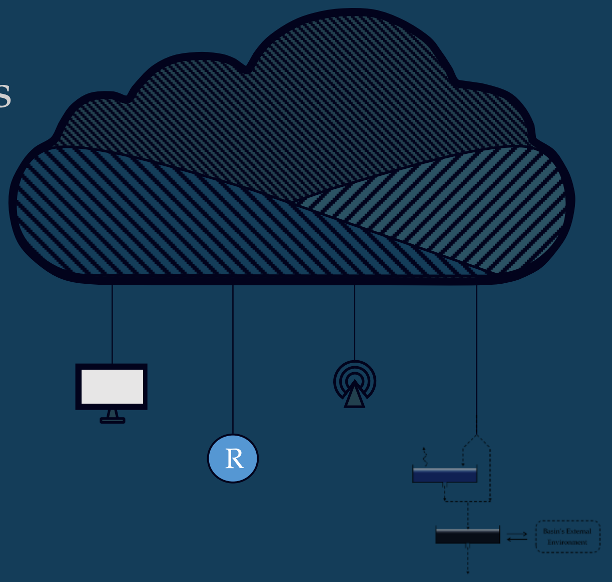
National Technical University of Athens, Greece
Department of Water Resources & Environmental Engineering

European Geosciences Union General Assembly, Online
19 - 30 April 2021
HS3.1: Hydroinformatics: computational intelligence, systems analysis, optimisation, data science, and innovative sensing techniques

Can hydrological model identifiability be improved? Stress-testing the concept of stochastic calibration

Vasileios Kourakos, Andreas Efstratiadis, and Ioannis Tsoukalas

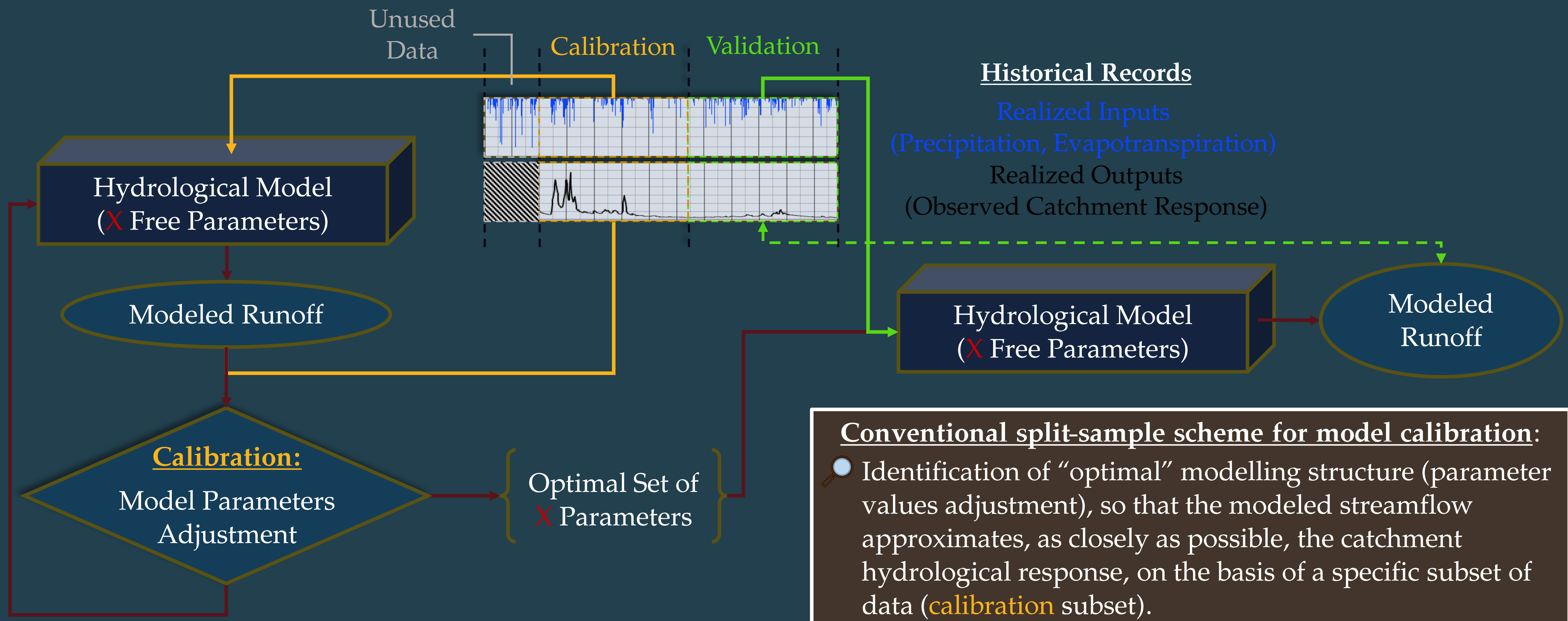


European Geosciences Union

Abstract

Hydrological calibrations with historical data are often deemed insufficient for deducing safe estimations about a model structure that imitates, as closely as possible, the anticipated catchment behaviour. In order to address this issue, we investigate a promising strategy, using as drivers synthetic time series, which preserve the probabilistic properties and dependence structure of the observed data. The key idea is **calibrating a model on the basis of synthetic rainfall-runoff data, and validating against the full observed data sample**. To this aim, we employed a proof of concept on a representative catchment, by testing several lumped conceptual hydrological models with alternative parameterizations and across two time-scales, monthly and daily. Next, we attempted to reinforce the validity of the recommended methodology by employing monthly stochastic calibrations in 100 MOPEX catchments. As before, a number of different hydrological models were used, for the purpose of proving that calibration with stochastic inputs is independent of the chosen model. The results highlight that in most cases the new approach leads to **stronger parameter identifiability and stable predictive capacity across different temporal windows**, since the model is trained over much extended hydroclimatic conditions.

Deterministic Hydrological Calibration Scheme



Conventional split-sample scheme for model calibration:

- Identification of "optimal" modelling structure (parameter values adjustment), so that the modeled streamflow approximates, as closely as possible, the catchment hydrological response, on the basis of a specific subset of data (calibration subset).
- Assessment of its predictive capacity on an independent subset (validation subset) of the historical dataset to ensure its adequacy for reproducing the hydroclimatic conditions of a period different than that for calibration.

Common Split-Sample Strategy Challenges

Discrepancies between the performance metrics for calibration and validation samples may indicate weak parameter identification. Thus, there is a need to extend the temporal horizon chosen for calibration, so that the model can be trained over longer periods and capture the catchment dynamics to a greater extent. **BUT**:

1. The available data may often not be representative of the catchment's hydrological regime and long-term hydroclimatic changes; a common pitfall in the case of data-scarce catchments.

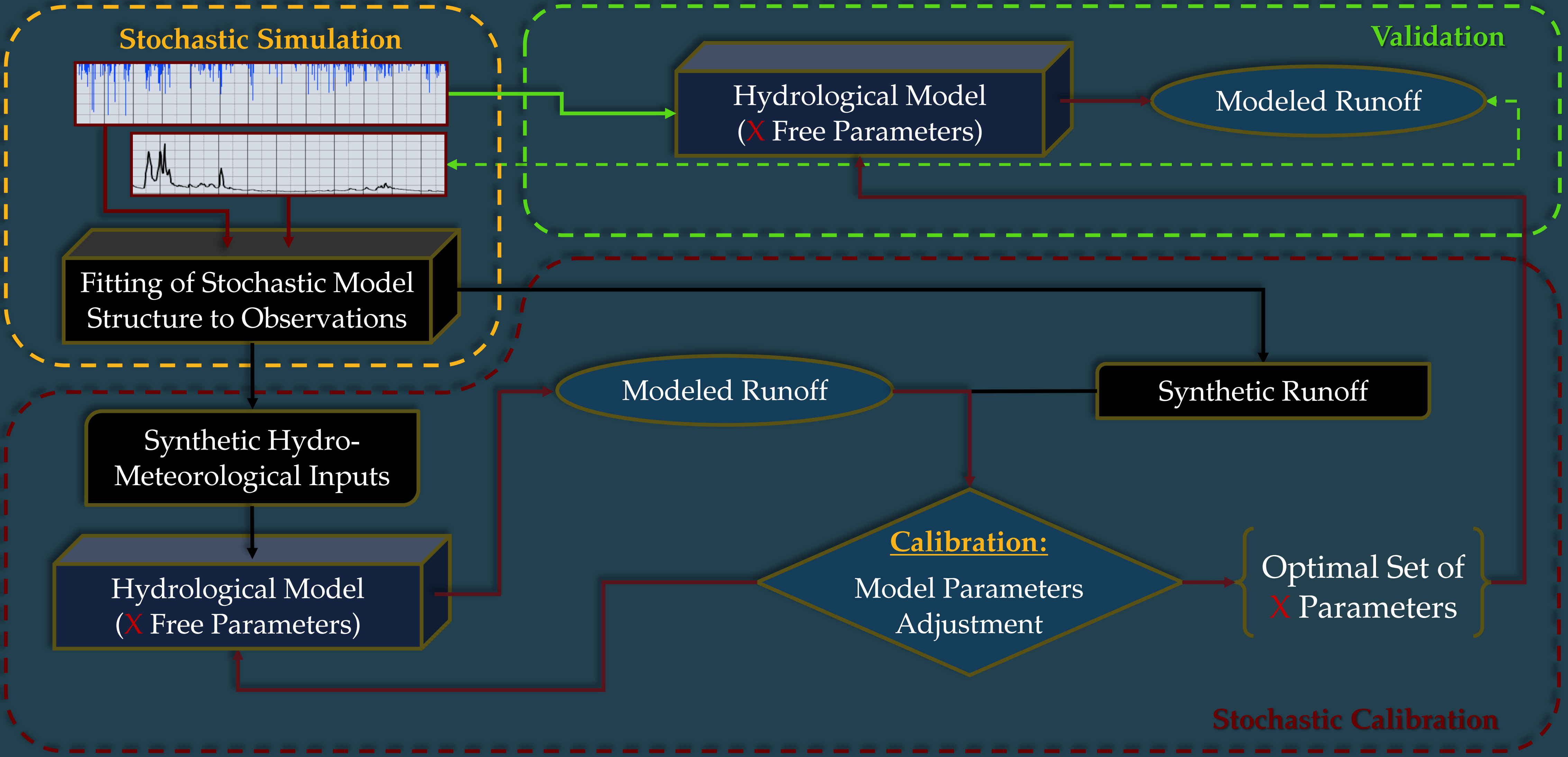
2. As stated before, a part (e.g. half) of the whole dataset must be allocated for validation purposes, hence precious hydrologic information encapsulated in the associated dataset is sacrificed.

3. As seen in the previous outline of this approach, only the overlapping periods of the observed input and output data can be exploited for model calibration and validation, ergo the remaining period of data is not accounted for. This is a commonly encountered scenario, since the length of rainfall records often exceeds that of the runoff observations.

In conclusion, the separation strategy, as well as the length of the calibration period, can pose a serious problem for hydrological modelling applied based on the split-sample rationale.

New approach for parameter identification: **Stochastic Calibration** [Efstratiadis *et al.*, 2021]

Conceptual Scheme for Stochastic Calibration



Stochastic Calibration Procedure

The proposed framework is a simple conceptual approach, which follows the aforementioned calibration-validation logic. Specifically, model calibration is accomplished by using **long synthetic data** as inputs, while the adjusted model structure is validated against the **full historical record**.

Thus, an additional component is important to precede the calibration procedure, a stochastic model generating synthetic time series, **stochastically resembling** the observed ones. Within those time series should be reproduced all statistical information regarding the full hydroclimatic regime of the basin. Such information is integrated in the realized input and output data and usually cannot be extracted in cases of length-limited data. What is more, this is further intensified since a significant part of this information is accounted for validation purposes within the split-sample calibration scheme.

Therefore, the drawbacks stemming from the application of the classical split-sample scheme are thereby eliminated, since:

- The substantially **longer** synthetic sample, that is expected to describe the **full hydroclimatic regime**, dispensed for model calibration will eventually lead to more robust parameters and stable predictive capacity.

- As stated above, the predictive capacity of the model is evaluated against the **full set** of observed data, hence the validation dataset is also extended.

Given the above, the **dilemma** of which part of data to allot for calibration and which for validation, does not exist any more.

Hydrological Modelling Tools

In order to test the functionality of the proposed framework for stochastic calibration, a number of hydrological models with conceptual structure of varying complexity have been either developed or used. Each one of them is a bucket-type model with a lumped schematization, which uses a set of mathematical equations to simulate the main hydrological mechanisms at the catchment scale. All hydrological fluxes are expressed in units of water depth per unit time (i.e. mm/month or mm/day), while storages are expressed in terms of water depths (i.e. mm). Each model requires either daily or monthly areal rainfall (P) and potential evapotranspiration (PET) as inputs. These models are:

1. **Zygos-4P**: a four-parameter lumped water balance model, based on a simplified version of Zygos model [Kozanis and Efstratiadis, 2006]
2. **Zygos-6P**: this six-parameter lumped model was also developed by adopting the structure of Zygos-4P scheme, by implementing a few modifications
3. **GR2M**: this monthly time step model belongs to the family of the GR rainfall-runoff models, a set of conceptual lumped hydrological models developed for specific time steps. For this study, it was selected the most recent version of GR2M model [Mouelhi *et al.*, 2006]
4. **GR4J**: another model of the GR family, a four-parameter daily lumped hydrological model proposed by Perrin *et al.* [2003]
5. **GR6J**: a progressively modified version of the GR4J model, applied at a daily time step with six free parameters [Le Moine, 2008; Pushpalatha *et al.*, 2011]

\# R-code

Both hydrological models Zygos-4P and Zygos-6P were configured in R-environment, whereas the aforementioned GR models have been already implemented within the R–package *airGR* [Coron *et al.*, 2017a; Coron *et al.*, 2017b].

