Hydrological Sciences Journal

ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: www.tandfonline.com/journals/thsj20

Assessing the performance of single and multi-criteria calibration approaches for hydrological modelling: a comparative analysis

Silvano Fortunato Dal Sasso, Alonso Pizarro, Beniamino Onorati, Maria Rosaria Margiotta, Yijian Zeng, Zhongbo Su, Salvatore Manfreda & Mauro Fiorentino

To cite this article: Silvano Fortunato Dal Sasso, Alonso Pizarro, Beniamino Onorati, Maria Rosaria Margiotta, Yijian Zeng, Zhongbo Su, Salvatore Manfreda & Mauro Fiorentino (25 Nov 2025): Assessing the performance of single and multi-criteria calibration approaches for hydrological modelling: a comparative analysis, Hydrological Sciences Journal, DOI: 10.1080/02626667.2025.2579162

To link to this article: https://doi.org/10.1080/02626667.2025.2579162

Published online: 25 Nov 2025.

HYDROLOGICAL SCIENCES JOURNAL
https://doi.org/10.1080/02626667.2025.2579162

Assessing the performance of single and multi-criteria calibration approaches for hydrological modelling: a comparative analysis

Silvano Fortunato Dal Sasso[a], Alonso Pizarro[b], Beniamino Onorati[c], Maria Rosaria Margiotta[c], Yijian Zeng[d], Zhongbo Su[d], Salvatore Manfreda[e] and Mauro Fiorentino[c]

[a]Department for Humanistic, Scientific, and Social Innovation, University of Basilicata, Matera, Italy; [b]Escuela de Ingeniería en Obras Civiles, Universidad Diego Portales, Santiago, Chile; [c]Department of Engineering, University of Basilicata, Potenza, Italy; [d]Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, the Netherlands; [e]Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, Naples, Italy

ABSTRACT

Accurate hydrological modelling is crucial for understanding natural processes and managing water resources. However, simulation accuracy depends on the availability of field observations for calibration and validation. It is therefore critical to develop effective calibration strategies to reduce prediction uncertainties. This study applies the DREAM model to the experimental basin of Fiumarella of Corleto in Southern Italy to assess the benefits of single and multicriteria calibration approaches. The former uses total runoff; the latter optimizes total runoff, baseflow, and annual water balance. The study also compares uniform or spatially-based parameterization, including correction factors and recession constants. Parameters were optimized through automatic calibration using a genetic algorithm (GA) and the Kling-Gupta efficiency (KGE) as the objective function. Results show that spatially distributed information improves model reliability compared to a uniform parameterization set-up. The multi-objective calibration constrained on baseflow and balance allowed us to optimize the model, reducing variability compared to mono-objective calibration.

ARTICLE HISTORY
Received 7 January 2025
Revised 4 July 2025
Accepted 26 September 2025

EDITOR
K. Soulis

ASSOCIATE EDITOR
T. Krueger

KEYWORDS
Hydrological modelling; model calibration; genetic algorithms; DREAM model; multi-criteria approach

1 Introduction

Hydrological models are widely used by the scientific community and in technical practice for various water resource management and decision making purposes. In the last few decades, several distributed hydrological models have been proposed to describe the physics of each component of the hydrological basin cycle (Kampf and Burges 2007). However, due to the growing number of parameters that are typical of complex models, and the scarcity of data, the reliability of these models is often conditioned by the availability of the observed data and by the calibration and validation procedures (Beven 2002). In this regard, in recent years, the increasing availability of remotely sensed data at large spatial and temporal scales, along with advancements in computational power enabled by modern computers (Blair *et al.* 2019), has significantly influenced approaches to data utilization and calibration techniques. Remote sensing data such as evapotranspiration (Immerzeel and Droogers 2008, Demirel *et al.* 2018), surface soil moisture (Echeverría *et al.* 2019, Yang *et al.* 2019), and Normalized Difference Vegetation Index (NDVI) (Ruiz-Pérez *et al.* 2017) allowed a great opportunity to overcome the calibration and validation issue induced by the lack of reliable in situ hydrometeorological data in poorly gauged basins, where the evaluation of hydrological models is quite challenging to achieve. This is even more common when there is a practical need to predict streamflow at certain internal locations and not only at the outlet of a basin.

This inevitably leads to the identification of parameter sets that can reproduce the overall behaviour of the basin. However, it also introduces complexity in selecting the most optimal solutions among them (Efstratiadis and Koutsoyiannis 2010). In this context, time-consuming4 manual techniques were substituted by a more efficient automatic procedure that uses optimization techniques to estimate proper parameters (sets) for hydrological models. Various techniques have been developed since the 1980s, including shuffled complex evolution (SCE), simulated annealing (SA), particle swarm optimization (see e.g. the HydroPSO package in R), Pareto optimization, and other algorithms (Reed *et al.* 2013). Stefnisdóttir *et al.* (2021) showed that genetic algorithms (GAs) outperformed the other methods and are particularly suitable for reproducible and consistent hydrological simulations. All were attempts to find a global optimum or Pareto set of optima in the case of multiple competing objective functions.

In parallel, a growing number of studies have introduced multi-objective calibration (MOC) frameworks to improve the identification of model parameters by simultaneously optimizing multiple aspects of the hydrological response (Seibert 2000, Shafii and De Smedt 2009). These frameworks generally involve either the simultaneous use of

CONTACT Silvano Fortunato Dal Sasso silvano.dalsasso@unibas.it Department for Humanistic, Scientific, and Social Innovation (DIUSS), University of Basilicata, Via Lanera 20, Matera 75100, Italy

multiple statistical objective functions (e.g. NSE, PBIAS, RMSE) or the application of a single performance metric to different components or conditions of the hydrological system. Common strategies include the use of multiple criteria for the same output variable, diversified according to flow regimes (Zhang *et al.* 2018), calibration at multiple gauging stations (Bai *et al.* 2017), or the incorporation of multiple output variables such as streamflow, soil moisture, and snow cover (Brocca *et al.* 2012, Manfreda *et al.* 2018). Other approaches have focused on integrating physically meaningful components, such as runoff ratio, various properties of the flow duration curve, and a simple index of watershed lag time (Yilmaz *et al.* 2008), or constraining the parameter space using soil and land cover information through pedo-transfer functions (Weber *et al.* 2024). Segment-based calibration schemes, such as those applied to different parts of the hydrograph (e.g. low, medium, and high flows), have also been tested to capture flow regime variability better (Tigabu *et al.* 2024). Furthermore, recent works have underscored the challenge of achieving parameter transferability and robustness in data-scarce or ungauged basins (Hrachowitz *et al.* 2013, Demirel *et al.* 2022), highlighting the need for calibration strategies that balance complexity and generalizability.

Although prior studies have provided valuable contributions, few have systematically assessed how progressively increasing physical realism across multiple model configurations affects calibration robustness and performance within a unified experimental framework. This study addresses this gap by offering a structured comparison of model configurations with increasing physical realism and by evaluating their impact on model performance and stability during both calibration and validation periods, using a consistent experimental set-up.

Specifically, the present study evaluates single- and multi-criteria calibration strategies applied to the Corleto basin in southern Italy, using the semi-distributed DREAM hydrological model and a GA optimization scheme. We specifically test a consistent, component-based multi-objective approach, where calibration targets runoff, baseflow, and long-term water balance using a single aggregated objective function (KGE). A series of model configurations with progressively stronger physical constraints is compared in terms of model performance under total flow and distinct flow regimes (low, medium, high), across both calibration and validation periods. This study ultimately seeks to provide practical guidance for designing calibration strategies that are both computationally efficient and physically consistent, particularly in data-limited contexts. To achieve this, we address the following research questions:

1. How beneficial is the inclusion of physical information in the calibration scheme with respect to a fully automated calibration procedure in terms of model performance and computation cost?

2. Is the multi-objective calibration scheme based on flow components appropriate in representing different flow conditions with respect to mono-objective calibration?

The study area and the DREAM model are introduced in Section 2 below, along with the datasets used for modelling. The calibration procedure, including the multi-objective calibration methods and their objective functions, is described in Section 3. The numerical results of the calibration and comparison of different methods are presented and discussed in Sections 4 and 5. The paper is concluded in Section 6.

2 Model and data

2.1 The DREAM model

The Distributed model for Runoff et Antecedent soil Moisture simulation (DREAM), introduced by Manfreda *et al.* (2005), is a semidistributed hydrological model, suitable for continuous simulations at daily and hourly temporal scales. The main hydrological processes are computed on a grid-based representation of the river basin that takes into account the spatial heterogeneity of hydrological variables using digital elevation models and soil and vegetation grid maps. The DREAM model was applied in several medium-sized basins, exhibiting considerable differences in climate and other physical characteristics (e.g. Manfreda *et al.* 2005, Fiorentino *et al.* 2007, Gigante *et al.* 2009) and was recently coupled with a shallow-water model for flood hazard modelling (Perrini *et al.* 2024).

The DREAM model conceptualization is briefly described below (Manfreda *et al.* 2005, Perrini *et al.* 2024) along with the main inputs required and model outputs (Fig. 1). Canopy cover determines the amount of rainfall intercepted by vegetation before it hits the soil surface. Throughfall (precipitation minus interception) is initially stored in surface depressions. The potential evapotranspiration was estimated using the Penman-Monteith equation modified by the FAO (Allen *et al.* 1998). Effects of basin morphology were incorporated in the computation using the analytical algorithm developed by Allen *et al.* (2006) for the estimation of the incident solar radiation, which affects evapotranspiration as well as snowmelt, taking into account both aspect and slope of the surface. Net precipitation (throughfall minus depression storage) is then sub-divided into surface runoff and infiltration into the soil; soil water content is redistributed within river sub-basins according to the morphological structure of the basin, exploiting the wetness index proposed by Beven and Kirby (1979). Groundwater is recharged by the percolation through the vadose zone and is routed as a global linear reservoir.

