



**HAL**  
open science

## Which objective function to calibrate rainfall–runoff models for low-flow index simulations?

F. Garcia, N. Folton, Ludovic Oudin

► **To cite this version:**

F. Garcia, N. Folton, Ludovic Oudin. Which objective function to calibrate rainfall–runoff models for low-flow index simulations?. *Hydrological Sciences Journal*, 2017, 62 (7), pp.1149-1166. 10.1080/02626667.2017.1308511 . hal-01734883

**HAL Id: hal-01734883**

**<https://hal.science/hal-01734883>**

Submitted on 15 Mar 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Which objective function to calibrate rainfall–runoff models for low–flow index simulations?

Florine Garcia<sup>1</sup>, Nathalie Folton<sup>1</sup>, Ludovic Oudin<sup>2</sup>

(1) RH Research Unit, Irstea, Aix-en-Provence, France

(2) UMR 7619 METIS, Sorbonne Université, UPMC Université Paris 6, Paris, France

## Abstract

Much has been written on the subject of objective functions to calibrate rainfall–runoff models. Many studies focus on the best choice for low–flow simulations or different multi–objective purposes. Only a few studies, however, investigate objective functions to optimise the simulations of low–flow indices that are important for water management. Here, we test different objective functions, from single objective functions with different discharge transformations or using low–flow indices to combinations of single objective functions and we evaluate their robustness and sensitivity to the rainfall–runoff model. We found that the Kling and Gupta efficiency (KGE) applied to a transformation of discharge was inadequate to fulfil all assessment criteria, whereas the mean of the KGE applied to the discharge and the KGE applied to the inverse of the discharge was sufficient. The robustness was dependent on the climate variability rather than the objective function and the results were not sensitive to the model.

# 1 Introduction

Studying low flows is of paramount importance to manage water resources, for water uses such as irrigation, water supply, navigation and hydroelectricity (Engeland et Hisdal, 2009 ; Lang Delus, 2011). Low flows can be characterised by different indices (Smakhtin, 2001 ; WMO, 2008) such as the mean annual minimum discharge or a percentile from the flow duration curve. When sites are gauged for long record periods, these indices can easily be calculated from river discharge values. Other strategies, however, are needed in the context of spatial or temporal extrapolation studies for which discharge data are not available. In these cases, the use of a rainfall–runoff model, taking into account climate variability and which parameters can be regionalised is a possible answer. Using a rainfall–runoff model to simulate long discharge records in order to derive low–flow indices is an uncertain task and raises questions about (i) the type of model and (ii) the calibration procedure. While the choice of the model to be used is often driven by data availability and/or the expert knowledge of the model user, the calibration procedure might be tricky since simulating low flows is not a common usage of the rainfall–runoff model. This point is discussed in detail hereafter.

The calibration step can make use of a visual comparison between observed and simulated hydrographs, but this technique is subjective, dependent on expert judgement and difficult to reproduce (Chiew et McMahon, 1993 ; Houghton-Carr, 1999 ; Crochemore *et al.*, 2015). Since numerical criteria are more objective because they are reproducible and are more easily explained (Krause *et al.*, 2005), the calibration step usually involves choosing an optimisation algorithm, an objective function and one or several criteria to assess the model's goodness of fit. The criteria used to assess the model's goodness of fit are often used as objective functions. However, if the target variables are low–flow indices, this potentially leads to difficulties in finding a robust set of parameters. Consequently, objective functions based on the goodness of fit of hydrographs are often preferred.

The relevance of an objective function depends on the aim of the study or a simulation but also on the time step of hydrological modelling. Most existing objective functions based on least–square errors generally favour the goodness of fit of the hydrograph for high flows. For low–flow simulations, many authors have suggested alternative objective functions or alternative optimization strategies and the three main types of approaches are listed below.

