https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

# Using evapotranspiration signatures to assess evapotranspiration realism in rainfall-runoff models

Hansini Gardiya Weligamage[1], Keirnan Fowler[1], Margarita Saft[2], Tim Peterson[3], Dongryeol Ryu[1], and Murray C Peel[1]

[1]Department of Infrastructure Engineering, The University of Melbourne, Parkville, Victoria, 3052, Australia
[2]Institute of Applied Geosciences, Technische Universität Berlin, 10587 Berlin, Deutschland, Germany
[3]Department of Civil Engineering, Monash University, Clayton, Victoria, 3168, Australia

*Correspondence to*: Hansini Gardiya Weligamage (h.gardiyaweligamage@unimelb.edu.au)

**Abstract.** Conceptual rainfall-runoff models are typically evaluated using streamflow data. Many studies have shown the benefits of moving from a conventional single-objective function to consideration of multiple signatures of streamflow behaviour, allowing more specific diagnosis of model deficiencies. In this study, we extend this approach beyond streamflow, using actual evapotranspiration (AET) signatures to assess AET dynamics in traditionally calibrated conceptual rainfall-runoff models. We calibrated models to streamflow only and separately to streamflow jointly with AET, then evaluated AET realism using AET signatures defined across various temporal scales. This was done for five models at 14 Australian sites spanning a variety of conditions, with each site co-located with a flux tower. Our results show that incorporating AET data into calibration significantly improves aspects of AET dynamics in models, particularly monthly variability and the degree to which AET and potential evapotranspiration are synchronous or asynchronous. This improvement extends even to the independent evaluation period in split sample testing. However, other signatures were not well improved, including aspects of seasonal and event-scale timing, in addition to interannual variability. Future research could explore a wider range of calibration strategies to assess whether these deficiencies can be calibrated away or are inherent to the models. Overall, these findings suggest that commonly used conceptual rainfall-runoff models struggle with many aspects of AET dynamics, even when AET information is included in the calibration. We recommend that future model evaluations examine a wider range of measures, aiming to characterise performance against non-streamflow variables in a more holistic manner.

## 1 Introduction

The unrealistic representation of water balance components in conceptual rainfall-runoff models may lead to bias and uncertainty in streamflow estimates, especially under climate change (Avanzi et al., 2020; Deb et al., 2019; Fowler et al., 2020; Gardiya Weligamage et al., 2023; Grigg & Hughes, 2018; Guo et al., 2017). For example, studies have shown that many commonly used conceptual rainfall-runoff models fail to capture long-term groundwater memory (i.e., long-term decline in groundwater storage) under multiyear drought (Fowler et al., 2020; Grigg & Hughes, 2018). Furthermore,

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

uncertainties associated with the representation of soil moisture and groundwater storage in models propagate into both streamflow and actual evapotranspiration (AET) estimates, as AET is often calculated based on storage and thus is sensitive to storage conceptualisation. Guo et al. (2017) revealed that simulated AET failed to match temporal patterns with flux tower AET, and the simulations of AET largely depended on the conversion method from potential evapotranspiration (PET) to AET within the model and on seasonal storage changes. Further, Avanzi et al. (2020) showed that models are incapable of estimating AET response time under climate change. These studies highlight the need for a comprehensive understanding of process representation especially in terms of AET within models and the potential for calibration (Zhang et al., 2009) and structural modifications (Fowler et al., 2021) to achieve realistic AET representation in conceptual rainfall-runoff models.

Hydrological (i.e., streamflow) signatures have been widely employed in identifying hydrological dynamics and for hydrological modelling (Araki et al., 2022; Kiraz et al., 2023; McMillan, 2020; McMillan, 2021; Westerberg et al., 2011). Hydrological signatures are statistics used to characterise aspects of the streamflow timeseries, such as flow magnitude, duration, frequency, timing, and rate of change, for ecohydrology and hydrological research (McMillan, 2021; Olden & Poff, 2003; Poff et al., 1997). The implementation of hydrological signatures in model calibration has shown some improvements in simulated streamflow dynamics (Refsgaard & Knudsen, 1996; Sugawara, 1979), integrating information on a wider range of hydrological dynamics into the calibration to obtain realistic estimates (Gupta et al., 1998; Wrede et al., 2015). In addition, hydrological signatures can be used in independent model evaluation to assess model performance by identifying the aspects of the flow regime that are (or are not) accurately captured by the model, including how such discrepancies vary under conditions different from model calibration (Gunkel et al., 2015; Hrachowitz et al., 2013).

Although there has been comprehensive advancement in the use of streamflow signatures in hydrological modelling, there is a notable gap in using such signatures for non-streamflow variables such as AET. Nearly all modelling studies that have considered AET have considered performance primarily via commonly used aggregate measures (for an exception, see Guo et al. 2017). Thus, an opportunity exists to examine simulated AET in a new way, using a recently introduced set of signatures specifically defined for AET (hereafter '*AET signatures*'), as described in Gardiya Weligamage et al. (2025). For modelling studies, these new signatures provide opportunities firstly to examine the performance of calibrated models in a new way; and secondly to undertake the calibration itself differently (e.g., by directly calibrating to signature values, see Hingray et al., 2010; Sadegh & Vrugt, 2014; Westerberg et al., 2011). Only the first of these two objectives is in focus here; as such, we are calibrating models in a way that is broadly similar to previous studies (including streamflow only calibration and streamflow jointly with AET, but never directly to the signatures), and then evaluating the realism of results through the lens of AET signatures. We employ the AET signatures defined in Gardiya Weligamage et al. (2025) to assess the realism of five commonly used conceptual rainfall-runoff models in 14 catchments in Australia that have nearby field-measured AET. This study systematically identifies deficiencies in AET representation, providing clues as to possible improvements in AET representation, with possible flow-on benefits to other components within rainfall-runoff models.

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

## 2 Materials and Methods

### 2.1 Choice of the best available AET data source

Since this study focuses on assessing the realism of AET simulations from conceptual rainfall-runoff models, a key question is from where we should source the observed AET data to best represent catchments' AET. We discuss this first because it also influences decisions such as catchment selection. The most commonly used source of information in modelling studies is remotely sensed AET. One key benefit of remotely sensed data is their ability to provide consistent spatial coverage. However, in a previous study, Gardiya Weligamage et al. (2025) assessed the quality of remotely sensed AET across Australia and identified several limitations of remotely sensed AET, including seasonal inconsistencies and variability issues. Here, these deficiencies make us question whether we should use remotely sensed AET in evaluating AET dynamics in conceptual rainfall-runoff models. The alternative source of AET data is flux tower AET data. However, there can be a spatial mismatch between flux tower location and the catchment of interest. For example, some flux towers are located outside the nearest catchments, and even when they are located within catchments, their footprint does not cover the entire catchment area. To investigate the impact of such spatial mismatch, we conducted an initial analysis (Supplementary Material, Section 1). This analysis was conducted at flux tower sites, using the locally measured information as "truth", and then comparing two alternative data sources; namely flux tower data sourced from some distance away (nearest neighbour flux tower, typically 10s to 100s of kilometers away), and remotely sensed data extracted at the site. The results show that the distant flux tower information is usually superior to the remotely sensed at-site information – a striking result given the large distances involved, in some cases. However, these test results still do not guarantee that the flux tower AET data are representative of the catchments in question. Given these considerations and findings, this study: 1) adopts flux tower AET information instead of remotely sensed AET, and 2) conducts additional tests on catchments where we are relatively more confident that the adopted flux tower AET data is representative of catchments. This question then allows us to question the representativeness of the flux tower AET data in the discussion section.

### 2.2 Study Area

This study focuses on 14 catchments and the nearby flux towers, based on the considerations outlined in Section 2.1, and they are shown in Figure 1. All adopted flux tower sites are part of the Ozflux program, the Australian component of the global FluxNet program. Each selected site has sufficient data availability (typically 7-20 years) (Beringer et al., 2016; Isaac et al., 2017), and the sites cover a wide variety of climatic conditions, with mean annual rainfall varying between 400 and 2500 mm/year (Figure 1). The catchment areas vary from around 90 $km^2$ to 4800 $km^2$. The distances between flux towers and catchment centroids vary from around 10 km to 80 km. The details are summarised in Table 1. Note that detailed information about the flux towers, including location (i.e., latitude, longitude), climate information, and eco-region classifications, is summarised in Table 1 in Gardiya Weligamage et al. (2025).

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

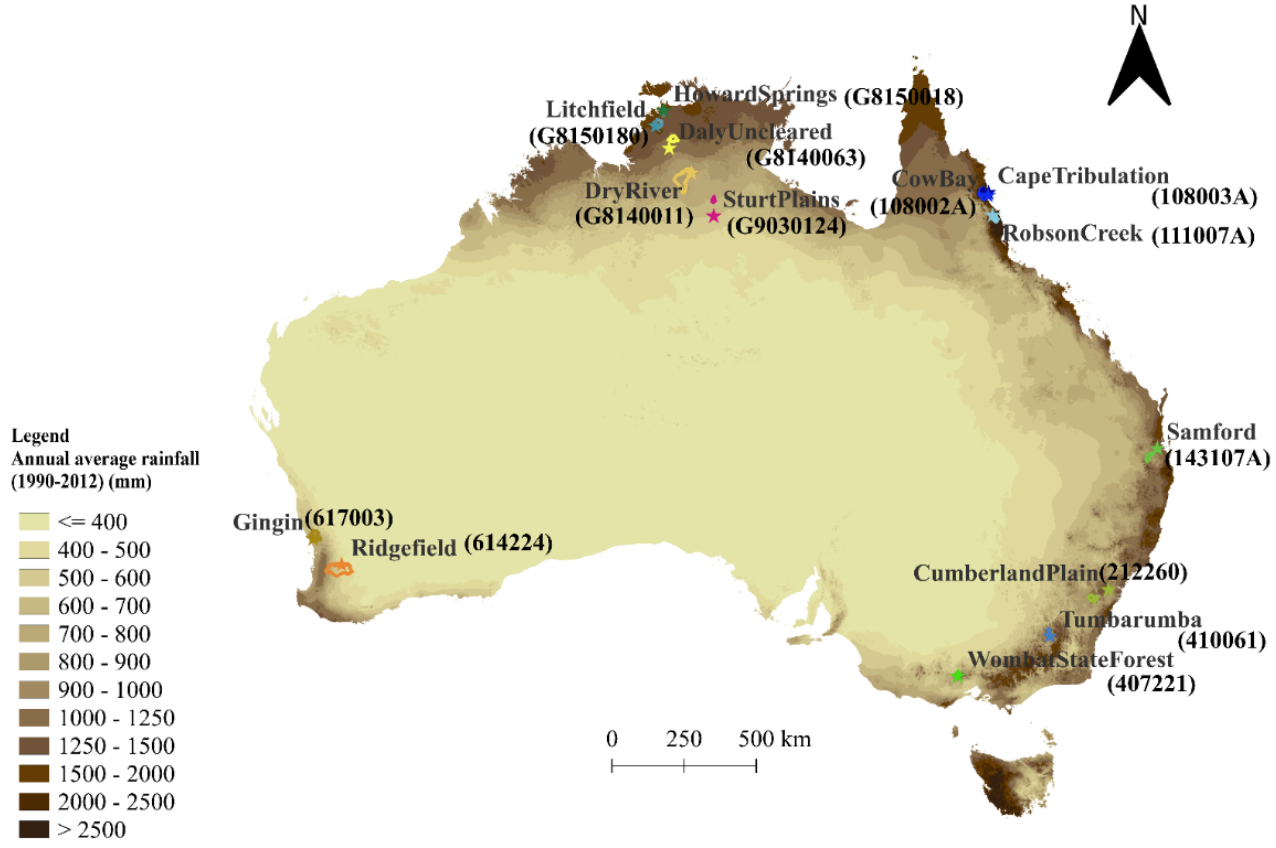



**Figure 1: Adopted catchments, each paired with a flux tower site (stars) shown within Australia. Catchments are shown as a polygon of the same colour as the flux tower; in cases where such a polygon is not visible, it is because the paired catchment is too small and/or close to the flux tower site to be distinguished at the continental scale. Paired catchment IDs are given within brackets next to flux tower names. Rainfall data are from the AWAP project** (Jones et al., 2009)**.**