2.1 Case study

The experimental basin "Fiumarella of Corleto," located in the Basilicata region (Southern Italy), is a tributary of the Sauro River (Agri basin) and has an area of 32.5 km^2. The basin's topography has a mean altitude of 1050 m (ranging from 650 to 1500 m). The basin's climate zone is classified as sub-humid, characterized by hot and humid summers and chilly to mild winters. The mean annual precipitation is 720 mm, and the mean annual temperature is 9.9°C. The basin's hydrology is continuously monitored through two permanent weather stations, with one on either slope to highlight the differences between them. Moreover, an additional rainfall station and

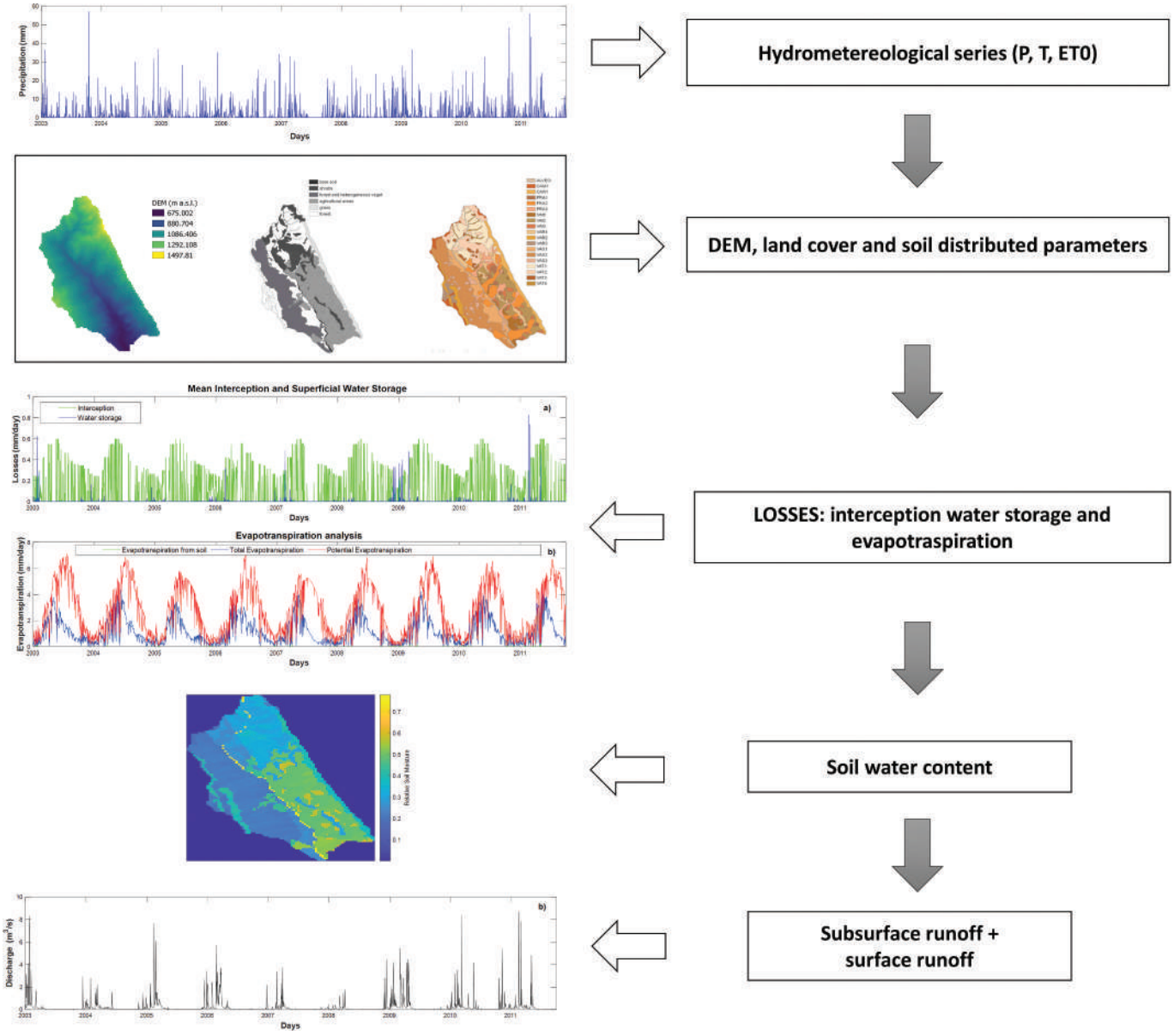

Figure 1. Conceptual scheme of the DREAM hydrological model with a description of the main inputs and outputs involved in its application.

streamflow gauge are located at the basin outlet. A general description of the basin is given in Fig. 2, where the geographical location of the basin (Fig. 2(a)), the sub-basin (Fig. 2(b)) and its experimental set-up are shown (e.g. the meteo-hydrological station (2b) and raingauge station located on two different slopes (2d), and the basin outlet (2c)). Station no. 1 (at 1115 m a.s.l.) is composed of a raingauge that becomes a snow gauge at low temperatures with the activation of heat resistance, a temperature sensor (active from September 2002), and sensors for the measurement of solar radiation, air relative humidity, wind velocity, and soil moisture (active from February 2006). Station no. 2 (at 680 m a.s.l.) is composed of a raingauge and a hydrometric ultrasound sensor, while station no. 3 (at 1040 m a.s.l.) is simply equipped with a raingauge. The recording period spans from September 2002 to the present. More details about the monitoring installations available on the basin are also provided in Manfreda *et al.* (2011).

2.2 Geographical database

2.2.1 Topography

The DREAM model was run at 80 m spatial resolution, upscaled from a 20 m resolution digital elevation model (DEM) in order to reduce computation time. The pre-processing consisted of filling sinks using the Wang and Liu algorithm (Wang and Liu 2006) and generating surface derivatives, including aspect, slope and topographic index maps.

2.2.2 Land cover and vegetation dynamics

As mentioned above, Fig. 1 shows the conceptual scheme of the DREAM model with the description of the main inputs employed. The land cover map results from data elaborated by Santini *et al.* (1999) that was thereafter used by Carriero *et al.* (2005) to define the soil hydraulic properties of each unit. The vegetation dynamics were taken into account through the use

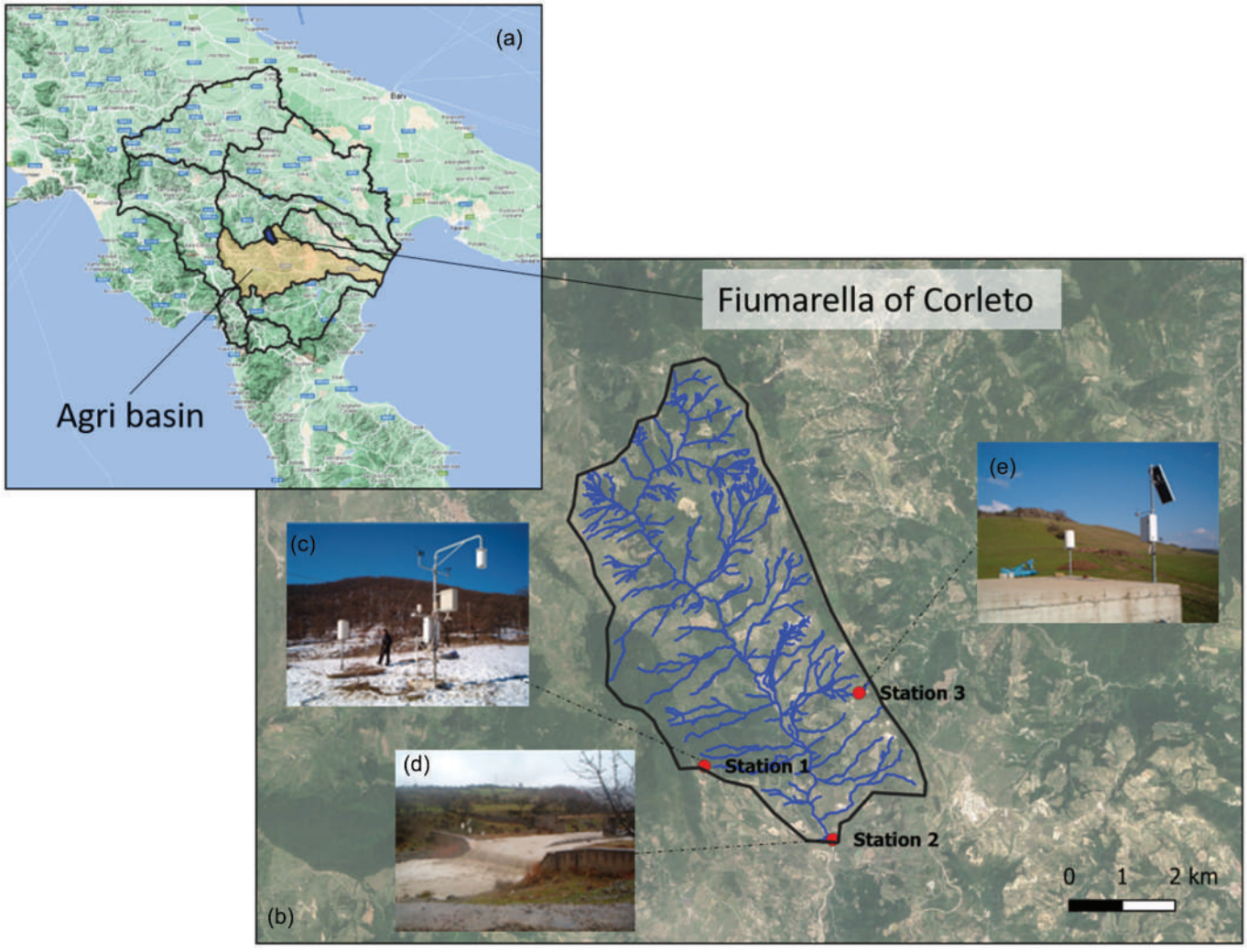

Figure 2. Description of the experimental area of "Fiumarella of Corleto" in the Agri basin (a) with the identification of the three permanent hydrological stations (b) devoted to the continuous monitoring of the basin: (c) meteo-hydrological station, (d) hydrometer and raingauge station, (e) raingauge station.

of the Leaf Area Index (LAI), which was determined monthly using NDVI values obtained from surface reflectance data acquired by the 1 km × 1 km NOAA-AVHRR sensor onboard National Oceanic and Atmospheric Administration satellites. The map was then resampled to a spatial resolution of 80 m to align it with the spatial resolution of the other variables used in the modelling process. It is worth noting that the Fiumarella basin is located in a mountainous area with minimal anthropogenic pressure, and satellite observations over the years (Land Cover 2006, 2012, 2018) have confirmed the absence of significant land use changes since 1999.

2.2.3 Soil physical and hydraulic properties

The basin's pedology was derived from the studies of Romano and Palladino (2002) and Romano and Santini (1997) that investigated the main soil-land units of the basin through field campaigns and laboratory measurements. For the zones without available information, the data was integrated with a land system map of the whole Agri basin, whose soil classification was realized according to the HYPRES (HYdraulic PRoperties of European Soil) database that also defines hydraulic characteristics (Carriero *et al.* 2005). The soil characterization is articulated in the following physical quantities: soil texture (according to the USDA classification), organic carbon, and thickness.

The use of simplified techniques allows us to define the hydrological quantities generally required by simulation models, such as field capacity and water content at saturation. These are the pedotransfer functions (PTFs), statistical functions that derive the hydraulic soil characteristics from the physical ones. In particular, for the estimation of soil water content at saturation (Saxton *et al.* 1986), we used:

$$\theta_s = 0.332 - 7.251 \cdot 10^{-4}(\%\text{sand}) + 0.1276 \cdot \log_{10}(\%\text{clay}) \quad (1)$$

For the estimation of residual water content θ_r (Vereecken *et al.* 1989), we used:

$$\theta_r = 0.015 + 0.005(\%\text{clay}) + 0.014\left(\%C_{\text{organic}}\right) \quad (2)$$

For the evaluation of the field capacity θ_c (Rawls *et al.* 1982), hypothesized as the value of the water content at the potential h = −33 kPa, we used:

$$\theta_c = 0.2576 + 0.002(\%\text{sand}) + 0.0036(\%\text{clay}) + 0.0299(\%\text{Corgnic}) \quad (3)$$

Finally, for the estimation of the hydraulic conductivity at saturation Ks, we used another PTF (Cosby *et al.* 1984):

$$k_s = 0.6096 \cdot 10 - 0.6 + 0.0126(\%\text{sand}) - 0.0064(\%\text{clay}) \quad (4)$$

The resulting physical and hydraulic characteristics of the soil are given in Table 1 alongside other model parameters.

