https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

SPatial Efficiency And Kmoments (SPEAK): Evaluating Spatial Consistency in (Semi)Distributed Rainfall–Runoff Models

Matías Moreno[1], Pablo Mendoza[2,3], Eduardo Muñoz-Castro[4,5,6], Mauricio Zambrano-Bigiarini[7], Alonso Pizarro[8]

[1] Independent Engineer and Researcher, Santiago, Chile
[2] Civil Engineering Department, Universidad de Chile, Santiago, Chile
[3] Advanced Mining Technology Centre (AMTC), Universidad de Chile, Santiago, Chile
[4] WSL Institute for Snow and Avalanche Research SLF, Davos Dorf, Switzerland
[5] Climate Change, Extremes and Natural Hazards in Alpine Regions Research Center CERC, Davos Dorf, Switzerland
[6] Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[7] Department of Civil Engineering, Universidad de la Frontera, Temuco, Chile
[8] Department of Civil and Environmental Engineering, Universidad del Bío-Bío, Concepción, Chile

Correspondence to: Alonso Pizarro (alonso.pizarro@ubiobio.cl)

Abstract. We introduce the Spatial Efficiency and Kmoments (SPEAK) metric, a novel objective function for the spatial calibration of hydrological models. SPEAK is built on Kmoment-based statistics, including a Kmoment-based: i) correlation, ii) coefficient-of-variation ratio, and iii) probability density function. This novel formulation is explicitly designed to overcome key limitations of existing spatial performance metrics, such as sensitivity to binning strategies, grid resolution, and sample heterogeneity. By relying on distributional properties rather than grid-to-grid correspondence, SPEAK provides a statistically robust framework for evaluating spatial patterns in gridded hydrological variables. The proposed metric is implemented in both semi-distributed and fully distributed configurations of the TUW hydrological model and tested across 99 near-natural Chilean catchments that encompass strong climatic and physiographic gradients. Actual evapotranspiration (ETa) from GLEAM v4.2a is used as an independent spatial benchmark, allowing the assessment of model performance beyond streamflow reproduction. Calibration using SPEAK is compared with a conventional streamflow-only calibration based on the Kling-Gupta Efficiency (KGE) and an ETa-only calibration based on the Spatial Efficiency metric (SPAEF). Model performance is evaluated using the normalised root-mean-square error (NRMSE), the spatial Pearson correlation coefficient, the Fraction Skill Score (FSS), and sensitivity to catchment attributes. Results demonstrate that while streamflow-only calibration leads to satisfactory runoff simulations ($KGE \geq 0.25$ for all catchments and cases analysed; whereas the mean and median KGE are 0.80 and 0.85, respectively), it fails to reproduce the spatial patterns of ETa. When ETa is used as a calibration target, SPEAK consistently outperforms SPAEF, exhibiting lower NRMSE (number of catchments with lower NRMSE: 85 and 92 in fully and semi-distributed configuration, respectively), reduced internal component dispersion, and improved representation of spatial patterns across seasons and hydroclimatic zones. Importantly, SPEAK shows limited dependence on catchment characteristics. These findings highlight SPEAK as a methodologically robust spatial performance metric, with clear potential for improving the calibration and diagnosis of distributed hydrological models and other gridded environmental variables.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

1 Introduction

Hydrological models are essential tools for understanding (e.g., Clark et al., 2011; Perrini et al., 2025) and predicting (e.g., Beck et al., 2020, Y. Guo et al., 2021, Wang et al., 2021, Guse et al., 2024) hydrological processes across different spatiotemporal scales (e.g., Huang et al., 2019, Aerts et al., 2022, Song et al., 2024), providing insights for effective water resources management (e.g., Baker et al., 2021; Hurkmans et al., 2023) and risk mitigation (e.g., Komma et al., 2008; Mendoza et al., 2012; Shi et al., 2024).

The calibration of model parameters plays a key role in the robustness of spatially distributed hydrological simulations, including runoff dynamics, spatial patterns and statistical distributions of internal model states and fluxes. Nevertheless, calibration methods often rely solely on streamflow observations or catchment-scale evaluations, dismissing internal dynamics and the correct reproduction of spatial patterns of hydrological variables (Rajib et al., 2018, Jin & Jin, 2020, Shah et al., 2021). Classic examples include the Nash-Sutcliffe Efficiency (NSE; Nash & Sutcliffe, 1970) and the Kling-Gupta Efficiency (KGE; Gupta et al., 2009), as well as subsequent variants (e.g., Fowler et al., 2018; Kling et al., 2012; Pool et al., 2018; Tang et al., 2021), which are widely applied to assess model performance based only on streamflow data. Hence, these metrics may yield acceptable streamflow simulations but may misrepresent the spatial variability of key hydrological processes (Beven, 2006; Hsu et al., 2025; Savenije and Hrachowitz, 2021).

Over the last few decades, the increasing availability of gridded products has provided spatially distributed estimates of key hydrological variables, including precipitation (Belay et al., 2022; Pradhan et al., 2022), evapotranspiration (Stisen et al., 2021; Tran et al., 2023), soil moisture (Dorigo et al., 2017; Karami et al., 2026; Miralles et al., 2025), and snow cover (Bonney and Zhang, 2026; Xiao et al., 2024). These datasets enable the explicit evaluation of spatial patterns and temporal dynamics at multiple scales, offering an unprecedented opportunity to complement traditional streamflow-based metrics. Consequently, gridded observations can be incorporated as additional calibration targets, allowing rainfall-runoff models to be constrained not only by streamflow at catchment outlets but also by the spatial distribution of internal states and fluxes (e.g., Dembélé et al., 2020a; Demirel et al., 2023; Hsu et al., 2025). However, achieving a robust calibration of hydrological models remains a challenging task (Kirchner, 2006; Klemeš, 1986), especially when attempting to simulate the spatial patterns of internal model states and fluxes (see, e.g., Pimentel et al., 2023; Rakovec et al., 2016). The latter due to their high spatiotemporal variability, shaped by factors such as climate, topography, soil characteristics, and land cover.

In recent years, the Spatial Efficiency (SPAEF; Koch et al., 2018) has become increasingly popular to assess the ability of hydrological models to capture spatial patterns (Gómez et al., 2024; Karpasitis et al., 2025). This metric simultaneously accounts for spatial correlation, the coefficient of variation, and histogram overlap between reference and simulated variables (Koch et al., 2018). Previous studies have shown that SPAEF may improve the representation of ETa spatial patterns in catchments characterised by diverse climatic and topographic conditions when used as an objective function in model parameter calibration (Koch et al., 2018, Stisen et al., 2021, Yorulmaz et al., 2023, Demirel et al., 2024, Yorulmaz et al., 2024). However, SPAEF has shown some limitations, such as high sensitivity to binning, issues with skewed data and outliers, among others. Table 1 shows a compilation of different SPAEF-based proposed metrics to overcome some of the mentioned limitations. It is worth mentioning that the SPAEF metric – and some of its proposed modifications – have been meticulously assessed by Yorulmaz et al. (2024).

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

Table 1: Metrics formulated to assess model performance in terms of the capability to replicate spatial patterns in simulated variables. The metrics are presented arbitrarily in chronological order.

Metric	Modification	Reference
SPatial efficiency (SPAEF)	Original	Koch et al. (2018)
SPatial pattern efficiency metric (E_{SP})*	Incorporates Spearman correlation and RMSE between standardised (z-scores) maps	Dembélé et al. (2020b)
SPatial efficiency prime (SPAEF')	Dynamic number of bins instead of n = 100	Yorulmaz et al. 2024)
SPatial count density efficiency (SPACD)	Replace histogram intersection with count-density normalisation	Yorulmaz et al. 2024)
SPatial hybrid 4 efficiency (SPAH4)	Consideration of a fourth component: the kurtosis ratio	Yorulmaz et al. 2024)
SPatial kurtosis efficiency (SPAK)	Replaces the histogram intersection with the kurtosis ratio	Yorulmaz et al. 2024)
SPatial hybrid 5 efficiency (SPAH5)	Adds both kurtosis and skewness ratios as fourth and fifth components	Yorulmaz et al. 2024)
SPatial histogram equalisation efficiency (SPAHE)	Applies histogram equalisation to the input maps before computing the intersection	Yorulmaz et al. 2024)
SPatial movers' distance efficiency (SPAMD)	Replaces the histogram intersection with Earth Mover's Distance (EMD)	Yorulmaz et al. 2024)
Wasserstein SPAtial Efficiency (WSPAEF)	Replaces the histogram intersection with the Wasserstein distance	Gómez et al. (2024)
Modified Spatial Efficiency (MSPAEF)	Bias-sensitive and at the same time scale-independent. Formulation with four terms	Karpasitis et al. (2025)

*also known as SPEM.

Although many improvements have been proposed to the original SPAEF formulation (Table 1), several structural limitations remain in current spatial performance metrics. Most variants still rely on histogram-based representations for comparing distributions, which are inherently sensitive to bin size, data resolution, and sample heterogeneity, thereby limiting their ability to fully capture continuous spatial and statistical variability. Such dependencies often introduce discontinuities and inconsistencies across catchments of different sizes or grid densities. Further, although incorporating higher-order statistical descriptors and distance metrics (e.g., kurtosis ratio, Earth Mover's Distance, or Wasserstein distance) has enhanced the original SPAEF formulation, these formulations still rely on discrete grid comparisons and therefore struggle to represent the spatial correlation and variability of key hydrological variables. In short, despite the valuable advances of existing metrics, a more flexible and theoretically consistent framework is still needed to robustly capture spatial patterns of simulated model states and fluxes.

In recent years, the limitations of classical statistical moments have been increasingly acknowledged in the context of hydroclimatic extremes (see Section 2.1). Classical estimators are highly sensitive to outliers and suffer from bias when using small samples (which are very common in hydrological records), or skewed and/or heavy-tailed hydrological records (Koutsoyiannis, 2019; Lombardo et al., 2014). To address these limitations, the knowable moments (Kmoments) were proposed as a promising alternative, offering less biased estimators for extreme-value analysis (Koutsoyiannis, 2022, 2023, 2025). Initially introduced in the context of stochastic hydrology, the Kmoments were defined as expectations of order statistics and possess desirable mathematical properties such as boundedness, convergence, and applicability to both continuous and discrete random variables (Koutsoyiannis, 2025). In the context of rainfall-runoff modelling, Pizarro and Jorquera (2024) introduced a new objective function – for calibration of lumped rainfall-runoff models – that combines the well-known Kling-Gupta Efficiency (KGE) and Kmoments, obtaining

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

promising results when compared with the original KGE formulation. However, the potential of Kmoments to enhance the ability to model spatial patterns has not been explored to date.

In this paper, we introduce the Spatial Efficiency And Kmoments (SPEAK), a new metric to be used as an objective function for calibrating hydrological models. This metric combines the strengths of SPAEF and Kmoments to enhance the calibration of (semi)distributed rainfall-runoff models. Specifically, we address the following research questions:

- To what extent does the integration of Kmoment-based statistics into SPAEF-type spatial performance metrics enhance the robustness and accuracy of simulated evapotranspiration (ETa) spatial patterns across semi-distributed and distributed hydrological models?
- How does the performance of a Kmoment-based spatial calibration metric vary across contrasting hydroclimatic and physiographic regimes, and to what degree is it sensitive to catchment characteristics and seasonal variability?

2 A new metric for spatial calibration

2.1 Problems of classical statistical moments in hydrological analysis

Classical statistical moments have long been used as fundamental descriptors of hydrological variability, particularly in the analysis of rainfall, streamflow, and flood extremes. Mean, variance, skewness, and kurtosis are routinely employed to summarise the magnitude, dispersion, and asymmetry of hydrological processes. However, despite their theoretical properties, the practical estimation of higher-order moments from finite hydrological records presents substantial difficulties. These difficulties arise not from bias in the estimators themselves – since sample moments are, under standard assumptions, unbiased estimators of their population counterparts – but from the inherently slow convergence of moment estimates and their strong sensitivity to extreme (in magnitude) observations.