**Table 1: Details of catchments, along with paired flux tower for each catchment**

| Ozflux tower site | Nearest catchment ID[a] | Catchment area ($km^2$) | Distance: flux tower to catchment centroid (km) | Flux tower coverage | Major vegetation cover in catchment[b] |
|---|---|---|---|---|---|
| Sturt Plains | G9030124 | 196 | 80.7 | Aug-08 to Feb-24 | Woody trees (Sparse) |
| Ridgefield | 614224 | 3964 | 27.2 | Jan-16 to Feb-24 | Rainfed crops |
| Gingin | 617003 | 1404 | 16.5 | Oct-11 to Feb-24 | Mixed vegetation |
| Dry River | G8140011 | 4786 | 49.6 | Oct-09 to Feb-24 | Woody trees (Sparse) |
| Daly Uncleared | G8140063 | 841 | 53.2 | Jan-08 to Dec-23 | Woody trees (Sparse) |
| Cumberland Plains | 212260 | 714 | 77.2 | Jan-14 to Dec-23 | Woody trees (Closed) |
| Samford | 143107A | 625 | 54.3 | Jun-10 to Dec-17 | Woody trees (Sparse) |

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

| | | | | | |
|---|---|---|---|---|---|
| Wombat State Forest | 407221 | 167 | 11.2 | Jan-10 to May-21 | Woody trees (Closed) |
| Howard Springs | G8150018 | 96 | 20.9 | Jan-02 to Aug-22 | Woody trees |
| Litchfield | G8150180 | 1044 | 17.2 | Jun-15 to Dec-23 | Woody trees (Sparse) |
| Robson Creek | 111007A | 583 | 17.7 | Aug-13 to Dec-23 | Woody trees (Closed) |
| Tumbarumba | 410061 | 147 | 24.4 | Jan-02 to Dec-22 | Mixed Vegetation |
| Cow Bay | 108002A | 910 | 30.6 | Jan-09 to Jul-23 | Woody trees (Closed) |
| Cape Tribulation | 108003A | 263 | 16.9 | Jan-10 to Nov-18 | Woody trees (Closed) |

- [a]This study used Catchment Attributes and Meteorology for Large-sample Studies (CAMELS, Fowler et al., 2020) catchment shapefiles and data.
- [b]Major vegetation coverage in catchments was obtained from Dynamic Land Cover Dataset Version 1.0 (DLCD V1.0, Lymburner et al. (2010)).

## 2.3 AET signatures

Table 2 lists the AET signatures defined by Gardiya Weligamage et al. (2025), which we use to assess the realism of AET representation in conceptual rainfall-runoff models. Most of the signature metrics presented in Table 2 are self-explanatory, except for two: *AET asynchronicity to PET (AAP)* and the *AET responsiveness to a rainfall event (R)*. For the benefit of readers who rely solely on this manuscript, these two signatures are elaborated upon in this section, with direct quotes from Gardiya Weligamage et al. (2025) where necessary.

The *AAP* is defined to quantify the asynchronicity between PET and AET. To calculate this metric, the monthly PET and AET timeseries are first normalised by their respective mean monthly values. The numerator of the metric is then computed by integrating the absolute difference between the normalised monthly PET and AET ($\int(|normPET_{month} - normAET_{month}|)\,dt$). The denominator is calculated by integrating the maximum value between the normalized PET and AET at each monthly step ($\int \max(normPET_{month}, normAET_{month})\,dt$). As stated by Gardiya Weligamage et al. (2025), "The rationale behind calculating the absolute difference between values is that a difference between the curves is indicative of asynchronicity, regardless of which of the curves happens to be greater at the given point in time."

The event scale *R* is defined to quantify how AET responds to a rainfall event. However, rainfall events may not only influence AET on the given day but also influence AET in the following days, which limits the applicability of standard correlation measures. As described by Gardiya Weligamage et al. (2025), "Seeking a generalisable metric, we select rainfall events greater than a threshold and identified the maximum daily AET value after the given rainfall event, up to a certain window duration (in days) after that event. We then apply a standard linear correlation equation to the anomalies of these ordered pairs of numbers (i.e., rainfall anomaly versus maximum AET anomaly in the window after the rainfall event)."

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

**Table 2: AET signature definitions. See text or Gardiya Weligamage et al. (2025) for more information on definitions**

| Temporal scale | Signature | Metric |
|---|---|---|
| Annual | 1. Long-term Median $(\widetilde{AET}_{annual})$ | $Median\ of\ annual\ AET$ |
| | 2.Interannual Variability $(CV_{annual})$ | $CV_{annual} = \frac{\sigma_{AET_{annual}}}{\overline{AET}_{annual}}$ |
| Seasonal (calculated using monthly timestep data) | 3. Periodicity $(P_{12month})$ | $Lag\ 12\ autocorrelation\ of\ monthly\ AET$ |
| | 4. Timing of Seasonal Peak $(TSP)$ | $month\ with\ maximum\ median\ of\ monthly\ AET$ |
| Monthly | 5. Monthly Variability $(CV_{monthly})$ | $CV_{monthly} = \frac{\sigma_{AET_{month}}}{\overline{AET}_{month}}$ |
| | 6. Water Stress $(WS)$ | $WS = \frac{\overline{PET}_{month} - \overline{AET}_{month}}{\overline{PET}_{month}}$ |
| | 7. [a]AET asynchronicity to PET $(AAP)$ | $AAP = \frac{\int(\lvert normPET_{month} - normAET_{month}\rvert)\,dt}{\int \max(normPET_{month}, normAET_{month})\,dt}$ |
| Event-scale | 8. AET responsiveness to a rainfall event $(R)$ | $R = \frac{\sum P_{anom_{event}} * AET_{anom_{event}}}{\sqrt{\sum P_{anom}{}^2 \sum AET_{anom}{}^2}}$ where, $P_{anom_{event}} = Rainfall_i - \overline{Rainfall_i}$ $AET_{anom_{event}} = AET_{i+j} - \overline{AET_{i+j}}$ $i$ is the position of the rainfall event, and j is the day with maximum AET after the rainfall event, $0 \leq j \leq 10$ |

[a]Note that the trapezoidal integration was conducted.

### 2.4 Data

For flux tower data, the quality-controlled, gap-filled Ozflux actual evapotranspiration data, version 1 of the year 2024 (2024_V1), were obtained from Terrestrial Ecosystem Research Network data portal at daily, monthly, and annual scales. Negative values occasionally arise in flux tower data, but we found that the negative value percentage assessed on a site-by-site basis is always less than 1% for all fourteen flux sites. All such values were excluded during model calibration and AET signature calculations. As per Table 1, the PET was required for some signatures. Morton's wet environment potential evapotranspiration (Morton, 1983) was adopted for this purpose, following previous studies (e.g., Gardiya Weligamage et al. (2025)). This was extracted from the SILO database (Scientific Information for Land Owners, (TERN,https://portal.tern.org.au/, last accessed on 30/04/2024), which is a gridded dataset on a 0.05 degree grid. For each flux tower, PET data were extracted for the grid cell containing the flux tower site.

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

In hydrological modelling, hydrometeorological timeseries data such as catchment average daily rainfall, PET, temperature, and streamflow data, along with catchment boundaries, were sourced from CAMELS-AUS (Fowler et al., 2021). The streamflow timeseries data within CAMELS-AUS are sourced from the Hydrological Reference Stations (HRS), Bureau of Meteorology (http://www.bom.gov.au/water/hrs/). CAMELS-AUS provides various options for climate data (precipitation and PET), and in each case we chose the SILO option (Jeffrey et al., 2001) mentioned above.

### 2.5 Hydrological modelling

In this study, five conceptual, lumped rainfall-runoff models were employed, namely SIMHYD (Chiew et al., 2002), IHACRES (Croke & Jakeman, 2004), VIC (Clark et al., 2008), SACRAMENTO (Burnash & Ferral, 1973) and GR4J (Perrin et al., 2003). These models are commonly used in hydrological studies in Australia and worldwide (Chiew et al., 2008; Dye & Croke, 2003; Fowler et al., 2020; Gou et al., 2020). The rationale for selecting these models was to encompass various model structure complexities with various evapotranspiration representations, and also to choose models that are commonly used so that the results are relatable to other studies. All simulations were done using version 2.1 (v2.1) of the MARRMoT framework (Trotter et al., 2022).

The optimisation algorithm implemented within v2.1 of the MARRMoT framework, namely the Covariance Matrix Adaptation Evolution Strategy (CMA-ES, Hansen et al., 2003), was used to calibrate models as it presents robust performance for hydrological model calibration (Arsenault et al., 2014). In this analysis, we calibrated models using two methods: 1) using discharge only and 2) using a composite objective function of both discharge and AET, as discussed below.

#### 2.5.1 Method 1: Model calibration using discharge only

We first calibrated models using observed streamflow, which is the conventional way of model calibration. The objective function ($OF_Q$, Eq. 1) was obtained from Trotter et al. (2023), and it is designed to reproduce high and low flows in hydrograph while minimizing volumetric bias:

$$OF_Q = \frac{1}{2}\left(KGE_Q + KGE_{Q^{0.2}}\right) - 5|\ln(B+1)|^{2.5} \quad (1)$$

The model efficiency of $OF_Q$ has two parts. The first part is the mean of two Kling Gupta Efficiencies (KGE, Gupta et al., 2009), with $KGE_Q$ calculated using all flow values and $KGE_{Q^{0.2}}$ calculated using the fifth root of those flows. Consideration of the fifth root favours predicting low flows and accommodates zero flows (unlike log transforms, which are not calculable for zero flows). The second part reduces bias by deducting a logarithmic function of volumetric bias (B) (Vaze et al., 2010).