2.2.4 Hydro-meteorological data

The hydro-meteorological data used cover 9 years (from January 2003 to September 2011). This period was sufficiently long to include representative dry and wet conditions and almost continuous rainfall registration spanning the entire period without gap-filling. The inclusion of more recent years beyond 2011 was not considered in this study due to the presence of data gaps and changes in the instrumentation set-up. As highlighted by Aung *et al.* (2025), the long-term maintenance of the Fiumarella experimental basin has faced recurring challenges, including damage to hydrometric stations from extreme flood events and the need for periodic reconstruction and recalibration. These operational constraints introduced uncertainty in the continuity and quality of recent observations, potentially compromising the reliability of model calibration and validation.

The average precipitation records were calculated using the Thiessen polygon map obtained from raingauge observations. During the monitoring period, the water discharge at the outlet was estimated using the rating curve at the hydrometric station. Figure 2 shows an overview of the different inputs needed for the application of the DREAM model.

3 Calibration procedure

We adopted single and multi-criteria calibration procedures that optimize: (A) the total runoff at the basin outlet (i.e. the total streamflow resulting from the combined contribution of surface runoff, subsurface flow, and baseflow); (B) both the surface runoff and baseflow at the basin outlet; and (C) the combination of (B) and the annual water balance components. In all cases, parameter optimization was carried out using an automatic calibration performed by GAs (Goldberg 1989). GAs are efficient automated tools for model calibration that use an iterative parameter search. Each parameter generation contributes to the next generation, using past experiences to produce efficient solutions. Through the genetic operators (selection, crossover, and mutation), the population evolves and searches for the optimal parameters set by selecting individuals randomly. For the present application, we set the population size at 20, the number of iterations/generation equal to 300, and the function tolerance equal to 10^{-8}. A single objective algorithm was used in approach A, considering 10 runs; for the rest of the configurations, a multi-objective approach was considered. In these cases, calibration was performed by targeting multiple hydrological components (surface runoff, baseflow, and water balance), using the Kling-Gupta efficiency (KGE) as the single objective function applied separately to each component.

Since all Pareto optimal solutions are equally important and it may be difficult to prefer one solution to another, every solution in the Pareto front was used to run the model. This allowed us to obtain 70 sets of simulated streamflow (7 solutions × 10 runs for each multi-objective approach) for a total of 280 runs. Finally, the model performance was assessed using root mean square error (RMSE) and Nash-Sutcliffe efficiency (NSE).

3.1 Calibration-validation configurations

The three different calibration approaches were carried out using the first four years of data (2004–2008), whereas validation was carried out during 2008–2011. A warm-up period from January 2003 to September 2004 was used to assign an initial value to the model state variables. The soil moisture content at the beginning of the calibration period was assumed to be equal to the last computed point of the warm-up period. In all configurations, the DREAM was run at 80 m and daily scale. For each approach (A, B, C), five different configurations were adopted in order to test the model performances during both the calibration and validation periods, described as follows:

- Configuration 1: uniform parameterization;
- Configuration 2: spatial parameterization obtained assigning the soil hydraulic parameters based on soil texture using PTFs;
- Configuration 3: as in configuration 2, including a scaling factor in all cells of the basin. This factor is the same for

Table 1. Calibration parameters and ranges considered.

Parameter	Description	Range
K_s	Soil permeability to saturation (mm/day)	0.1–5
β	Dimensionless exponent beta (-)	0.1–5
coef_red	Sub-surface runoff redistribution coefficient	0–0.4
K_g	Baseflow recession coefficient (1/day)	0–0.4
M_{fr}	Melting factor (mm/°C * day)	0.02–0.06
T_n	Precipitation to snow	−3–+1
T_0	Melting temperature	0–4
ΔT	°C	0.01–5
D	Soil depth (mm)	100–1500
θ_c	Field capacity	0–0.5
θ_s	Soil water content at saturation	0–0.7

all cells in the basin, so the variability and distribution of the original parameter estimation were respected;
- Configuration 4: as in configuration 3, assigning the recession constant (Kg) based on the observed streamflow recession curves.

Variable outputs were weighted in the same way to avoid prioritizing any of them. The efficiency of simulating these three output variables was assessed using the single objective function Kling-Gupta efficiency (KGE) from Gupta *et al.* (2009), exploring the multiple sub-optimal parameter sets according to the Pareto front. The KGE is considered herein for model calibration, notably because the KGE has proven to be efficient in reproducing the magnitude and variability of hydrological extremes (Mizukami *et al.* 2019).

Figure 3 displays the workflow followed. In addition, Table 1 summarizes the variables used for calibration in the DREAM model and the lower and upper variation bound for each parameter derived from experience and literature references (Manfreda and Mancusi 2013, Manfreda *et al.* 2018).

The calibration methodology exploits a physically-based filter to decompose the streamflow time series into two time series referring to the surface and baseflow components. The baseflow component was estimated by interpolation using a cubic spline applied to the recorded local minima (Manfreda *et al.* 2018). Figure 4(b) shows the linear groundwater reservoir's recession constant, estimated by log-linear regression (Fig. 4(c)) on the recession curve of daily streamflow records (Fig. 4(a)).

Moreover, a comparison with three grouped flow indicators was performed in order to analyse the model performance with respect to different flow conditions: low (LQ), medium (MQ), and high (HQ). These indicators were selected as reported in Table 2 (Gnann *et al.* 2021, Sikorska-Senoner 2021).

3.2 Main error metrics adopted

The RMSE is given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - S_i)^2} \tag{5}$$

where M and S represent the measured and simulated time series, respectively, and N is the number of components in the series. An RMSE value of 0 would describe a perfect fit to observed data.

The KGE incorporates three different statistical measures (the correlation coefficient, r; the variability error, a = σS/σm; and the bias error, b = μS/μM) of the relation between measured and simulated data into one objective function. μS and μM are the mean values of measured and simulated data, while σS and σM are the standard deviations.

$$\mathrm{KGE} = \sqrt{1 - (r-1)^2 + (a-1)^2 + (b-1)^2} \tag{6}$$

The KGE can range between −∞ and 1, where 1 is the optimal value.

The NSE is a normalized efficiency coefficient (Nash and Sutcliffe 1970). It determines the relative magnitude of the residual variance compared with the measured data variance.

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(S_i - M_i)^2}{\sum_{i=1}^{n}\left(M_i - \overline{M_i}\right)^2} \tag{7}$$

where Si and Mi are the predicted and observed values at a given time step. The NSE varies from −∞ to 1, where 1 corresponds to the maximum agreement between predicted and observed values.

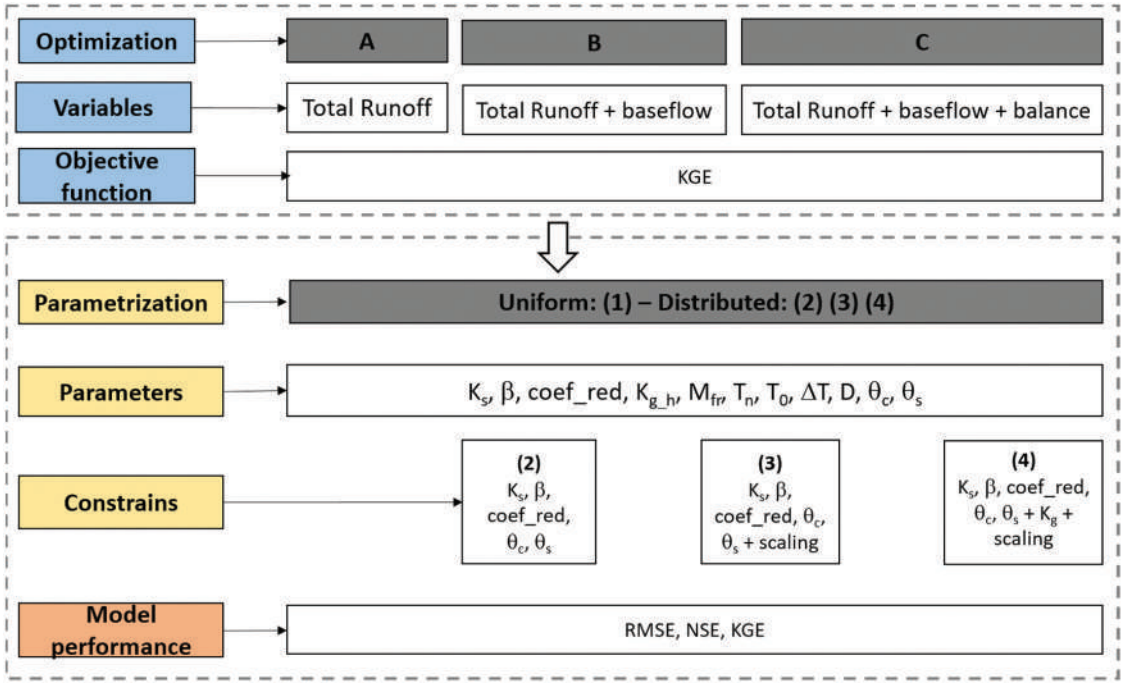

Figure 3. Summary of the calibration configuration.