In hydrological systems, the probability distribution of key variables is often heavy-tailed. Under such conditions, high-order moments (i.e., $q > 2$) are dominated by the largest observations in the sample (i.e., the maximum norm is the limit of the q-norm as q tends to infinite). As the order of the moment increases, the contribution of moderate values becomes negligible relative to that of extreme values, effectively rendering the moment estimator a proxy for the maximum observed value. Consequently, the uncertainty associated with high-order moment estimates remains large even for relatively long records, because the probability of observing sufficiently large extremes within the available sample remains low. This phenomenon is particularly pronounced in hydrology, where observations are typically limited to a few decades, whereas the processes of interest may operate on centennial or millennial time scales. Additionally, for this type of scaling in state (i.e., power-type tail), moments are theoretically infinite beyond a certain order $q_{\max}$, but moment estimators are always finite. The latter implies that the estimation is infinite biased. Taken together, these considerations highlight that the principal difficulty with classical statistical moments in hydrological analysis is not statistical bias but rather the uncertainty associated with finite-sample estimation in heavy-tailed and state-scaling behaviours. These limitations motivate the exploration of alternative statistical frameworks that are better suited to the intrinsic variability and tail behaviour characteristic of hydrological processes. For a detailed exploration of these ideas, the reader is referred to Koutsoyiannis (2019, 2025) and Lombardo et al. (2014).

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

2.2 Kmoments, KCorrelation, and KPDF

The Kmoment of order p is defined as:

$$K'_p=\frac{p}{n}\sum_{i=p}^{n}\frac{(i\text{-}1)!\,(n\text{-}p)!}{(i\text{-}p)!\,(n)!}x_{(i)}=\sum_{i=p}^{n}b_{i,n,p}x_{(i:n)} \tag{1}$$

$$b_{i,n,p}=\frac{p}{n}\frac{(i\text{-}1)(i\text{-}2)\cdots(i\text{-}p\text{+}1)}{(n\text{-}1)(n\text{-}2)\cdots(n\text{-}p\text{+}1)} \tag{2}$$

where, $x_{(i:n)}$ denotes the i^{th} order statistic (i.e., i^{th} smallest value) of a sample of size n. This formulation ensures that the Kmoments are computable up to the sample size and are directly tied to the behaviour of the tails, making them particularly effective for quantifying extremes. Table 2 shows a conceptual comparison between classical moments and their Kmoment counterparts for the first four statistical moments:

Table 2: Kmoments' estimators for the first four statistical moments.

Moment Order	Classical Moment	Description	Coefficient $b_{i,n,p}$	Kmoment Equivalent
1	$\mu'_1=\frac{1}{n}\sum_{i=1}^{n}x_i$	Central tendency	$b_{i,n,1}=\frac{1}{n}$	$K'_1=\sum_{i=1}^{n}b_{i,n,1}x_{(i:n)}$
2	$\mu'_2=\frac{1}{n}\sum_{i=1}^{n}(x_i\text{-}M'_1)^2$	Dispersion	$b_{i,n,2}=\frac{2}{n}\frac{i\text{-}1}{n\text{-}1}$	$K'_2=\sum_{i=1}^{n}b_{i,n,2}x_{(i:n)}$
3	$\mu'_3=\frac{1}{n}\sum_{i=1}^{n}(x_i\text{-}M'_1)^3$	Asymmetry	$b_{i,n,3}=\frac{3}{n}\frac{i\text{-}1}{n\text{-}1}\frac{i\text{-}2}{n\text{-}2}$	$K'_3=\sum_{i=1}^{n}b_{i,n,3}x_{(i:n)}$
4	$\mu'_4=\frac{1}{n}\sum_{i=1}^{n}(x_i\text{-}M'_1)^4$	Tail heaviness	$b_{i,n,4}=\frac{4}{n}\frac{i\text{-}1}{n\text{-}1}\frac{i\text{-}2}{n\text{-}2}\frac{i\text{-}3}{n\text{-}3}$	$K'_4=\sum_{i=1}^{n}b_{i,n,4}x_{(i:n)}$

Using the above formulation, we define the Kcorrelation (r_{km}) as the correlation coefficient based on Kmoments. r_{km} quantifies the agreement based on normalised ranks of order statistics between two variables:

$$r_{km}=1\text{-}\left(E\left[\max\left(\frac{\text{ref}\text{-}K'_1}{K'_2\text{-}K'_1},\frac{\text{sim}\text{-}K'_1}{K'_2\text{-}K'_1}\right)\right]\right)^2 \tag{3}$$

where K'_1 and K'_2 are the first and second Kmoments of the respective variables in Table 2, and ref and sim represent the reference and simulated values, respectively. A higher r_{km} value indicates closer agreement, especially in the tails of the distribution. This formulation is particularly robust under dependence and, therefore, it is more suitable than traditional correlations, such as Pearson or Spearman, for performance evaluation in hydrological modelling.

Additionally, we define a smooth alternative to histograms based on Kmoments, following the framework proposed by Koutsoyiannis (2022), in which histograms are replaced by a smooth empirical probability density function (KPDF) derived directly from Kmoments. Following the previous K-based definitions, the cumulative distribution function (KCDF) can be computed as:

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

$$F_p = 1 - \left(\frac{1}{\Lambda_\infty * p + \Lambda_1 - \Lambda_\infty}\right) \quad (4)$$

where, $\Lambda_\infty = e^\gamma \approx 1.78$ (with γ the Euler–Mascheroni constant), and $\Lambda_1 = 2$. These values were proposed in the theoretical framework of the Λ-coefficients (see Koutsoyiannis (2022). The KPDF is then reconstructed by numerically differentiating Eq. (4) in terms of the Kmoment order p, i.e.:

$$\hat{f}(K_p) = \begin{cases} \frac{\hat{F}(\hat{K}'_{p+1}) - \hat{F}(\hat{K}'_p)}{\hat{K}'_{p+1} - \hat{K}'_p}, \hat{K}'_p < x < \hat{K}'_{p+1} \\ \frac{\hat{F}\left(\hat{\hat{K}}'_p\right) - \hat{F}\left(\hat{\hat{K}}'_{p+1}\right)}{\hat{\hat{K}}'_p - \hat{\hat{K}}'_{p+1}}, \hat{\hat{K}}'_{p+1} < x < \hat{\hat{K}}'_p \end{cases} \quad (5)$$

This approach yields a smooth and normalised representation of the empirical density function, constructed through a Kmoment-based estimation.

2.3 Introducing the SPatial Efficiency And KMoments (SPEAK)

The Spatial Efficiency and Kmoments (SPEAK) is a new metric, specifically designed to assess the capability of (semi)distributed hydrological models to reproduce spatial patterns in rainfall-runoff modelling. SPEAK also evaluates the statistical consistency of simulated and reference values of the analysed variable. The formulation of SPEAK is as follows:

$$\text{SPEAK} = 1 - \sqrt{(r_{km}-1)^2 + (\beta_{km}-1)^2 + (\gamma_{km}-1)^2} \quad (6)$$

where r_{km} is the Kmoment-based correlation coefficient (between the reference and simulated ETa maps, concatenated on time), β_{km} is the ratio between simulated and reference coefficient of variations ($CV = K_2/K_1$ and $\beta = CV_{\text{sim}}/CV_{\text{ref}}$); and, γ_{km} is the similarity between reference and simulated KPDFs assessed through the Hellinger distance $H(P, Q)$:

$$H(P,Q) = \frac{1}{\sqrt{2}}\left(\sum_{x \in X}\left(\sqrt{P_{\text{ref}}} - \sqrt{P_{\text{sim}}}\right)^2\right)^{\frac{1}{2}} \quad (7)$$

where, P_{ref} and P_{sim} represent the normalised probability densities of the reference and simulated data, respectively, evaluated at a shared set of points x. We define γ_{km} as a bounded similarity metric in Eq. (8):

$$\gamma_{km} = 1 - H(P,Q) \quad (8)$$

γ_{km} ranges from 0 to 1, where values approaching 1 reflect a closer agreement between the reference and simulated KPDFs, and values near 0 indicate substantial divergence. The Hellinger distance is computed over a common support using the square-root transformation of the interpolated densities to ensure numerical stability and a consistent geometric interpretation (see Cha (2007)). By construction, this approach preserves continuity, avoids discretisation artefacts, and provides a robust basis for quantifying similarity. By aggregating these terms, SPEAK provides a novel similarity score ranging from -∞ to 1, with higher values indicating a stronger agreement.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

3 Example application

3.1 Study area and data

Our study domain is continental Chile, which is characterised by a large hydroclimatic and physiographic diversity. Specifically, we selected 99 near-natural catchments (see Fig. 1) from the CAMELS-CL database based on four criteria: (1) less than 25% of missing daily streamflow records during the period 1980-2020, allowing only non-consecutive gaps; (2) absence of large dams, minimising the influence of reservoir operations on streamflow; (3) land-use and water-use indicators (i.e., consumptive water withdrawals $<10\%$ of mean streamflow, urban land cover $\leq 5\%$ of the catchment area, irrigated agriculture $\leq 20\%$ of catchment area; and forest plantations $<20\%$); (4) exclusion of catchments with more than 5% glacier coverage or with any clear evidence of artificial modifications in the river network. Collectively, these thresholds are aligned with previous studies (e.g., Acuña and Pizarro, 2023; Baez-Villanueva et al., 2021; Jorquera and Pizarro, 2023; Pizarro and Jorquera, 2024) and enable the explicit quantification of human influence, ensuring that the selected catchments exhibit near-natural hydrological regimes. The analysis period spans from January 1, 1980, to March 31, 2020. The selected catchments exhibit a broad range of characteristics, highlighting substantial variability across the study domain. Catchment areas vary considerably, from as small as 35 km² to as large as 11,137 km², with a median area of 672 km². Mean annual precipitation also shows pronounced variability, ranging from 94 to 3,660 mm yr^{-1}, with a median of 1,393 mm yr^{-1}. The aridity index spans a wide range, from 0.3 in southern Chile to 31.6 in northern Chile, with a median of 0.69. Mean catchment elevations extend from 118 m a.s.l. in western areas near the Pacific Ocean to 4,270 m a.s.l. in the eastern Andes Mountains, with a median elevation of 1,052 m a.s.l. Regarding precipitation seasonality, most catchments are characterised by a winter-dominated rainfall regime; however, some northern catchments are exceptions, with precipitation mainly concentrated during the summer months. Furthermore, a marked latitudinal gradient is observed across Chile, with precipitation generally increasing toward the south while temperatures tend to decrease.

Figure 1 shows the spatial distribution of the catchment area, mean elevation, mean precipitation, and mean annual potential evapotranspiration (PET) of the selected catchments. One can note the strong latitudinal and altitudinal gradients of Chile, which are crucial in shaping hydrological responses. The selected catchments are representative of different hydroclimatic zones in continental Chile, such as Far and Near North (N=4 and N=2, respectively), Central (N=29), South (N=45), and Austral (N=19). This stratification relies on latitude, following previous studies (e.g., Segovia et al. 2025), to facilitate comparative analysis across regions.