- More often than not, it is suggested using common objective functions with an appropriate transformation applied to the discharge in order to put more emphasis on low-flow goodness of fit. Oudin *et al.* (2006) used a modified Nash and Sutcliffe efficiency (NSE) criterion applied to logarithmic transformed discharge and Pushpalatha *et al.* (2012) suggested using the inverse of the discharge as a criteria of efficiency. Another option consists in calculating model errors under a certain threshold discharge (Deckers *et al.*, 2010) or weighting the errors depending on their flow ranges (Krause *et al.*, 2005 ; Le Moine, 2008).
- Other studies (Westerberg *et al.*, 2011) use an objective function, based on the flow-duration curve (FDC). The rationale underlying this usage in the case of low-flow index simulations is that these indices are derived in the frequency domain and the temporal goodness of fit of the hydrographs is not a prerequisite of good low-flow index simulations (see e.g. the results obtained by Price *et al.*, 2012). Putting this idea further, Olsen *et al.* (2013) suggested calibrating the model using directly the flow indices that are targeted by the model user. They obtained interesting results on a set of 49 catchments in Denmark but here the choice of the low-flow indices used in the objective function can also be a thorny issue.
- Some authors advocate the use of multi-objective techniques. Multi-objective techniques provide the key advantage to possibly combine objective functions in a tailor-made framework. Multi-objective technique can be implemented in the theoretical framework of multi-objective optimization such as a Pareto front (Khu et Madsen, 2005 ; Fenicia *et al.*, 2007), or more simply by deriving aggregated (combined) objective function (Madsen, 2003 ; Merz et Blöschl, 2004 ; Oudin *et al.*, 2006 ; Booij et Krol, 2010 ; Nicolle *et al.*, 2014). Vis *et al.* (2015) assessed the ability of seven objective functions to simulate flow statistics at 27 catchments in the United States. They showed that combined objective functions generally gave the best results for flow statistics but the choice of the combined objective varied according to the desired flow statistic.

This study intends first to define an appropriate objective function for the simulations of low-flow indices at the scale of the French territory, but also the simulations of flow

seasonality and mean annual runoff. Second, we wish to evaluate the robustness of the calibrated set of parameters.

To meet these objectives, we will test a large panel of existing approaches and objective functions, including objective functions with different discharge transformations, objective functions based on the FDC, objective functions directly using low-flow indices and combined objective functions. The validation criteria to assess the relative merits of these objective functions are based on low-flow indices, seasonality and the mean annual runoff, most of them being used by national water management agencies.

The next section presents the data and the catchment set, the rainfall-runoff models used in this study and the methodology. Then the results are presented and discussed before the conclusions.

## 2 Study area, models and methodology

This section presents the dataset, the hydrological models used for this study and the methodology to test different objective functions.

### 2.1 Dataset

#### 2.1.1 Catchment set

A set of 691 catchments throughout France was collected for this study (see Figure 1 for the catchment locations). They were chosen following different criteria: the type of catchment, the availability of flow time series and the diversity of hydro–meteorological regimes.

Therefore, we first retained natural catchments with limited human influence, and therefore small in size (less than 8 000 km<sup>2</sup>). Then we intended to have long series of data from gauging stations, at least 25 years available between 1970 and 2013 for the cross–validation and the calculation of the low–flow indices. These data were deemed good by the station managers and a threshold of 10% missing data per year was allowed. Finally, we evaluated the diversity of the set so that it would be representative of all hydrological regimes in France (Sauquet *et al.*, 2008): pluvial (77% of the catchment set), Mediterranean (9%) and snowmelt–fed (14%) hydrological regimes.