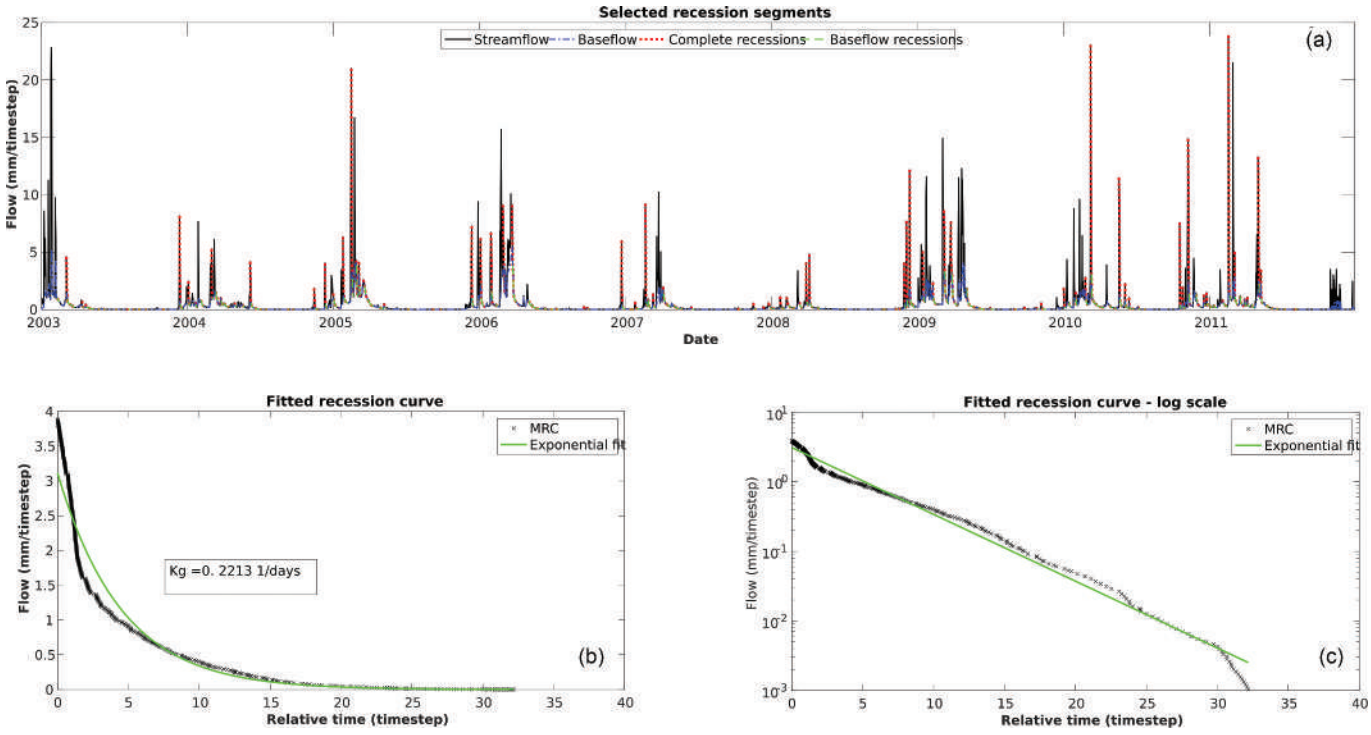

Figure 4. Streamflow and baseflow recessions for the period 2003–2011 (a); master recession curve (b) and exponential fitting (c).

4 Results

4.1 Comparison between single and multi-criteria calibration schemes

Results of DREAM simulations are displayed in Fig. 5, comparing simulated vs. measured time series (Fig. 5(a)) and flow duration curves (Fig. 5(b)) at the basin outlet, both relative to the entire record of observations. Generally, we obtained satisfactory results in representing flow regimes, with RMSE between 0.25 and 0.34 m^3/s, while NSE and KGE values were 0.43–0.67 and 0.46–0.82, respectively. Table 3 shows the best results for the entire set of runs performed for each configuration using the three adopted metrics.

Generally, considering the overall configuration results, it is worth observing that distributed configurations (2, 3, 4) provided, on average, better results compared to the uniform configuration (1). This result is appreciable in both mono- and multi-criteria approaches and corroborates the value of the distributed parameterization of the model when available data is consistent in representing the spatial variability characteristics of the basin.

Figures 6–8 present the box plots of RMSE, NSE and KGE for all configurations B and C during calibration (left panels) and validation (right panels). All this information is analytically explicated in the Appendix. Starting with uniform configuration 1, the results show that the use of multi-objective calibration approaches (B and C) provides a substantial improvement in model performance compared to the single criterion based on total runoff. This was primarily observed in the calibration phase, whereas no improvement in model performance is observed moving from approaches B and C to A in the validation phase, where the model realizes an independent quality check in a period different from the one used for calibration.

Considering the spatial distributed configuration 2, the inclusion of spatial information related to PTF offered a strategy to obtain reliable results in all the approaches considered. In this case, no significant improvement in model performance was observed for the multi-objective approaches B and C compared to A that in this configuration provides the best results.

This was particularized in configuration 3, with the inclusion of scaling factors for the correction of soil physical properties. In this case, no improvement of the overall metrics was observed during the calibration process in comparison to configuration 1.

Moving to configuration 4, the inclusion of the recession constant together with physical soil characteristics in calibration allowed us to further reduce the number of free parameters, thus reducing the scattering between Pareto optimal solutions. In particular, this was observed for all the approaches considered in both calibration and validation periods.

The inclusion of physical constraints, coupled with the scaling correction in all cells of the basin and the recession constant, ensured greater reliability and robustness of the model, especially verified in the validation phase. This approach improved the model performances in calibration and validation periods compared to configuration 2.

The reduced variability in model performance observed in configuration 4 is consistent across all calibration approaches and is mainly due to the inclusion of physical constraints – specifically, the pedotransfer-based soil parameterization and the recession constant derived from observed streamflow data.

Table 2. Hydrological signatures selected for the evaluation of model performance.

Cat	Signature	Description	Unit
LQ	BFI	Baseflow index	–
	Q5	5th streamflow percentile	(m^3/s)
	low_Q_ frequency	High Q duration	–
	low_Q_duration	Low flow duration	Time step
	BaseflowRecessionK	Exponential recession constant	(1/d)
	BaseflowMagnitude	Difference between maximum and minimum of annual baseflow regime	(mm)
	RecessionK_early	Recession constant of early recessions	1/time step
	Q_7_day_min	7-day minimum streamflow	mm/time step
	Q_var	Variance of streamflow	mm^2/time $step^2$
	Spearmans_rho	Non-uniqueness in the storage-discharge relationship	–
MQ	Q_mean	Mean streamflow	mm/time step
	Q_mean_monthly	Mean monthly streamflow	mm/time step
	Q_CoV	Coefficient of variation	–
	Q_skew	Skewness of streamflow	mm^3/time $step^3$
	VariabilityIndex	Variability index of flow	–
	Recession_a_Seasonality	Seasonal variations in recession parameters	–
	TotalRR	Total runoff ratio	–
	RR_Seasonality	Runoff ratio seasonality	–
HQ	Q95	95th streamflow percentile	(m^3/s)
	high_Q_duration	High flow duration	Time step
	high_Q_frequency	High flow frequency	–
	PeakDistribution	Slope of distribution of peaks	–
	FlashinessIndex	Richards-Baker flashiness idex	–
	EventRR	Event runoff ratio	–
	HFD_mean	Half flow date	Day of year
	HFI_mean	Half flow interval	Days
	RLD	Rising limb density	1/time step
	Q_7_day_max	7-day maximum streamflow	mm/time step
	AC1	Lag-1 autocorrelation (-)	–
	FDC_slope	Slope of the flow duration curve	–
	QP_elasticity	Streamflow-precipitation elasticity	–

These constraints reduce model flexibility and limit deviations from physically plausible behaviour. It is also important to highlight that, for all configurations, the variability of performance metrics increases in the validation phase. This behaviour, more pronounced in configurations with lower physical realism, can be attributed to the reduced model adaptability under hydrological conditions that differ from those of the calibration period. This highlights the importance of structural robustness and physically-based constraints to maintain performance stability over time.

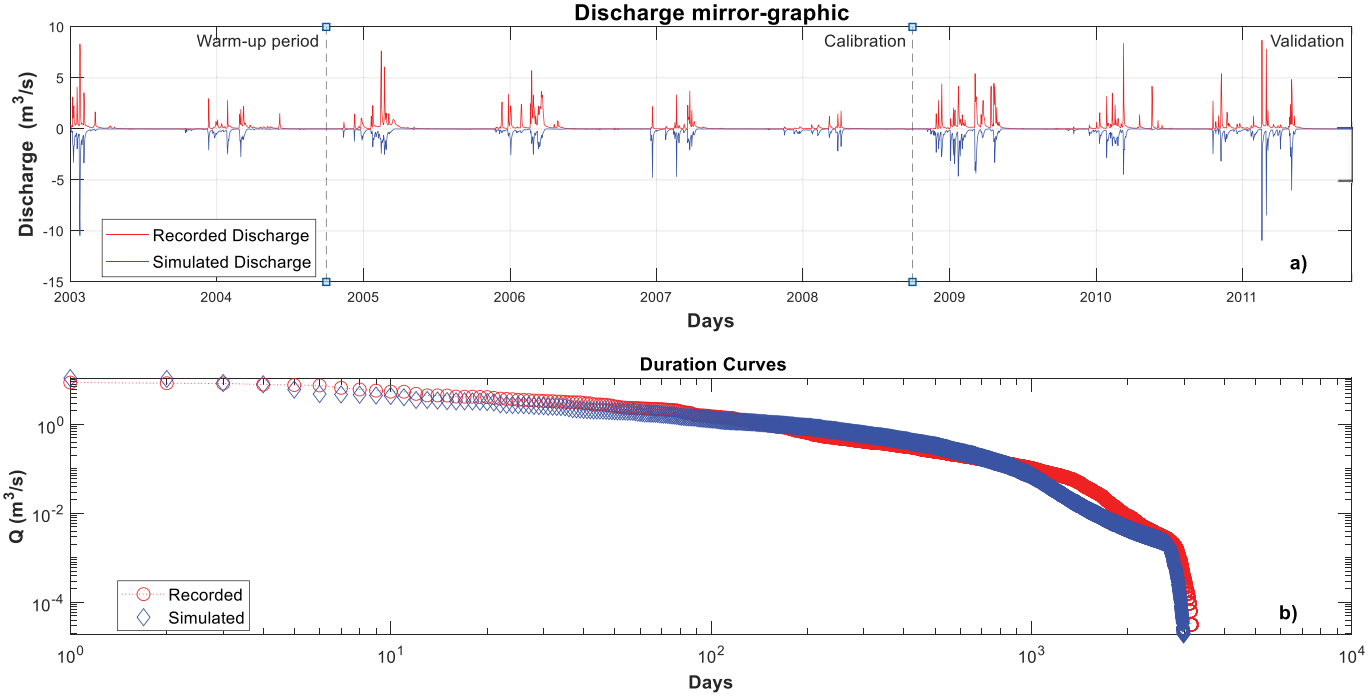

Figure 5. Comparison between the simulated and recorded daily streamflow for the "Fiumarella of Corleto" basin (a) and flow duration curves (b) for the entire period.

Table 3. Summary of the best results obtained for all the configurations considered.

	A			B			C		
	RMSE	NSE	KGE	RMSE	NSE	KGE	RMSE	NSE	KGE
(1) Uniform parameterization									
Calibration (2004–2008)	0.32	0.49	0.49	0.34	0.43	0.42	0.32	0.49	0.45
Validation (2008–2011)	0.30	0.53	0.75	0.28	0.58	0.79	0.29	0.57	0.78
(2) PTF									
Calibration (2004–2008)	0.32	0.49	0.57	0.31	0.51	0.59	0.31	0.53	0.58
Validation (2008–2011)	0.27	0.63	0.77	0.27	0.63	0.77	0.27	0.63	0.76
(3) PTF + scaling									
Calibration (2004–2008)	0.33	0.46	0.51	0.32	0.48	0.50	0.33	0.46	0.49
Validation (2008–2011)	0.25	0.67	0.82	0.29	0.57	0.65	0.25	0.67	0.82
(4) PTF + scaling + Kg									
Calibration (2004–2008)	0.32	0.47	0.49	0.32	0.48	0.51	0.32	0.48	0.51
Validation (2008–2011)	0.25	0.67	0.81	0.26	0.66	0.81	0.25	0.66	0.81

Table 4. Computation time for each calibration scheme adopted.