Hydrological data were obtained from the CAMELS-CL database (Alvarez-Garreton et al., 2018), which compiles daily streamflow records from hydrometric stations maintained by the Chilean Water Directorate (DGA). Daily time series of precipitation and maximum/minimum temperatures were retrieved from the gridded meteorological product CR2METv2.5 (Boisier, 2023), which provides data at a horizontal resolution of 0.05 ° × 0.05°. PET was estimated using the gridded temperatures and the Hargreaves-Samani equation (Hargreaves and Allen, 2003; Hargreaves and Samani, 1985). Additionally, actual evapotranspiration (ETa) was retrieved from the Global Land Evaporation Amsterdam Model (GLEAM), a satellite-based framework that estimates global-scale terrestrial evaporation and transpiration. GLEAM combines microwave remote sensing observations with reanalysis meteorological data and a soil water balance model to provide spatially distributed fields of evaporation, transpiration, interception loss, and root-

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

zone soil moisture. The most recent version, GLEAMv4.2a (Miralles et al., 2025), provides daily estimates at a 0.10° horizontal resolution, making it particularly suitable for large-scale hydrological modelling studies (see, e.g., Ding and Zhu, 2022; Guo et al., 2024; Jiang et al., 2020; Sirisena et al., 2020a). We use ETa data from the GLEAMv4.2a product, initially available at a 0.10° x 0.10° horizontal resolution but regridded to 0.05° x 0.05° through bilinear interpolation to ensure the same horizontal resolution available at CR2MET and, therefore, ensuring spatial consistency with the input meteorological forcings. The temporal resolution of all datasets is maintained daily to facilitate seamless integration into the modelling framework. It is worth mentioning that reference ETa data – from the GLEAM v4.2a product – were used without applying any bias correction or temporal adjustment procedures. This decision was made to preserve the intrinsic variability and spatial structure of the satellite-derived fields, ensuring that the evaluation reflected the raw performance of the calibration metrics.

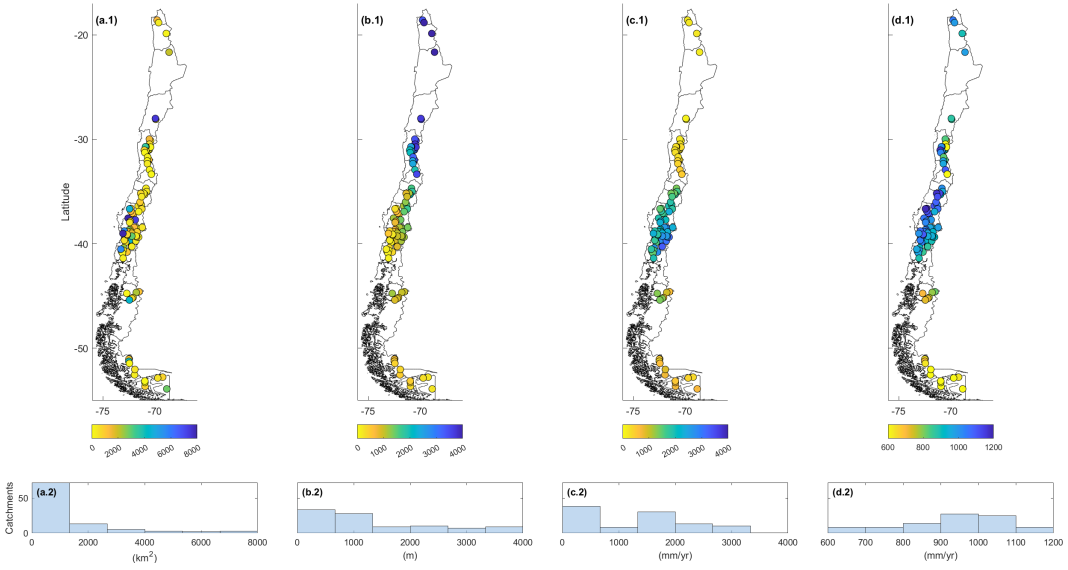

Figure 1: Analysed case studies and catchment attributes. (a) Area (km^2), (b) mean elevation (m), (c) mean annual precipitation (mm), (d) mean annual PET (mm). The coloured dots represent the location of catchment outlets. A histogram of the analysed variable is shown below each subplot.

3.2 Hydrological model

We used the Technische Universität Wien (TUW) hydrological model, which is a conceptual, bucket-style, and semi-distributed model developed by the Technical University of Vienna (Parajka et al., 2007). Its structure is based on the Hydrologiska Byråns Vattenbalansavdelning (HBV, Bergström, 1976) rainfall-runoff model, and has been widely applied in hydrological modelling and water resources management studies (Araya et al., 2023; Baez-Villanueva et al., 2021; Parajka et al., 2016; Sleziak et al., 2018, 2021; Széles et al., 2020; Segovia et al., 2025). The TUW model requires precipitation, temperature, and PET as input meteorological forcings, which are distributed across elevation bands to capture altitudinal controls on snowfall and melt. Within each band, the snow module partitions precipitation into rainfall or snowfall based on temperature thresholds and updates the snowpack using a degree-day formulation.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

Meltwater and rainfall enter the soil module, where soil storage, ETa, and nonlinear runoff generation control the amount of water transferred to the subsurface reservoirs, where all infiltrated water is collected. The response module then separates this flux into fast, very fast, and slow components using two conceptual stores, each governed by distinct storage coefficients and a constant percolation rate. Finally, the combined outflows are routed to the catchment outlet through a transfer function, producing the simulated runoff. Figure 2 shows the model structure, and Table 3 includes a brief description of each parameter.

In this study, we adopted three configurations in each catchment regarding the spatial distribution of hydrological model inputs:

a. Lumped configuration: Catchment-averaged model inputs with uniform-in-space parameters' values (i.e., not depending on the spatial dimension).
b. Semi-distributed model configuration: Model inputs are spatially disaggregated to 200-m elevation bands with uniform-in-space parameters' values (i.e., not depending on the spatial dimension).
c. Fully distributed model configuration: Model inputs are provided at the grid-cell (0.05° x 0.05°) level across each catchment with uniform-in-space parameters' values (i.e., not depending on the spatial dimension).

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

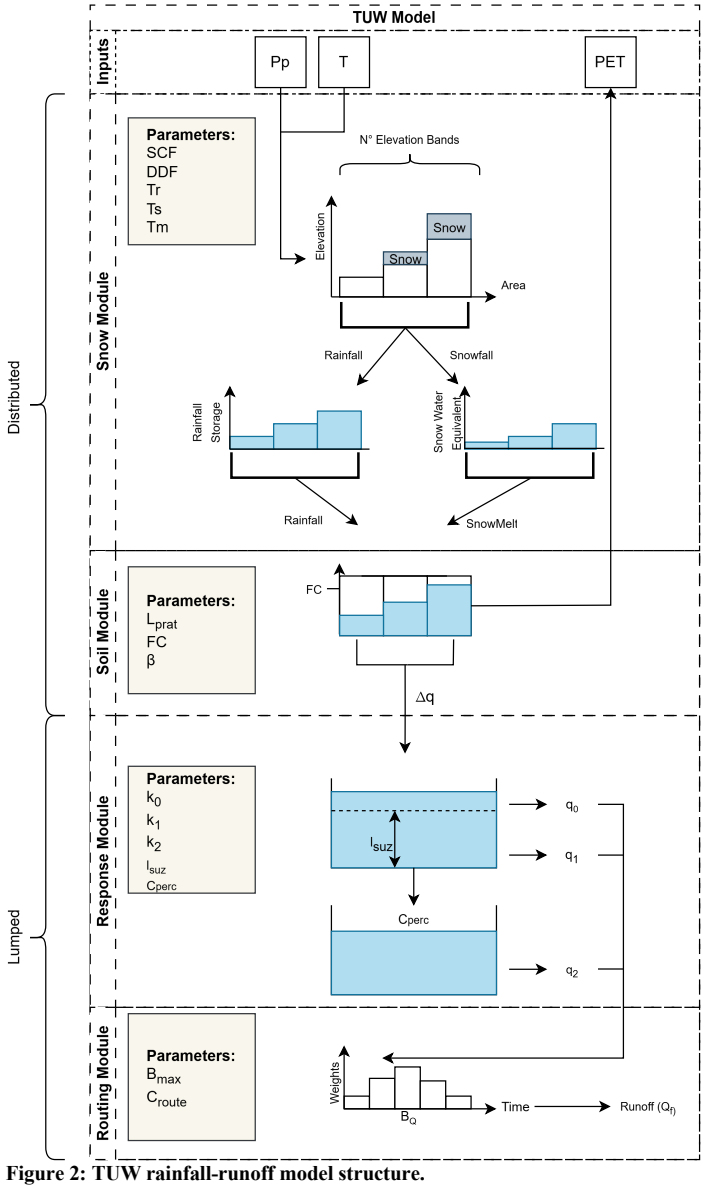

Figure 2: TUW rainfall-runoff model structure.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

Table 3: TUW model parameters, their description, and the range of adopted values.

Parameter	Symbol	Unit	Description	Range
Snow Correlation Factor	SCF	-	Empirical factor adjusting solid precipitation for snow accumulation.	0.9-1.5
Degree Day Factor	DDF	mm/degC/timestep	Controls the rate of snowmelt based on accumulated degree days above the threshold T°.	0.0-5.0
Threshold Temperature Rain	Tr	°C	The temperature above which precipitation is considered entirely as rainfall.	1.0-3.0
Threshold Temperature Snow	Ts	°C	The temperature below which precipitation is considered entirely as snowfall.	-3.0-1.0
Melt Temperature	Tm	°C	Threshold temperature controlling the onset of snowmelt. When the air temperature exceeds Tm.	-2.0-2.0
Limit Potential Evaporation	L_{Prat}	-	The fraction of potential evapotranspiration limiting actual evapotranspiration under dry conditions.	0.0-1.0
Field Capacity	FC	mm	Maximum soil moisture storage capacity before percolation and surface runoff.	Catchment-dependent
Non-Linear Runoff Production	BETA	-	Control the non-linearity of runoff generation from excess soil moisture.	0.0-20.0
Storage Coefficient for very fast response	k_0	Timestep	Recession parameter for quickflow routing through the very fast response reservoir.	0.0-2.0
Storage Coefficient for fast response	k_1	Timestep	Recession parameter for routing through the fast response storage.	2.0-30.0
Storage Coefficient for slow response	k_2	Timestep	Recession parameter governing slow flow or baseflow routing.	30.0-250.0
Threshold Storage State	l_{suz}	mm	Storage level threshold in the upper zone triggering quick runoff generation.	1.0-100.0
Constant Percolation Rate	c_{perc}	mm/timestep	Maximum percolation rate from the upper to the lower soil zone under saturated conditions.	0.0-8.0
Maximum Base at Low Flow	B_{max}	Timestep	Maximum storage coefficient during low-flow conditions.	0.0-30.0
Free Scaling Parameter	C_{route}	$\mathrm{Timestep}^2/\mathrm{mm}$	Calibration parameter scaling the routing time of surface runoff through the basin.	0.0-50.0

3.3 Calibration and evaluation strategies

The parameters of the TUW model were calibrated using a Particle Swarm Optimisation (PSO)-based algorithm, implemented in the R package hydroPSO (Zambrano-Bigiarini and Rojas, 2013), designed for hydrological model parameter estimation. PSO is a metaheuristic optimisation technique inspired by the collective behaviour of natural swarms, such as birds in flight. Originally developed by Kennedy and Eberhart (1995), PSO has gained popularity in optimisation due to its efficiency at exploring complex search spaces.

The calibration and evaluation periods range from April 1, 1982, to March 31, 2002, and from April 1, 2002, to March 31, 2020, respectively. Additionally, a warm-up period, from January 1, 1980, to March 31, 1982, is used to stabilise internal model states and reduce the influence of initial hydrologic conditions before formal calibration begins. It is worth mentioning that, as the catchments are located in Chile (in the Southern Hemisphere), the water year starts on April 1 in most of the case study catchments. Therefore, we use April 1 and March 31 as the beginning and end of the water year for all the catchments for simplicity.