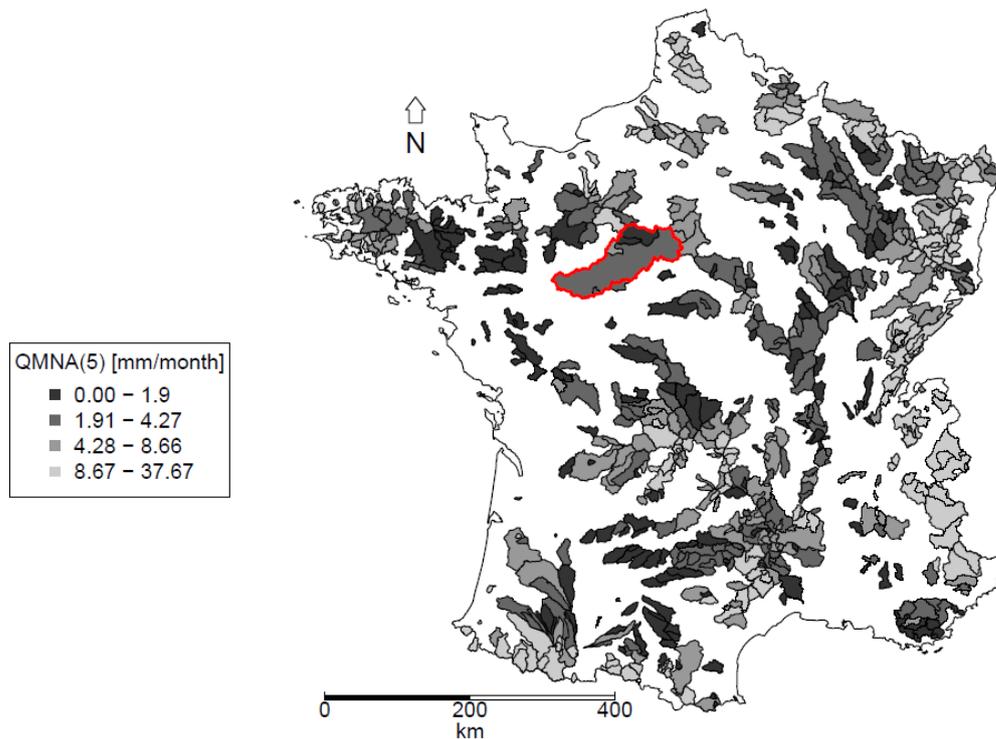


Figure 1: Spatial distribution of the QMNA(5) over the set of 691 French catchments. The red contour line delimits the catchment of the Loir River at Durtal.

### 2.1.2 Data

Meteorological data came from the distributed mesoscale atmospheric analysis system SAFRAN developed by Météo-France (Quintana-Seguí *et al.*, 2008 ; Vidal *et al.*, 2010), which provides daily solid and liquid precipitation and temperature data throughout France. Daily potential evapotranspiration (PE) was estimated using a temperature-based formula (Oudin *et al.*, 2005). These data were available on the 1959–2013 period over France.

Daily discharge data came from the French Hydro database ([www.hydro.eaufrance.fr](http://www.hydro.eaufrance.fr)). For the catchment set of interest in this study, the length of the discharge records for each catchment varied but they were generally available for the 1970–2013 period, which includes very humid years (e.g. 1977) and very dry years (e.g. summers 1976, 1989, 2003 and 2005).

Table 1 summarises the hydro-meteorological diversity of our set and some catchment characteristics.

Table 1: Percentiles of the distributions of climatic and catchment characteristics on the set of 691 study catchments.

	Min	25%	Median	75%	Max
Mean annual runoff (mm/year)	44	256	390	609	2 211
Mean annual precipitations (mm/year)	634	867	988	1 181	2 094
Mean annual PE (mm/year)	237	623	660	703	864
Catchment area (km <sup>2</sup> )	4	98	211	478	7 935
Mean elevation (m a.s.l.)	29	188	362	743	2 866
Base-Flow Index (–)	0.12	0.47	0.57	0.67	0.98

### 2.1.3 Low-flow indices

Low flows in France mostly occur in summer and at the beginning of autumn, due to a lack of precipitation and an increased evaporative demand, except over the mountainous area where they occur in winter. A multitude of low-flow indices can be estimated (Smakhtin, 2001 ; Laaha et Blöschl, 2006 ; WMO, 2008 ; Lang Delus, 2011). Since our focus was on low-flow severity, at different time steps, and low-flow indices used by water management agencies, the following indices were selected: Q95 and Q75, MAM3, MAM10 and QMNA at the recurrence interval of 5 years (called MAM3(5), MAM10(5) and QMNA(5)).

Q95 and Q75 are derived from the flow-duration curve (Vogel et Fennessey, 1994 ; 1995), calculated over all the available years. Q95 is the discharge exceeded 95% of the time and Q75, the discharge exceeded 75% of the time.

MAM3 and MAM10 are the mean annual 3- and 10-day minimum discharge, respectively, and QMNA is the annual minimum monthly discharge. The latter is widely used in France for low-flow management and drought management plans.

Since these three indices were estimated for each year, we calculated the 5-year return period of these three indices. Therefore, for each catchment, the distribution of each index was fitted to a log-normal distribution. The parameters of the log-normal distribution were estimated by the maximum likelihood method since Catalogne (2012) showed that this distribution and this estimation method are the most appropriate for low flows in France. For some catchments, the rivers dry up for several years; a conditional probability model is used in this case (Stedinger *et al.*, 1993). In Figure 1, the different catchments are represented with their QMNA(5) value and Table 2 summarises the diversity of these low-flow indices for the dataset.