Calibration scheme	Time (h)
A1	14.74
A2	11.09
A3	11.72
A4	11.03
B1	33.81
B2	34.50
B3	34.07
B4	32.76
C1	32.84
C2	33.02
C3	32.22
C4	32.95

4.2 Comparison with hydrological signatures

Model performances compared to the hydrological signatures representing flow conditions at the water outlet station were evaluated considering different hydrological signatures for each low, medium and high streamflow during the entire period (Figs. 9–11). The results show that, for all signatures, the model can reach similar satisfying performances in all three approaches considered, with a decrease of relative error from uniform to spatially distributed configurations. On average, results show a general increase in model robustness with a significative reduction of variability, moving to configuration 4 for approaches B and C. In particular, the model constrained by the recession constant (B4 and C4) proved to be sufficiently robust to simulate some hydrological signatures related, respectively, to low conditions (BFI, low Q duration, baseflow magnitude), medium (i.e. skewness of streamflow) and high

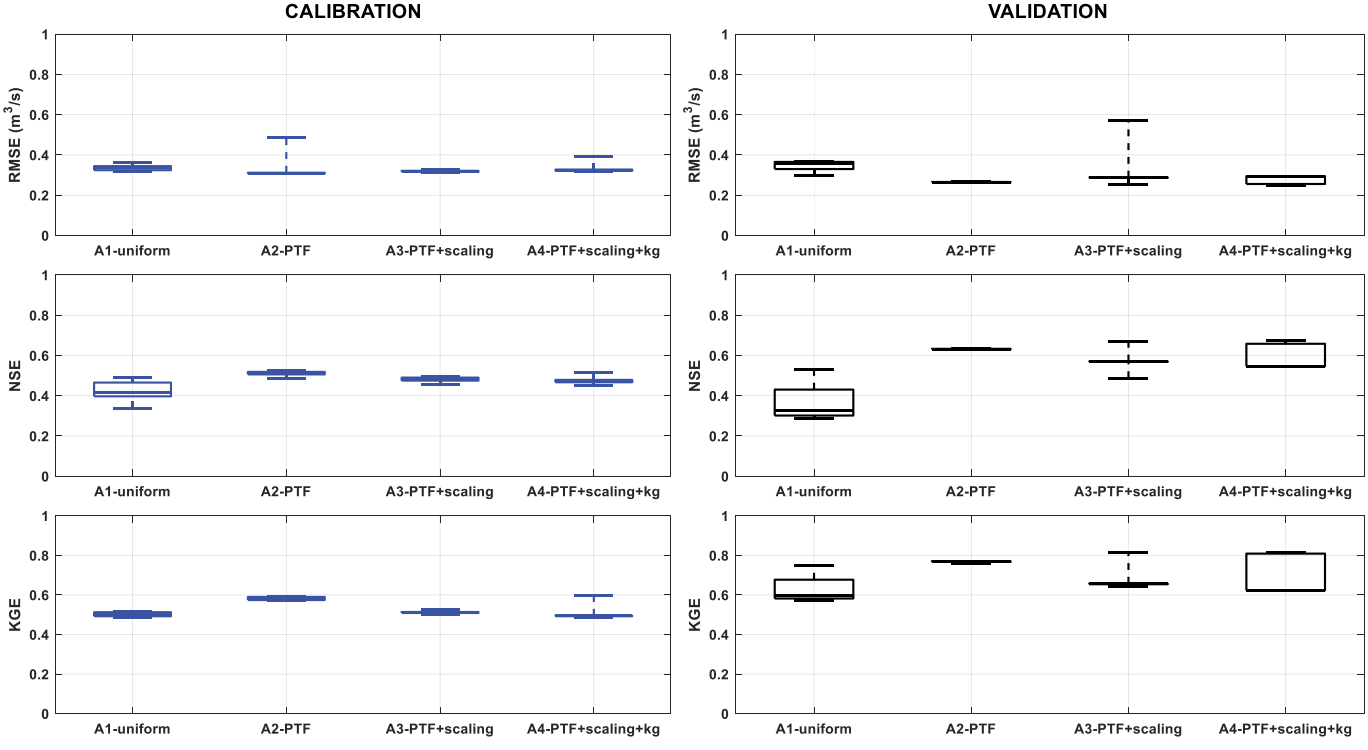

Figure 6. Box plots of RMSE, NSE, KGE values considering configuration A.

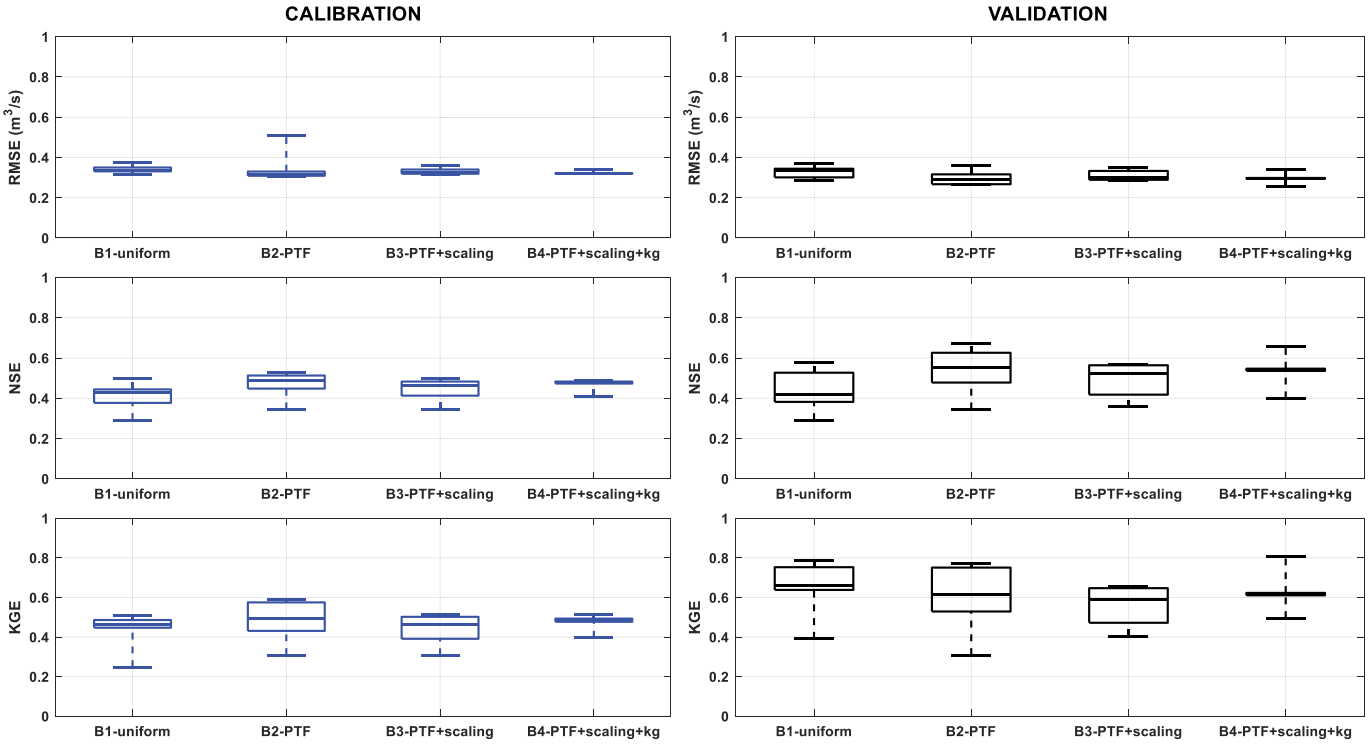

Figure 7. Box plots of RMSE, NSE, KGE values considering configuration B.

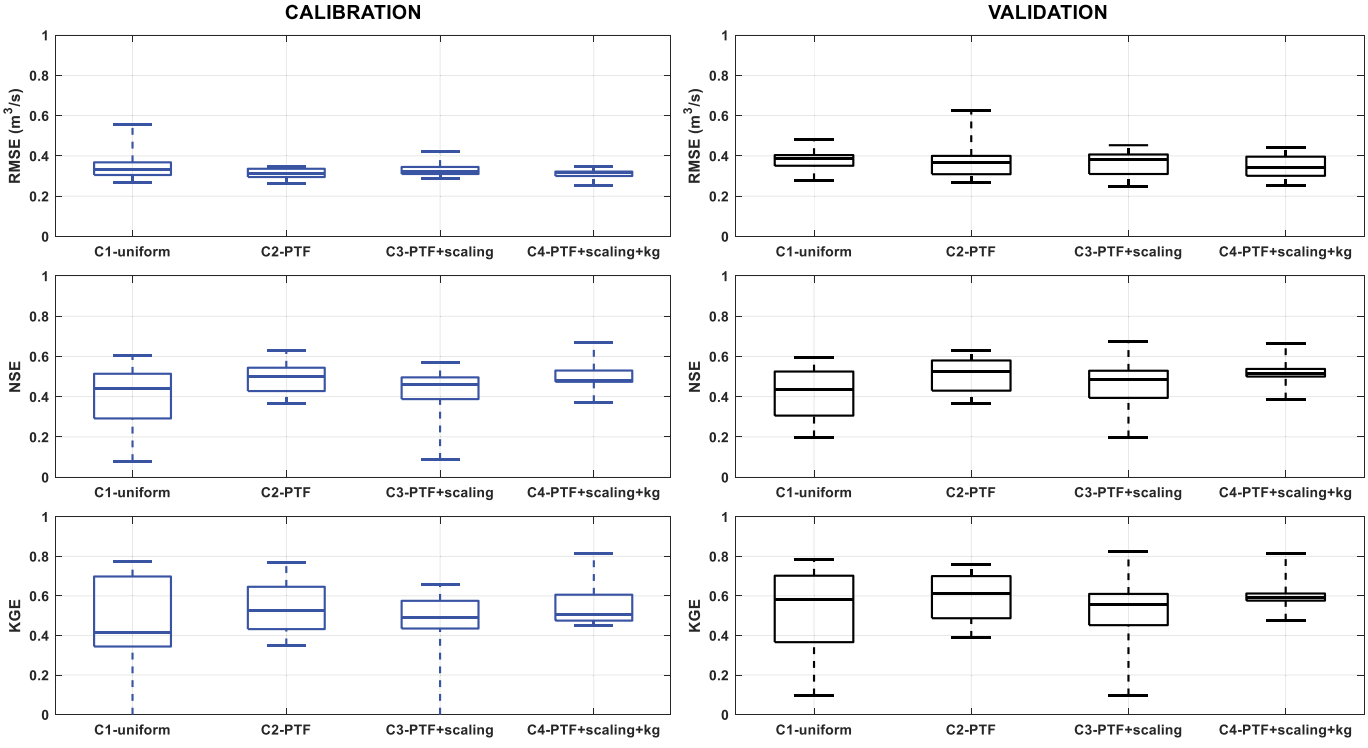

Figure 8. Box plots of RMSE, NSE, KGE values considering configuration C.