The conceptual structure for each of the five models, as well as the equations which regulate each mathematical model, are presented in the following sections of this presentation.

Hydrological Model Zygos-4P

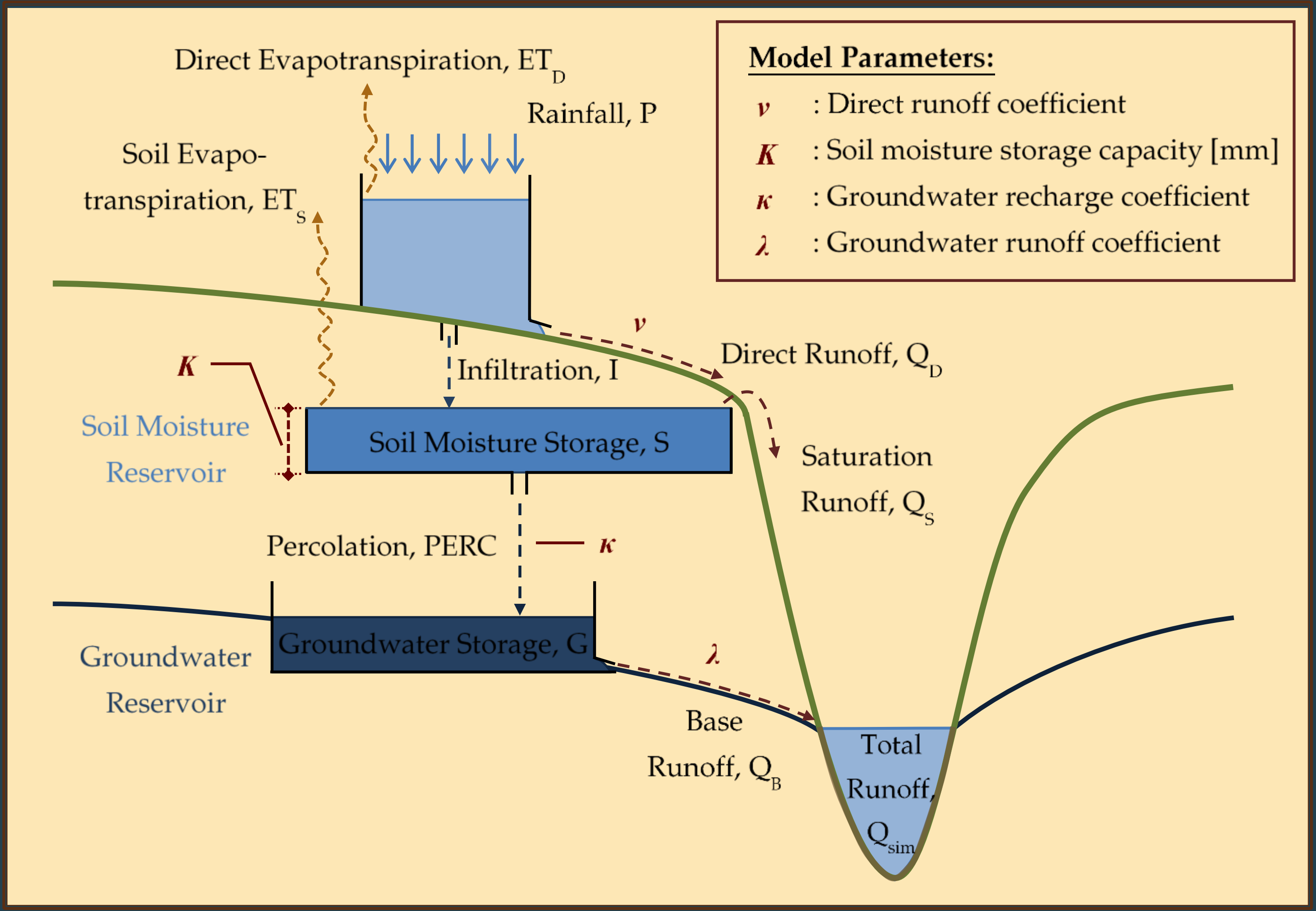


Figure 1 | Schematic representation of the conceptual structure of the Zygos-4P model

Zygos-4P Equations for each time step t:

$$Q_D(t) = \nu\, P(t)$$

$$ET_D(t) = \min[\, PET(t)\, ;\, P(t) - Q_D(t)\,]$$

$$I(t) = P(t) - Q_D(t) - ET_D(t)$$

$$S(t) = S(t-1) + I(t)$$

$$ET_S(t) = S(t)\left(1 - e^{-\frac{PET(t) - ET_D(t)}{K}}\right)$$

$$PERC(t) = \kappa\, S(t)$$

$$Q_S(t) = \max[\, S(t) - K\, ;\, 0\,]$$

$$G(t) = G(t-1) + PERC(t)$$

$$Q_B(t) = \lambda\, G(t)$$

$$Q_{sim}(t) = Q_D(t) + Q_S(t) + Q_B(t)$$

$$ET_{act}(t) = ET_D(t) + ET_S(t)$$

$$S(t) = S(t-1) + I(t) - ET_S(t) - PERC(t) - Q_S(t)$$

$$G(t) = G(t-1) + PERC(t) - Q_B(t)$$

Hydrological Model Zygos-6P

Zygos-6P Equations for each time step t:
(additional or modified equations)

$$ET_D(t) = \min\left[\, \frac{i^{[P(t) \neq 0]}}{N} P(t)\ ;\ PET(t) \,\right]$$

where $i^{[P(t) \neq 0]}$ indicates the number of time steps t that a rainfall event occurred; N is the length of the rainfall data sample

$$Q_D(t) = (\, P(t) - ET_D(t) \,) \left(\frac{S(t-1)}{K}\right)^{\nu}$$

$$Q_B(t) = \max[\, \lambda\, (\, G(t) - H \,)\ ;\ 0 \,]$$

$$L(t) = a\, G(t)$$

With the exception of aforementioned components, all of the remaining hydrological processes are calculated in an identical way to that of the Zygos-4P scheme.

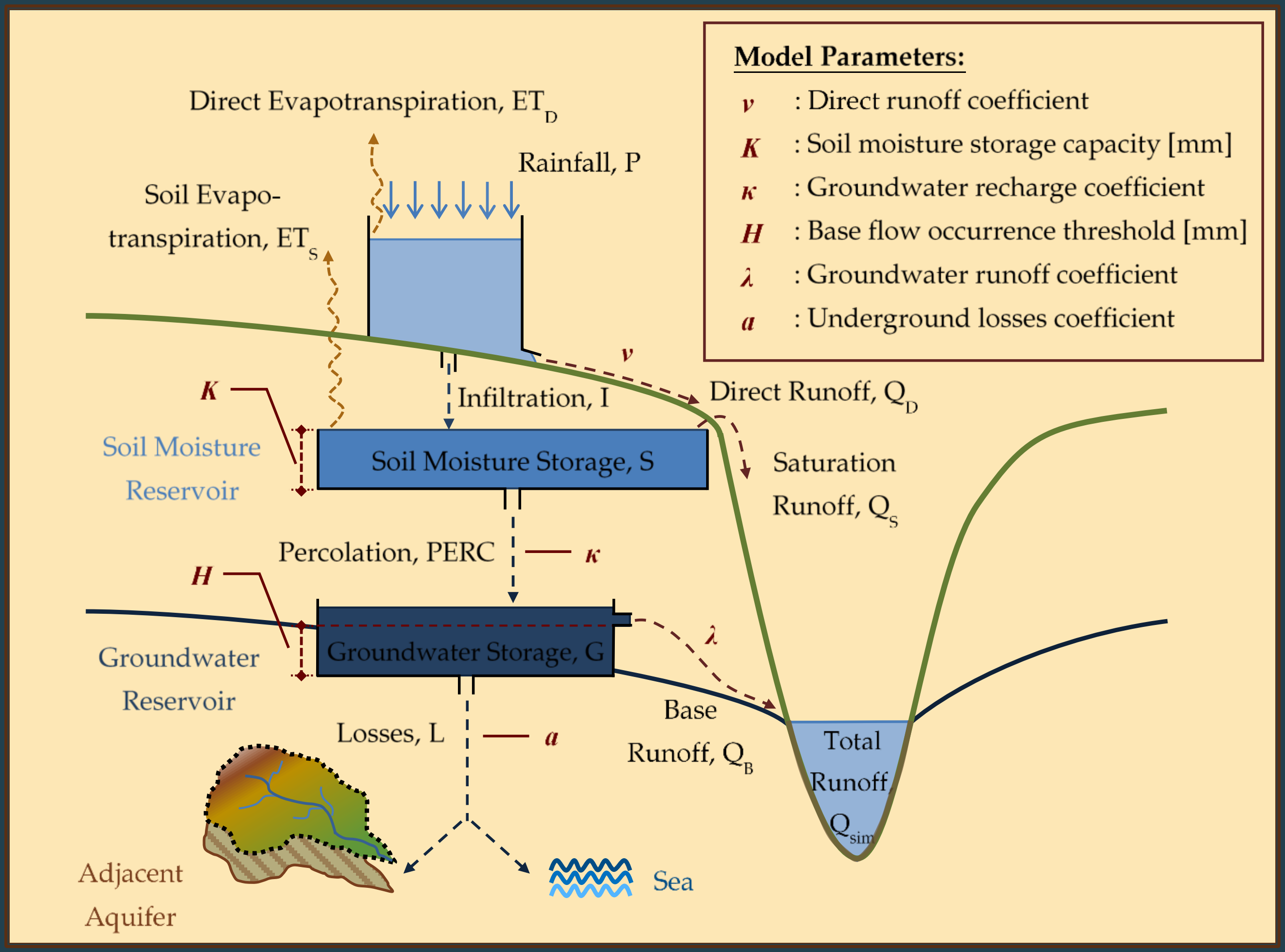


Figure 2 | Schematic representation of the conceptual structure of the Zygos-6P model

Monthly Hydrological Model GR2M

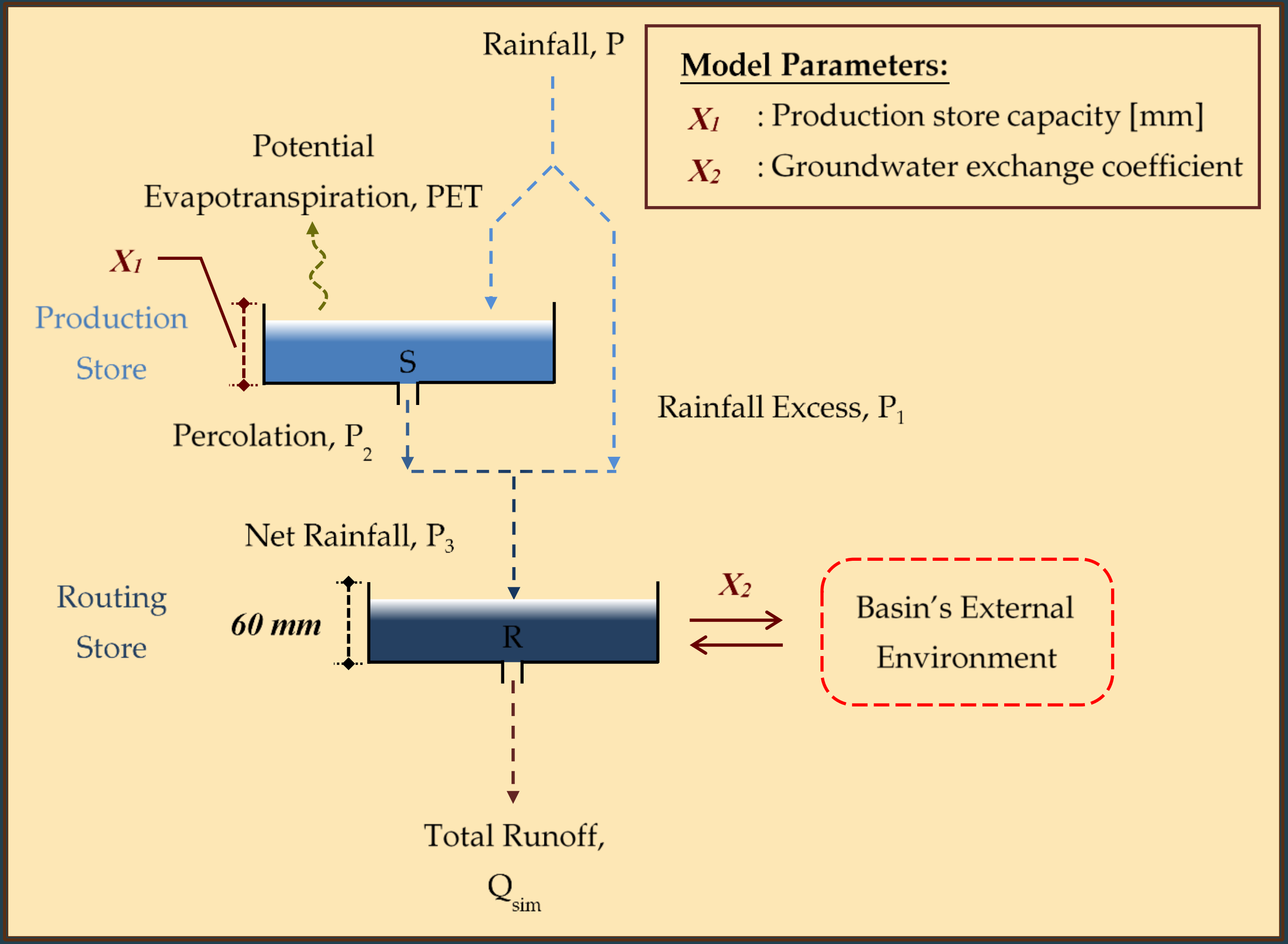


Figure 3 | Schematic representation of the conceptual structure of the GR2M model