In this analysis, the simulation period was from January 1, 1980, to April 30, 2022. The end date of the simulation period was decided based on the daily rainfall and PET coverage, as the full record of daily data in months is unavailable after April 2022 in CAMELS-AUS. The calibration period was selected from 1982-2022, leaving two years to warm up. The reason for

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

selecting almost the entire period in the calibration was to give the models every possibility to learn from a variety of dry and wet events, so that any observed failure cannot be due to withheld data.

### 2.5.2 Method 2: Model calibration using discharge and actual evapotranspiration

For the second calibration method, we first developed an objective function ($OF_{AET}$, Eq. 2) to calculate performance against observed AET data.

$$OF_{AET} = \frac{1}{2}(KGE_{AET^{0.5},NoMean,Daily} + KGE_{AET^{0.5},NoMean,Monthly}) \quad (2)$$

This objective function blends information across two different timesteps, calculating the mean of an adapted version of the KGE calculated separately on daily and monthly timestep, respectively. $KGE_{AET^{0.5},NoMean,Daily}$ and $KGE_{AET^{0.5},NoMean,Monthly}$ were calculated using the square root of daily and monthly AET, respectively. The reason for using the square root of AET values in both temporal scales is to reduce the influence of extreme values in the data. The use of two separate components places emphasis on different information in the timeseries, with the monthly version focussed on seasonal variations, while the daily version is more sensitive to the day-to-day changes resulting from shorter-term weather activity. In these two KGE components, the bias component in the original KGE has been removed because it could be misleading to try to match the mean AET between catchment average AET with smaller scale AET from a flux tower, since biases can arise from differences in location. Even in the ideal case where the flux tower is located within the catchment boundary (such as Gingin, Dry River, and Litchfield flux towers), the spatial support of a flux tower typically extends hundreds of meters, rather than kilometers, from the tower (and is thus significantly smaller than a typical gauged catchment). It should be acknowledged that some inconsistency in AET variability will remain due to spatial differences in AET across the landscape.

Finally, the above objectives for different response variables were combined together into a single composite objective function ($OF_{Q+AET}$, Eq. 3) to calibrate the rainfall-runoff models using both streamflow and flux tower AET by calculating the mean of $OF_Q$ and $OF_{AET}$.

$$OF_{Q+AET} = \frac{1}{2}(OF_Q + OF_{AET}) \quad (3)$$

Here, similar to Method 1 in section 2.5.1, the simulation period was selected from 1980-2022, and model calibration using discharge was conducted from 1982-2022. However, model calibration using flux tower AET was conducted over time periods that varied from site to site, depending on the availability of flux tower data.

## 2.6 Acceptance of models and evaluation using evapotranspiration signatures

Regardless of the calibration method, the simulated AET was only evaluated in this study if the model met pre-set performance criteria. The idea is to focus on instances where performance (judged by traditional metrics) is broadly acceptable and use the signatures to gain insight into failings that are not apparent from those traditional metrics. For this

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

purpose, a threshold of 0.6 was adopted for the optimised objective function in both calibration methods. The optimised objective function values of models for all the study catchments are available in Tables S2 & S3 in the Supplementary material.

Model simulations were evaluated by calculating the AET signatures mentioned in Section 2.3. AET signatures calculated from model outputs were compared against those calculated using flux tower AET, to assess the realism of AET representation within the models, and also to evaluate the extent to which incorporating AET information into the model calibration improves AET simulations across a variety of different aspects of AET dynamics.

In addition, as mentioned in Section 2.1, we defined an additional test focusing on sites considered to contain relatively higher quality information (i.e., flux tower data more representative of the catchments). For this purpose, a subset of study sites was further selected based on the closer proximity of flux towers (either very close to or within the catchments) and similarity in dominant land cover for a similar comparison of AET signatures between models and flux towers but to get insights into whether the results change when focusing on a more representative subset of the data. The results are available in Section S3 in the Supplementary material.

### 2.7 Independent model evaluation using split sampling calibration

In addition to assessing performance over the calibration period (Method 2 above), it is also important to assess improvements over an independent period not included in calibration. Therefore, to address this, the analysis was extended using a split-sample approach, where half of the dataset, covering both streamflow and AET, was used for calibration and the other half for independent evaluation, and vice versa. The evaluation was done for both AET and streamflow, by reporting objective function values (Section 3.4) for each. In addition, we report results for AET dynamics via the AET signatures, but for reasons of brevity, we only evaluated the signatures that showed improvements in Method 2 of model calibration.

## 3 Results

### 3.1 Evaluation of model performance using traditional metrics

Given our focus on showing the additional value of signatures over traditional metrics, we first express simulation performance using traditional metrics. Figure 2 shows the performance values using traditional metrics – specifically, using the metrics against which the models were calibrated, which are themselves variants of the Kling Gupta Efficiency (see Section 2.5). As expected, Figure 2a shows that AET performance is lower in models calibrated using streamflow only than in models calibrated using both streamflow and AET. Adding the AET into the model calibration causes the scores to increase (medians of AET performance across the 14 catchments increased by between 0.06 and 0.41, depending on model), but at some cost to the streamflow objective function (medians of streamflow performance reduced by between 0.02 and 0.08). Without further information, it would seem reasonable to conclude that, in general:

- Given the increase in performance, the realism of the model has improved; and

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

- Given the final scores are in a range that is typically considered "good", the performance of the AET simulations should be considered acceptable.

Both these statements can now be assessed using the alternative lens of signatures, and this is done in the following sub-section.

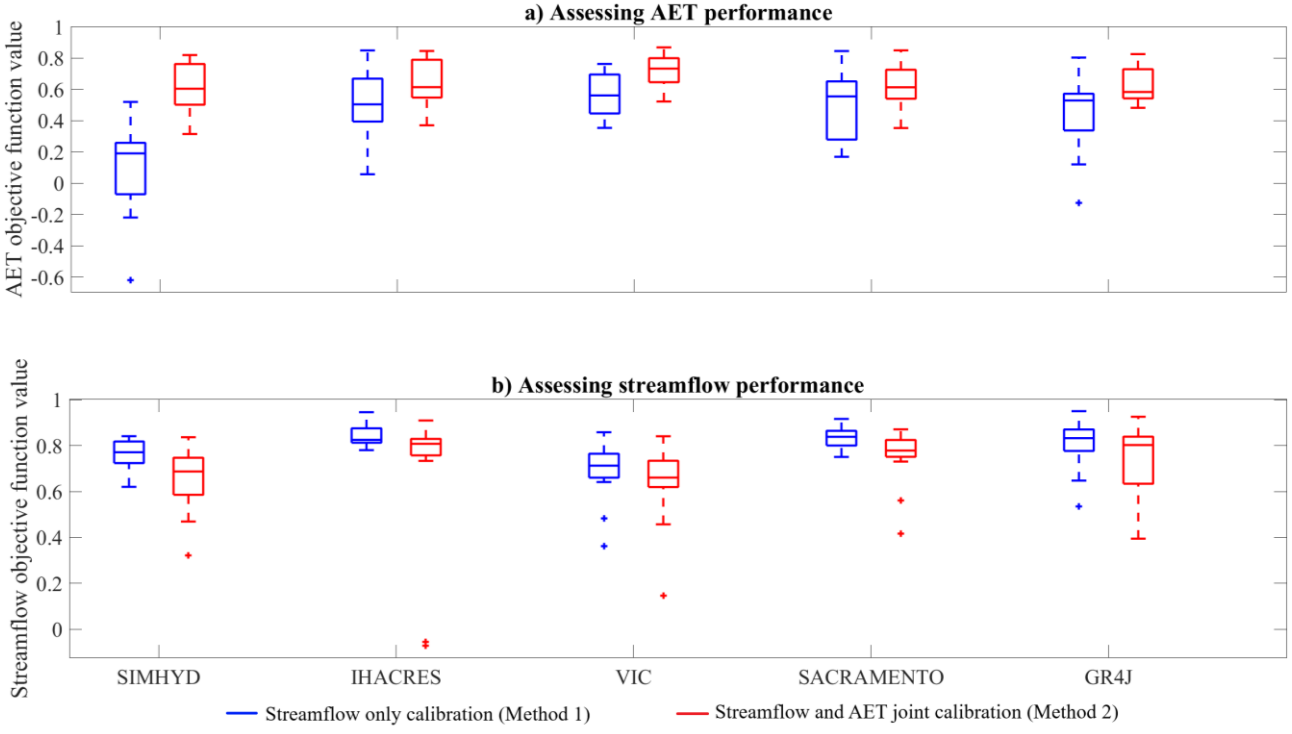



**Figure 2: Assessing the performance of models using traditional calibration metrics. Each boxplot shows the distribution of objective function values across the set of 14 catchments.**

### 3.2 Using signatures to assess simulation performance of AET

#### 3.2.1 Comparison of AET signatures at annual scale

Figure 3 shows the comparison of the first of the signatures, namely the long-term median annual AET ($\widehat{AET}_{annual}$). Signature values calculated from model simulations calibrated using discharge only (Figure 3a) and those calibrated using discharge and AET (Figure 3b) are compared with signature values calculated from flux towers. According to Figure 3a, $\widehat{AET}_{annual}$ calculated from model simulations are relatively close or tend to overestimate the flux tower AET at most of the catchments (except 407221, G8150018, G8150180, and 108003A catchments near Wombat State Forest, Howard Springs,

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

Litchfield, and Cape Tribulation sites, respectively). Similar results are observed when we incorporated flux tower AET data with streamflow into the calibration (Figure 3b). Overall, these results are consistent with expectation given that long-term AET is set by the water balance, and thus we would expect models to match this value even if calibrated to streamflow only. The interannual variability (expressed as the coefficient of variation, $CV_{annual}$) shows overestimation compared to flux tower estimates (Figure 4); in other words, the variability in annual AET is higher in reality than in the simulations. Moreover, the incorporation of flux tower AET information into the calibration does not improve the $CV_{annual}$, similar to Figure 4b. These annual scale results suggest that including daily and monthly AET information in the model calibration does not improve AET dynamics at the annual scale (recall that the adopted objective function ignores the annual timescale, which is typical of commonly used objective functions).

### 3.2.2 Comparison of AET signatures at seasonal scale

At the seasonal scale, incorporating AET information into the calibration does not significantly improve the seasonal AET signatures, similar to the findings at the annual scale. Figure 5 shows models both overestimate and underestimate AET periodicity ($P_{12month}$) compared to flux tower $P_{12month}$, regardless of the calibration method used.