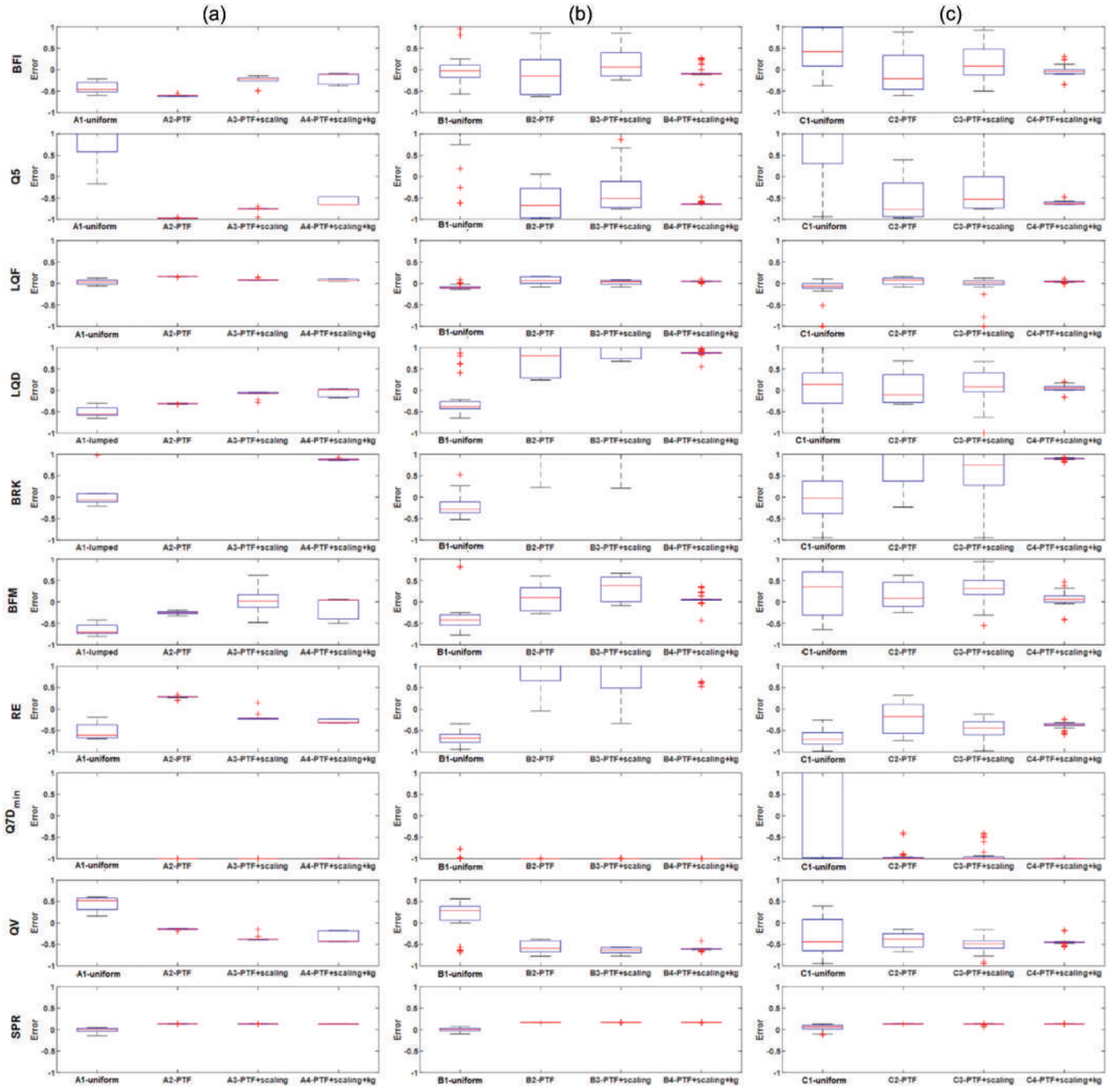

Figure 9. Box plots of relative error metrics considering all configurations (A, B and C) for each LQ signature.

conditions (i.e. Event RR and HFD mean). It can also be seen that respecting the high signature (HQ), all the approaches led to slightly lower model performance.

5 Discussion

The calibration schemes proposed in this work were compared to model performances representing runoff components during 9 years of hydrological simulations. The analyses evidenced the added value of spatially distributed configurations compared to the uniform ones for both mono-objective and multi-objective approaches. Configurations 2, 3 and 4 (based on physical constraint information related to physical soil properties) outperformed configuration 1 for all the approaches considered (A, B, C) in representing total runoff, the latter for low flow and mean flow hydrological regimes. These results are in line with previous studies that demonstrated the benefits of incorporating physically meaningful constraints during calibration (e.g. Yilmaz *et al.* 2008, Rosero *et al.* 2010). However, our study extends this knowledge by systematically comparing the effect of increasing physical realism within a unified experimental framework. The added value of multi-criteria calibration approaches (B and C)

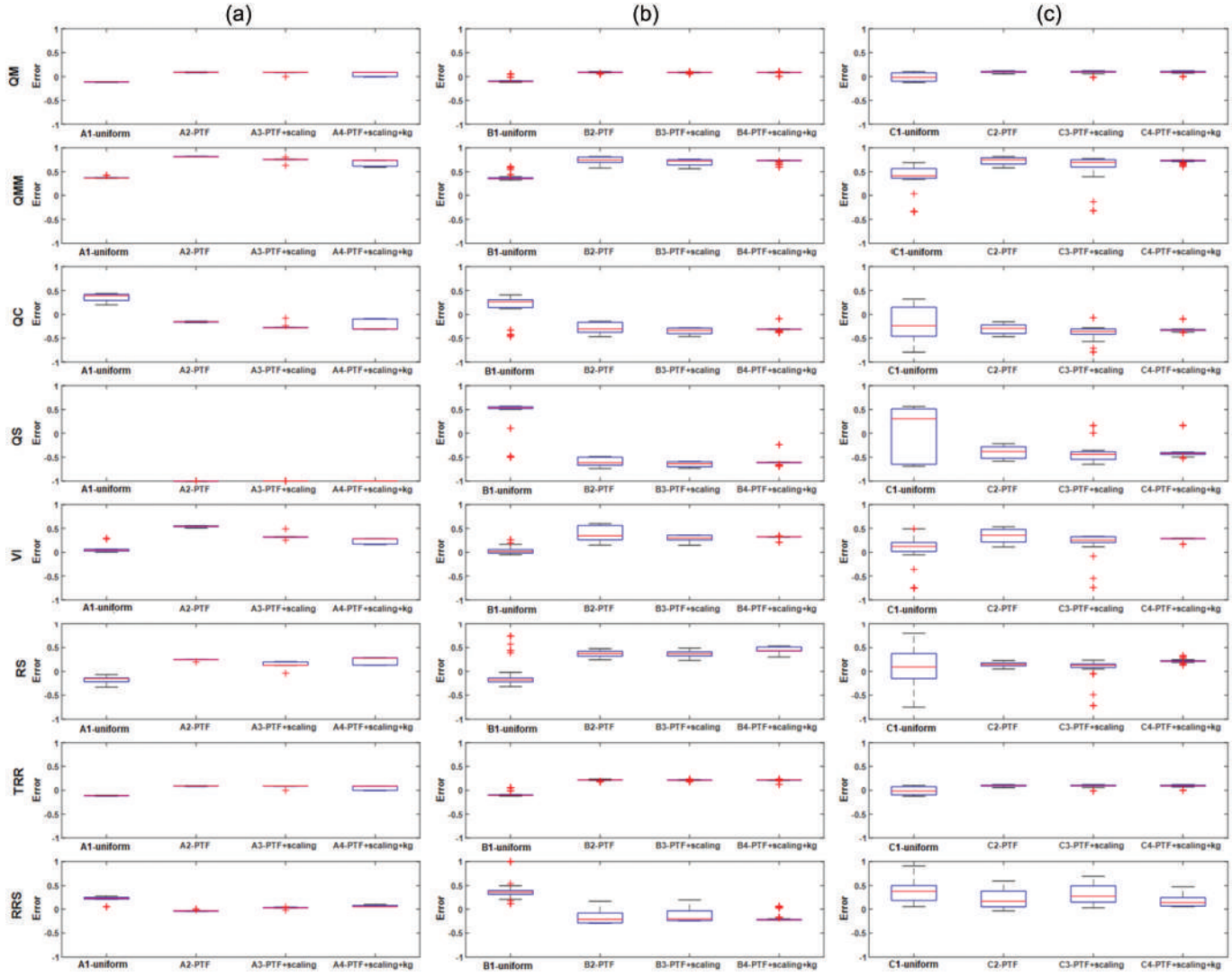

Figure 10. Box plots of relative error metrics considering all configurations (A, B and C) for each MQ signature.

compared to A is evaluable for configurations 3 and 4, where simulations reach the best performance. In addition, the inclusion of water balance as additional constraint in calibration (approach C) allowed us to optimize the model, reducing model variability (configuration 4). With regard to total flow comparison, the mono- and multi-objective strategies impacted in a similar way, and differences were mostly not statistically significant. However, including the additional calibration step focused on baseflow and balance significantly improved the partitioning of simulated flow components to match low, medium and high flow signatures. Therefore, the findings of this study suggest that adding physical information in calibration should be performed, when possible, to ensure a more physically meaningful calibration result. This is consistent with findings by Tigabu *et al.* (2024), who emphasize the importance of adding hydrologically meaningful constraints to reduce parameter equifinality. These findings suggest that targeting process-based components, rather than optimizing statistical metrics alone, can lead to more physically consistent simulations. This approach also avoids the arbitrariness of hydrograph segmentation into flow classes, as used in some recent studies (e.g. Tigabu *et al.* 2024), offering a more internally coherent calibration strategy.

Regarding high flow regimes, the calibration scheme constrained on runoff components is not enough to represent extreme wet conditions, when it does not couple with other variables such as the snow cover or soil moisture (Manfreda *et al.* 2018, Gomis-Cebolla *et al.* 2022). Optionally, further datasets could be included in multi-data approaches such as evapotranspiration data, soil moisture and other datasets contributing to the runoff generation.

With regard to the processing times, considering the same hardware (i7-11700 CPU at 2.50 GHz and 32 GB RAM) for all the configurations, we obtained the values reported in Table 4.