Table 2: Percentiles of the distribution of low-flow characteristics on the set of 691 study catchments.

	Min	25%	Median	75%	Max
Q95 (mm/d)	0	0.06	0.13	0.26	1.32
Q75 (mm/d)	0	0.15	0.29	0.51	2.14
MAM3(5) (mm/d)	0	0.03	0.09	0.18	1.14
MAM10(5) (mm/d)	0	0.04	0.1	0.21	1.16
QMNA(5) (mm/month)	0	1.83	3.96	8.22	37.67

## 2.2 Hydrological models

The rainfall-runoff model used for this study was GR4J (Perrin *et al.*, 2003), a continuous lumped model with four parameters. This model has been used in many studies, especially in France for management and operational purposes. Figure 2 shows the structure of the model. Effective rainfall and actual evapotranspiration are calculated as functions of the soil moisture store level ( $S$ ), the net rainfall ( $P-PE$ ) and the parameter  $X_1$  (mm), the maximum capacity of the soil moisture store. The percolation of the store is also a function of the store filling rate.

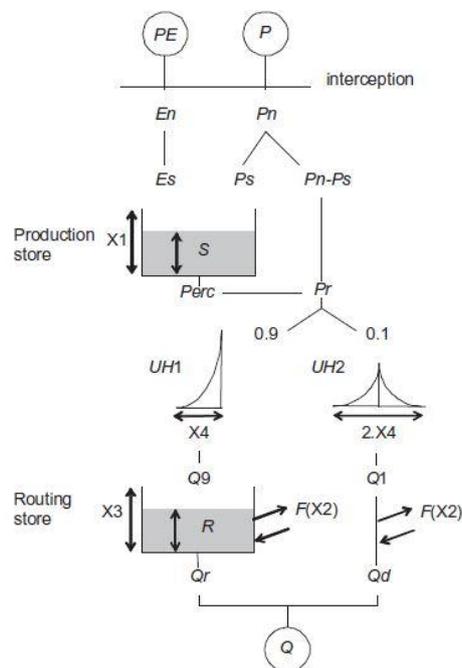


Figure 2: Schematic diagram of the GR4J model structure (PE: potential evapotranspiration, P: rainfall, Q: streamflow,  $X_i$ : parameter i, other notations correspond to internal state variables).

Inter-basin groundwater flows are controlled by a second parameter ( $X_2$ , mm/d). If  $X_2$  is positive, the catchment gains water and if  $X_2$  is negative, there is a water loss.

Effective rainfall is then divided into two flow components in the routine function: 90% are routed by a unit hydrograph, whose time-base parameter is  $X_4$  (d), and a nonlinear routing store, and the remaining 10% are routed through a unit hydrograph. The last parameter,  $X_3$  (mm), is the maximum level of the routing store. The simulated discharge is the sum of these two components.

Since our catchment set includes mountainous catchments, GR4J was used with a two-parameter snow module called CemaNeige (Valéry, 2010 ; Valéry *et al.*, 2014), using the regionalised parameters for each French mountain massif that were established by Brigode *et al.* (2014).

To assess whether the results depend on the model structure, a six-parameter version of TOPMODEL (Beven et Kirkby, 1979), named TOPMO, was also used. This version differs from the original TOPMODEL structure since it is a lumped version of the model with a parameterized expression of the soil-topographic distribution (see Edijatno *et al.*, 1999 and Michel *et al.*, 2003 for more details). This model was also used with the snow module CemaNeige.

## **2.3 Calibration and validation methodologies**

### **2.3.1 Cross-validation and calibration algorithm**

The calibration algorithm used for this study is a two-step search procedure. First, the parameter space is screened and then a local search algorithm is performed. This approach will not be discussed here but has been proved efficient for parsimonious models like GR4J (Edijatno *et al.*, 1999 ; Mathevet, 2005).

The evaluation of the different objective functions was based on classical split-sample-test scheme (Klemeš, 1986). The discharge records were divided into two independent sub-periods of equivalent lengths (P1: 1970–1991 and P2: 1992–2013). We first calibrated the model parameters on P1 and validated it on P2, then exchanged the two sub-periods, i.e. we calibrated on P2 and validated it on P1. Five years before each period (1965–1969 for P1 and 1987–1991 for P2) were used to initialise the model and to avoid any problem with

catchments with long-term memory. Then we evaluated the model's performance on the two validation periods.