GR2M Equations for each time step:

$$S_1 = \frac{S + X_1\varphi}{1+\varphi\frac{S}{X_1}} \quad \text{where } \varphi = \tanh(P / X_1)$$

$$P_1 = P + S - S_1$$

$$S_2 = \frac{S_1(1-\psi)}{1+\psi\left(1-\frac{S_1}{X_1}\right)} \quad \text{where } \psi = \tanh(PET / X_1)$$

$$S = \frac{S_2}{\left[1+\left(\frac{S_2}{X_1}\right)^3\right]^{1/3}} \quad \& \quad P_2 = S_2 - S$$

$$P_3 = P_1 + P_2 \quad \& \quad R_1 = R + P_3$$

$$R_2 = X_2 R_1$$

$$Q_{sim} = \frac{R_2^2}{R_2 + 60} \quad \& \quad R = R_2 - Q_{sim}$$

Daily Hydrological Model GR4J

GR4J Equations for each time step t:

$$P_N = \max[\, P - PET \,;\, 0 \,]$$

$$ET_N = \max[\, PET - P \,;\, 0 \,]$$

$$P_S = \frac{X_1 \left(1 - \left(\frac{S}{X_1}\right)^2\right) \varphi}{1 + \frac{S}{X_1} \varphi}$$

where $\varphi = \tanh(\, P_N / X_1 \,)$

$$ET_S = \frac{S \left(2 - \frac{S}{X_1}\right) \psi}{1 + \left(1 - \frac{S}{X_1}\right) \psi}$$

where $\psi = \tanh(\, ET_N / X_1 \,)$

$$S_1 = S + P_S - ET_S$$

$$Perc = S\left\{1 - \left[1 + \left(\frac{4}{9}\frac{S_1}{X_1}\right)^4\right]^{-1/4}\right\}$$

$$S = S_1 - Perc$$

$$P_R = Perc + (P_N - P_S)$$

$$Q_9(t) = 0.9 \sum_{k=1}^{l} [\, UH_1(k) \cdot P_R(t - k + 1) \,]$$

$$Q_1(t) = 0.1 \sum_{k=1}^{m} [\, UH_2(k) \cdot P_R(t - k + 1) \,]$$

$$\text{where } \begin{cases} l = \mathrm{int}[\, X_4 \,] + 1 \\ m = \mathrm{int}[\, 2\, X_4 \,] + 2 \end{cases}$$

$$F = X_2 \left(\frac{R}{X_3}\right)^{7/2}$$

$$R_1 = \max[\, R + Q_9 + F \,;\, 0 \,]$$

$$Q_R = R_1 \left\{1 - \left[1 + \left(\frac{R_1}{X_3}\right)^4\right]^{-1/4}\right\}$$

$$R = R_1 - Q_R$$

$$Q_D = \max[\, Q_1 + F \,;\, 0 \,]$$

$$Q_{sim} = Q_R + Q_D$$

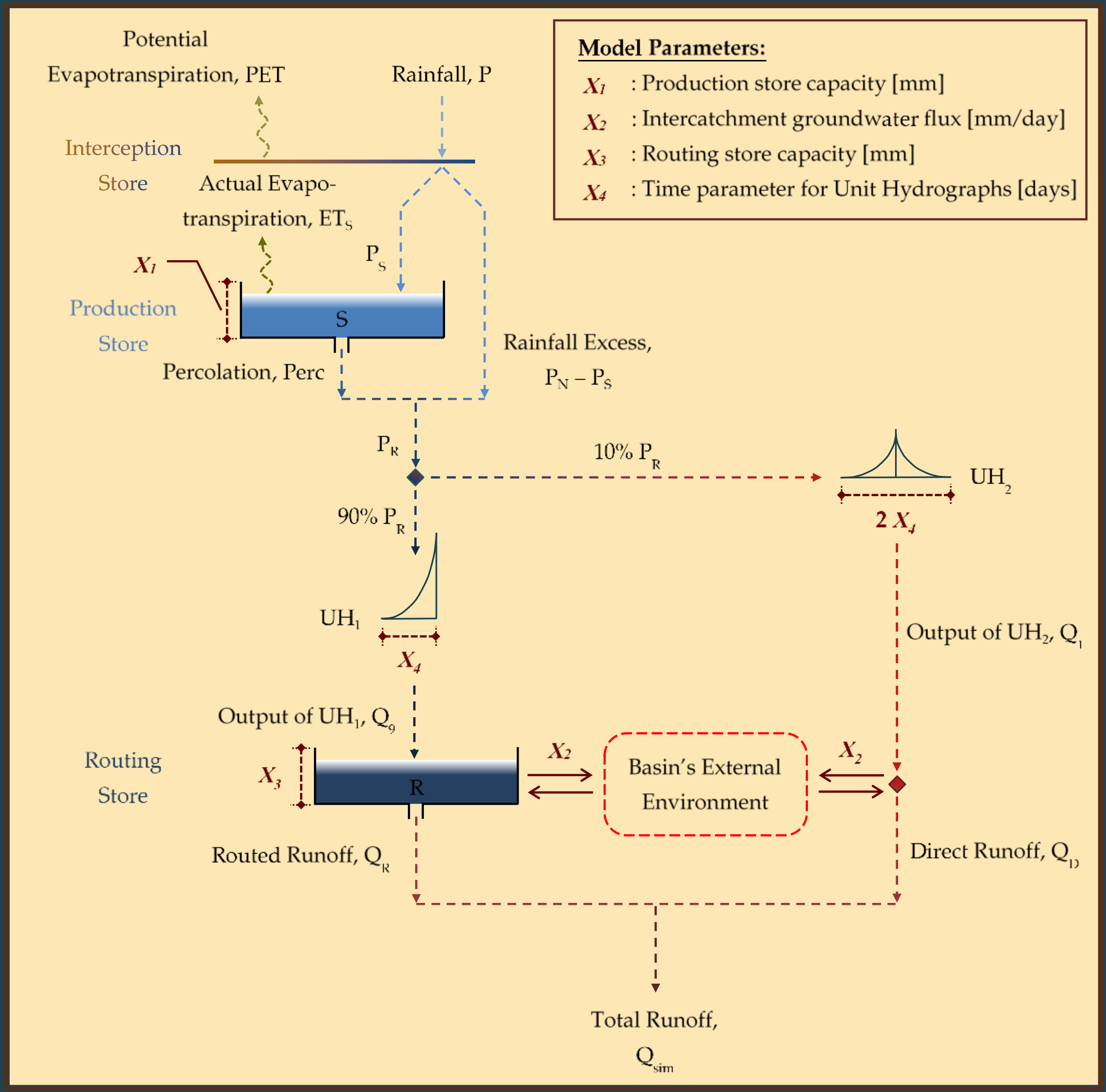


Figure 4 | Schematic representation of the conceptual structure of the GR4J model

Daily Hydrological Model GR6J

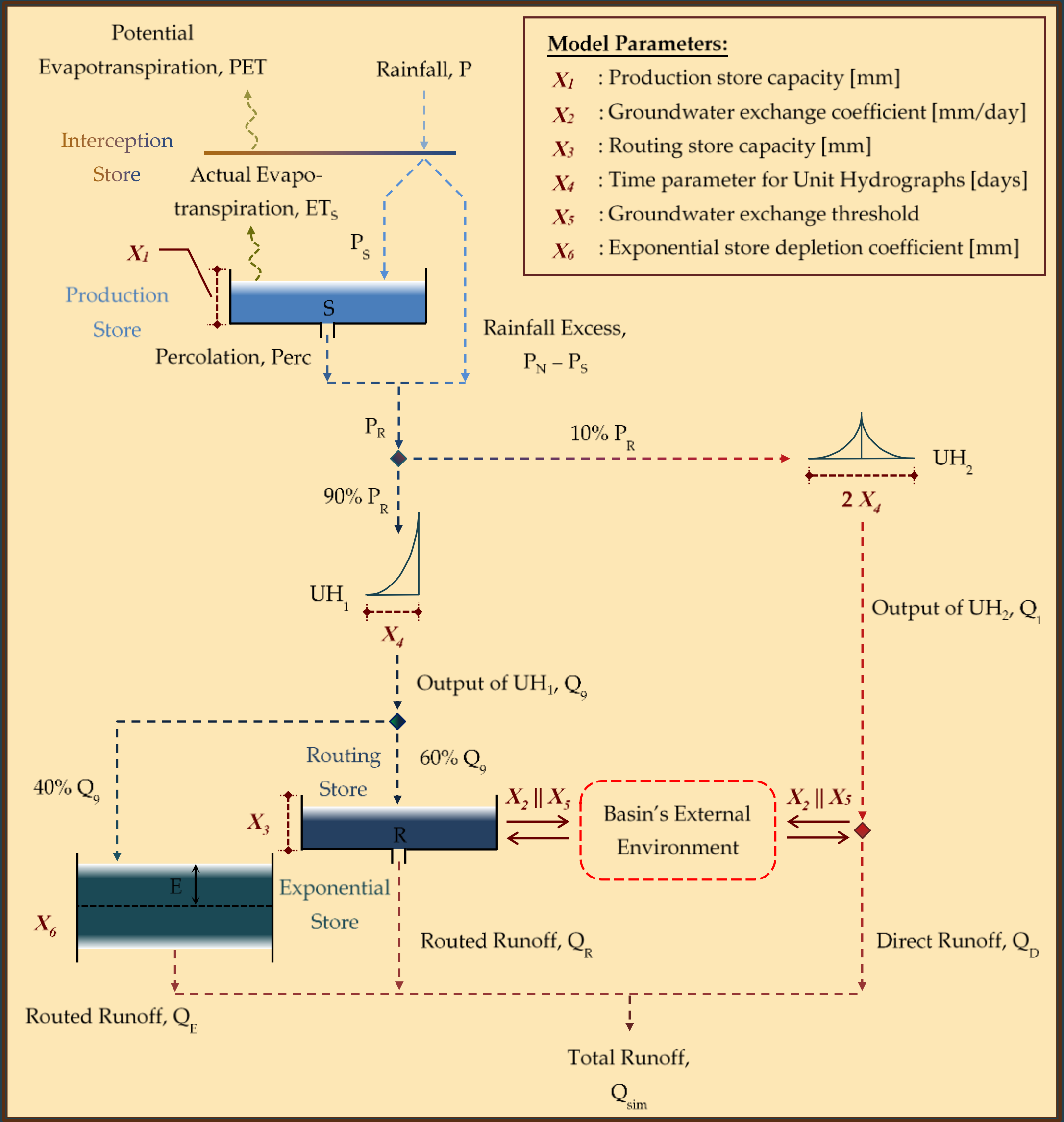


Figure 5 | Schematic representation of the conceptual structure of the GR6J model

GR6J Equations for each time step:

(additional or modified equations)

The mathematical equations developed by Perrin *et al.* [2003] to describe the hydrological processes until the point of two flow components Q_1 and Q_9, resulted from the two unit hydrographs (UH_1 & UH_2), continue being valid for the GR6J model.

$$F = X_2 \left(\frac{R}{X_3} - X_5 \right)$$

$$R_1 = \max[\, R + 0.6\, Q_9 + F \,;\, 0 \,]$$

$$E_1 = E + 0.4\, Q_9$$

The routed runoff Q_R and the routing reservoir level R at the end of each daily time step are calculated according to the respective equations of GR4J.

The exponential store is controlled by a model parameter, X_6, which acts as a base level in the reservoir. The routed runoff Q_E is estimated as a function of the current store level E_1 and this parameter.

$$Q_{sim} = Q_E + Q_R + Q_D$$

Calibration Methods & Criteria

- The parameter sets for both Zygos-4P and Zygos-6P models are estimated for all the study catchments using the Evolutionary Annealing-Simplex (EAS) optimization method [Efstratiadis and Koutsoyiannis, 2002], a hybrid scheme that merges the strengths of both local and global search.
- The calibration of all three of the GR models was employed with the technique proposed by Michel [1991], an algorithm that also combines a local and a global approach and is already implemented in *airGR* R–package [Coron *et al.*, 2017a; Coron *et al.*, 2017b].

Table 1 summarizes the ranges of variation for each model parameter, as obtained from the literature.

Model Performance Index:
[Nash and Sutcliffe, 1970]

$$\mathbf{NSE} = 1 - \frac{\sum_{i=1}^{N}\left(Q_{obs,i} - Q_{sim,i}\right)^2}{\sum_{i=1}^{N}\left(Q_{obs,i} - \overline{Q_{obs}}\right)^2}$$

where **NSE** is the efficiency metric,
$Q_{obs,i}$ is the observed runoff values,
$Q_{sim,i}$ is the modeled runoff values,
$\overline{Q_{obs}}$ stands for the mean runoff

Table 1 | Parameters of conceptual rainfall-runoff models and their feasible ranges

Model	Parameter	Unit	Range	Reference
Zygos-4P	***v***	-	[0 , 1]	Ranges deduced from numerous simulations
	K	mm	[0 , 1000]	
	κ	-	[0 , 1]	
	λ	-	[0 , 1]	
Zygos-6P	***v***	-	[0.1 , 2.5]	
	H	mm	[0 , 300]	
	a	-	[0 , 1]	
GR2M	X_1	mm	[0 , 1500]	Mouelhi *et al.*, 2006
	X_2	– or mm/day	[-10 , 10]	
GR4J	X_3	mm	[1 , 500]	Perrin *et al.*, 2003
	X_4	days	[0.5 , 8]	
GR6J	X_5	-	[-4 , 4]	Le Moine, 2008
	X_6	mm	[0.5 , 20]	Pushpalatha *et al.*, 2011

Proof-of-Concept – Study Area

Before we proceed to a large scale analysis, the proposed framework is established upon a specific study basin, the Loing river basin. The geographical representation of the basin is given in Figure 6 (catchment boundary was taken from European Catchments and Rivers Network System (ECRINS)).