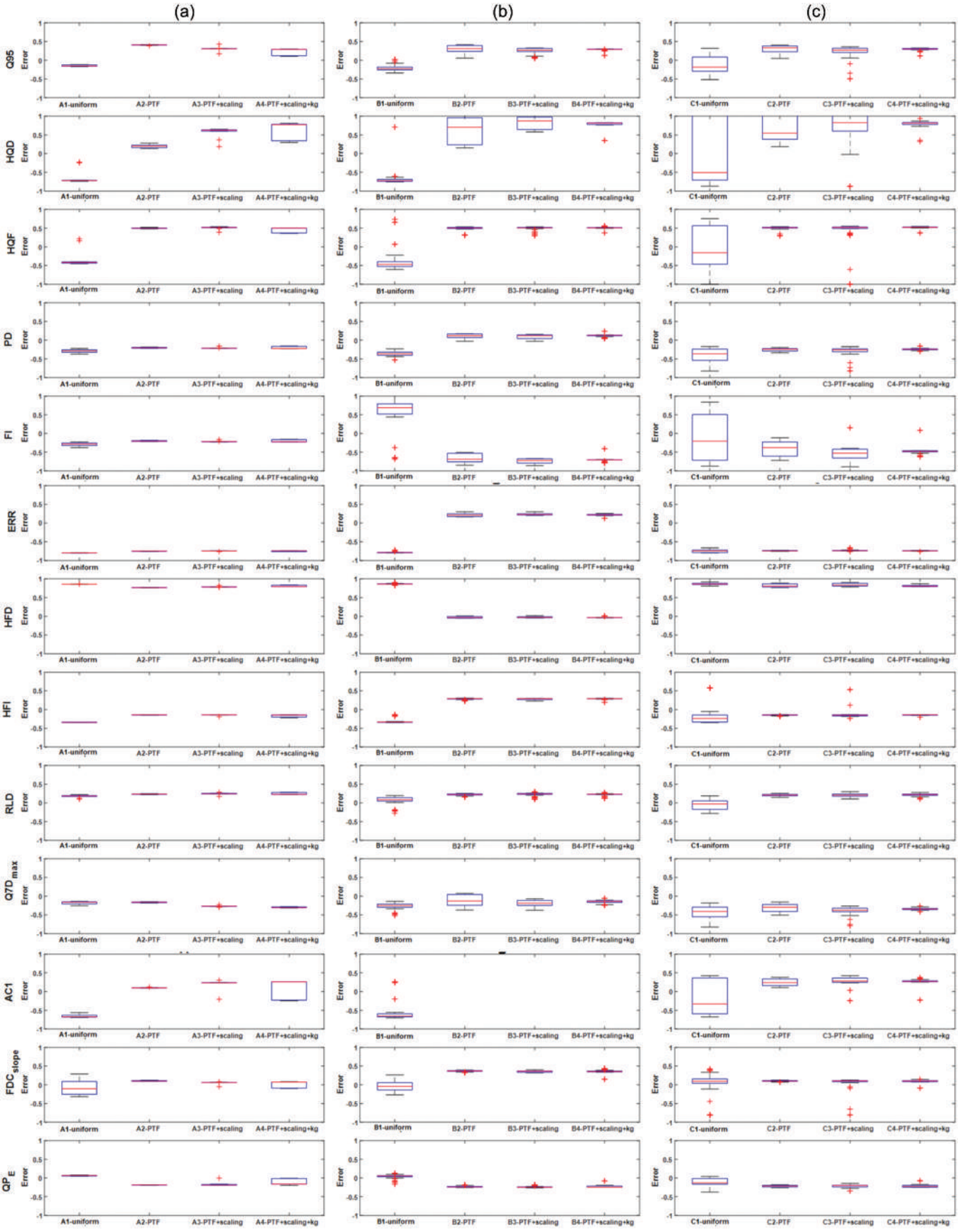

Figure 11. Box plots of relative error metrics considering all configurations (A, B and C) for each HQ signature.

Multi-objective computing time was on average almost three times higher than mono-objective under the circumstances considered in this study. The increasing inclusion of physical parameters leads to the advantage of less computational time in order to identify such parameter configurations among the many that are possible. However, the progressive introduction of physical constraints facilitated a faster convergence towards hydrologically consistent parameter sets, partially compensating for this computational cost. This is a key consideration when designing calibration frameworks that aim for both efficiency and robustness.

Overall, our results support the conclusion that physically informed, component-based calibration strategies can offer a better trade-off between performance, consistency, and transferability. This is particularly relevant in data-scarce or ungauged basins, where traditional calibration approaches may yield non-robust solutions.

6 Conclusion

This work explores a methodology that considers the use of single- and multi-criteria approaches for the calibration of the semi-distributed hydrological DREAM model. The framework consisted in performing three different calibration schemes based on the optimization of different flow components (herein denoted as A, B, and C). In addition, we explored the use of a uniform vs. a spatial parameterization derived from the soil type characteristics of the river basin. This latter included the physical information related to soil characteristics and recession constant as crucial variables to reduce the range of choices among the parameters to calibrate.

The different experiments performed showed that the distributed configuration provided better performances – on average – compared to the configuration with uniform parameterization. For spatially distributed configurations, the inclusion of physical information during the calibration process (e.g. Pedotransfer function and recession curves) allowed us to obtain less variable results, reducing the range of choices to identify the correct solution among the several possible ones identified by an automated optimization procedure. Multi-objective calibration schemes with ad hoc configurations allowed us to obtain statistically similar results in terms of variability and consistency of total runoff. However, the inclusion of other calibration steps focused on baseflow and water balance allow us to to ensure a more physically meaningful calibration results. Additional information for constraining the model parameters is necessary to obtain high flow time series reliability.

Acknowledgements

The present research was conducted as part of the following projects: NODES (funded by MUR-M4C2 1.5 of the PNRR – European Union – NextGenerationEU, grant agreement no. ECS00000036), "OurMED: Sustainable water storage and distribution in the Mediterranean" (part of the PRIMA Programme supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 2222), PRIN PNRR 2022 project "An integrated modeling approach for mitigating climate change effects through enhanced weathering in Southern Italy" (CHANCES – CUP E53D23021850001) and the iAqueduct (financed under the 2018 Joint call of the Water Works 2017 ERA-NET Cofund, project number: ENWWW.2018.5). AP was supported by the National Research and Development Agency of the Chilean Ministry of Science, Technology, Knowledge and Innovation (ANID), grant no. FONDECYT Iniciación 11240171.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Silvano Fortunato Dal Sasso http://orcid.org/0000-0003-1376-7764
Alonso Pizarro http://orcid.org/0000-0002-7242-6559
Salvatore Manfreda http://orcid.org/0000-0002-0225-144X

References

Allen, R.G., *et al.*, 1998. Crop evapotranspiration – guidelines for computing crop water requirements – fao irrigation and drainage paper 56. Rome, Italy: Food and Agriculture Organization of the United Nations.

Allen, R.G., *et al.*, 2006. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) – Model. *Journal of Irrigation & Drainage Engineering*, 132 (4), 380–394.

Aung, H.H., *et al.*, 2025. Long-term evolution and challenges of hydrological observations at the Fiumarella Basin in Southern Italy, *EGU General Assembly 2025*, 27 Apr–2 May 2025, Vienna, Austria, EGU25–20109. doi:10.5194/egusphere-egu25-20109

Bai, J., Shen, Z., and Yan, T., 2017. A comparison of single- and multi-site calibration and validation: a case study of SWAT in the Miyun Reservoir watershed, China. *Frontiers of Earth Science*, 11 (3), 592–600. doi:10.1007/s11707-017-0656-x

Beven, K., 2002. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes*, 16 (2), 189–206. doi:10.1002/hyp.343

Beven, K.J. and Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24 (1), 43–69.

Blair, G.S., *et al.*, 2019. Data science of the natural environment: a research roadmap. *Frontiers in Environmental Science*, 7, 121. doi:10.3389/fenvs.2019.00121

Brocca, L., *et al.*, 2012. Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall–runoff modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 50 (7), 2542–2555. doi:10.1109/TGRS.2011.2177468

Carriero, D., Romano, N., and Fiorentino, M., 2005. A simplified approach for characterizing soil hydrologic behavior and depth of soils at basin scale. *Journal of Agricultural Engineering*, 38 (2), 1–10.

Copernicus Land Monitoring Service (CLMS), 2018. *CORINE Land Cover (CLC) datasets*. European Environment Agency (EEA). https://land.copernicus.eu/pan-european/corine-land-cover

Cosby, B.J., *et al.*, 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research*, 20 (6), 682–690.

Demirel, M.C., *et al.*, 2018. Combining satellite data and appropriate objective functions for improved spatial pattern performanceof a distributed hydrologic model. *Hydrology and Earth System Sciences*, 22 (2), 1299–1315. doi:10.5194/hess-22-1299-2018

Demirel, M.C., *et al.*, 2022. Trade-offs between temporal and spatial pattern calibration and their impacts on robustness and transferability of hydrologic model parameters to ungauged basins. *Water Resources Research*, 58 (9), e2021WR031604.

Echeverría, C., *et al.*, 2019. Assessment of remotely sensed near-surface soil moisture for distributed eco-hydrological model implementation. *Water*, 11 (12), 2613. doi:10.3390/w11122613

Efstratiadis, A. and Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal*, 55 (1), 58–78. doi:10.1080/02626660903526292

Fiorentino, M., Manfreda, S., and Iacobellis, V., 2007. Peak runoff contributing area as hydrological signature of the probability distribution of floods. *Advances in Water Resources*, 30 (10), 2123–2134. doi:10.1016/j.advwatres.2006.11.017

Gigante, V., *et al*., 2009. Influences of leaf area index estimations on water balance modeling in a mediterranean semi-arid basin. *Natural Hazards and Earth System Sciences*, 9 (3), 979–991. doi:10.5194/nhess-9-979-2009

Gnann, S.J., *et al*., 2021. TOSSH: a toolbox for 551 streamflow signatures in hydrology. *Environmental Modelling & Software*, 138, 104983.

Goldberg, D.E., 1989. *Genetic algorithms in search, optimization, and machine learning*. Reading, MA, USA: Addison-Wesley.

Gomis-Cebolla, J., *et al*., 2022. Integrating snow and soil moisture remote sensing data to improve high-flow simulations in Mediterranean catchments. *Remote Sensing of Environment*, 280, 113145. doi:10.1016/j.rse.2022.113145

Gupta, H.V., *et al*., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology*, 377 (1–2), 80–91.

Hrachowitz, M., *et al*., 2013. A decade of predictions in ungauged basins (PUB)—a review. *Hydrological Sciences Journal*, 58 (6), 1198–1255.

Immerzeel, W.A. and Droogers, P., 2008. Calibration of a distributed hydrological model based on satellite evapotranspiration. *Journal of Hydrology*, 349 (3–4), 411–424.

Kampf, S.K. and Burges, S.J., 2007. A framework for classifying and comparing distributed hillslope and catchment hydrologic models. *Water Resources Research*, 43, W05423. doi:10.1029/2006WR005370

Manfreda, S., *et al*., 2011. On the use of AMSU-based products for the description of soil water content at basin scale. *Hydrology and Earth System Sciences*, 15, 2839–2852. doi:10.5194/hess-15-2839-2011

Manfreda, S., *et al*., 2018. Exploiting the use of physical information for the calibration of lumped hydrological model. *Hydrological Processes*, 32 (10), 1420–1433.

Manfreda, S., Fiorentino, M., and Iacobellis, V., 2005. DREAM: a distributed model for runoff, evapotranspiration, and antecedent soil moisture simulation. *Advances in Geosciences*, 2. doi:10.5194/adgeo–2–31–2005.