In order to systematically assess how the choice of objective function influences the model's ability to reproduce observed ETa's spatial patterns, we perform three model calibrations at each catchment:

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

- KGE-based calibration, which aims at maximising the KGE computed with daily flows at the catchment outlet:

$$\mathrm{KGE} = 1 - \sqrt{(r_Q\text{-}1)^2+(\beta_Q\text{-}1)^2+(\gamma_Q\text{-}1)^2} \quad (9)$$

where r_Q is the Pearson correlation between simulated and observed streamflow values, β_Q is the bias ratio between simulated and observed streamflow values, and γ_Q is the ratio of coefficients of variation between simulated and observed streamflow values. KGE ranges from -∞ to 1, with 1 being the optimal value.

- SPAEF-based calibration, which aims at maximising the SPAEF metric computed with ETa:

$$\mathrm{SPAEF} = 1 - \sqrt{(r\text{-}1)^2+(\beta\text{-}1)^2+(\gamma\text{-}1)^2} \quad (10)$$

where r denotes the Pearson correlation coefficient between reference and simulated spatial fields (concatenated on time); $\beta(=(\sigma_{\mathrm{sim}}/\mu_{\mathrm{sim}})/(\sigma_{\mathrm{ref}}/\mu_{\mathrm{ref}}))$ represents the fraction of the coefficient of variation, assessing relative variability; and, $\gamma\left(=\frac{\sum_{j=1}^{n}\min(K_j,L_j)}{\sum_{j=1}^{n}K_j}\right)$ quantifies histogram similarity (intersection for the given histogram K of the observed pattern and histogram L of the simulated pattern, each containing n bins), thereby comparing the underlying distributional shapes. By aggregating these terms, SPAEF provides a bounded similarity score in the range]-∞,1] with higher values indicating stronger agreement.

- SPEAK-based calibration, which aims at maximising the SPEAK metric. Computations were performed between the reference and simulated ETa maps, concatenated on time.

Note that the KGE-based calibration was performed using the lumped model configuration, whereas the SPAEF- and the SPEAK-based calibrations were conducted using the semi-distributed and the fully distributed model configurations. To compare the performance provided by the calibration experiments, we use the root mean square error normalised by the mean of the referenced data, which provides a dimensionless indicator of model accuracy (Willmott, 1981; Willmott and Matsuura, 2005):

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - R_i)^2}}{\overline{R}} \quad (11)$$

where S_i and R_i are the simulated and referenced values, $\overline{R}$ denotes the mean of the reference data, and n is the number of grid cells. Lower NRMSE values indicate a closer agreement between simulations and reference values, enabling meaningful comparison across catchments or spatial domains with differing magnitudes.

Additionally, the Fraction Skill Score (FSS) is also used for performance comparison tasks. FSS is usually used in precipitation forecasts (Gaur et al., 2022; Necker et al., 2024; Roberts and Lean, 2008). FSS takes values from zero to one, with one indicating perfect agreement. Computational steps involve the identification of values exceeding certain thresholds (under a given window size) and, as a consequence, the creation of binarised maps (relying on the selected thresholds) and the computation of fraction cells above such threshold. The computation of mean-square error (MSE) between observed and simulated fractions as well as its normalisation by the worst-case MSE ($\mathrm{MSE_{WC}}$) are essential steps in FSS quantification:

$$\mathrm{FSS} = 1 - \frac{\mathrm{MSE}}{\mathrm{MSE_{WC}}} \quad (12)$$

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

Here, FSS was computed using the top four ETa percentiles (i.e., 75th, 90th, 95th, and 99th) and a pixel-scale window size.

4 Results and discussion

4.1 Streamflow-based calibration: Highlighting issues of spatial distribution of internal states and fluxes

Fig. 3a shows the KGE values obtained with the streamflow-based calibration over the full analysis period. Results show that the streamflow-based calibration effectively adjusted parameter values to reproduce runoff at catchment outlets under diverse hydroclimatic regimes. The mean and median KGE are 0.80 and 0.85, respectively, with a standard deviation of 0.15. The 5th and 95th percentiles are 0.46 and 0.94, respectively, with minimum and maximum values of 0.25 and 0.95. The interquartile range is 0.11, and only six catchments exhibited KGE values below 0.5, indicating a generally robust performance across the study region. However, ETa results (Fig. 3b-f) show persistent mismatches between simulated and reference values, with error patterns that differ notably across hydroclimatic zones, similar to those obtained by Mei et al. (2023) and Sirisena et al. (2020b).

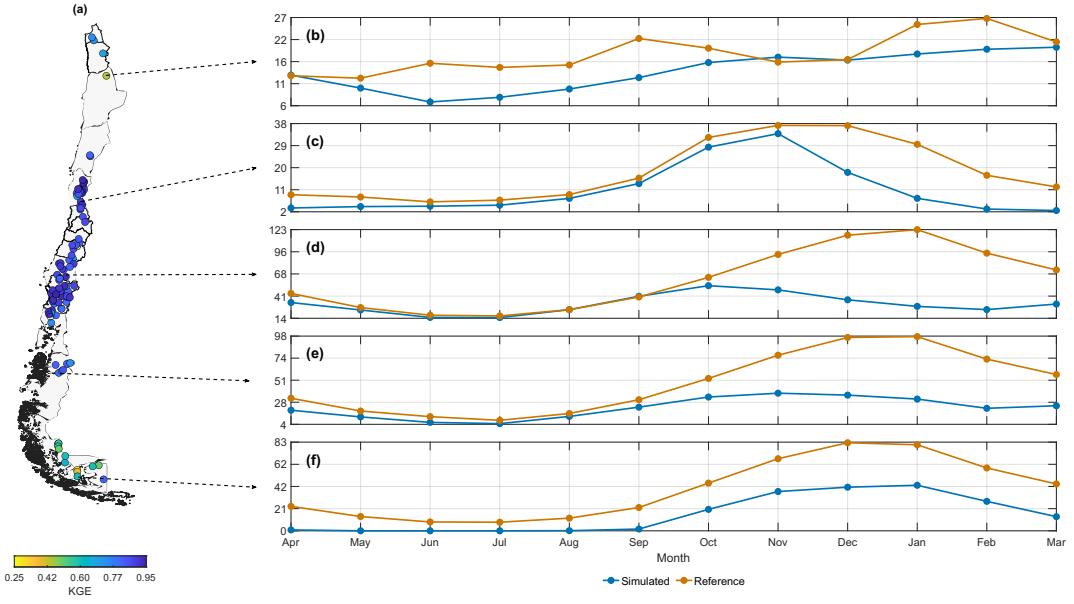

Figure 3: Results obtained from the streamflow-based calibration (period 1982-2002). Panel (a) shows KGE values for 99 catchments, considering only streamflow data during calibration. Panels (b) to (f) show the monthly mean actual evapotranspiration (ETa) obtained during calibration, considering only streamflow data. Panel (b), Río Loa Antes Represa Lequena (ID: 2101001); panel (c), Río Chalinga en la Palmilla (ID: 4712001); panel (d), Río BioBío en Coihue (ID: 8334001); panel (e), Río Mañihuales Antes Junta Río Simpson (ID: 11308001); panel (f), Río San Juan en Desembocadura (ID: 12582001).

For instance, in the Loa River catchment (Fig. 3b), the model fails to reproduce the observed seasonality of ETa (i.e., timing and magnitude). Additionally, substantial deficiencies are obtained towards the end of the water year (December–March) in the Chalinga, Biobío and Mañihuales catchments (see Fig. 3c-e), where simulated ETa fails to follow the sharp increase and sustained peak of the reference dataset. Lastly, an inverse behaviour is obtained in the southern

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

catchment (Fig. 3f): the model remains unresponsive, primarily during the early months (April–August), underestimating ETa considerably, and subsequently fails to reach the reference peak values during summer, resulting in a consistent bias. These discrepancies emphasise the structural limitations of a streamflow-based calibration strategy in capturing the spatial and temporal complexity of ETa dynamics under diverse climatic regimes. While the KGE optimisation generally yields good streamflow simulations at the catchment outlet, it does not necessarily ensure an adequate reproduction of other water fluxes such as ETa. The latter highlights the importance of integrating additional hydrological variables and fluxes into the calibration, alongside streamflow.

4.2 ETa-based calibration: SPEAK performance

Fig. 4 shows the spatial distribution of seasonal ETa amounts over one of the analysed catchments (Catchment ID: 11302001, Rio Ñirehuao En Villa Mañihuales). The catchment in question was arbitrarily selected to show the spatially distributed ETa results. Fig. 4a, 4b, and 4c show seasonal ETa for the reference data, SPEAK-based, and SPAEF-based simulations, respectively. Interestingly, SPEAK- and SPAEF-based ETa simulations yielded coherent, reproducible spatial ETa patterns across all seasons. The catchment-average absolute errors per season were 10.17, 12.81, 39.54, 21.78 and 13.9, 15.04, 41.94, 47.86 – for SPEAK and SPAEF, respectively – for autumn, winter, spring, and summer; respectively. Modelling issues arise for warmer seasons (OND and JFM), evidencing slightly better results for SPEAK-based simulations.

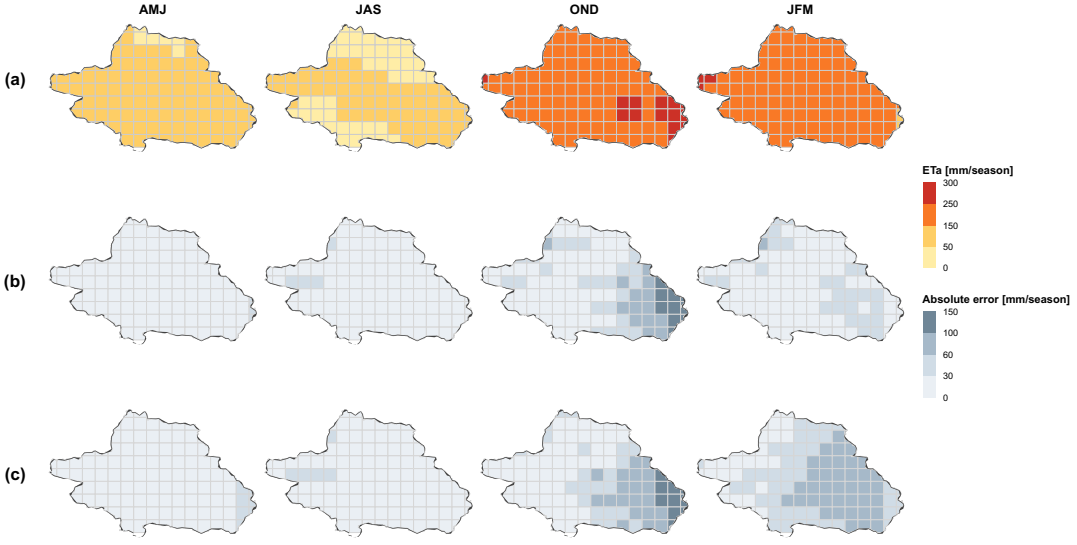

Figure 4: (a) Reference ETa from GLEAM for each three-month season (AMJ, JAS, OND, JFM), and spatial distribution of relative errors in simulated ETa with the TUW model with (b) the best SPEAK and (c) the best SPAEF calibrations for catchment ID: 11302001 (period 2001-2020). The red colour bar indicates the ETa in (mm/season), whereas the grey colour bar indicates the absolute error (in mm/season) in simulated ETa with respect to the reference data.

Fig. 5 shows the fraction skill score maps associated with SPEAK- and SPAEF-based simulations for four different ETa thresholds taken into consideration (Fig. 5a: 75th; Fig. 5b: 90th; Fig. 5c: 95th; and, Fig. 5d: 99th mean annual

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

ETa percentiles) over the same catchment. Overall, Fig. 5 shows a number of low-FSS-value pixels for both SPEAK- and SPAEF-based simulations. It is also evident that SPEAK-based simulations have more pixels with higher FSS values (compared with SPAEF-based simulations) and outperform SPAEF-based simulations across all analysed thresholds. Additionally, there is a spatial FSS gradient, with higher values from west to east. The catchment-average FSS was 0.17, 0.06, 0.05, 0.02, and 0.05, 0.03, 0.03, 0.00 for the SPEAK- and SPAEF-based simulations, respectively, and for the 75th, 90th, 95th, and 99th ETa percentiles, respectively.