The available hydrological data (precipitation, potential evapotranspiration, runoff) are of daily time scale and extend over the period 01/08/1958 to 31/07/2016 (Fig. 7). The data were taken from a previous study conducted by Rebolho *et al.* [2018] in the same watershed.

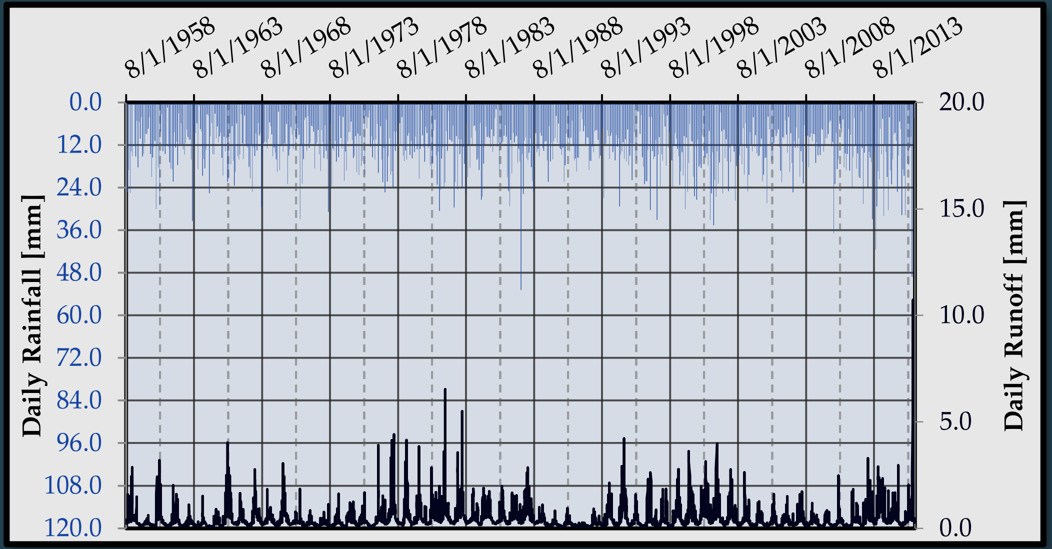


Figure 7 | Historical daily rainfall and runoff time series

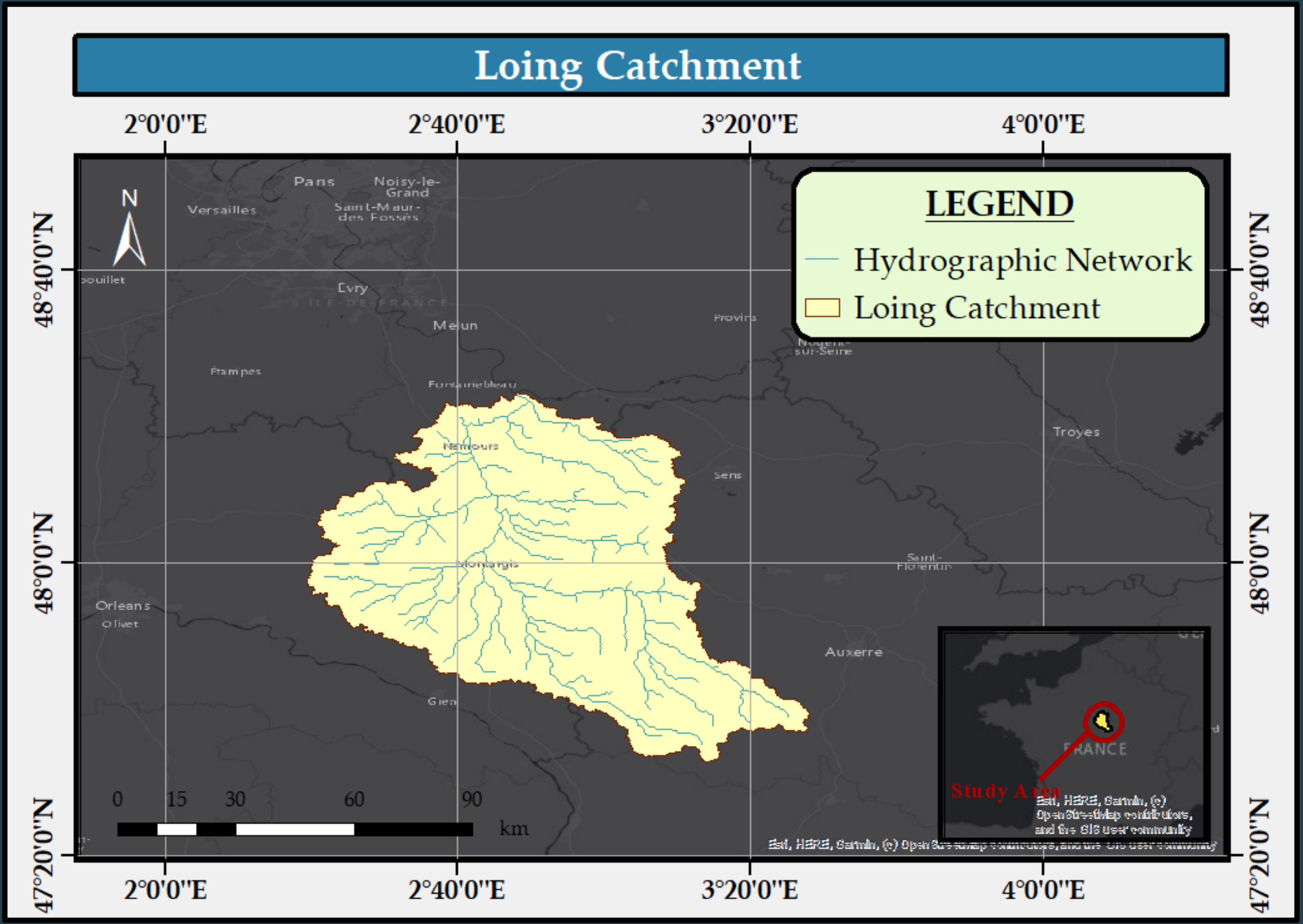


Figure 6 | Geographical representation of the Loing catchment

Proof-of-Concept – Analysis Setup

At first, stochastic calibration for this specific study case was employed at a monthly time-scale using the hydrological models GR2M and Zygos-6P. Thus, the daily observed data are aggregated at the monthly scale (Fig. 8). Subsequently, we proceed to a further analysis, testing the validity of the framework at the daily scale, by employing GR4J and GR6J rainfall-runoff models.

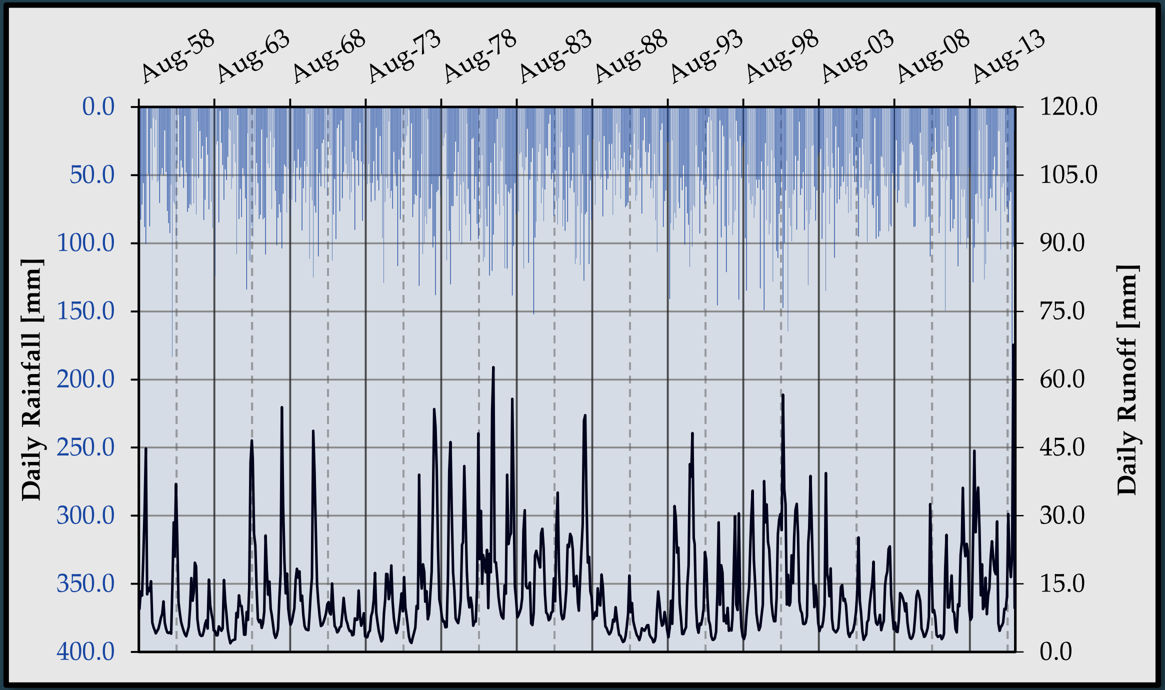
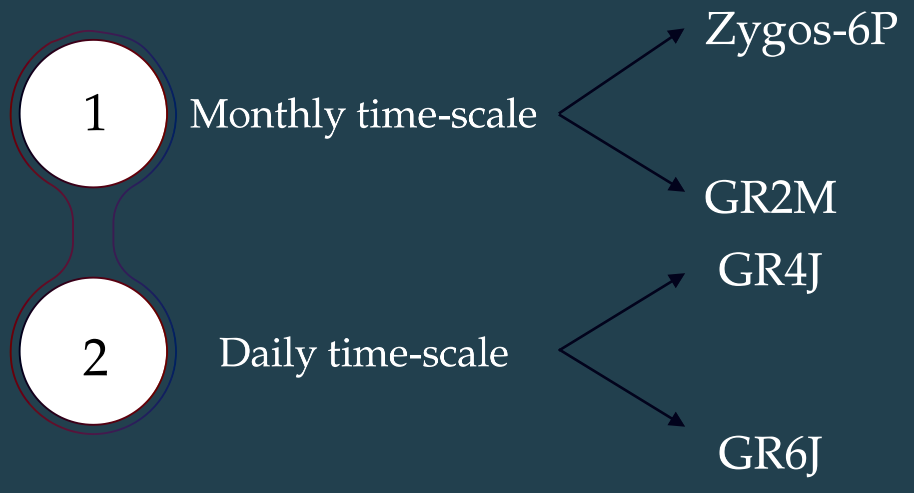


Figure 8 | Historical monthly rainfall and runoff time series



For each model calibration, the classical split-sample scheme is initially employed, by dividing the historical records into two equal subsets of length 29 years each, for calibration and validation. Consequently, the stochastic calibration procedure is employed, and the models are fitted against a synthetic time series of length 1000 years, and their structure is validated against the full sample of observations.

Stochastic Simulation – Conditions on the Synthesis Procedure

A consistent and robust stochastic calibration framework dictates a few **specifications** for the data synthesis procedure:

- Representation of the **full probabilistic behavior of the input and output processes**, through careful assignment of suitable distribution models across processes and scales, instead of a blind reproduction of the observed statistical characteristics *per se* [Tsoukalas *et al.*, 2019, 2020].

- Representation of **auto- and cross-dependencies**, which are statistical descriptors of the major drivers and cause-effect phenomena across the hydrological cycle.

- **Multi-scale consistency**, through effective coupling of stochastic models across multiple time scales, to ensure the reproduction of the marginal and dependence behavior of the modelled processes not only at the temporal resolution of the hydrological simulation (e.g., daily, monthly), but also across coarser levels of aggregation (e.g., annual).

Regarding the synthetic time series generation procedure, these were produced via the ***anySim*** R-package [Tsoukalas *et al.*, 2020], specifically designed for the simulation of non-Gaussian behavior, which characterizes hydrometeorological processes, apart from other significant peculiarities such as periodicity, intermittency, and auto- and cross-dependence.

Concerning the evapotranspiration synthetic input data, we simply apply the mean monthly values of the historical sample.

Monthly Rainfall & Runoff Synthetic Time Series Generation

Table 2 demonstrates the key statistics of historical rainfall and runoff data (mean, standard deviation, skewness, auto- and cross-correlations), to be reproduced within the synthetic time series that are used in stochastic calibration.

As shown in **Figure 9**, the synthesis scheme ensured the reproduction of **dependency patterns that are much extended than the observed ones**. These are expected to represent the full hydroclimatic regime of the basin, which cannot be traced in the case of the observed data, due to their limited length. Actually, while using the split-sample calibration approach, only half of this information is accounted for.