Manfreda, S. and Mancusi, L., 2013. Previsione idrologica per la produzione idroelettrica. *L'Acqua*, 4, 41–48.

Mizukami, N., *et al*., 2019. On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrology and Earth System Sciences*, 23, 2601–2614.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models: part I. A discussion of principles. *Journal of Hydrology*, 10, 282–290.

Perrini, P., *et al*., 2024. A runoff-on-grid approach to embed hydrological processes in shallow water models. *Water Resources Research*, 60, e2023WR036421. doi:10.1029/2023WR036421

Rawls, W.J., Brakensiek, D.L., and Saxton, K.E., 1982. Estimation of soil water properties. *Transactions of the ASAE*, 25 (5), 1316–1320.

Reed, P.M., *et al*., 2013. Advances in the evolutionary multi-objective optimization of environmental models. *Advances in Water Resources*, 51, 249–250.

Romano, N. and Palladino, M., 2002. Prediction of soil water retention using soil physical data and terrain attributes. *Journal of Hydrology*, 265 (1), 56–75.

Romano, N. and Santini, A., 1997. Effectiveness of using pedo-transfer functions to quantify the spatial variability of soil water retention characteristics. *Journal of Hydrology*, 202, 137–157.

Rosero, E., *et al*., 2010. Incorporating parametric uncertainty into evaluations of hydrologic model parameters and predictions. *Journal of Hydrology*, 387 (1–2), 48–60.

Ruiz-Pérez, G., *et al*., 2017. Calibration of a parsimonious distributed ecohydrological daily model in a data-scarce basin by exclusively using the spatio-temporal variation of NDVI. *Hydrology and Earth System Sciences*, 21 (12), 6235–6251.

Santini, A., *et al*., 1999. Interpretation of the spatial variability of soil hydraulic properties using a land system analysis. *Modelling of Transport Processes in Soils*, 1, 491–500.

Saxton, K.E., *et al*., 1986. Estimating generalized soil-water characteristics from texture. *Soil Science Society of America Journal*, 50 (4), 1031–1036.

Seibert, J., 2000. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences*, 4 (2), 215–224. doi:10.5194/hess-4-215-2000

Shafii, M. and De Smedt, F., 2009. Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm. *Hydrology and Earth System Sciences*, 13, 2137–2149.

Sikorska-Senoner, A.E., 2021. Delineating modelling uncertainty in river flow indicators with representative parameter sets. *Advances in Water Resources*, 156, 104024.

Stefnisdóttir, S., *et al*., 2021. Improving the pareto frontier in multi-dataset calibration of hydrological models using metaheuristics, hydrol. *Earth System Science Data Discussions* [preprint]. doi:10.5194/hess-2021-325

Tigabu, T.B., *et al*., 2024. Optimization of the SWAT+ model to adequately predict different segments of a managed streamflow hydrograph. *Hydrological Sciences Journal*, 69 (9), 1198–1217. doi:10.1080/02626667.2024.2364714

Vereecken, H., *et al*., 1989. Estimating the soil moisture retention characteristic from texture, bulk density and carbon content. *Soil Science*, 148 (6), 389–403.

Wang, L. and Liu, H., 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *International Journal of Geographical Information Science*, 20, 193–213.

Weber, T.K.D., *et al*., 2024. Hydro-pedotransfer functions: a roadmap for future development, hydrol. *Journal of Earth System Science*, 28, 3391–3433. doi:10.5194/hess-28-3391-2024

Yang, H., *et al*., 2019. Utilizing satellite surface soil moisture data in calibrating a distributed hydrological model applied in humid regions through a multiobjective Bayesian hierarchical framework. *Remote Sensing*, 11 (11), 1335.

Yilmaz, K.K., Gupta, H.V., and Wagener, T., 2008. A process-based diagnostic approach to model evaluation: application to the NWS distributed hydrologic model. *Water Resources Research*, 44 (9), W09417.

Zhang, R., *et al*., 2018. Can multiobjective calibration of streamflow guarantee better hydrological model accuracy? *Journal of Hydroinformatics*, 20 (3), 687–698.

Appendix

APPROACH A	APPROACH B	APPROCH C
RMSE Calibration:	**RMSE Calibration:**	**RMSE Calibration:**
A1-lumped: Average = 0.337, Std Dev = 0.014, Max = 0.361	B1-lumped: Average = 0.338, Std Dev = 0.013, Max = 0.373	C1-lumped: Average = 0.344, Std Dev = 0.051, Max = 0.560
A2-PTF: Average = 0.327, Std Dev = 0.056, Max = 0.485	B2-PTF: Average = 0.325, Std Dev = 0.028, Max = 0.507	C2-PTF: Average = 0.313, Std Dev = 0.026, Max = 0.350
A3-PTF+scaling: Average = 0.320, Std Dev = 0.004, Max = 0.328	B3-PTF+scaling: Average = 0.329, Std Dev = 0.013, Max = 0.359	C3-PTF+scaling: Average = 0.328, Std Dev = 0.028, Max = 0.420
A4-PTF+scaling+kg: Average = 0.330, Std Dev = 0.022, Max = 0.393	B4-PTF+scaling+kg: Average = 0.323, Std Dev = 0.006, Max = 0.340	C4-PTF+scaling+kg: Average = 0.312, Std Dev = 0.016, Max = 0.350
RMSE Validation:	**RMSE Validation:**	**RMSE Validation:**
A1-lumped: Average = 0.347, Std Dev = 0.028, Max = 0.369	B1-lumped: Average = 0.326, Std Dev = 0.024, Max = 0.368	C1-lumped: Average = 0.381, Std Dev = 0.059, Max = 0.480
A2-PTF: Average = 0.265, Std Dev = 0.001, Max = 0.266	B2-PTF: Average = 0.298, Std Dev = 0.028, Max = 0.359	C2-PTF: Average = 0.358, Std Dev = 0.064, Max = 0.630
A3-PTF+scaling: Average = 0.314, Std Dev = 0.091, Max = 0.569	B3-PTF+scaling: Average = 0.309, Std Dev = 0.022, Max = 0.350	C3-PTF+scaling: Average = 0.367, Std Dev = 0.055, Max = 0.450
A4-PTF+scaling+kg: Average = 0.282, Std Dev = 0.020, Max = 0.294	B4-PTF+scaling+kg: Average = 0.300, Std Dev = 0.015, Max = 0.339	C4-PTF+scaling+kg: Average = 0.350, Std Dev = 0.052, Max = 0.440
NSE Calibration:	**NSE Calibration:**	**NSE Calibration:**
A1-lumped: Average = 0.421, Std Dev = 0.047, Max = 0.488	B1-lumped: Average = 0.416, Std Dev = 0.046, Max = 0.496	C1-lumped: Average = 0.396, Std Dev = 0.145, Max = 0.610
A2-PTF: Average = 0.509, Std Dev = 0.013, Max = 0.523	B2-PTF: Average = 0.472, Std Dev = 0.049, Max = 0.526	C2-PTF: Average = 0.494, Std Dev = 0.078, Max = 0.630
A3-PTF+scaling: Average = 0.480, Std Dev = 0.014, Max = 0.496	B3-PTF+scaling: Average = 0.449, Std Dev = 0.042, Max = 0.496	C3-PTF+scaling: Average = 0.439, Std Dev = 0.096, Max = 0.570
A4-PTF+scaling+kg: Average = 0.475, Std Dev = 0.018, Max = 0.517	B4-PTF+scaling+kg: Average = 0.470, Std Dev = 0.022, Max = 0.487	C4-PTF+scaling+kg: Average = 0.498, Std Dev = 0.046, Max = 0.670
NSE Validation:	**NSE Validation:**	**NSE Validation:**
A1-lumped: Average = 0.366, Std Dev = 0.096, Max = 0.533	B1-lumped: Average = 0.440, Std Dev = 0.082, Max = 0.578	C1-lumped: Average = 0.412, Std Dev = 0.121, Max = 0.590
A2-PTF: Average = 0.632, Std Dev = 0.002, Max = 0.634	B2-PTF: Average = 0.534, Std Dev = 0.091, Max = 0.674	C2-PTF: Average = 0.506, Std Dev = 0.084, Max = 0.630
A3-PTF+scaling: Average = 0.571, Std Dev = 0.042, Max = 0.667	B3-PTF+scaling: Average = 0.497, Std Dev = 0.073, Max = 0.569	C3-PTF+scaling: Average = 0.460, Std Dev = 0.097, Max = 0.670
A4-PTF+scaling+kg: Average = 0.582, Std Dev = 0.057, Max = 0.672	B4-PTF+scaling+kg: Average = 0.529, Std Dev = 0.049, Max = 0.659	C4-PTF+scaling+kg: Average = 0.516, Std Dev = 0.043, Max = 0.660
KGE Calibration:	**KGE Calibration:**	**KGE Calibration:**
A1-lumped: Average = 0.502, Std Dev = 0.010, Max = 0.515	B1-lumped: Average = 0.450, Std Dev = 0.059, Max = 0.509	C1-lumped: Average = 0.451, Std Dev = 0.233, Max = 0.770
A2-PTF: Average = 0.582, Std Dev = 0.008, Max = 0.592	B2-PTF: Average = 0.490, Std Dev = 0.083, Max = 0.592	C2-PTF: Average = 0.542, Std Dev = 0.125, Max = 0.770
A3-PTF+scaling: Average = 0.511, Std Dev = 0.006, Max = 0.524	B3-PTF+scaling: Average = 0.448, Std Dev = 0.060, Max = 0.514	C3-PTF+scaling: Average = 0.470, Std Dev = 0.147, Max = 0.660
A4-PTF+scaling+kg: Average = 0.506, Std Dev = 0.033, Max = 0.595	B4-PTF+scaling+kg: Average = 0.477, Std Dev = 0.030, Max = 0.515	C4-PTF+scaling+kg: Average = 0.541, Std Dev = 0.079, Max = 0.810
KGE Validation:	**KGE Validation:**	**KGE Validation:**
A1-lumped: Average = 0.628, Std Dev = 0.070, Max = 0.748	B1-lumped: Average = 0.668, Std Dev = 0.091, Max = 0.785	C1-lumped: Average = 0.523, Std Dev = 0.203, Max = 0.780
A2-PTF: Average = 0.768, Std Dev = 0.004, Max = 0.771	B2-PTF: Average = 0.618, Std Dev = 0.128, Max = 0.771	C2-PTF: Average = 0.594, Std Dev = 0.113, Max = 0.760
A3-PTF+scaling: Average = 0.670, Std Dev = 0.051, Max = 0.816	B3-PTF+scaling: Average = 0.565, Std Dev = 0.086, Max = 0.657	C3-PTF+scaling: Average = 0.518, Std Dev = 0.145, Max = 0.820
A4-PTF+scaling+kg: Average = 0.680, Std Dev = 0.091, Max = 0.815	B4-PTF+scaling+kg: Average = 0.609, Std Dev = 0.051, Max = 0.809	C4-PTF+scaling+kg: Average = 0.593, Std Dev = 0.052, Max = 0.810