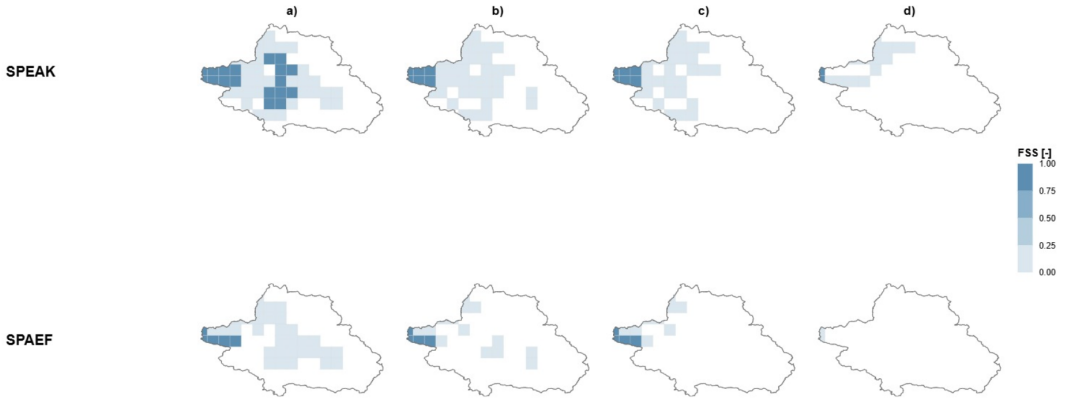

Figure 5: FSS for SPEAK- and SPAEF-based mean annual ETa simulation (catchment ID: 11302001; period: 2001-2020). From a) to d): 75th, 90th, 95th, and 99th ETa percentiles taken as thresholds for FSS computation.

Four additional case studies are shown in Fig. 6, comparing reference and simulated annual mean ETa by SPEAK and SPAEF. It is possible to observed that SPEAK-based simulations reproduce the ETa spatial fields with closer agreement to the reference GLEAM data. In the Río Itata catchment (Fig. 6a), SPEAK accurately captures the west-to-east moisture gradient, reproducing the attenuation of ETa toward the interior valley, aligning well with regional ETa magnitudes. In the Biobío River catchment (Fig. 6b), the SPEAK-based simulation preserves the topographic contrasts that govern ETa under semi-humid conditions, demonstrating the metric's capability to maintain elevation-dependent variability (spatial correlation between reference and simulated maps: 0.289 and 0.097 for SPEAK and SPAEF, respectively). Further south, in Río Cautín and Río Trancura (Figs. 6c and 6d), SPEAK maintains the internal heterogeneity of the observed fields and avoids excessive smoothing, a particularly valuable feature in humid catchments where vegetation and orographic effects modulate spatial fluxes. In contrast, the SPAEF-calibrated outputs exhibit higher ETa magnitudes, particularly in lowland regions, with a loss of spatial definition across all catchments. These deviations are more pronounced in Fig. 6a, where SPAEF yields spatially homogeneous fields that fail to capture the catchment's gradual spatial characteristics. Similarly, in Biobio (Fig. 6b), the simulated field is smoothed over elevation gradients, limiting its representativeness in systems with complex terrain. Such a pattern likely stems from the histogram-based formulation of SPAEF's γ component, which introduces rigidity in cases of strong physiographic variability. The average values for the four presented catchments, in terms of spatial correlation between the reference and simulated maps, were 0.272 and 0.160 for SPEAK and SPAEF, respectively.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

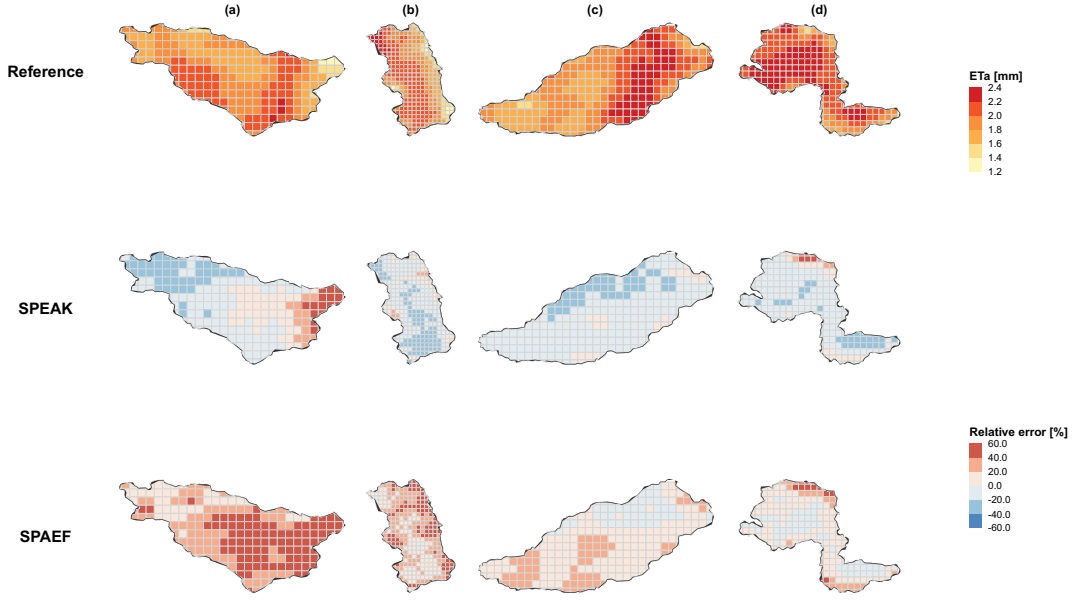

Figure 6: Comparison between (a) GLEAM actual evapotranspiration fields (annual mean per pixel) and simulated ETa using parameters obtained by calibrating with (b) SPEAK and (c) SPAEF against GLEAM reference data. Each column corresponds to a specific catchment: (a) Río Itata at Balsa Nueva Aldea (ID: 8135002); (b) Río Biobío at Rucalhue (ID: 8317001); (c) Río Cautín at Almagro (ID: 9140001); and (d) Río Trancura Antes Río Llafenco (ID: 9414001).

The spatial results displayed in Figures 4, 5, and 6 highlight the hydrological implications of adopting a Kmoment-based formulation for spatial calibration. By employing continuous, bin-free KPDF representations, SPEAK enhances both the magnitude and the agreement in variability with the reference fields, ensuring that the simulated spatial patterns remain consistent across distinct hydroclimatic regimes. Complementing these findings, five ETa hydrologic signatures (mean, min, max, P05, and P95) were computed for the four catchments analysed above, considering reference data as well as SPEAK- and SPAEF-based simulations. These findings are shown in Table 4, alongside evidence indicating which simulation had closer agreement (green colours indicate that SPEAK-based simulations had closer agreement with the reference data than SPAEF-based simulations).

Figure 7 shows the spatial evaluation of model performance in reproducing ETa, considering all case studies. Fig. 7a-c show NRMSE values, while Fig. 7d-f show the Pearson correlation between simulations and reference data. NRMSE and Pearson correlation were used as extra benchmark assessment metrics. We use the NRMSE because the mathematical formulations of SPEAK and SPAEF differ, so their values (magnitude) cannot be directly compared. Additionally, we use the Pearson correlation because it is insensitive to bias. Our results reveal clear differences in NRMSE between SPEAK- and SPAEF-based calibrations. In general, SPEAK provides more consistent spatial patterns of ETa across central and southern Chile, as reflected by lower NRMSE values in a large fraction of catchments (see Table 5), indicating that the inclusion of the Kmoment formulation improves its ability to represent the overall spatial behaviour of ETa. Conversely, SPAEF tends to outperform SPEAK in a limited number of catchments, primarily in

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

areas where both metrics yield similar performance. Pearson correlation results were similar, indicating superior performance to SPEAK compared to SPAEF. These results highlight that, although both formulations are based on spatial correspondence principles, SPEAK provides a more stable and accurate characterisation of model reference agreement under diverse conditions. Table 5 shows the statistical summary of NRMSE values for both analysed configurations (fully and semi-distributed) across the 99 Chilean catchments. Our findings show that SPEAK provided a lower NRMSE for ETa in 85 (fully distributed configuration) and 92 (semi-distributed configuration) catchments, in comparison with SPAEF.

Table 4: Catchment-average ETa hydrologic signatures. Mean: average in time ETa; Min: minimum ETa value; Max: Maximum ETa value; P05: 5th percentile; P95: 95th percentile.

Catchment ID	ETa Signature	Reference [mm]	SPEAK [mm]	SPAEF [mm]	Closer agreement
8135002	Mean	685.99	658.71	950.94	SPEAK
	Min	449.65	440.03	696.65	SPEAK
	Max	812.96	951.46	1247.79	SPEAK
	P05	572.27	468.86	724.83	SPEAK
	P95	788.00	882.50	1165.20	SPEAK
8317001	Mean	692.59	611.88	914.80	SPEAK
	Min	460.84	455.10	700.13	SPEAK
	Max	841.97	827.54	1233.23	SPEAK
	P05	516.69	501.83	783.22	SPEAK
	P95	813.82	740.84	1094.12	SPEAK
9140001	Mean	709.02	631.78	777.93	SPAEF
	Min	511.44	470.44	642.03	SPEAK
	Max	852.63	808.35	936.34	SPEAK
	P05	609.30	514.08	671.03	SPAEF
	P95	831.52	759.63	891.61	SPAEF
10111001	Mean	758.91	688.52	831.68	SPEAK
	Min	536.44	501.09	638.86	SPEAK
	Max	877.90	928.45	1085.07	SPEAK
	P05	623.07	587.01	734.83	SPEAK
	P95	848.12	825.52	971.62	SPEAK

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

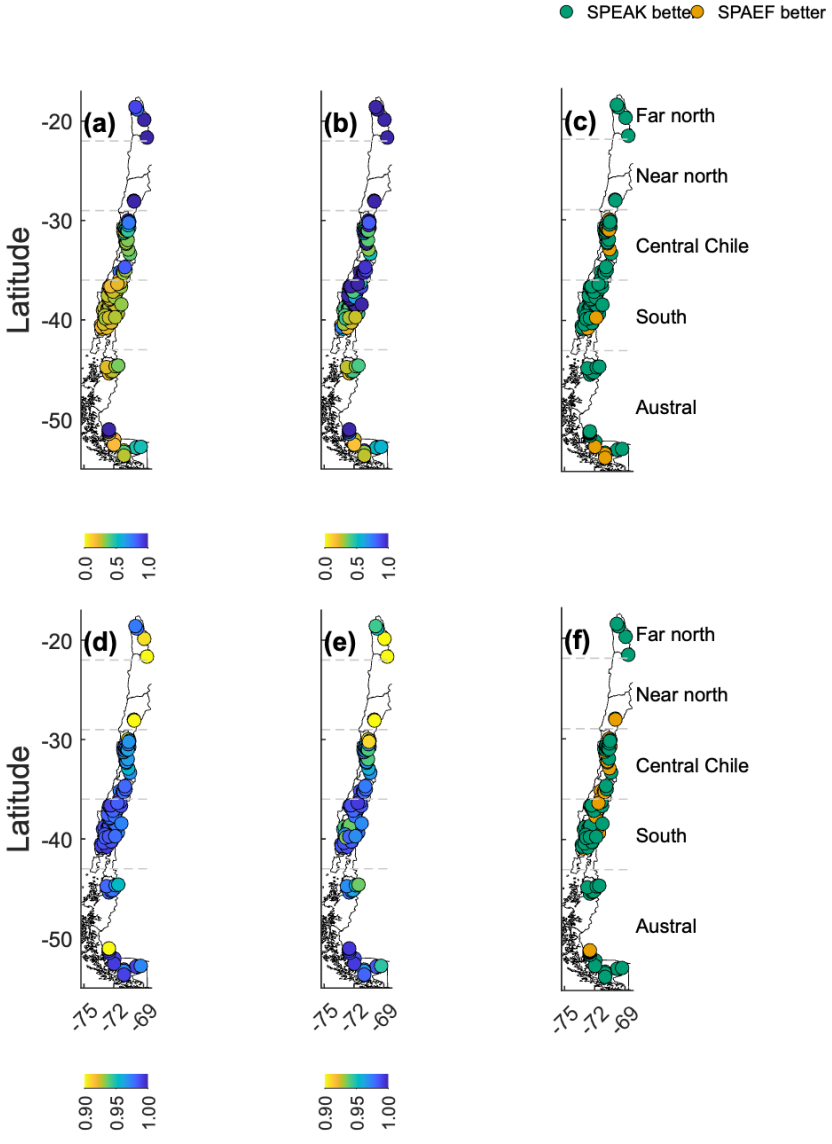

Figure 7: Spatial distribution comparison between SPEAK and SPAEF performance in terms of additional performance scores: NRMSE (a, b, c) and Pearson correlation (d, e, f). The fully distributed TUWmodel was adopted. NRMSE (a) and Pearson correlation (d) values for simulations calibrated with the SPEAK metric. NRMSE (b) and Pearson correlation (e) values from simulations calibrated with the SPAEF metric. Lighter colours indicate lower NRMSE (Pearson correlation) values and, therefore, a better (worse) simulation. One-to-one comparison for NRMSE and Pearson correlation values obtained from both calibration metrics (Fig. 7c-f), where green markers indicate catchments with better SPEAK performance and orange markers denote better SPAEF performance.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