### 2.3.2 Assessment criteria

The objective was to evaluate the quality of the simulation of low-flow indices, mean annual runoff and seasonality. The low-flow indices selected were Q95, Q75, MAM3(5), MAM10(5) and QMNA(5). We obtained one observed and one simulated value of these five indices and also for mean annual runoff (MAR) for each catchment and for each simulation. To obtain a general assessment of each simulation over the entire set of catchments, we calculated for each index a spatial criteria based on the NSE as a transposition of the NSE from a temporal to a spatial analysis, called "Spatial Index Efficiency" (SIE):

$$SIE = 1 - \frac{\sum_{i=1}^n (FI_i^{obs} - FI_i^{sim})^2}{\sum_{i=1}^n (FI_i^{obs} - m_{obs})^2} \quad (1)$$

where  $FI_i^{obs}$  and  $FI_i^{sim}$  are the observed and simulated indices for the catchment  $i$ ,  $n$  the number of catchments (here 691) and  $m_{obs}$  the mean of the observed indices. A SIE close to unity means that the indices are well simulated over France.

To assess the goodness of fit of flow seasonality, we computed a NSE (Nash et Sutcliffe, 1970) over the 12 observed and simulated mean monthly discharges and then we calculated the median of the 691 NSE values. To assess the impact of the different objective functions on the continuous simulations, we computed a bounded-version of the NSE (Mathevet *et al.*, 2006), called NSE\*:

$$NSE^* = \frac{NSE}{2 - NSE} \quad (2)$$

$$\text{with } NSE = 1 - \frac{\sum_{i=1}^M (Q_i^{obs} - Q_i^{sim})^2}{\sum_{i=1}^M (Q_i^{obs} - \mu_{obs})^2} \quad (3)$$

where  $\mu_{obs}$  is the mean of the observed and simulated discharge series.

We computed this criterion on the discharge, the square-root of the discharge and the inverse of the discharge and then calculated the median of each distribution. This provides assessment criteria on seven indices and three continuous simulations to evaluate each

simulation; radial plots were used to present the different results and ease the comparisons. The larger the polygon is, the better the simulations are.

Table 3 summarises the list of criteria for the test evaluation.

Table 3: Selection of the assessment criteria.

Name	Description
Low-flow indices	
SIE(Q95)	SIE calculated on Q95, discharge exceeded 95% of the time
SIE(MAM3(5))	SIE calculated on MAM3(5), mean annual 3-day minimum at T = 5 years
SIE(MAM10(5))	SIE calculated on MAM10(5), mean annual 10-day minimum at T = 5 years
SIE(QMNA(5))	SIE calculated on QMNA(5), annual minimum monthly flow at T = 5 years
SIE(Q75)	SIE calculated on Q75, discharge exceeded 75% of the time
Seasonality	
q <sub>50</sub> (NSE(QMM))	Median of the distribution of the 691 NSE on the mean monthly flows (QMM)
Mean annual runoff	
SIE(MAR)	SIE calculated on MAR, mean annual runoff
Continuous simulations	
q <sub>50</sub> (NSE*(Q))	Bounded version of the NSE calculated on discharges
q <sub>50</sub> (NSE*(Q <sup>0.5</sup> ))	Bounded version of the NSE calculated on root-squared transformed discharges
q <sub>50</sub> (NSE*(1/Q))	Bounded version of the NSE calculated on inverse transformed discharges

### 2.3.3 Objective functions analysed

First, this study evaluated the impact of choosing the assessment criteria on the seven indices as an objective function. To this aim, an objective function derived from the RMSE was computed on the seven assessment criteria and used as the objective function. This objective function is noted OF<sub>AC</sub> hereafter:

$$OF_{AC} = \frac{1}{N} \sum_{k=1}^N \left( 1 - \frac{AC_k^{sim} + \varepsilon}{AC_k^{obs} + \varepsilon} \right)^2 \quad (4)$$

where  $AC_k^{obs}$  and  $AC_k^{sim}$  are the observed and simulated values of the assessment criterion  $k$ , respectively,  $N$  is the number of assessment criteria and  $\varepsilon$  is a small quantity in case  $AC_k^{obs}$  equals zero (taken equal to one one-hundredth of the median value over the catchment set). The assessment criteria are the five low-flow indices, Q95, MAM3(5), MAM10(5), QMNA(5) and Q75, the twelve mean monthly flows, QMM, and the mean annual discharge, MAR. All indices are calculated in mm/d. An OF<sub>AC</sub> close to zero means that all assessment criteria are well simulated for the catchment.