Figure 6a shows that models calibrated using only discharge tend to align with flux tower timing of seasonal peak AET ($TSP$) month or are mostly offset by one month (except for catchments 108002A, and 108003A near Cow Bay and Cape Tribulation flux sites, respectively). Figure 6b shows largely unchanged $TSP$ in models. However, SIMHYD and GR4J tend to show some improvements in estimating $TSP$ when incorporating AET information into the calibration.

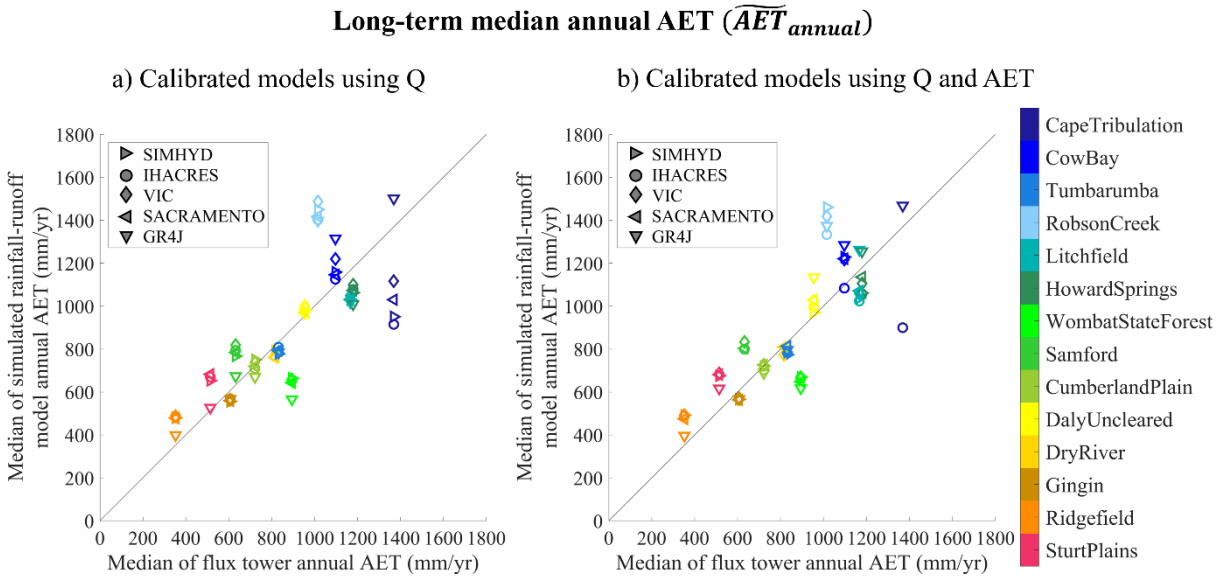



**Figure 3: Comparison of long-term median annual AET ($\widetilde{AET}_{annual}$) between models a) calibrated using streamflow (Q) only, and b) calibrated using Q and flux tower AET, and flux towers.**

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

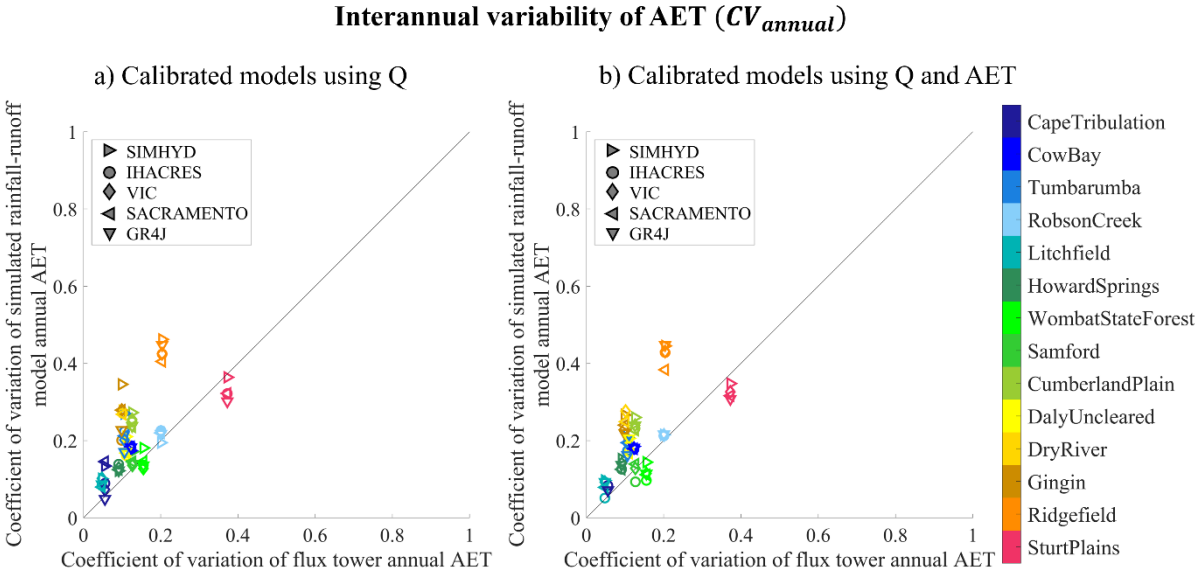



**Figure 4: Comparison of interannual variability of AET ($CV_{annual}$) between models a) calibrated using Q only, and b) calibrated using Q and flux tower AET, and flux towers.**

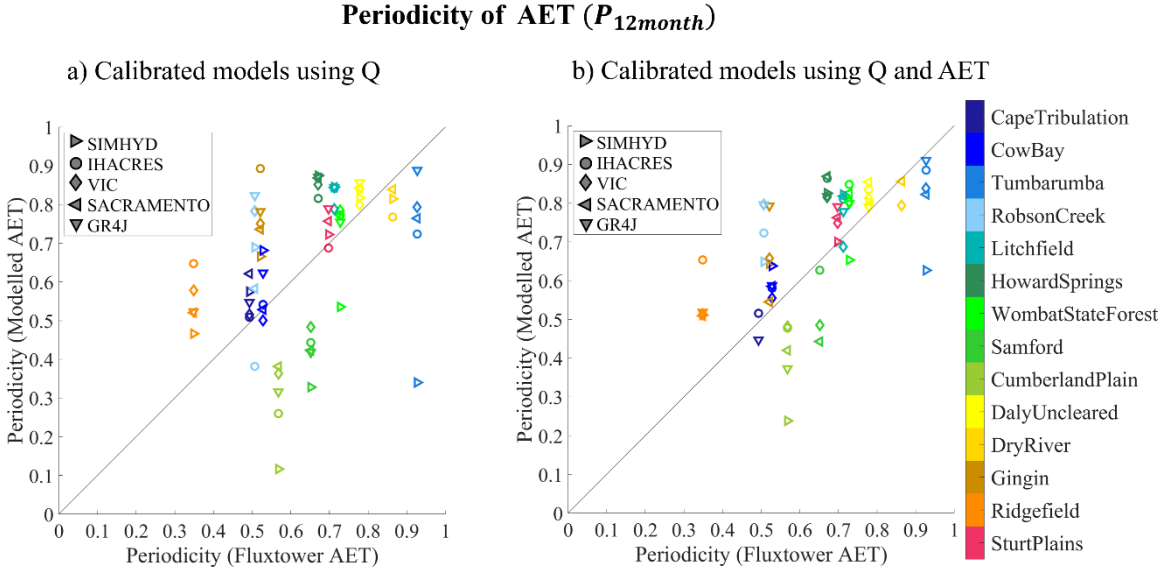



**Figure 5: Comparison of periodicity of AET ($P_{12month}$) between models a) calibrated using Q only, and b) calibrated using Q and flux tower AET, and flux towers.**