Table 5: NRMSE summary statistics from the sample of 99 catchments for all the spatial calibrations conducted in this study, using semi-distributed and fully-distributed model configurations and SPEAK and SPAEF as calibration metrics across the 99 Chilean catchments. Bold numbers mean a better score value.

	Fully Distributed SPAEF	**Fully Distributed SPEAK**	**Semi-Distributed SPAEF**	**Semi-Distributed SPEAK**
Mean	0.76	**0.46**	0.62	**0.36**
Median	0.56	**0.28**	0.40	**0.26**
Standard Deviation	0.55	**0.49**	0.57	**0.42**
IQR	0.63	**0.17**	0.39	**0.14**
P5%	0.23	**0.17**	0.21	**0.13**
P95%	**1.72**	1.79	1.57	**1.14**
Min	0.15	**0.14**	0.16	**0.10**
Max	2.90	**2.50**	3.44	**3.13**
Range	2.72	**2.40**	3.27	**3.02**
CV	0.72	**1.05**	**0.90**	1.17
Skewness	1.56	2.62	2.60	4.29
Number of catchments with lower NRMSE	14	**85**	7	**92**

Fig. 8 shows the spatial distribution of SPEAK and SPAEF values across all considered case studies. The results show that the spatial pattern of SPEAK-based simulations is similar to that obtained for SPAEF-based ones. The latter is highlighted by the high Pearson correlation coefficient between SPEAK and SPAEF values of 0.84 and 0.81 for semi-distributed and fully distributed configurations, respectively.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

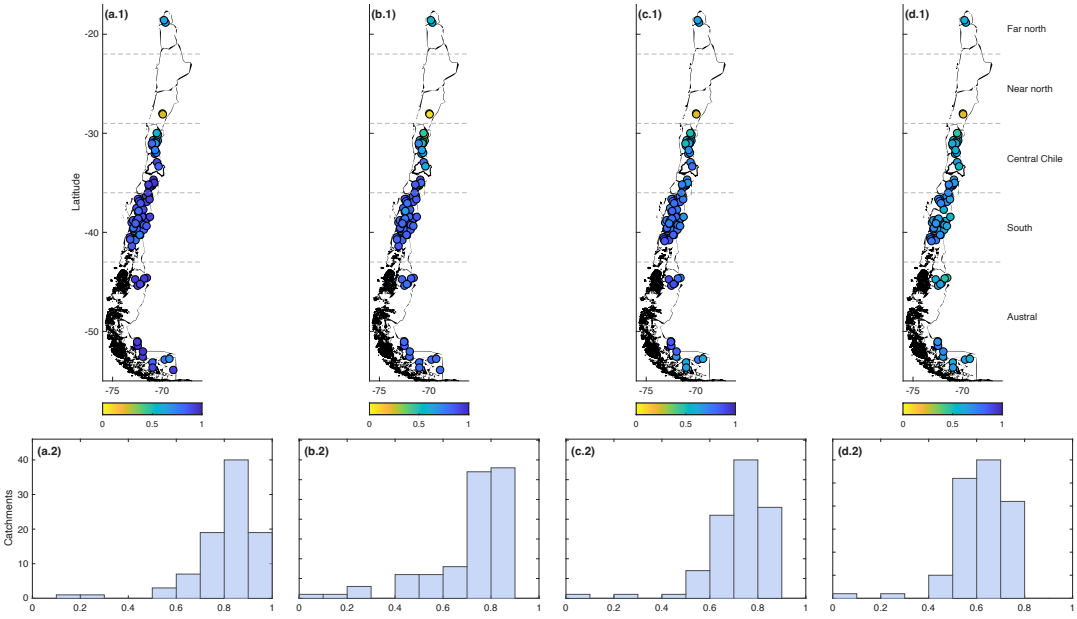

Figure 8: SPEAK and SPAEF values for 99 catchments, computed at a daily temporal scale. (a) SPEAK for semi-distributed configuration (max. value: 0.95); (b) SPAEF for semi-distributed configuration (max. value: 0.88); (c) SPEAK for fully distributed configuration (max. value: 0.87); (d) SPAEF for fully distributed configuration (max. value: 0.76). The bottom row displays the histograms with metric values.

We also explored potential links between catchment descriptors and SPEAK and SPAEF values obtained after calibration, for both the fully distributed and semi-distributed configurations, by calculating their Pearson correlation (see Fig. 9). Overall, both metrics display similar relationships with geomorphological, land-cover, and hydroclimatic descriptors, confirming that no single catchment attribute dominates their performance. The attributes mentioned above were adopted from CAMELS-CL. For elevation- and slope-related variables, negative correlations are observed for both metrics, consistent with the hydrological gradients imposed by topography and with the model's design, which initially incorporated elevation bands as a structural element.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

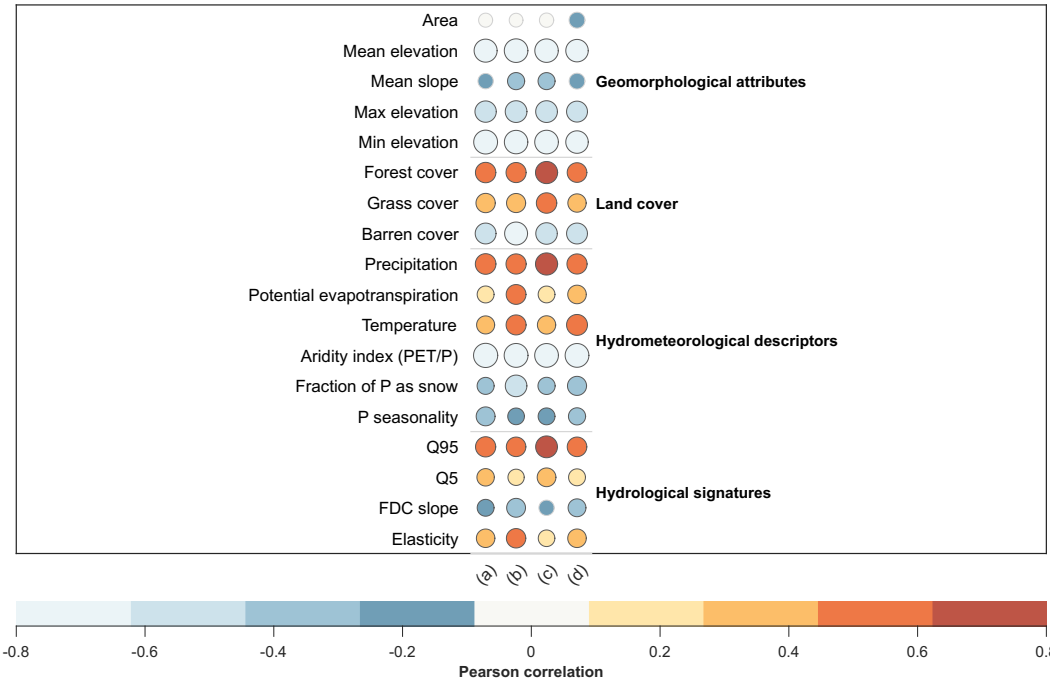

Figure 9: Pearson correlation coefficient between the value of different performance metrics and catchment attributes. (a) Semi-distributed SPEAK; (b) Semi-distributed SPAEF; (c) Fully distributed SPEAK; (d) Fully distributed SPAEF. Circles with thick borders represent statistically significant correlations (p-value < 0.05). Each column shows results for (a) SPEAK with elevation bands (b) SPAEF with elevation bands (c) SPEAK fully distributed, and (d) SPAEF full distributed (Fig. 9d).

The correlation analysis indicates that the spatial performance of both metrics is not strongly controlled by geomorphological attributes or land-cover descriptors, which generally show weak to moderate Pearson correlations. In contrast, stronger associations are observed for hydrometeorological descriptors and hydrological signatures, particularly precipitation, temperature, and Q95. These relationships suggest that the values attained by SPEAK and SPAEF tend to vary systematically across catchments with different hydroclimatic conditions. However, since the objective functions were evaluated using ETa, these correlations should not be interpreted as direct evidence that either metric captures rainfall-driven processes or streamflow variability more effectively. Rather, they indicate that the spatial efficiency scores are partly associated with the prevailing climatic and hydrological regimes of the catchments. In this context, the lower correlation between SPEAK and PET compared with SPAEF is noteworthy, because PET represents atmospheric evaporative demand rather than the actual water flux used as a calibration target. This result suggests that SPEAK may be less dependent on PET-related spatial gradients, while still evaluating the spatial agreement of simulated ETa. Overall, SPEAK shows a correlation structure broadly comparable to that of SPAEF, preserving a similar degree of structural neutrality while exhibiting sensitivity to selected hydroclimatic controls.

Finally, the internal components of SPEAK and SPAEF – and their performance across the simulations – are shown in Fig. 10. Our findings revealed apparent structural differences between these metrics. It is worth noticing that the mathematical formulation of the components is different and therefore, a one-to-one comparison is impossible to perform. Despite the latter, and to provide a panoramic view of our results, a smaller spread is observed across all

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

SPEAK components, with a particularly marked improvement in γ, especially under the semi-distributed configuration. This behaviour results from SPEAK's dependence on smooth empirical density functions derived from Kmoments, which provide a continuous, bin-free estimate of the ETa probability density function (KPDF). By contrast, SPAEF's γ component depends on histogram-based similarity, which is inherently sensitive to bin size and data resolution, often introducing discontinuities or instability across catchments of varying scale. The use of KPDFs in SPEAK mitigates this limitation, ensuring consistent and scalable performance in both small and large catchments. Moreover, the r_{km} (Kmoment-based spatial correlation) and β_{km} (Kmoment-based CVs' ratio) components of SPEAK are closer to 1, showing that a SPEAK-based calibration enables to effectively achieve a simultaneous improvement of the three components, as opposed to SPAEF.