Table 2 | Key statistical information of observed rainfall and runoff data and their lag-0 cross-correlation coefficient at the monthly and annual time scales

Rainfall [mm]	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Annual
Mean	67.5	66.6	69.1	61.8	53.0	55.3	53.5	68.3	57.4	56.7	57.0	56.4	722.5
St. deviation	36.9	27.0	30.2	29.6	28.2	30.5	31.2	31.5	26.2	30.7	29.9	35.6	132.3
Skewness	0.603	0.554	0.488	0.508	0.537	0.856	0.558	1.019	0.458	1.111	0.217	0.865	0.131
Lag-1 correl.	0.162	-0.016	-0.173	0.035	-0.058	0.197	0.102	-0.031	0.296	0.135	0.108	-0.231	0.126
Runoff [mm]													
Mean	7.9	10.7	17.0	22.2	21.3	20.1	16.1	13.6	9.6	6.6	5.8	5.4	156.4
St. deviation	5.0	7.3	10.5	13.8	12.7	11.3	9.9	7.2	8.8	3.6	3.6	2.3	66.2
Skewness	1.769	2.31	1.177	0.962	0.82	1.345	1.659	0.95	5.284	1.402	2.404	1.044	0.357
Lag-1 correl.	0.713	0.786	0.576	0.672	0.514	0.717	0.683	0.746	0.438	0.455	0.833	0.703	0.483
Lag-0 correl.	0.616	0.392	0.61	0.681	0.748	0.758	0.625	0.436	0.394	0.532	0.330	0.202	0.841

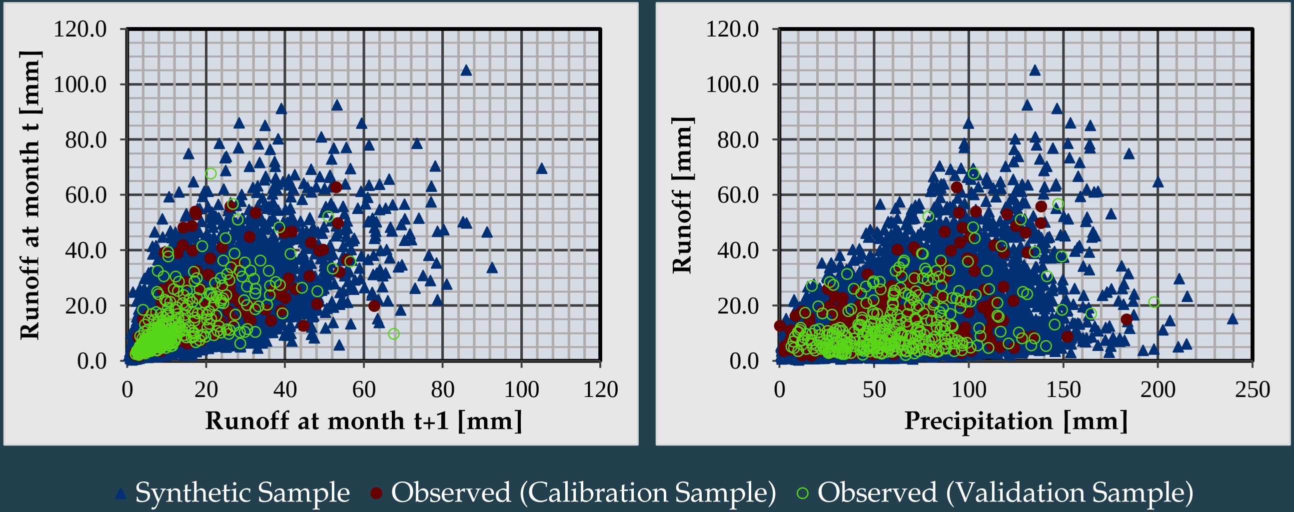


Figure 9 | Auto-dependency patterns among runoff data between subsequent months (left) and cross-dependency patterns between rainfall and runoff (right), derived from the observed data (split into two periods) and the synthetic ones (12 000 values)

Monthly Time-Scale Analysis – Zygos-6P Model

Table 3 | Model performance evaluation

Split-Sample Approach	NSE_{Cal}	NSE_{Val}	NSE_{Tot}
	0.837	0.836	0.836
Stochastic Calibration	NSE		
	0.836		

NSE_{Cal} : NSE in Calibration

NSE_{Val} : NSE in Validation

NSE_{Tot} / NSE : Overall NSE

Table 4 | Optimized Zygos-6P parameter values

Split-Sample Approach	v	K [mm]	κ	H [mm]	α	λ
	1.763	235.0	0.184	60.8	0.120	0.142
Stochastic Calibration	v	K [mm]	κ	H [mm]	α	λ
	1.835	234.2	0.180	44.2	0.157	0.173

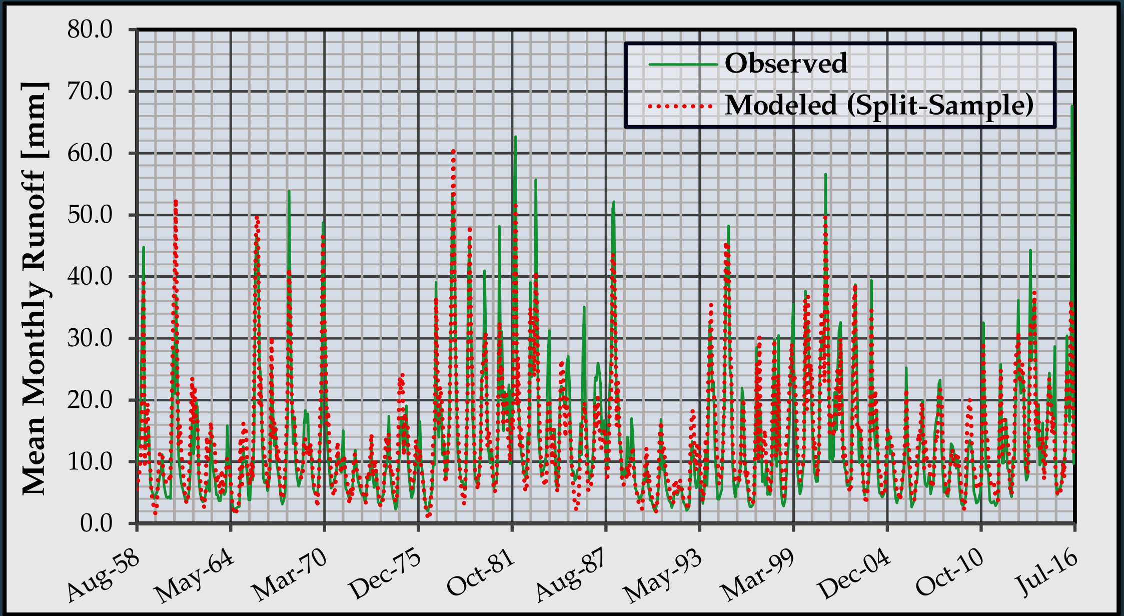


Figure 10 | Visual inspection of the agreement between observations and model predictions (**Split-Sample** approach)

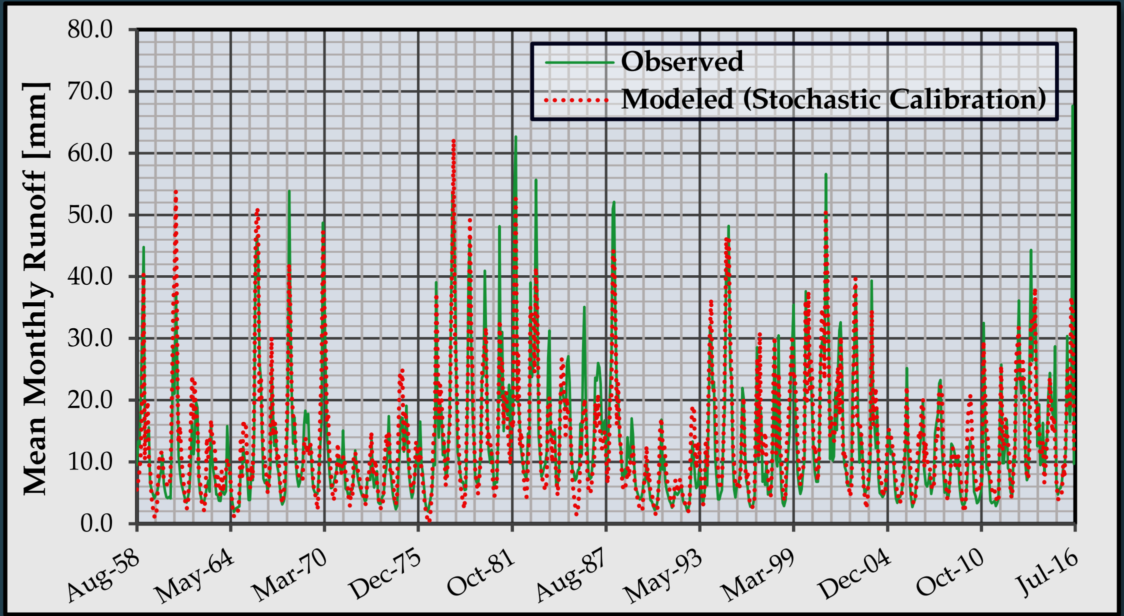


Figure 11 | Visual inspection of the agreement between observations and model predictions (**Stochastic Calibration** approach)

Monthly Time-Scale Analysis – GR2M Model

Table 5 | Model performance evaluation

Split-Sample Approach	NSE_{Cal}	NSE_{Val}	NSE_{Tot}
	0.817	0.776	0.798
Stochastic Calibration	**NSE**		
	0.798		

Table 6 | Optimized GR2M parameter values

Split-Sample Approach	X_1 [mm]	X_2
	395.4	0.77
Stochastic Calibration	X_1 [mm]	X_2
	399.4	0.77

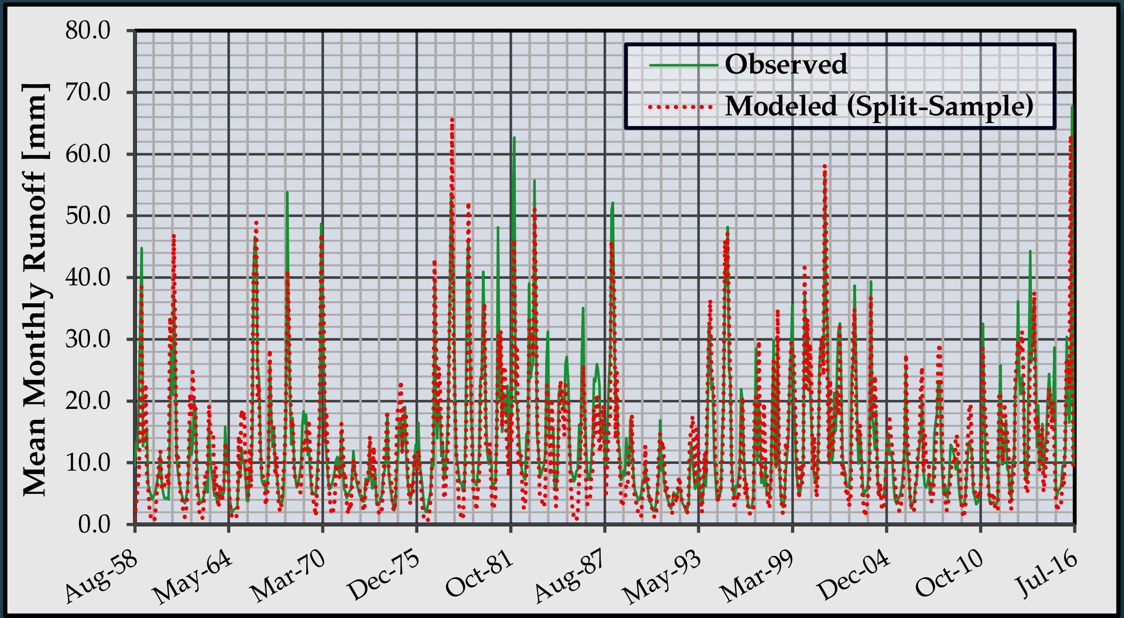


Figure 12 | Visual inspection of the agreement between observations and model predictions (**Split-Sample** approach)

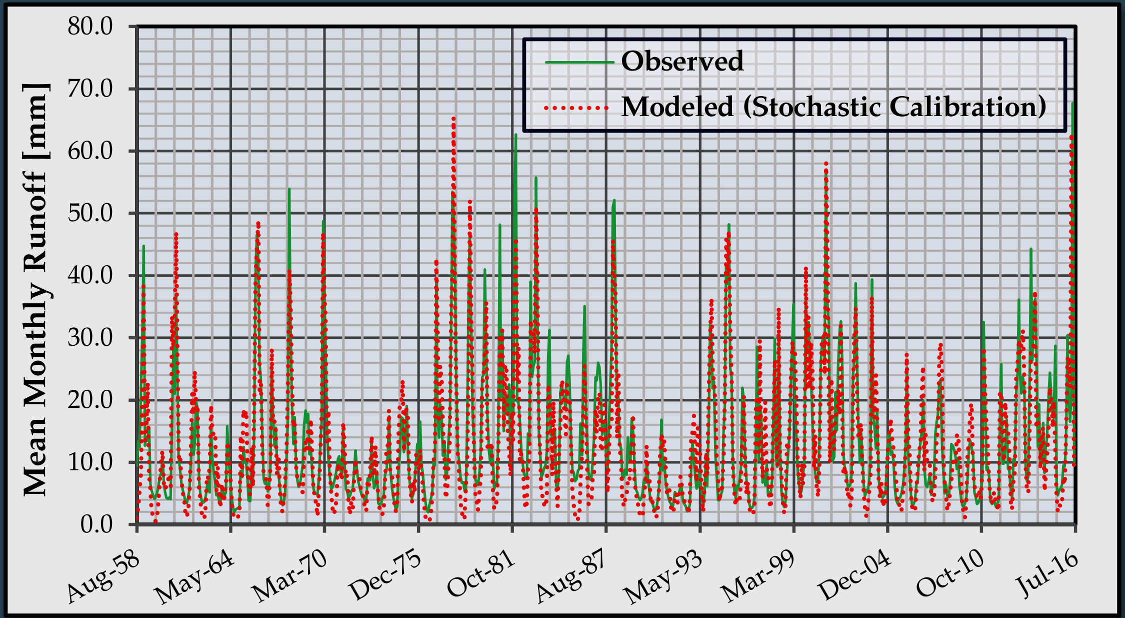


Figure 13 | Visual inspection of the agreement between observations and model predictions (**Stochastic Calibration** approach)

Daily Time-Scale Analysis – GR4J Model

Table 7 | Model performance evaluation

Split-Sample Approach	NSE_{Cal}	NSE_{Val}	NSE_{Tot}
	0.848	0.895	0.871
Stochastic Calibration	**NSE**		
	0.805		

Table 8 | Optimized GR4J parameter values

Split-Sample Approach	X_1 [mm]	X_2 [mm/d]	X_3 [mm]	X_4 [days]
	520.1	-0.70	35.9	4.20
Stochastic Calibration	X_1 [mm]	X_2 [mm/d]	X_3 [mm]	X_4 [days]
	1011.6	-0.48	38.8	4.39