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

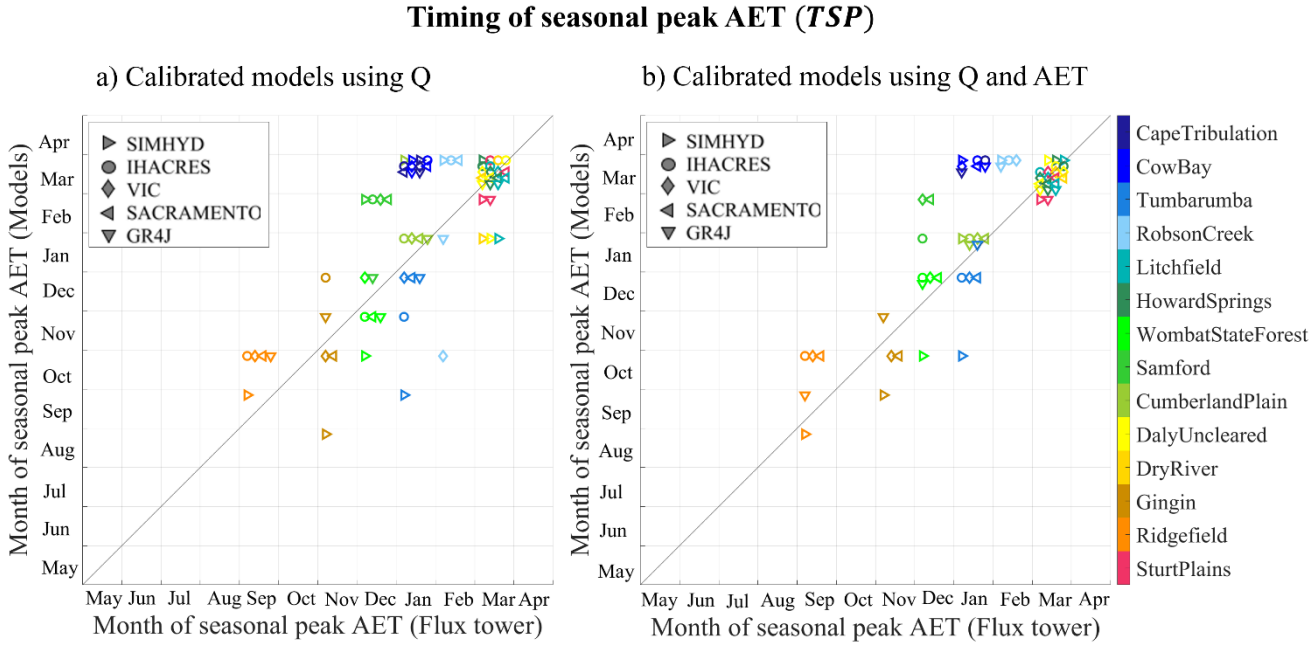


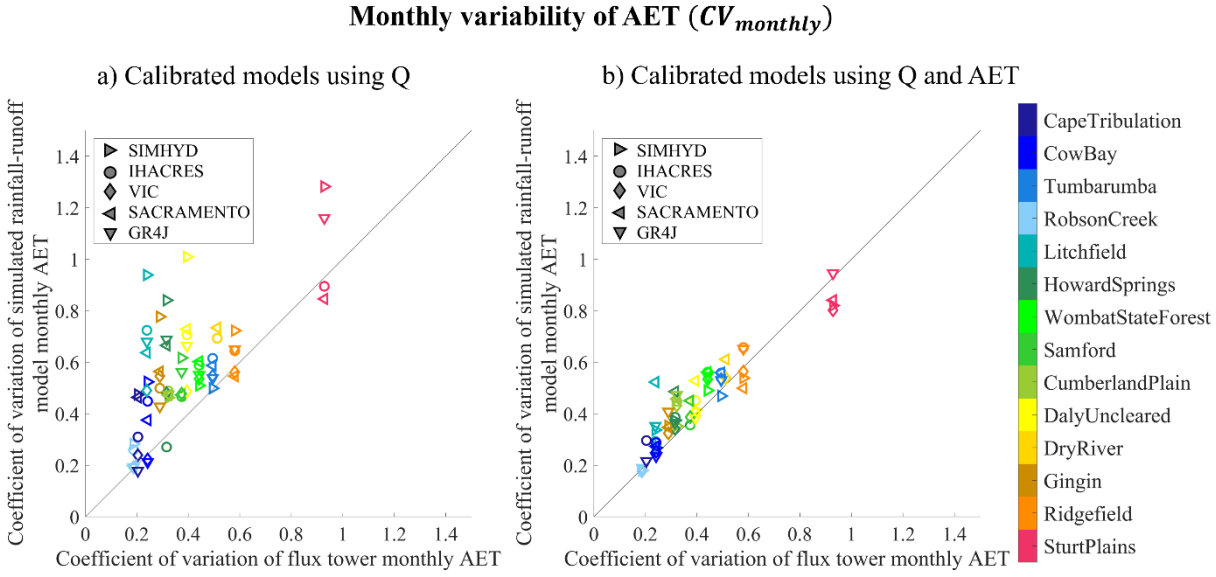

**Figure 6: Comparison of timing of seasonal peak (*TSP*) between models a) calibrated using Q only, and b) calibrated using Q and flux tower AET, and flux towers (Note that this signature takes integer values only (i.e., either one calendar month or the next), leading to several points overlying the same plotting position; to make every point visible we subject each point to a jitter (i.e., a unique offset within the same grid cell)).**

### 3.2.3 Comparison of AET signatures at monthly scale

Monthly scale AET signatures, such as monthly variability of AET ($CV_{monthly}$) and AET asynchronicity to PET from models ($AAP$), show significant improvement when AET information is incorporated into the calibration with streamflow. For example, models calibrated only with discharge data tend to significantly overestimate the $CV_{monthly}$ compared to flux tower estimates, as seen in Figure 7a. However, when AET information is included in the calibration process with streamflow data, the models capture $CV_{monthly}$ more realistically, as shown in Figure 7b.

Similar outcomes are observed in Figure 9, which shows improved $AAP$ when incorporating AET information into the calibration. Whereas the water stress signature ($WS$), which consider both the magnitude and asynchronicity between PET and AET in monthly timeseries, has similar estimates from models calibrated using both methods (Figure 8a, 8b). This is because the water stress signature quantifies the difference between AET and PET, and this is in turn a function of the long-term water balance. Since models calibrated to streamflow only must do a reasonable job of the long-term water balance, it is unsurprising that they can replicate this signature. The inclusion of additional AET information in the calibration in Figure 8b does not yield further improvement in modelling $WS$.

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.




280 **Figure 7: Comparison of monthly variability of AET ($CV_{monthly}$) between models a) calibrated using Q only, and b) calibrated using Q and flux tower AET, and flux towers.**

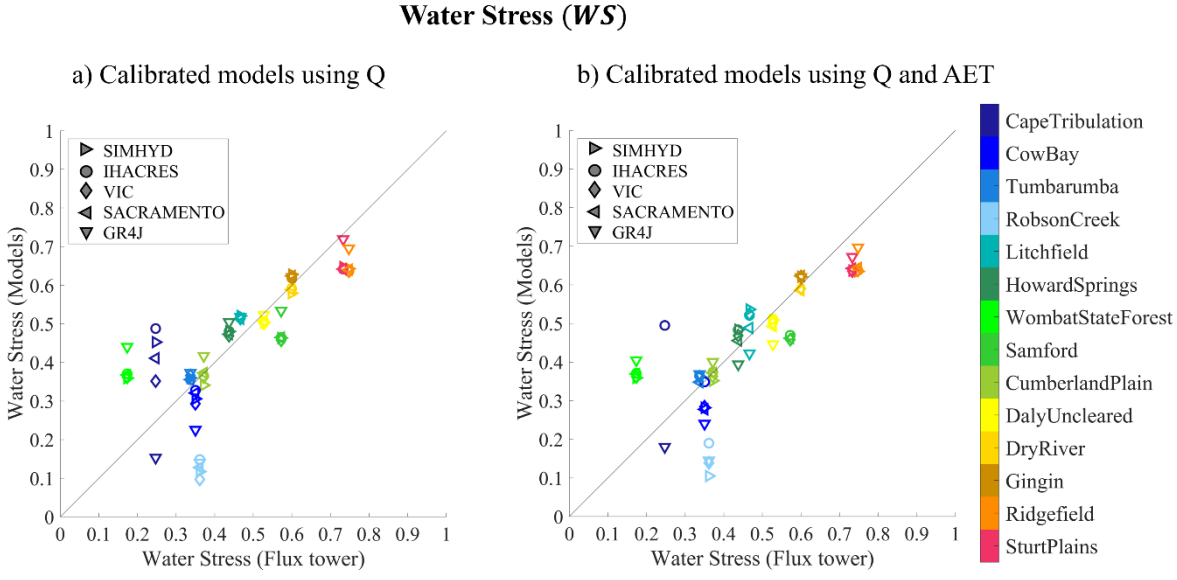



**Figure 8: Comparison of water stress ($WS$) estimated between models a) calibrated using Q only, and b) calibrated using Q and flux tower AET, and flux towers.**

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

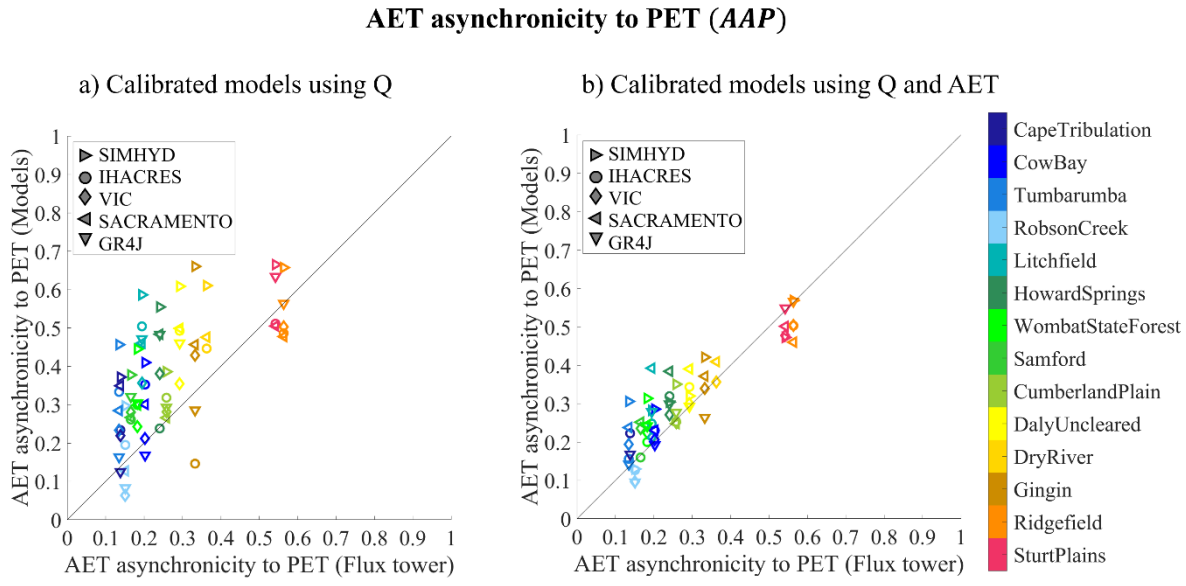



**Figure 9: Comparison of AET asynchronicity to PET (*AAP*) between models a) calibrated using Q only, and b) calibrated using Q and flux tower AET, and flux towers.**

### 3.2.4 Comparison of AET signatures at event scale

At the event scale, Figure 10 shows that models display a wide range of values in AET responsiveness to a rainfall event ($R$), both overestimating and underestimating the same signature calculated at flux tower sites in both the calibration methods. Moreover, SIMHYD tends to highly overestimate whereas GR4J tends to highly underestimate AET responsiveness to a rainfall event compared to flux tower values. The contrast among these models is likely to be related to the inclusion (or not) of an interception store. These results indicate that models exhibit unrealistic AET dynamics at the daily scale, and even incorporating daily AET information does not improve model simulations.

## 3.3 Synthesis of AET signatures performance across the sample of catchments

In this section, we characterise signatures according to the following idealised categories:

1. Signatures which models were already matching, even in streamflow-only calibration;
2. Signatures which models could match, but only after including AET in the calibration; and
3. Signatures which were not matched, even after including AET in the calibration.