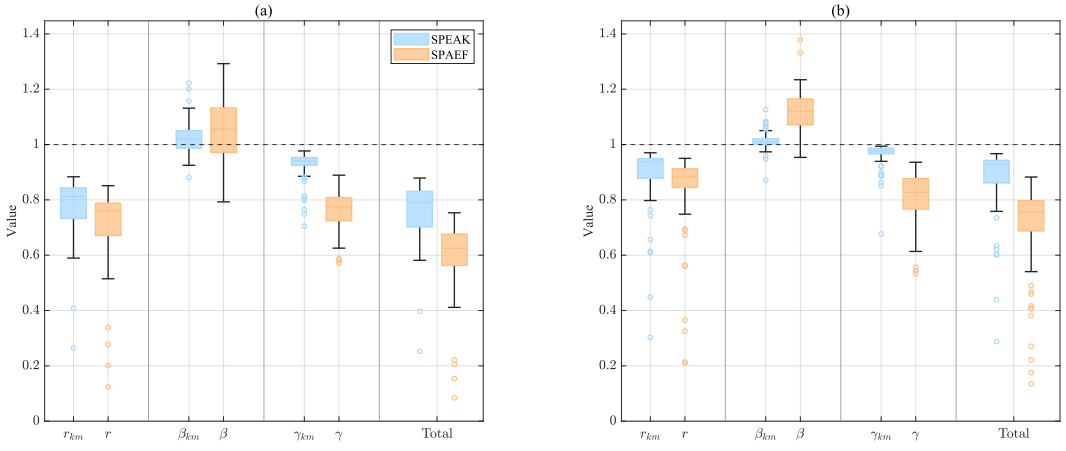

Figure 10: Boxplots of SPEAK and SPAEF components obtained with the (a) fully distributed and (b) semi-distributed configurations. The horizontal dashed line represents the optimal value. Each boxplot comprises results from the 99 case study catchments. The boxes represent the interquartile range (IQR; 25th to 75th percentiles), the horizontal line indicates the median, whiskers extend to 1.5 × IQR, and circles denote outliers falling outside the whiskers.

5 Conclusions

We introduced the Spatial Efficiency And Kmoments (SPEAK), a novel spatial calibration metric that integrates Kmoment-based statistics into a SPAEF-type framework to improve the representation of spatial hydrological processes in (semi)distributed rainfall-runoff models. Across 99 Chilean catchments characterised by strong hydroclimatic and physiographic gradients, SPEAK consistently improved the simulation of actual evapotranspiration (ETa) spatial patterns compared with conventional SPAEF-based calibration. The integration of Kmoment-based statistics (e.g., KPDF, Kcorrelation, and Kmoments) substantially enhanced the robustness and accuracy of ETa simulations in both semi-distributed and fully distributed model configurations. SPEAK systematically produced lower NRMSE values, stronger spatial correlation, and reduced dispersion among calibration components than benchmark metrics. The performance of SPEAK remained generally robust across contrasting hydroclimatic and physiographic regimes. Although some sensitivity to precipitation seasonality was identified, most geomorphological, land-cover, and hydrological descriptors exhibited weak correlations with metric performance, indicating limited dependence on catchment

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

characteristics. This suggests that SPEAK maintains stable behaviour across catchments with diverse elevations, aridity conditions, precipitation regimes, and spatial scales. In addition, the metric showed consistent performance across seasons, although larger modelling discrepancies persisted during warmer periods, which were characterised by stronger ETa dynamics. Despite these promising results, some limitations should be acknowledged. The spatial evaluation relied exclusively on ETa from the GLEAM v4.2a product, which remains subject to uncertainties associated with remote sensing retrievals and forcing data. Furthermore, the TUW model employed spatially uniform parameter values, potentially limiting its ability to represent local heterogeneity. Future work should therefore evaluate SPEAK using additional hydrological variables, modelling frameworks, and climatic regions. Key findings as follows:

- SPEAK improves the spatial calibration of hydrological models by integrating Kmoment-based statistics.
- The proposed metric enhances the robustness and accuracy of simulated ETa spatial patterns in both semi-distributed and distributed model configurations.
- SPEAK consistently achieved lower NRMSE values and stronger spatial correlation than benchmark metrics across 99 Chilean catchments.
- SPEAK exhibited limited dependence on most catchment descriptors, indicating robust behaviour across contrasting hydroclimatic and physiographic regimes.
- Seasonal variability influenced model performance, with larger discrepancies observed during warmer seasons, when ETa dynamics are stronger.
- Incorporating spatially distributed information is essential for improving the representation of internal hydrological processes beyond streamflow-only calibration approaches.

Code and Data availability

Codes and data are available in Moreno et al. (2026): https://doi.org/10.17605/OSF.IO/86VQ3

Author contribution

MM and AP designed the initial methodology. MM developed the codes and performed formal analyses. MM and AP prepared the paper with contributions from all co-authors.

Competing Interest

The authors declare that they have no conflict of interest.

Financial support

This work was carried out with the support of The National Research and Development Agency of the Chilean Ministry of Science, Technology, Knowledge and Innovation (ANID) through grant no. FONDECYT Iniciación

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

11240171. EMC thanks the support from the Swiss National Science Foundation through grant 200021_214907. MZB thanks the support from ANID PCI 190018 and ANID Fondecyt 1212071.

References

Acuña, P. and Pizarro, A.: Can continuous simulation be used as an alternative for flood regionalisation? A large sample example from Chile, J. Hydrol., 626, 130118, https://doi.org/10.1016/j.jhydrol.2023.130118, 2023.

Aerts, J. P. M., Hut, R. W., Van De Giesen, N. C., Drost, N., Van Verseveld, W. J., Weerts, A. H., and Hazenberg, P.: Large-sample assessment of varying spatial resolution on the streamflow estimates of the wflow_sbm hydrological model, Hydrol. Earth Syst. Sci., 26, 4407–4430, https://doi.org/10.5194/hess-26-4407-2022, 2022.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset, Hydrol. Earth Syst. Sci., 22, 5817–5846, https://doi.org/10.5194/hess-22-5817-2018, 2018.

Araya, D., Mendoza, P. A., Muñoz-Castro, E., and McPhee, J.: Towards robust seasonal streamflow forecasts in mountainous catchments: impact of calibration metric selection in hydrological modeling, Hydrol. Earth Syst. Sci., 27, 4385–4408, https://doi.org/10.5194/hess-27-4385-2023, 2023.

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Mendoza, P. A., McNamara, I., Beck, H. E., Thurner, J., Nauditt, A., Ribbe, L., and Thinh, N. X.: On the selection of precipitation products for the regionalisation of hydrological model parameters, Hydrol. Earth Syst. Sci., 25, 5805–5837, https://doi.org/10.5194/hess-25-5805-2021, 2021.

Baker, S. A., Rajagopalan, B., and Wood, A. W.: Enhancing ensemble seasonal streamflow forecasts in the Upper Colorado River Basin using multi-model climate forecasts, JAWRA J. Am. Water Resour. Assoc., 57, 906–922, 2021.

Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I. J. M., and Wood, E. F.: Global Fully Distributed Parameter Regionalization Based on Observed Streamflow From 4,229 Headwater Catchments, J. Geophys. Res. Atmospheres, 125, https://doi.org/10.1029/2019JD031485, 2020.

Belay, H., Melesse, A. M., and Tegegne, G.: Merging satellite products and rain-gauge observations to improve hydrological simulation: a review, Earth, 3, 1275–1289, 2022.

Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments, 1976.

Beven, K.: On undermining the science?, Hydrol. Process., 20, 3141–3146, https://doi.org/10.1002/hyp.6396, 2006.

Boisier, J. P.: CR2MET: A high-resolution precipitation and temperature dataset for the period 1960-2021 in continental Chile. (v2.5), https://doi.org/10.5281/zenodo.7529682, 2023.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

Bonney, M. T. and Zhang, Y.: Monitoring snow cover dynamics at 30-m resolution in higher latitude regions using Harmonized Landsat Sentinel-2, ISPRS J. Photogramm. Remote Sens., 233, 360–382, 2026.

Cha, S.-H.: Comprehensive Survey on Distance/Similarity Measures Between Probability Density Functions, Int J Math Model Meth Appl Sci, 1, 2007.

Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resour. Res., 47, 2011.

Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., and Schaefli, B.: Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets, Water Resour. Res., 56, 1–26, https://doi.org/10.1029/2019WR026085, 2020a.

Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., and Schaefli, B.: Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets, Water Resour. Res., 56, e2019WR026085, https://doi.org/10.1029/2019WR026085, 2020b.

Demirel, M. C., Koch, J., Rakovec, O., Kumar, R., Mai, J., Müller, S., and Thober, S.: Tradeoffs Between Temporal and Spatial Pattern Calibration and Their Impacts on Robustness and Transferability of Hydrologic Model Parameters to Ungauged Basins Water Resources Research, Water Resour. Res., 60, 1–24, https://doi.org/https://doi.org/10.1029/2022WR034193, 2023.

Demirel, M. C., Koch, J., Rakovec, O., Kumar, R., Mai, J., Müller, S., Thober, S., Samaniego, L., and Stisen, S.: Tradeoffs Between Temporal and Spatial Pattern Calibration and Their Impacts on Robustness and Transferability of Hydrologic Model Parameters to Ungauged Basins, Water Resour. Res., 60, e2022WR034193, https://doi.org/10.1029/2022WR034193, 2024.

Ding, J. and Zhu, Q.: The accuracy of multisource evapotranspiration products and their applicability in streamflow simulation over a large catchment of Southern China, J. Hydrol. Reg. Stud., 41, 101092, https://doi.org/10.1016/j.ejrh.2022.101092, 2022.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, Remote Sens. Environ., 203, 185–215, https://doi.org/10.1016/j.rse.2017.07.001, 2017.

Fowler, K., Peel, M., Western, A., and Zhang, L.: Improved Rainfall-Runoff Calibration for Drying Climate: Choice of Objective Function, Water Resour. Res., 54, 3392–3408, https://doi.org/10.1029/2017WR022466, 2018.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

Gaur, S., Singh, B., Bandyopadhyay, A., Stisen, S., and Singh, R.: Spatial pattern-based performance evaluation and uncertainty analysis of a distributed hydrological model, Hydrol. Process., 36, e14586, 2022.

Gómez, M. J., Barboza, L. A., Hidalgo, H. G., and Alfaro, E. J.: Comparison of indicators to evaluate the performance of climate models, Int. J. Climatol., 44, 4907–4924, https://doi.org/10.1002/joc.8619, 2024.

Guo, X., Wu, Z., Fu, G., and He, H.: A multi-variable calibration framework at the grid scale for integrating streamflow with evapotranspiration data to improve the simulation of distributed hydrological model, J. Hydrol. Reg. Stud., 55, 101944, https://doi.org/10.1016/j.ejrh.2024.101944, 2024.

Guo, Y., Zhang, Y., Zhang, L., and Wang, Z.: Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review, WIREs Water, 8, e1487, https://doi.org/10.1002/wat2.1487, 2021.

Guse, B., Han, L., Kumar, R., Rakovec, O., Luedtke, S., Herzog, A., Thober, S., Samaniego, L., and Wagener, T.: Spatio-Temporal Consistency and Variability in Parameter Dominance on Simulated Hydrological Fluxes and State Variables, Water Resour. Res., 60, e2023WR036822, https://doi.org/10.1029/2023WR036822, 2024.

Hargreaves, G. H. and Allen, R. G.: History and evaluation of Hargreaves evapotranspiration equation, J. Irrig. Drain. Eng., 129, 53–63, 2003.

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

Hsu, S.-C., de Lavenne, A., Perrin, C., and Andréassian, V.: Extra constraint on actual evaporation in a semi-distributed conceptual model to improve model physical realism, Hydrol. Sci. J., 70, 1143–1156, 2025.

Huang, Y., Bárdossy, A., and Zhang, K.: Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data, Hydrol. Earth Syst. Sci., 23, 2647–2663, https://doi.org/10.5194/hess-23-2647-2019, 2019.

Hurkmans, R. T., Van Den Hurk, B., Schmeits, M., Wetterhall, F., and Pechlivanidis, I. G.: Seasonal streamflow forecasting for fresh water reservoir management in the netherlands: an assessment of multiple prediction systems, J. Hydrometeorol., 24, 1275–1290, 2023.