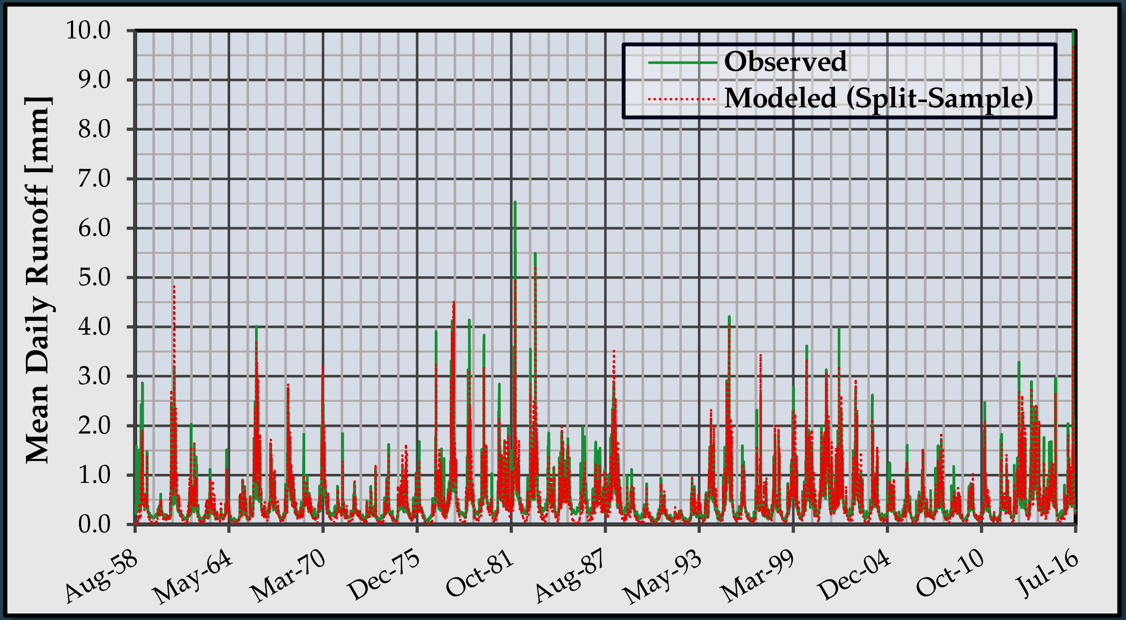


Figure 14 | Visual inspection of the agreement between observations and model predictions (**Split-Sample** approach)

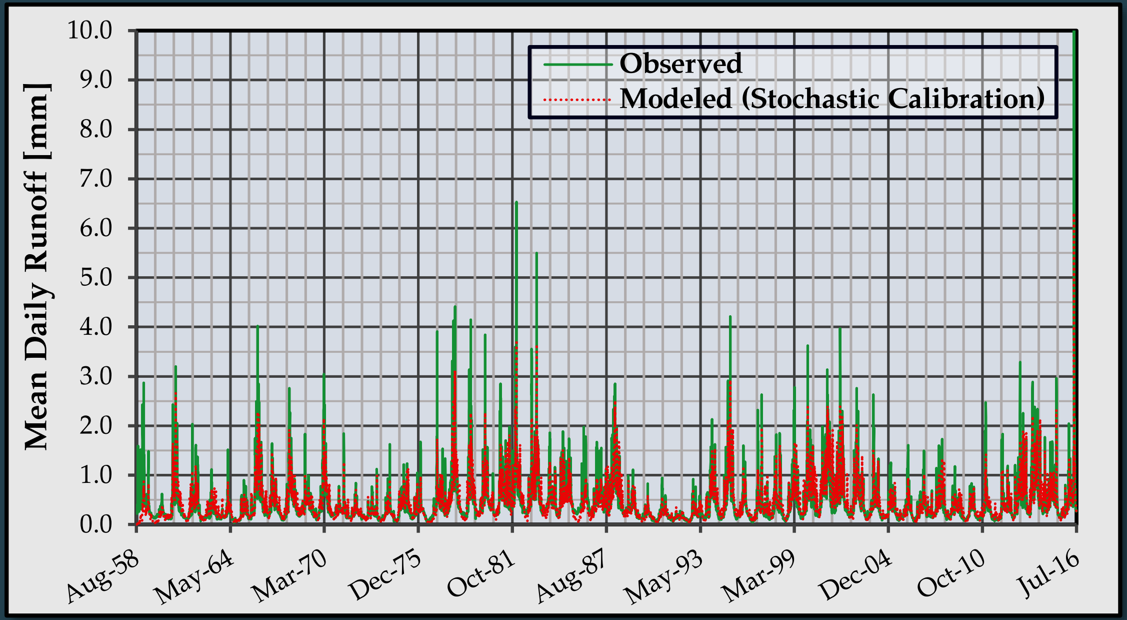


Figure 15 | Visual inspection of the agreement between observations and model predictions (**Stochastic Calibration** approach)

Daily Time-Scale Analysis – GR6J Model

Table 9 | Model performance evaluation

Split-Sample Approach	NSE_{Cal}	NSE_{Val}	NSE_{Tot}
	0.849	0.802	0.858
Stochastic Calibration	NSE		
	0.815		

Table 10 | Optimized GR6J parameter values

Split-Sample Approach	X_1 [mm]	X_2 [mm/d]	X_3 [mm]	X_4 [days]	X_5 [-]	X_6 [mm]
	242.1	-1.63	88.54	4.40	0.22	1.49
Stochastic Calibration	X_1 [mm]	X_2 [mm/d]	X_3 [mm]	X_4 [days]	X_5 [-]	X_6 [mm]
	296.6	-2.38	52.1	4.38	0.41	3.04

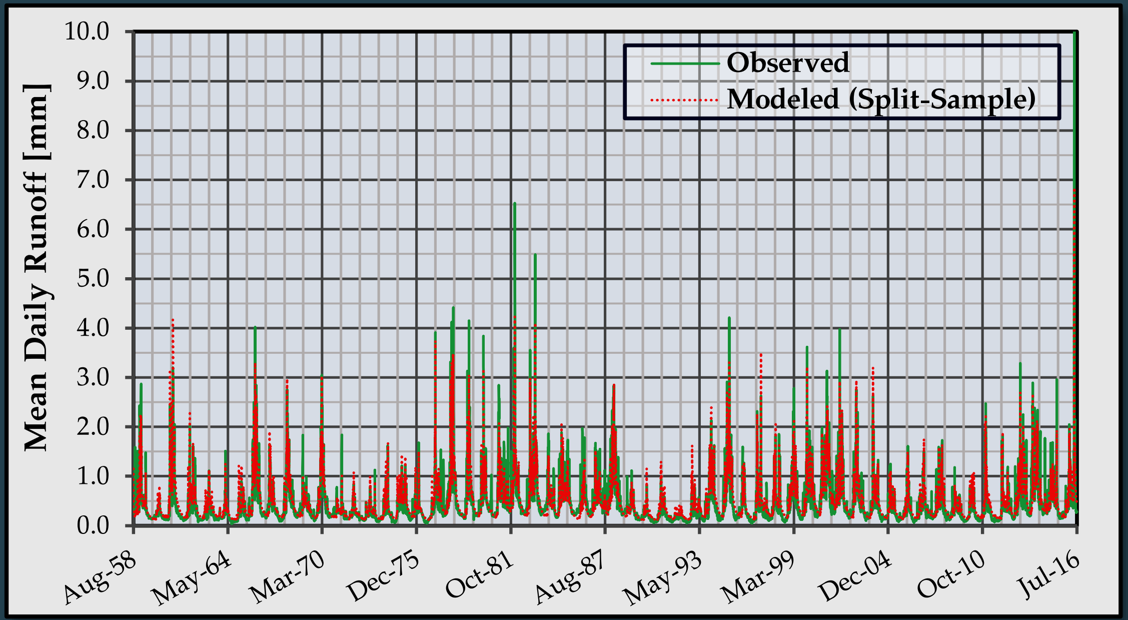


Figure 16 | Visual inspection of the agreement between observations and model predictions (**Split-Sample** approach)

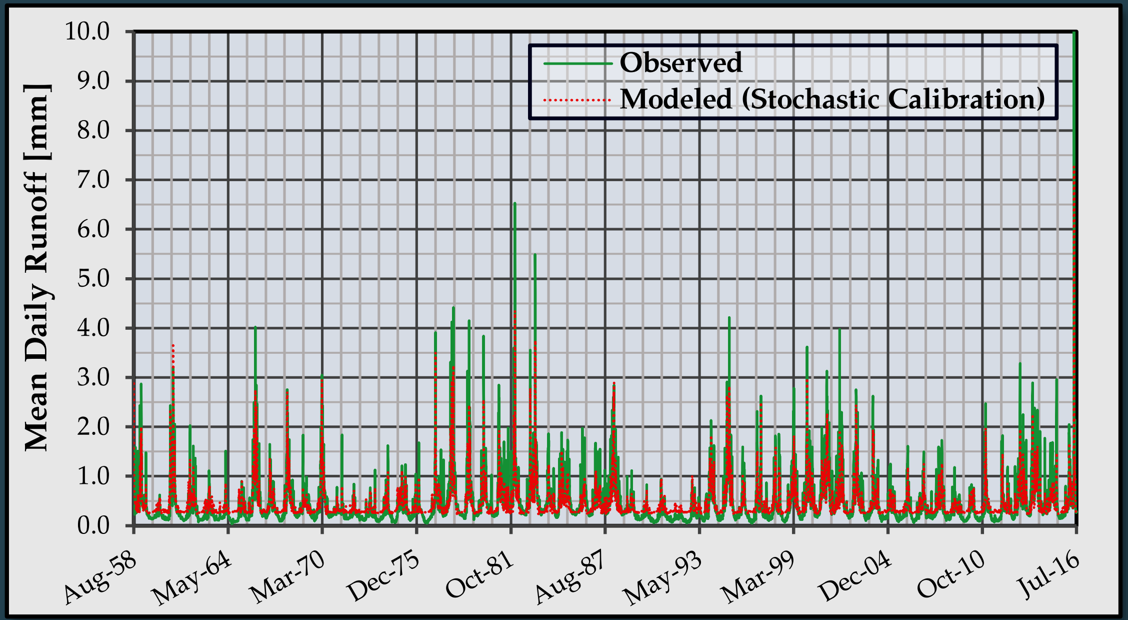


Figure 17 | Visual inspection of the agreement between observations and model predictions (**Stochastic Calibration** approach)

Conclusive Remarks Regarding the Proof-of-Concept Case

- In regard to the **monthly time-scale analysis**, it is evident that the proposed methodology for calibration against synthetic rainfall-runoff data is functional. Specifically, the Zygos-6P and GR2M models **performed equally well** for calibration against historical data (split-sample approach) and against synthetic data (stochastic calibration approach). Moreover, the model parameter values for each calibration method are really close.

- Concerning the **daily time-scale analysis**, for the application of the split-sample scheme the overall efficiency NSE_{tot} for the two daily hydrological models (GR4J and GR6J) is slightly higher than the efficiency metric NSE that was estimated for the stochastic calibration case. Hence, the convergence between the above-mentioned metrics indicate that the **stochastic calibration framework has certain potentials for application also to the daily scale**. It is also worth noticing that the optimized parameter values against the synthetic rainfall-runoff data are quite different with respect to the ones derived by calibrating against the half of historical data, especially for the case of GR6J model.

- The results of this initial investigation consist some first evidence to assume that the implementation of the stochastic calibration framework is independent of the chosen hydrological model.

Large Scale Analysis

After proving the functionality of the proposed framework for the previously presented study area, the validity of the proposed framework is tested against **a large set of catchments at a monthly scale**. Specifically, this investigation was conducted by selecting 100 catchments, accompanied by their respective datasets, from the MOPEX database (**Fig. 18**).

MOPEX (**MO**del **P**arameter **E**stimation **EX**periment) is a project developed for the enhancement of *a priori* parameter estimation methodologies for hydrological models and land surface parameterization schemes [Schaake *et al.*, 2006].

A combination of criteria was taken into account for the selection of optimal set of catchments. Specifically, it resulted from:

- the percentage of stream flow missing values and
- cross-correlation between rainfall and streamflow.