Obviously, the results are unique to each model, so some signatures will not fit neatly into one of these categories. A seemingly obvious way to quantify signature matching would be to calculate an error metric, say, the root mean square error (RMSE), to quantify the deviation in signature value for simulated signatures compared to observed. However, each signature can take a different range of values, so it's difficult to know what a good or bad RMSE might be, a priori. Instead,

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

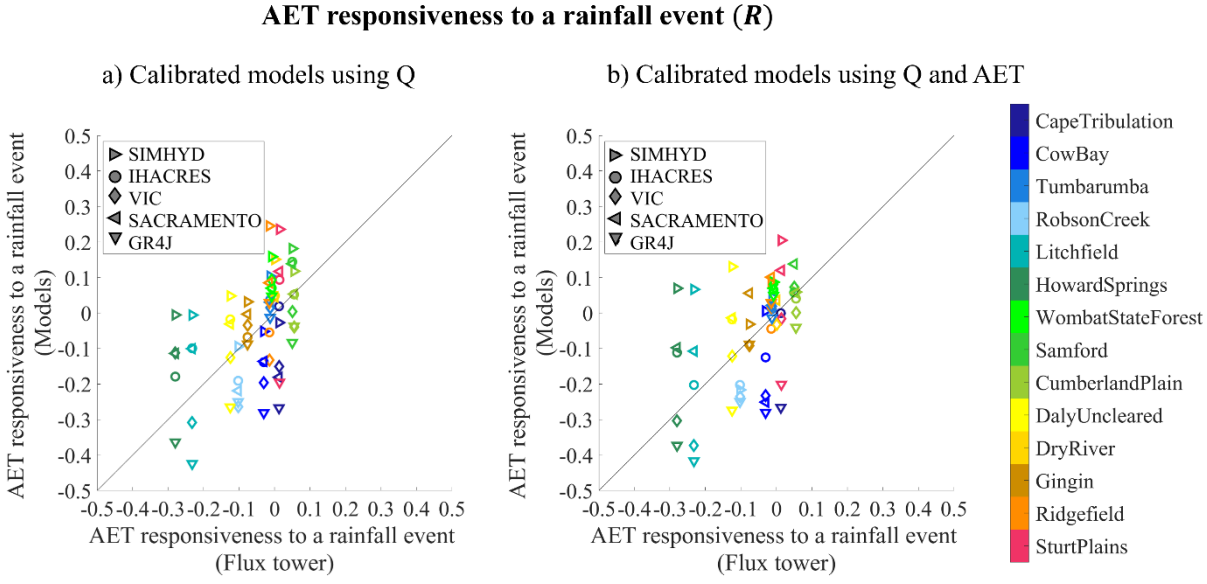



**Figure 10: Comparison of AET responsiveness to a rainfall event ($R$) between models a) calibrated using Q only, and b) calibrated using Q and flux tower AET, and flux towers.**

we report the Nash Sutcliffe Efficiency (NSE), noting that these NSE values should not be confused with commonly reported NSEs (i.e., those calculated based on observed and simulated streamflow on each simulation timestep); in contrast, these are calculated by comparing the observed and simulated signature values in each case. To get a high NSE, the error between simulated and observed signature values must be relatively small compared to the variance among the observed signatures for the 14 catchments (i.e., the spatial variance).

Following this definition, signatures in category 1 can be identified as those with high NSE in the first panel (Figure 11a) below. Figure 11a shows that all 5 models exhibit moderately high NSE values for the annual median (NSE range from 0.5-0.65) and water stress signatures (NSE range from 0.45 – 0.6), indicating moderate match with those calculated from observations. Moreover, GR4J shows an NSE of 0.44 for the timing of seasonal peak between simulations and observations.

Likewise, signatures in category 3 can be identified as those with low NSE in the second panel (Figure 11b) below. In Figure 11b, the models show at least a slight improvement in AET signatures after incorporating AET into the model calibration. However, all the models still prominently show poor $CV_{annual}$ values (NSE range from -2.16 - -0.34), even after those slight improvements. Moreover, SIMHYD, SACRAMENTO, and GR4J poorly estimate the AET responsiveness to rainfall events (NSE range from -2.23 - -0.34).

Lastly, signatures in category 2 can be identified as those with a large change in NSE in the third panel (Figure 11c). In Figure 11c, all models show a noticeable change in NSE for $CV_{monthly}$ (Change in NSE range from 0.94 – 3.6) and AET asynchronicity to PET ($AAP$, change in NSE range from 0.5 – 3.0), reconfirming the improvements in these two monthly

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

signatures. Moreover, the $P_{12month}$ of all models shows positive change in NSE (range from 0.28 – 1.67). Although these models did not significantly improve $P_{12month}$ in Method 2, which incorporated AET into the calibration, the improvement is comparatively higher than in Method 1. Therefore, it is important to note that a positive change in NSE does not necessarily mean the better match of AET dynamics between simulations and observations in Method 2 but rather the degree of improvement relative to Method 1. Moreover, the observed mismatch in AET signatures may not be necessarily inherent to the models chosen but could have arisen from the adopted calibration method. This is discussed further in Section 4.2.

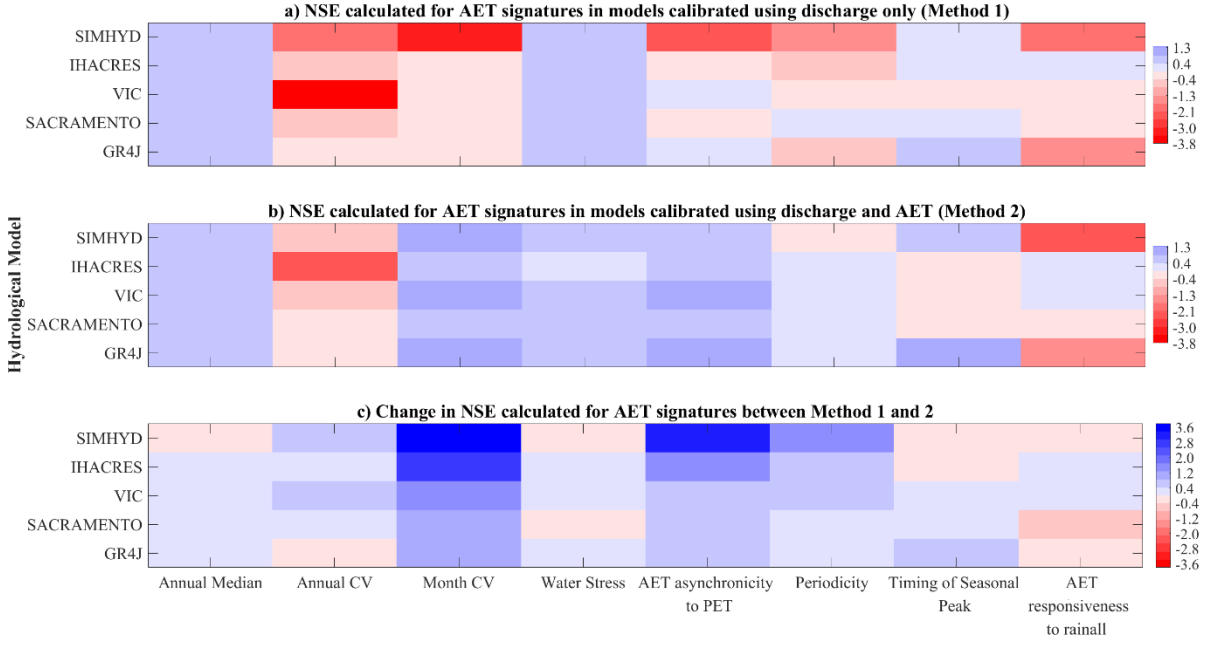



**Figure 11: Performance of models for each signature, assessed across the entire sample of 14 catchments, and expressed as a Nash Sutcliffe Efficiency (NSE) score calculated on observed and simulated signature values. (a) For calibration Method 1 (streamflow only); (b) for calibration Method 2 (streamflow and AET, but using traditional objective functions); (c) Change in NSE in moving from (a) to (b). Note, these Nash Sutcliffe Efficiencies should not be confused with the commonly reported NSEs which are calculated based on observed and simulated streamflow on each simulation timestep; in contrast, these are calculated by comparing the 14 observed and 14 simulated signature values in each case.**

### 3.4 Signature results for split sample testing

Figure 12a shows split sample results, showing the distribution of performance values calculated using the objective functions used in model calibration (Method 2) for streamflow (i.e., Eq. 1) and AET (i.e., Eq. 2) across the five models over the calibration and independent evaluation periods. Note that, similar to Section 2.5, only models with an optimised objective function value (i.e., Eq. 3) above a pre-defined threshold are shown here. Moreover, as mentioned in Section 2.7, the independent evaluation was based on AET signatures (Figure 12b and 12c), which only showed improvements in AET dynamics under Method 2, which used the full record of AET with streamflow in model calibration.

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

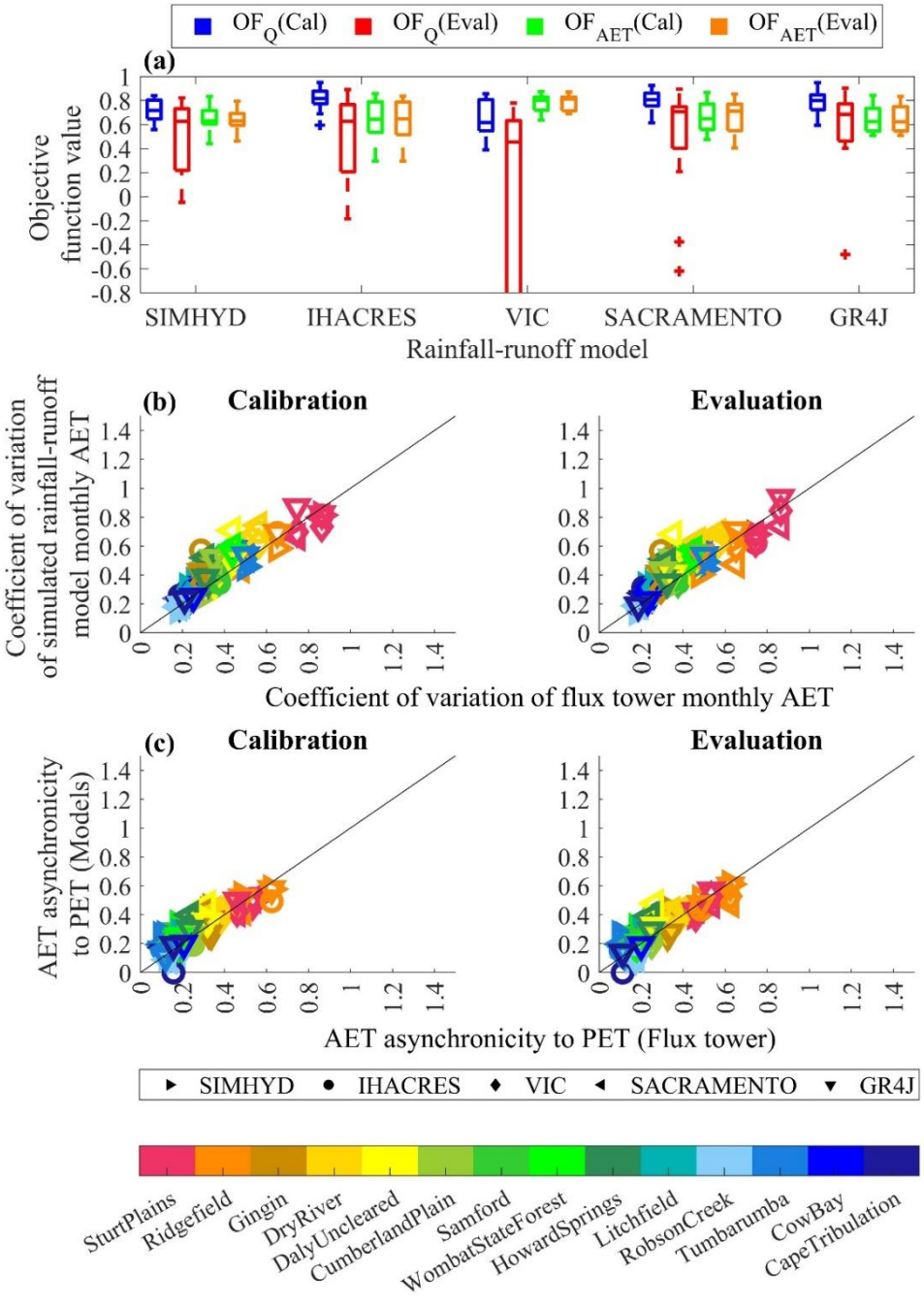



**Figure 12: a) Distribution of objective function values for streamflow and AET during the calibration and evaluation periods in models used for AET signature calculations, b) Coefficient of variation (CV) of monthly simulated AET vs CV of monthly flux tower AET over calibration and independent evaluation periods, c) AET aysnchroncity to PET (AAP) calculated using simulated monthly AET vs AAP calculated using monthly flux tower over the calibration and independent evaluation periods. (Note that two separate calibrations and evaluations were conducted using split sampling data, and the objective function values and signature results are combined in each category. Independent plots are available in Section S4 of Supplementary material).**

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

According to Figure 12a, all models exhibit significantly reduced model performance in simulating streamflow over independent evaluation, a phenomenon commonly observed in split-sample calibrations, even for model calibrations using only streamflow.