Jiang, L., Wu, H., Tao, J., Kimball, J. S., Alfieri, L., and Chen, X.: Satellite-Based Evapotranspiration in Hydrological Model Calibration, Remote Sens., 12, 428, https://doi.org/10.3390/rs12030428, 2020.

Jin, X. and Jin, Y.: Calibration of a Distributed Hydrological Model in a Data-Scarce Basin Based on GLEAM Datasets, Water, 12, 897, https://doi.org/10.3390/w12030897, 2020.

Jorquera, J. and Pizarro, A.: Unlocking the potential of stochastic simulation through Bluecat: Enhancing runoff predictions in arid and high-altitude regions, Hydrol. Process., 37, https://doi.org/10.1002/hyp.15046, 2023.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

Karami, A., Baghdadi, N., Bazzi, H., Nasrallah, Y., Zribi, M., Najem, S., and Wigneron, J.-P.: Soil moisture estimation at 1-km resolution over croplands and grasslands using sentinel-1/2 and SMOS-IC data: algorithm and validation, Eur. J. Remote Sens., 59, 2622132, 2026.

Karpasitis, A., Hadjinicolaou, P., and Zittis, G.: A new efficiency metric for the spatial evaluation and inter-comparison of climate and geoscientific model output, EGUsphere, 1–27, https://doi.org/10.5194/egusphere-2025-1471, 2025.

Kennedy, J. and Eberhart, R.: Particle swarm optimization (PSO), Proc. IEEE international conference on neural networks, Perth, Australia, 1942–1948, 1995.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42, https://doi.org/10.1029/2005WR004362, 2006.

Klemeš, V.: Dilettantism in hydrology: Transition or destiny?, Water Resour. Res., 22, 177S-188S, 1986.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, J. Hydrol., 424–425, 264–277, https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

Koch, J., Demirel, M. C., and Stisen, S.: The SPAtial EFficiency metric (SPAEF): multiple-component evaluation of spatial patterns for optimization of hydrological models, Geosci. Model Dev., 11, 1873–1886, https://doi.org/10.5194/gmd-11-1873-2018, 2018.

Komma, J., Blöschl, G., and Reszler, C.: Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting, J. Hydrol., 357, 228–242, 2008.

Koutsoyiannis, D.: Knowable moments for high-order stochastic characterization and modelling of hydrological processes, Hydrol. Sci. J., 64, 19–33, https://doi.org/10.1080/02626667.2018.1556794, 2019.

Koutsoyiannis, D.: Replacing Histogram with Smooth Empirical Probability Density Function Estimated by K-Moments, Sci, 4, 50, https://doi.org/10.3390/sci4040050, 2022.

Koutsoyiannis, D.: Knowable Moments in Stochastics: Knowing Their Advantages, Axioms, 12, 590, https://doi.org/10.3390/axioms12060590, 2023.

Koutsoyiannis, D.: Stochastics of Hydroclimatic Extremes—A Cool Look at Risk, 2025.

Lombardo, F., Volpi, E., Koutsoyiannis, D., and Papalexiou, S.: Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology, Hydrol. Earth Syst. Sci., 18, 243–255, 2014.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

Mei, Y., Mai, J., Do, H. X., Gronewold, A., Reeves, H., Eberts, S., Niswonger, R., Regan, R. S., and Hunt, R. J.: Can Hydrological Models Benefit From Using Global Soil Moisture, Evapotranspiration, and Runoff Products as Calibration Targets?, Water Resour. Res., 59, e2022WR032064, https://doi.org/10.1029/2022WR032064, 2023.

Mendoza, P. A., McPhee, J., and Vargas, X.: Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment, Water Resour. Res., 48, 2012.

Miralles, D. G., Bonte, O., Koppa, A., Baez-Villanueva, O. M., Tronquo, E., Zhong, F., Beck, H. E., Hulsman, P., Dorigo, W., Verhoest, N. E. C., and Haghdoost, S.: GLEAM4: global land evaporation and soil moisture dataset at 0.1° resolution from 1980 to near present, Sci. Data, 12, 416, https://doi.org/10.1038/s41597-025-04610-y, 2025.

Necker, T., Wolfgruber, L., Kugler, L., Weissmann, M., Dorninger, M., and Serafin, S.: The fractions skill score for ensemble forecast verification, Q. J. R. Meteorol. Soc., 150, 4457–4477, 2024.

Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, Hydrol. Process., 21, 435–446, https://doi.org/10.1002/hyp.6253, 2007.

Parajka, J., Blaschke, A. P., Blöschl, G., Haslinger, K., Hepp, G., Laaha, G., Schöner, W., Trautvetter, H., Viglione, A., and Zessner, M.: Uncertainty contributions to low-flow projections in Austria, Hydrol. Earth Syst. Sci., 20, 2085–2101, https://doi.org/10.5194/hess-20-2085-2016, 2016.

Perrini, P., Iacobellis, V., Gioia, A., Cea, L., Savenije, H. H., and Fenicia, F.: Can dominant runoff generation mechanisms be disentangled through hypothesis testing? insights from integrated hydrological-hydrodynamic modeling, Water Resour. Res., 61, e2024WR039394, 2025.

Pimentel, R., Crochemore, L., Andersson, J. C., and Arheimer, B.: Assessing robustness in global hydrological predictions by comparing modelling and Earth observations, Hydrol. Sci. J., 68, 2357–2372, 2023.

Pizarro, A. and Jorquera, J.: Advancing objective functions in hydrological modelling: Integrating knowable moments for improved simulation accuracy, J. Hydrol., 634, 131071, https://doi.org/10.1016/j.jhydrol.2024.131071, 2024.

Pool, S., Vis, M., and Seibert, J.: Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency, Hydrol. Sci. J., 63, 1941–1953, https://doi.org/10.1080/02626667.2018.1552002, 2018.

Pradhan, R. K., Markonis, Y., Godoy, M. R. V., Villalba-Pradas, A., Andreadis, K. M., Nikolopoulos, E. I., Papalexiou, S. M., Rahim, A., Tapiador, F. J., and Hanel, M.: Review of GPM IMERG performance: A global perspective, Remote Sens. Environ., 268, 112754, 2022.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

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

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

Roberts, N. M. and Lean, H. W.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, Mon. Weather Rev., 136, 78–97, 2008.

Savenije, H. and Hrachowitz, M.: Behind the scenes of streamflow model performance, Hydrol Earth Syst Sci, 25, 645, 2021.

Segovia, S., Mendoza, P. A., Lagos-Zúñiga, M., Scaff, L., and Prein, A.: Benchmarking convection-permitting climate simulations for hydrological applications: A comparative study of WRF-SAAG and observation-based products, EGUsphere, 2025, 1–33, 2025.

Shah, S., Duan, Z., Song, X., Li, R., Mao, H., Liu, J., Ma, T., and Wang, M.: Evaluating the added value of multi-variable calibration of SWAT with remotely sensed evapotranspiration data for improving hydrological modeling, J. Hydrol., 603, 127046, https://doi.org/10.1016/j.jhydrol.2021.127046, 2021.

Shi, P., Wu, H., Qu, S., Yang, X., Lin, Z., Ding, S., and Si, W.: Advancing real-time error correction of flood forecasting based on the hydrologic similarity theory and machine learning techniques, Environ. Res., 246, 118533, 2024.

Sirisena, T. A. J. G., Maskey, S., and Ranasinghe, R.: Hydrological Model Calibration with Streamflow and Remote Sensing Based Evapotranspiration Data in a Data Poor Basin, Remote Sens., 12, 3768, https://doi.org/10.3390/rs12223768, 2020a.

Sirisena, T. A. J. G., Maskey, S., and Ranasinghe, R.: Hydrological Model Calibration with Streamflow and Remote Sensing Based Evapotranspiration Data in a Data Poor Basin, Remote Sens., 12, 3768, https://doi.org/10.3390/rs12223768, 2020b.

Sleziak, P., Szolgay, J., Hlavčová, K., Duethmann, D., Parajka, J., and Danko, M.: Factors controlling alterations in the performance of a runoff model in changing climate conditions, J. Hydrol. Hydromech., 66, 381–392, https://doi.org/10.2478/johh-2018-0031, 2018.

Sleziak, P., Výleta, R., Hlavčová, K., Danáčová, M., Aleksić, M., Szolgay, J., and Kohnová, S.: A Hydrological Modeling Approach for Assessing the Impacts of Climate Change on Runoff Regimes in Slovakia, Water, 13, 3358, https://doi.org/10.3390/w13233358, 2021.

https://doi.org/10.5194/egusphere-2026-2912
Preprint. Discussion started: 5 June 2026

Song, Y., Bindas, T., Shen, C., Ji, H., Knoben, W. J. M., Lonzarich, L., Clark, M. P., Liu, J., Werkhoven, K. Van, Lemont, S., Denno, M., Pan, M., Yang, Y., Rapp, J., Kumar, M., Rahmani, F., Thébault, C., Sawadekar, K., and Lawson, K.: High-resolution national-scale water modeling is enhanced by multiscale differentiable physics-informed machine learning, https://doi.org/10.22541/essoar.172736277.74497104/v1, September 2024.

Stisen, S., Soltani, M., Mendiguren, G., Langkilde, H., Garcia, M., and Koch, J.: Spatial Patterns in Actual Evapotranspiration Climatologies for Europe, Remote Sens., 13, 2410, https://doi.org/10.3390/rs13122410, 2021.

Széles, B., Parajka, J., Hogan, P., Silasari, R., Pavlin, L., Strauss, P., and Blöschl, G.: The Added Value of Different Data Types for Calibrating and Testing a Hydrologic Model in a Small Catchment, Water Resour. Res., 56, https://doi.org/10.1029/2019wr026153, 2020.

Tang, G., Clark, M. P., and Papalexiou, S. M.: SC-earth: A station-based serially complete earth dataset from 1950 to 2019, J. Clim., 34, 6493–6511, https://doi.org/10.1175/JCLI-D-21-0067.1, 2021.

Tran, B. N., Van Der Kwast, J., Seyoum, S., Uijlenhoet, R., Jewitt, G., and Mul, M.: Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps, Hydrol. Earth Syst. Sci., 27, 4505–4528, 2023.

Wang, H., Cao, L., and Feng, R.: Hydrological Similarity-Based Parameter Regionalization under Different Climate and Underlying Surfaces in Ungauged Basins, Water, 13, 2508, https://doi.org/10.3390/w13182508, 2021.

Willmott, C. and Matsuura, K.: Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance, Clim. Res., 30, 79, https://doi.org/10.3354/cr030079, 2005.

Willmott, C. J.: ON THE VALIDATION OF MODELS, Phys. Geogr., 2, 184–194, https://doi.org/10.1080/02723646.1981.10642213, 1981.

Xiao, X., He, T., Liang, S., Liang, S., Liu, X., Ma, Y., and Wan, J.: Towards a gapless 1 km fractional snow cover via a data fusion framework, ISPRS J. Photogramm. Remote Sens., 215, 419–441, 2024.

Yorulmaz, E. B., Kartal, E., and Demirel, M. C.: Benchmarking multi-component spatial metrics for hydrologic model calibration using MODIS AET and LAI products, https://doi.org/10.22541/essoar.169290537.78371628/v1, August 2023.

Yorulmaz, E. B., Kartal, E., and Demirel, M. C.: Toward robust pattern similarity metric for distributed model evaluation, Stoch. Environ. Res. Risk Assess., 38, 4007–4025, https://doi.org/10.1007/s00477-024-02790-4, 2024.

Zambrano-Bigiarini, M. and Rojas, R.: A model-independent Particle Swarm Optimisation software for model calibration, Environ. Model. Softw., 43, 5–25, https://doi.org/10.1016/j.envsoft.2013.01.004, 2013.