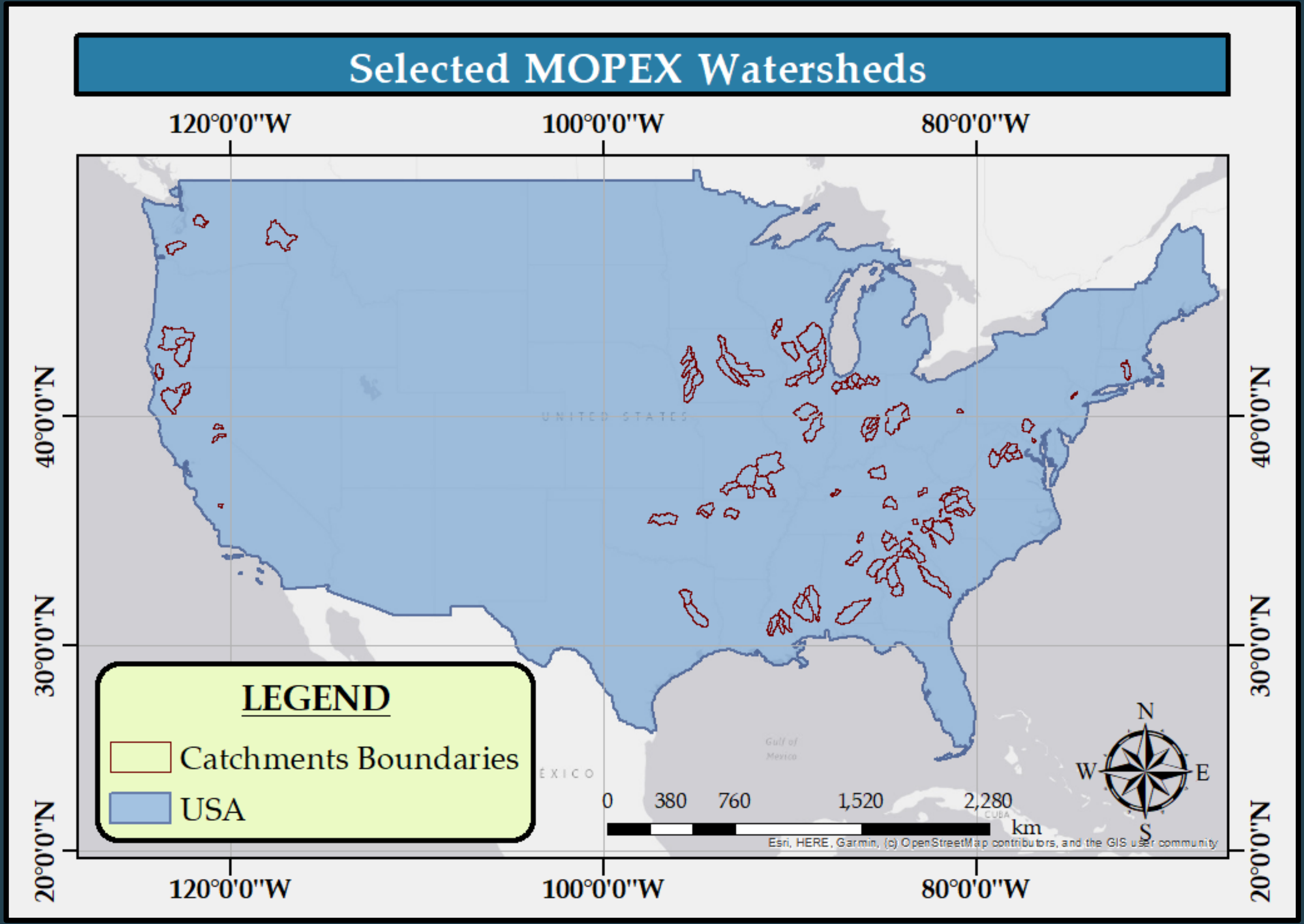


Figure 18 | Geographical location of the 100 selected MOPEX watersheds

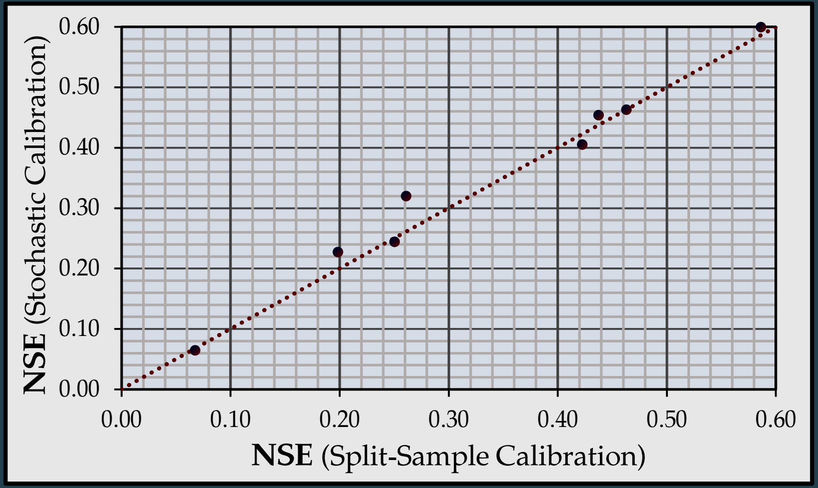
Analysis Setup

The spatially averaged rainfall, evapotranspiration and streamflow data, provided for each of the 100 catchments, are observed at the daily time interval, thus they are aggregated at the monthly scale.

As in the previous study, a number of different hydrological models were used, to further fortify the assumption that calibration with stochastic inputs is independent of the chosen rainfall-runoff model structure. Specifically, for this study we use the GR2M, Zygos-4P and Zygos-6P models, which have been already presented in a former section. It should be noted that for the case of the split-sample approach, once more half of the total historical sample is used for calibration purposes.

Aiming at an <u>efficient</u> comparison between the performance for each calibration methodology, we presented graphically the model efficiency metrics (NSE) concerning the historical sample allocated for <u>validation</u> purposes. This decision stemmed from the notion that the validation period of observed records contains the only data over which, <u>in both calibration methodologies</u>, the rainfall-runoff model has not been trained. On the contrary, model efficiency metrics regarding the model calibration period of historical data by employing the split-sample calibration approach would be biased since the selected model has been over-fitted on this period.

Large Scale Analysis – GR2M Model



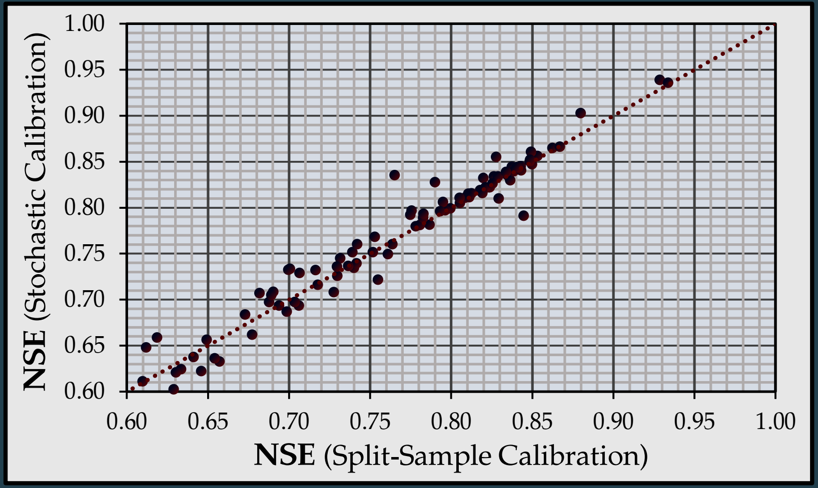


Figure 19 | Scatter plots of poor/low (above) or good/high (below) model performance (NSE) by employing the Split-Sample approach and the Stochastic Calibration

Hist. Calibration Sample: NSE (Stoch Cal.) > NSE (Split-Sample Cal.) : 1/100 Basins

Hist. Validation Sample: NSE (Stoch Cal.) > NSE (Split-Sample Cal.) : 62/100 Basins

Total Hist. Sample: NSE (Stoch Cal.) > NSE (Split-Sample Cal.) : 42/100 Basins

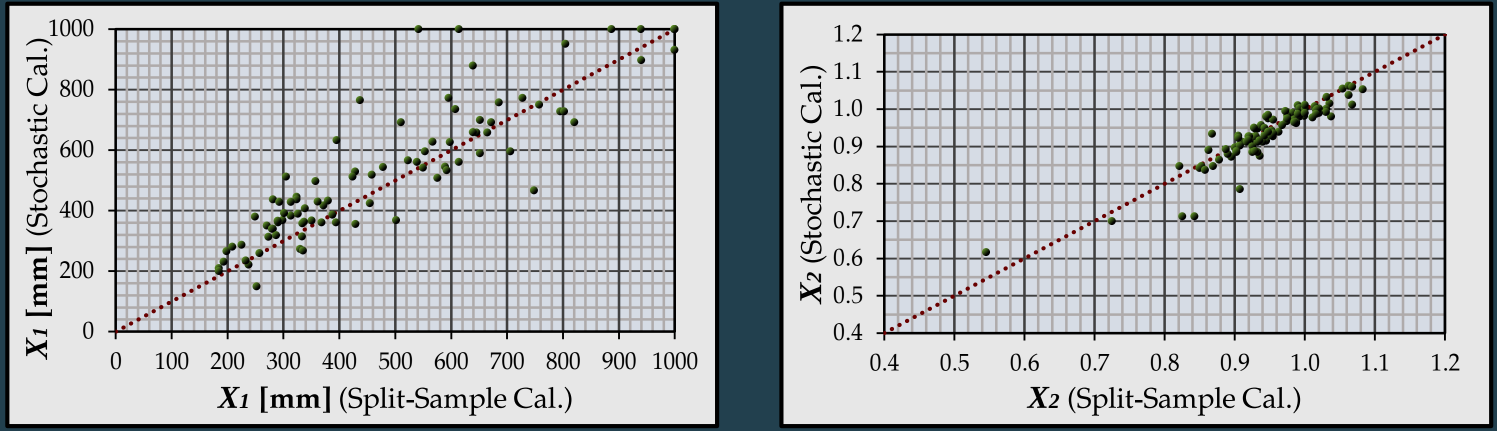


Figure 20 | Scatter plots of GR2M model parameters estimated through Split-Sample approach or Stochastic Calibration

Large Scale Analysis – Zygos-4P Model

Hist. Calibration Sample: NSE (Stoch Cal.) > NSE (Split-Sample Cal.) : 0/100 Basins

Hist. Validation Sample: NSE (Stoch Cal.) > NSE (Split-Sample Cal.) : 67/100 Basins

Total Hist. Sample: NSE (Stoch Cal.) > NSE (Split-Sample Cal.) : 38/100 Basins

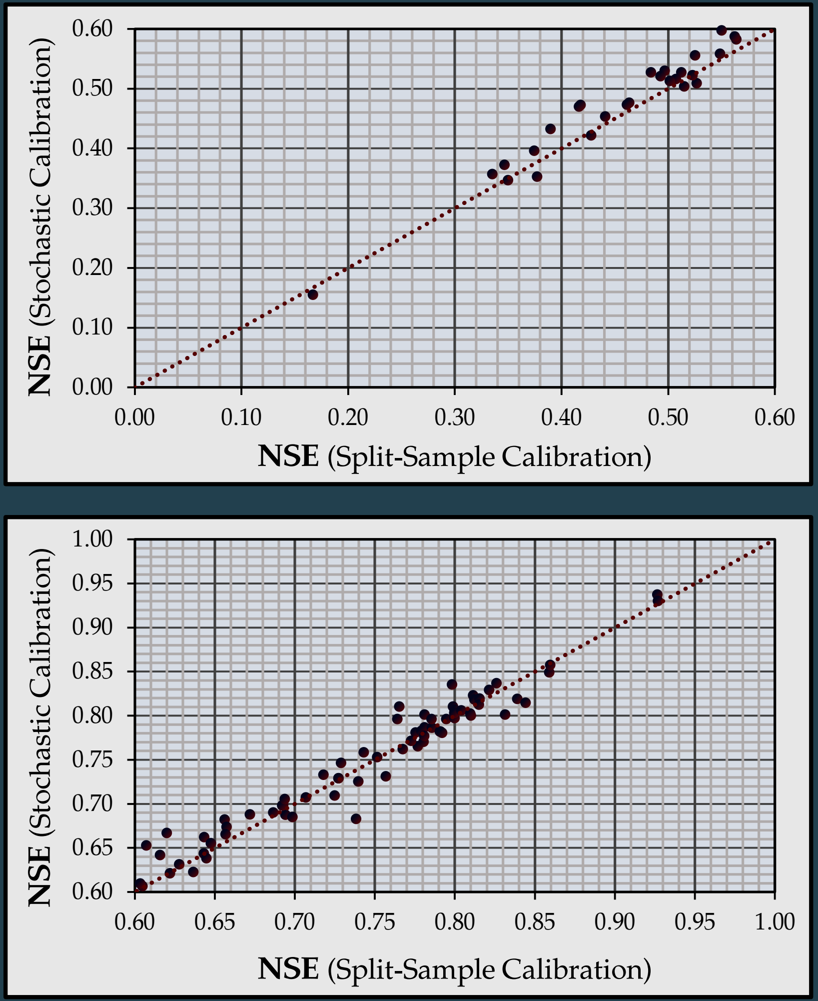


Figure 21 | Scatter plots of poor/low (above) or good/high (below) model performance (NSE) by employing the Split-Sample approach and the Stochastic Calibration

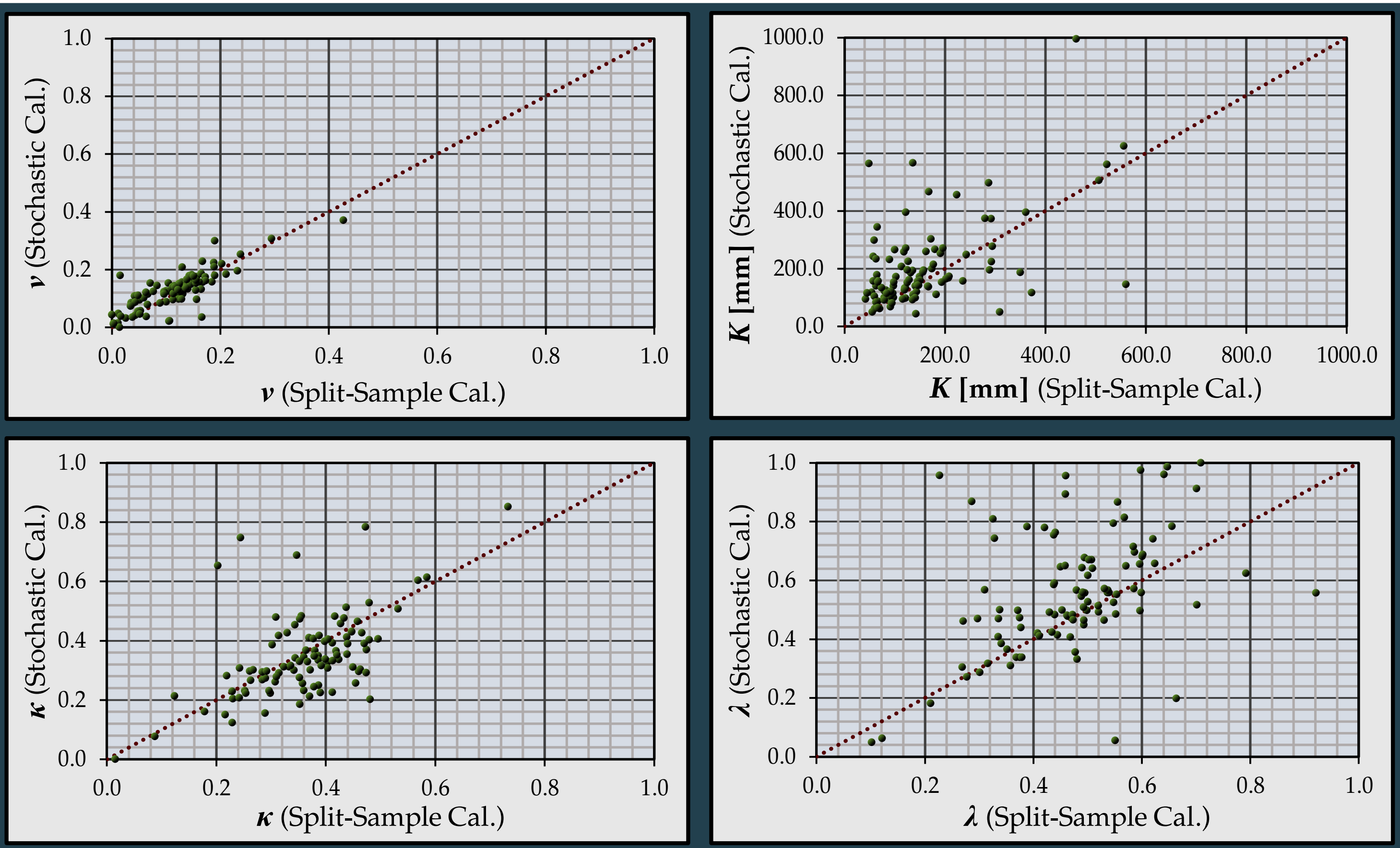


Figure 22 | Scatter plots of Zygos-4P model parameters estimated through Split-Sample approach or Stochastic Calibration

