K. K. and Gupta, H. V.: Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, J. Hydrol., 418–419, 49–60, doi:10.1016/j.jhydrol.2008.12.004, 2012.

Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Mon. Weather Rev., 100(2), 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

Razavi, S. and Gupta, H. V.: A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2.
15 Application, Water Resour. Res., 52(1), 440–455, doi:10.1002/2015WR017559, 2016.

Roy, T., Serrat-Capdevila, A., Gupta, H. and Valdes, J.: A platform for probabilistic Multimodel and Multiproduct Streamflow Forecasting, Water Resour. Res., 53, doi:10.1002/2016WR019752, 2017.

Savenije, H. H. G.: *HESS Opinions* "Topography driven conceptual modelling (FLEX-Topo)," Hydrol. Earth Syst. Sci., 14(12), 2681–2692, doi:10.5194/hess-14-2681-2010, 2010.

20 Schaefli, B., Harman, C. J., Sivapalan, M. and Schymanski, S. J.: *HESS Opinions*: Hydrologic predictions in a changing environment: behavioral modeling, Hydrol. Earth Syst. Sci., 15(2), 635–646, doi:10.5194/hess-15-635-2011, 2011.

Trambauer, P., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., van Beek, L. P. H. and Uhlenbrook, S.: Comparison of different evaporation estimates over the African continent, Hydrol. Earth Syst. Sci., 18(1), 193–212, doi:10.5194/hess-18-193-2014, 2014.

25 Troch, P. A., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J. W., Roy, T. and Valdés-Pineda, R.: Catchment coevolution: A useful framework for improving predictions of hydrological change?, Water Resour. Res., 51(7), 4903–4922, doi:10.1002/2015WR017032, 2015.

Vrugt, J. A., Gupta, H. V., Bouten, W. and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resour. Res., 39(8), n/a-n/a,
30 doi:10.1029/2002WR001642, 2003.

Vrugt, J. A., Gupta, H. V., Dekker, S. C., Sorooshian, S., Wagener, T. and Bouten, W.: Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting model, J. Hydrol., 325(1–4), 288–307, doi:10.1016/j.jhydrol.2005.10.041, 2006.

Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V. and Robinson, B. A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, Stoch. Environ. Res. Risk Assess., 23(7), 1011–1026, doi:10.1007/s00477-008-0274-y, 2009.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V. and Sorooshian, S.: A framework for development and application of hydrological models, Hydrol. Earth Syst. Sci., 5(1), 13–26, doi:10.5194/hess-5-13-2001, 2001.

Wang, D., Chen, Y. and Cai, X.: State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter, Water Resour. Res., 45(11), n/a-n/a, doi:10.1029/2008WR007401, 2009.

Winsemius, H. C., Savenije, H. H. G. and Bastiaanssen, W. G. M.: Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins?, Hydrol. Earth Syst. Sci., 5(4), 2293–2318, doi:10.5194/hessd-5-2293-2008, 2008.

WREM: Mara River Basin Monograph. [online] Available from: http://nileis.nilebasin.org/system/files/Mara Monograph.pdf, 2008.

Zhang, Y., Chiew, F. H. S., Zhang, L. and Li, H.: Use of Remotely Sensed Actual Evapotranspiration to Improve Rainfall–Runoff Modeling in Southeast Australia, J. Hydrometeorol., 10(4), 969–980, doi:10.1175/2009JHM1061.1, 2009.