In contrast, the distribution of AET objective function values remains relatively consistent between calibration and evaluation periods across all models, suggesting more stable AET simulations. This is also apparent from the signature results. Figures 12b and 12c compare the coefficient of variation of monthly AET ($CV_{monthly}$) and AET asynchronicity to PET ($AAP$) between rainfall-runoff models and flux towers, respectively, over both calibration and independent evaluation periods. Both results confirm that the improvements in AET dynamics observed using Method 2 of calibration, using both streamflow and AET in Section 3.2.3, are even valid over split sampling calibration and over the independent evaluation. Overall, it is clear that AET performance is more stable in conceptual rainfall-runoff models than streamflow performance, and the possible reasons for this are discussed under Section 4.4.

## 4 Discussion

This study assessed five conceptual rainfall-runoff models' abilities to simulate actual evapotranspiration realistically by employing evapotranspiration signatures across a set of sites with varying climatic conditions. To our knowledge, this is the first study that has used AET signatures in conceptual rainfall-runoff model assessment. Overall, the study demonstrated that commonly used rainfall-runoff models are capable of improving evapotranspiration dynamics, such as monthly variability of AET and AET asynchronicity to PET by integrating additional AET data into the model calibration, but the other aspects of AET performance are not improved and thus require further actions to ensure realism.

### 4.1 AET signature-based assessment of modelled AET dynamics and calibration strategies

The results of this study show how realistically AET is simulated in commonly used conceptual rainfall-runoff models. According to the findings, these models fail to realistically capture most of the AET dynamics at different temporal scales, especially when calibrated conventionally with daily streamflow data. For example, we found significant errors associated with predicting AET dynamics at the seasonal scale, such as periodicity and month of seasonal peak, and these specific aspects support the findings of Guo et al. (2017). Similarly, challenges exist in representing AET variability, water stress, and AET asynchronicity to PET on the monthly scale. Our signature-based evaluation method shows that the common practice of calibrating to a single objective function (in this case jointly with streamflow) (e.g., Zhang et al., 2009) improves only some aspects of AET dynamics in models. For example, only the monthly variability and AET asynchronicity to PET were found to be significantly improved after incorporating AET into the calibration. These findings hold regardless of whether we use the full dataset or a more representative subset of flux towers and catchments. Moreover, the improvement in AET dynamics resulting from the combined streamflow and AET calibration is further supported by independent evaluation

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

using split-sample calibration. These findings suggest that either the calibration strategy or deficiencies in model structures may hamper these other aspects of AET representation in models.

### 4.2 Potential for improvements with better calibration

One could argue that directly incorporating AET signatures into the calibration process might better enhance certain aspects of AET behaviour that were not improved with the joint calibration of streamflow and AET. While this is a valid argument, it is out of the scope of this study. A common challenge in this context is determining which signatures to calibrate, as if we choose multiple signatures, there is a risk of diluting the beneficial effect. Two possible approaches are the method developed by Sadegh & Vrugt (2014), which employs the sampling technique called Differential Evolution Adaptive Metropolis (DREAM) within signature-based Approximate Bayesian Computation to delineate a behavioural model space. Alternatively, Shafii & Tolson (2015) develop a scoring method that converts hydrological signatures into a calibration objective, aiming to maximise hydrological consistency in models.

### 4.3 AET conceptualisation in conceptual rainfall-runoff models - Potential for model structural improvements

We could argue that if the model cannot match AET dynamics with observations, even when directly calibrated to AET signatures, then there must be deficiencies in the model structure. As mentioned in the introduction, AET is often calculated based on storage; hence, poor representation of storage in models can lead to errors in AET representation. For example, this logic has been applied to some existing structural modifications, such as those by Grigg & Hughes (2018), which showed improved model performance in GR4J by modifying more flexibility in storage-dependent AET through the creation of catchment memory. A similar approach is seen Fowler et al. (2021b), where they modified the simulated soil moisture in GR4J, converting it to a deficit system with no upper limit to catchment moisture deficit. This modification was mainly focused on improving the problem related to multiannual storage change but also allowed for a more realistic representation of AET dynamics. Future studies could focus on improving AET representation in existing model frameworks and could explore options among existing AET equations in models as a start to better capture AET behaviour in rainfall-runoff models.

### 4.4 Stability of AET and streamflow performance during split sample testing

The multi-objective test during split sample testing revealed that AET performance is more stable than streamflow performance. This AET stability in models is also supported by other studies (e.g., Arciniega-Esparza et al. (2022); Meyer Oliveira et al. (2021)), who employed AET in model calibration and evaluation. However, the underlying reasons for this are not fully understood.

One possible explanation could be the greater influence of error structure on streamflow simulations compared to AET simulations. For example, objective functions used in calibration are often sensitive to flood events and the associated errors. As a result, during calibration, streamflow peaks may be matched through overfitting, leading to poor performance during

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

independent evaluation periods. In contrast, AET appears to lack significant error structures that would influence its performance across evaluation periods.

### 4.5 Limitations and further studies

While this study advances the understanding and identifies deficiencies in the representation of AET in models employing AET signatures in the assessment, the limitations and areas of future studies can be highlighted as follows:

- *No direct model calibration to the AET signatures*: The scope of this study was to employ AET signatures in the evaluation of models that had been calibrated in a way that is broadly typical of common practice; that is, calibrated to a single AET objective function (jointly with a single streamflow objective function). Future studies can consider using these signatures directly in the model calibration to assess improvements in AET simulations before making any structural changes.

- *Possibility of other models providing different implications for modelled AET dynamics*: This study used five commonly used conceptual rainfall-runoff models to assess their AET representation. However, there is a possibility that other models could yield different implications regarding AET representation, as we did not explore a broader range of models. Future studies could extend this analysis by incorporating a wide range of models to derive more generalized implications for AET representation in conceptual rainfall-runoff models.

- *Examining implications of using flux tower versus remote sensing information in model calibration.* Here, we used the best available AET information, namely, flux tower data. For most studies, flux towers are not available for the site of interest, meaning that remotely sensed information is the only option for model calibration. Given the discrepancies noted by Gardiya Weligamage et al. (2025) between remotely sensed and flux tower estimates, future studies could aim to determine whether calibrating to remotely sensed information actually adds value to the overall process.

## 5 Conclusion

This study evaluated the realism of AET representation in five commonly used conceptual rainfall-runoff models using AET signatures. To achieve this, we selected 14 study catchments near flux towers in Australia to simulate AET by calibrating models using discharge data alone, and in combination with flux tower AET data. AET signatures derived from both calibration methods were then compared with those calculated using flux tower AET. The results indicate that the models fail to realistically represent most of the AET dynamics at different temporal scales when calibrated with discharge only. However, the joint calibration of streamflow with AET improves some aspects of AET dynamics in models, particularly the monthly variability of AET and AET asynchronicity to PET, which extends even to the independent evaluation period. Despite this, other AET signatures derived from models did not show significant improvement in AET dynamics, even with

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

the inclusion of AET information into the calibration. Future investigations could examine AET representation in models using a wider range of calibration strategies, such as calibrating models directly to AET signatures to determine if models can reproduce the AET dynamics expressed in these signatures; in addition, future work may also consider conceptual rainfall-runoff model improvements. We recommend future studies employ these AET signatures to evaluate models and improve AET model components for more realistic process representations in rainfall-runoff models.

## Data availability

The simulated actual evapotranspiration data, derived from five conceptual rainfall-runoff models and their input data, are publicly available at https://doi.org/10.5281/zenodo.14228809

## Author contribution

HGW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft preparation, Writing – review and editing.
KF: Conceptualization, Methodology, Funding acquisition, Supervision, Writing – review and editing
MS: Conceptualization, Methodology, Funding acquisition, Supervision, Writing – review and editing
TP: Conceptualization, Methodology, Funding acquisition, Supervision, Project administration, Writing – review and editing
DR: Conceptualization, Methodology, Supervision, Writing – review and editing
MP: Conceptualization, Methodology, Funding acquisition, Supervision, Project administration, Writing – review and edit

## Competing interests

The authors declare that they have no conflict of interest.

## Acknowledgements

This research was funded by the Australian Research Council (ARC) Linkage Project (LP180100796) with partner organisations Victorian Department of Environment, Land, Water and Planning (DELWP), and Melbourne Water.

## References

Araki, R., Branger, F., Wiekenkamp, I., & McMillan, H.: A signature-based approach to quantify soil moisture dynamics under contrasting land-uses. *Hydrological Processes*, *36*(4), 1–21. https://doi.org/10.1002/hyp.14553, 2022.

Arciniega-Esparza, S., Birkel, C., Chavarría-Palma, A., Arheimer, B., & Breña-Naranjo, J. A.: Remote sensing-aided rainfall-runoff modeling in the tropics of Costa Rica. *Hydrology and Earth System Sciences*, 26(4), 975–999.

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

https://doi.org/10.5194/hess-26-975-2022,2022.

Arsenault, R., Poulin, A., Côté, P., & Brissette, F.: Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration. *Journal of Hydrologic Engineering*, *19*(7), 1374–1384. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000938, 2014.

Avanzi, F., Rungee, J., Maurer, T., Bales, R., Ma, Q., Glaser, S., & Conklin, M.: Climate elasticity of evapotranspiration shifts the water balance of Mediterranean climates during multi-year droughts. *Hydrology and Earth System Sciences*, *24*(9), 4317–4337. https://doi.org/10.5194/hess-24-4317-2020, 2020.