Large Scale Analysis – Zygos-6P Model

Hist. Calibration Sample: NSE (Stoch Cal.) > NSE (Split-Sample Cal.) : 4/100 Basins

Hist. Validation Sample: NSE (Stoch Cal.) > NSE (Split-Sample Cal.) : 60/100 Basins

Total Hist. Sample: NSE (Stoch Cal.) > NSE (Split-Sample Cal.) : 31/100 Basins

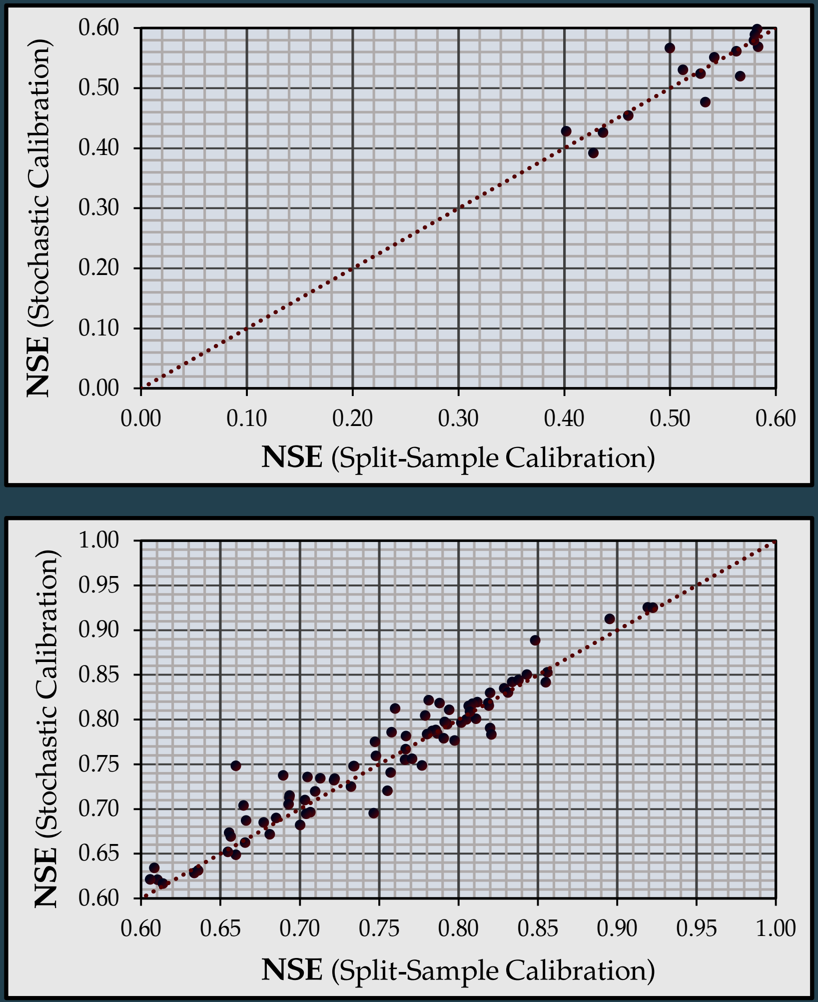


Figure 23 | Scatter plots of poor/low (above) or good/high (below) model performance (NSE) by employing the Split-Sample approach and the Stochastic Calibration

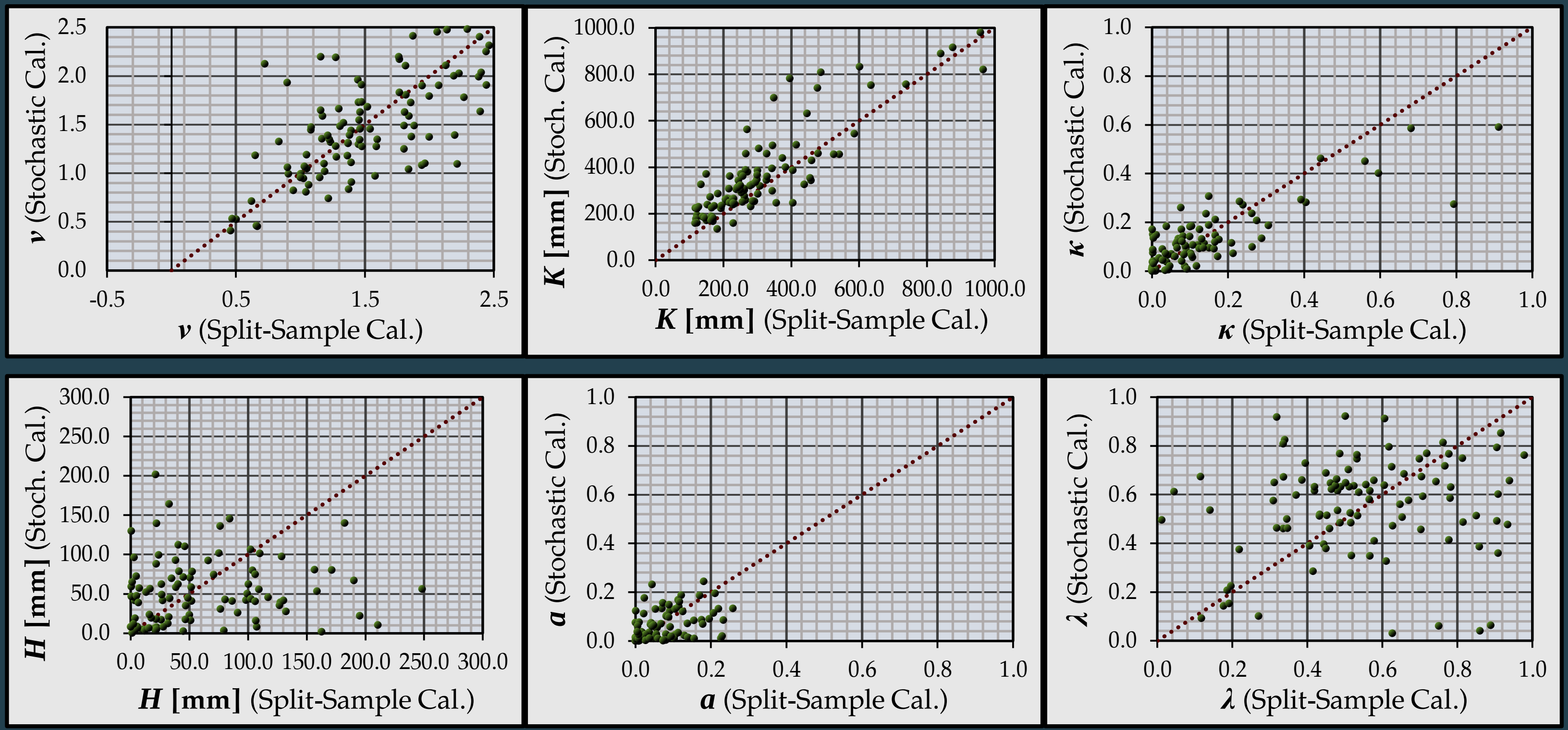


Figure 24 | Scatter plots of Zygos-6P model parameters estimated through Split-Sample approach or Stochastic Calibration

Conclusive Remarks Regarding the Large Scale Analysis

- On the basis of our previous view concerning efficient comparison between the tested calibration methodologies, it is evident that **the stochastic calibration outperforms the conventional split-sample approach**.

- It is evident that in each model analysis some parameters estimated by stochastic calibration **differ** from the ones stemmed from the split-sample approach. Concerning the analysis using the **GR2M model**, parameter values for X_1 differs noticeably for calibration employed by a split-sample approach and for stochastic calibration. Regarding the analysis using the **Zygos-4P model**, parameter values for K, κ and λ derived from the use of the two calibration methodologies disagree in most of the tested basins, with this disagreement being more intense in the case of parameter λ. Discrepancy between the parameter values estimated by each calibration methodology appears for every parameter also in the case of the **Zygos-6P model**.

Conclusions & Proposals for Future Research

This study showed the strengths of stochastic calibration against conventional split-sample approaches. Initially, this novel methodology **proved equally sufficient as the classical split-sample scheme** when we employed a **monthly-scale analysis** by means of proof-of-concept, whereas at a **daily scale** the stochastic calibration methodology **demonstrated slightly lower performance**.

Afterwards, the proposed method was stressed-tested against a set of 100 catchments and **outperformed the split-sample approach for model calibration**, regardless of the chosen model for hydrological simulation.

Moreover, calibration with the use of synthetic inputs and outputs proved to be **independent of the hydrological model complexity**.

Undoubtedly, this is just a first attempt to prove the validity of stochastic calibration, and calls for further studies. In particular:

- It may not always be efficient to characterize the different aspects of model performance for a particular rainfall-runoff model with only one performance metric. Thus, more efficiency metrics should be compared to evaluate the predictive capacity of each method.
- It is also worth exploring the potentials of this methodology by employing rainfall-runoff models of more complex structure.
- Another suggestion would be to further test the applicability of stochastic calibration at the daily time scale.

References

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V. (2017). The suite of lumped GR hydrological models in an R package. *Environmental Modelling and Software*, 94, 166 – 171, doi: 10.1016/j.envsoft.2017.05.002.

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., and Michel, C. (2021). *airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling*. R package version 1.6.10.4, doi: 10.15454/EX11NA, https://CRAN.R-project.org/package=airGR.

Efstratiadis, A., and Koutsoyiannis, D. (2002). An evolutionary annealing-simplex algorithm for global optimisation of water resource systems. *Proceedings of the Fifth International Conference on Hydroinformatics*. Cardiff, UK, 1423 – 1428, International Water Association.

Efstratiadis, A., Tsoukalas, I., and Kossieris, P. (2021). *Improving model identifiability by embedding stochastic simulation within hydrological calibration*. *In:* Solomatine, D., and Corzo, G. (editors), *Advances in Hydroinformatics: Artificial Intelligence and Optimization for Water Resources*. Edited volume of AGU Books / Wiley (in review).

Kozanis, S., and Efstratiadis, A. (2006). Zygos: A basin processes simulation model. *21st European Conference for ESRI Users*, Athens, Greece.

Le Moine, N. (2008). *Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performance et du réalisme des modéles pluie-débit?* Doctoral dissertation, Doctorat Géosciences et Ressources Naturelles, Université Pierre et Marie Curie Paris VI.

Nash, J.E. and Sutcliffe, J.V., (1970). River flow forecasting through conceptual models. Part 1: a discussion of principles. *Journal of Hydrology*, 10 (3), 282 – 290. doi:10.1016/ 0022-1694(70)90255-6.

Michel, C. (1991). Hydrologie appliquée aux petits bassins ruraux, *Hydrology Handbook* (in French), Cemagref, Antony, France.

References

Mouelhi, S., Michel, C., Perrin, C., and Andréassian, V. (2006). Stepwise development of a two-parameter monthly water balance model. *Journal of Hydrology*, 318 (1–4), 200 – 214. doi:10.1016/j.jhydrol.2005.06.014.

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

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V. (2011). A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. *Journal of Hydrology*, 411 (1–2), 66 – 76. doi:10.1016/j.jhydrol.2011.09.034.

Rebolho, C., Andréassian, V., Tsoukalas, I., et Efstratiadis, A. (2018). La crue du Loing de Juin 2016 était-elle exceptionnelle? *De la prévision des crues à la gestion de crise,* Avignon, Société Hydrotechnique de France.

Schaake, J., Cong, S., and Duan, Q. (2006). The US MOPEX data set. IAHS-AISH Publication. 9 – 28.

Tsoukalas, I., Efstratiadis, A., and Makropoulos, C. (2019). Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure. *Journal of Hydrology*, 575, 354 – 380. doi:10.1016/j.jhydrol.2019.05.017.

Tsoukalas, I., Kossieris, P., and Makropoulos, C. (2020). Simulation of non-Gaussian correlated random variables, stochastic processes and random fields: Introducing the anySim R-package for environmental applications and beyond. *Water,* 12 (6), doi:1645. 10.3390/w12061645.