Beringer, J., Hutley, L. B., McHugh, I., Arndt, S. K., Campbell, D., Cleugh, H. A., Cleverly, J., Resco de Dios, V., Eamus, D., Evans, B., Ewenz, C., Grace, P., Griebel, A., Haverd, V., Hinko-Najera, N., Huete, A., Isaac, P., Kanniah, K., Leuning, R., … Wardlaw, T.: An introduction to the Australian and New Zealand flux tower network – OzFlux. *Biogeosciences*, *13*(21), 5895–5916. https://doi.org/10.5194/bg-13-5895-2016, 2016.

Burnash, R. J. C., & Ferral, R. L.: *A generalized streamflow simulation system: Conceptual modeling for digital computers*. US Department of Commerce, National Weather Service, and State of California, Department of Water Resources, 1973.

Chiew, F H S, Peel, M. C., & Western, A. W.: Application and testing of the simple rainfall-runoff model SIMHYD. In V. P. Singh & D. Frevert (Eds.), *Mathematical models of small watershed hydrology and applications* (pp. 335–367). Water Resources Publications, 2002.

Chiew, Francis H. S., Vaze, J., Viney, N. R., Perraud, J., Teng, J., & Post, D. A.: Modelling Runoff and Climate Change Impact on Runoff in 178 Catchments in the Murray-Darling Basin Using Sacramento and SIMHYD Rainfall-Runoff Models. *Water Down Under 2008*, 87–97, 2008.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., & Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*(12), 1–14. https://doi.org/10.1029/2007WR006735, 2008.

Croke, B. F. W., & Jakeman, A. J. (2004). *A catchment moisture deficit module for the IHACRES rainfall- runoff model*. *19*, 1–5. https://doi.org/10.1016/j.envsoft.2003.09.001, 2004.

Deb, P., Kiem, A. S., & Willgoose, G.: A linked surface water-groundwater modelling approach to more realistically simulate rainfall-runoff non-stationarity in semi-arid regions. *Journal of Hydrology*, *575*(May), 273–291. https://doi.org/10.1016/j.jhydrol.2019.05.039, 2019.

Dye, P. J., & Croke, B. F. W.: Evaluation of streamflow predictions by the IHACRES rainfall-runoff model in two South African catchments. *Environmental Modelling & Software*, *18*(8–9), 705–712. https://doi.org/10.1016/S1364-8152(03)00072-0, 2003.

Fowler, K., Acharya, S. C., Addor, N., Chou, C., & Peel, M.: CAMELS-AUS v1: Hydrometeorological time series and landscape attributes for 222 catchments in Australia. [Dataset]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.921850, 2020.

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

Fowler, K. J. A., Acharya, S. C., Addor, N., Chou, C., & Peel, M. C.: CAMELS-AUS: Hydrometeorological time series and landscape attributes for 222 catchments in Australia. *Earth System Science Data*, *13*(8), 3847–3867. https://doi.org/10.5194/essd-13-3847-2021, 2021a.

Fowler, K. J. A., Coxon, G., Freer, J. E., Knoben, W. J. M., Peel, M. C., Wagener, T., Western, A. W., Woods, R. A., & Zhang, L.: Towards more realistic runoff projections by removing limits on simulated soil moisture deficit. *Journal of Hydrology*, *600*(May 2021), 126505. https://doi.org/10.1016/j.jhydrol.2021.126505, 2021b.

Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., Seo, K., & Western, A.: Many Commonly Used Rainfall-Runoff Models Lack Long, Slow Dynamics: Implications for Runoff Projections. *Water Resources Research*, *56*(5), e2019WR025286. https://doi.org/10.1029/2019WR025286, 2020.

Gardiya Weligamage, H., Fowler, K., Saft, M., Peterson, T., Ryu, D., & Peel, M.: Characterising evapotranspiration signatures for improved behavioural insights. *Hydrology and Earth System Sciences Discussions*, 1-23, https://doi.org/10.5194/hess-2024-373, 2025.

Gardiya Weligamage, H., Fowler, K., Peterson, T. J., Saft, M., Peel, M. C., & Ryu, D.: Partitioning of Precipitation Into Terrestrial Water Balance Components Under a Drying Climate. *Water Resources Research*, *59*(5), e2022WR033538. https://doi.org/10.1029/2022WR033538, 2023.

Gou, J., Miao, C., Duan, Q., Tang, Q., Di, Z., Liao, W., Wu, J., & Zhou, R.: Sensitivity Analysis-Based Automatic Parameter Calibration of the VIC Model for Streamflow Simulations Over China. *Water Resources Research*, *56*(1), 1–19. https://doi.org/10.1029/2019WR025968, 2020.

Grigg, A. H., & Hughes, J. D.: Nonstationarity driven by multidecadal change in catchment groundwater storage: A test of modifications to a common rainfall–run-off model. *Hydrological Processes*, *32*(24), 3675–3688. https://doi.org/10.1002/hyp.13282, 2018.

Gunkel, A., Shadeed, S., Hartmann, A., Wagener, T., & Lange, J.: Model signatures and aridity indices enhance the accuracy of water balance estimations in a data-scarce Eastern Mediterranean catchment. *Journal of Hydrology: Regional Studies*, *4*, 487–501. https://doi.org/10.1016/j.ejrh.2015.08.002, 2015.

Guo, D., Westra, S., & Maier, H. R.: Impact of evapotranspiration process representation on runoff projections from conceptual rainfall-runoff models. *Water Resources Research*, *53*(1), 435–454. https://doi.org/10.1002/2016WR019627, 2017.

Gupta, H. V., Sorooshian, S., & Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763. https://doi.org/10.1029/97WR03495, 1998.

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

Hansen, N., Müller, S. D., & Koumoutsakos, P.: Reducing the Time Complexity of the Derandomized Evolution Strategy

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

with Covariance Matrix Adaptation (CMA-ES). *Evolutionary Computation*, *11*(1), 1–18. https://doi.org/10.1162/106365603321828970, 2003.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., … Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)-a review. *Hydrological Sciences Journal*, *58*(6), 1198–1255. https://doi.org/10.1080/02626667.2013.803183, 2013.

Isaac, P., Cleverly, J., McHugh, I., van Gorsel, E., Ewenz, C., & Beringer, J.: OzFlux data: network integration from collection to curation. *Biogeosciences*, *14*(12), 2903–2928. https://doi.org/10.5194/bg-14-2903-2017, 2017.

Jeffrey, S. J., Carter, J. O., Moodie, K. B., & Beswick, A. R.: Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling and Software*, *16*(4), 309–330. https://doi.org/10.1016/S1364-8152(01)00008-1, 2001.

Kiraz, M., Coxon, G., & Wagener, T.: A Signature-Based Hydrologic Efficiency Metric for Model Calibration and Evaluation in Gauged and Ungauged Catchments. *Water Resources Research*, *59*(11). https://doi.org/10.1029/2023WR035321, 2023.

Lymburner, L., Tan, P., Mueller, N., Thackway, R., Lewis, A., Thankappan, M., Randall, L., Islam, A., & Senarath, U.: *250 metre Dynamic Land Cover Dataset (1st Edition)*. Geoscience Australia, 2010.

Meyer Oliveira, A., Fleischmann, A. S., & Paiva, R. C. D.: On the contribution of remote sensing-based calibration to model hydrological and hydraulic processes in tropical regions. *Journal of Hydrology*, 597. https://doi.org/10.1016/j.jhydrol.2021.126184, 2021.

McMillan, H.: Linking hydrologic signatures to hydrologic processes: A review. *Hydrological Processes*, *34*(6), 1393–1409. https://doi.org/10.1002/hyp.13632, 2020.

McMillan, H. K.: A review of hydrologic signatures and their applications. *WIREs Water*, *8*(1), 1–23. https://doi.org/10.1002/wat2.1499, 2021.

Morton, F. I.: Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. *Journal of Hydrology*, *66(1-4)*, 1–76. https://doi.org/10.1016/0022-1694(83)90177-4, 1983.

Peel, M. C., McMahon, T. A., Finlayson, B. L., & Watson, F. G. R.: Identification and explanation of continental differences in the variability of annual runoff. *Journal of Hydrology*, *250*(1–4), 224–240. https://doi.org/10.1016/S0022-1694(01)00438-3, 2001.

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

Poff, N. L. R., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E., & Stromberg, J. C.: The natural flow regime: A paradigm for river conservation and restoration. *BioScience*, *47*(11), 769–784. https://doi.org/10.2307/1313099, 1997.

Refsgaard, J. C., & Knudsen, J.: Operational Validation and Intercomparison of Different Types of Hydrological Models.

https://doi.org/10.5194/egusphere-2025-3373
Preprint. Discussion started: 18 July 2025
© Author(s) 2025. CC BY 4.0 License.

*Water Resources Research*, *32*(7), 2189–2202. https://doi.org/10.1029/96WR00896, 1996.

Roy, T., Gupta, H. V., Serrat-Capdevila, A., & Valdes, J. B.: Using satellite-based evapotranspiration estimates to improve the structure of a simple conceptual rainfall-runoff model. *Hydrology and Earth System Sciences*, *21*(2), 879–896. https://doi.org/10.5194/hess-21-879-2017, 2017.

Sugawara, M.: Automatic calibration of the tank model. *Hydrological Sciences Bulletin*, *24*(3), 375–388. https://doi.org/10.1080/02626667909491876, 1979.

Trotter, L., Knoben, W. J. M., Fowler, K. J. A., Saft, M., & Peel, M. C.: Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v2.1: an object-oriented implementation of 47 established hydrological models for improved speed and readability. *Geoscientific Model Development*, *15*(16), 6359–6369. https://doi.org/10.5194/gmd-15-6359-2022, 2022.

Trotter, L., Saft, M., Peel, M. C., & Fowler, K. J. A.: Symptoms of Performance Degradation During Multi-Annual Drought: A Large-Sample, Multi-Model Study. *Water Resources Research*, *59*(2), 1–23. https://doi.org/10.1029/2021WR031845, 2023.

Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J., Viney, N. R., & Teng, J.: Climate non-stationarity – Validity of calibrated rainfall – runoff models for use in climate change studies. *Journal of Hydrology*, *394*(3–4), 447–457. https://doi.org/10.1016/j.jhydrol.2010.09.018, 2010.

Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., & Xu, C. Y.: Calibration of hydrological models using flow-duration curves. *Hydrology and Earth System Sciences*, *15*(7), 2205–2227. https://doi.org/10.5194/hess-15-2205-2011, 2011.

Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H. H. G., Uhlenbrook, S., Kavetski, D., & Pfister, L.: Towards more systematic perceptual model development: a case study using 3 Luxembourgish catchments. *Hydrological Processes*, *29*(12), 2731–2750. https://doi.org/10.1002/hyp.10393, 2015.

Zhang, Y., Chiew, F. H. S., Zhang, L., & Li, H.: Use of remotely sensed actual evapotranspiration to improve rainfall-runoff modeling in Southeast Australia. *Journal of Hydrometeorology*, *10*(4), 969–980. https://doi.org/10.1175/2009JHM1061.1, 2009.