Then, this study focused on the Kling and Gupta efficiency, KGE (Gupta *et al.*, 2009), since it has been analysed and recommended in many studies (Lobligeois, 2014 ; Magand, 2014 ; Osuch *et al.*, 2015). KGE is given by:

$$\text{KGE} = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (5)$$

$$\text{with} \begin{cases} r = \frac{1}{M} \sum_{j=1}^M \frac{(Q_j^{\text{obs}} - \mu_{\text{obs}})(Q_j^{\text{sim}} - \mu_{\text{sim}})}{\sigma_{\text{obs}}\sigma_{\text{sim}}} \\ \alpha = \frac{\sigma_{\text{sim}}}{\sigma_{\text{obs}}} \\ \beta = \frac{\mu_{\text{sim}}}{\mu_{\text{obs}}} \end{cases} \quad (6)$$

where  $r$  is the Pearson product–moment correlation coefficient,  $\mu$  and  $\sigma$  are the mean and the standard deviation of the observed and simulated discharge series, respectively,  $\alpha$  is the ratio between the simulated and observed standard deviation values and  $\beta$  is the ratio between the simulated and observed mean values. The objective was to tend towards the ideal value of  $r$ ,  $\alpha$  and  $\beta$ , which is 1, so that the Euclidean distance tends towards 0 and the KGE criterion is maximised.

The first part of this study evaluated if a single objective function provided fair results to simulate low flow, mean annual runoff and seasonality. Since the simplest objective function that can be used for all catchments was sought, we decided to evaluate the KGE with discharge transformations, choosing the following:

- KGE(Q), which puts more weight on high flow,
- KGE(Q<sup>0.5</sup>), which does not favour high flow nor low flow,
- KGE(1/Q), which puts more weight on low flow.

For the third discharge transformation, to avoid any problems with zero discharge, a small constant,  $\varepsilon$ , was added, which was one one–hundredth of the mean observed discharge as advised by Pushpalatha *et al.* (2012). Since the objective here was primarily to simulate low–flow indices, we chose to evaluate the KGE on the FDC, with the same discharge transformations, leading to the three following objective functions:

- KGE(Q<sub>sort</sub>), which puts more weight on high flow,
- KGE(Q<sub>sort</sub><sup>0.5</sup>), which does not favour high flow nor low flow,
- KGE(1/Q<sub>sort</sub>), which puts more weight on low flow.

Table 4 summarises these six objective functions and their specificities.

Table 4: Selection of the objective functions tested.

	Criterion	Names	Weights on
Temporal criteria	KGE(Q)	KGE calculated on discharges	High flows
	KGE(Q <sup>0.5</sup> )	KGE calculated on root-squared transformed discharges	Neither high nor low flows
	KGE(1/Q)	KGE calculated on inverse transformed discharges	Low flows
Frequency criteria	KGE(Q <sub>sort</sub> )	KGE calculated on the FDC	High flows
	KGE(Q <sub>sort</sub> <sup>0.5</sup> )	KGE calculated on the root-squared transformation of the FDC	Neither high nor low flows
	KGE(1/Q <sub>sort</sub> )	KGE calculated on the inverse transformation of the FDC	Low flows

The second part evaluated all the different combinations of two of these functions. The combination was the mean of these two functions and the different weights that can be applied between the two functions were evaluated.

#### 2.3.4 Temporal robustness assessment

For this study, we evaluated the temporal robustness of the calibrated set of the model parameters. The robustness of the parameter sets was analysed by dividing the discharge records into two different periods: one with the 22 driest years and one with the 22 wettest years. To split these 44 years, an annual aridity index (Budyko, 1974) was calculated and the 22 lowest values of this aridity index corresponded to the 22 driest years and the 22 highest values to the 22 wettest years. Then we calibrated the model parameters on the dry period and validated it on the wet period and then exchanged the two sub-periods.

### 3 Results and discussions

This section first presents the results obtained with assessment criteria used as objective functions, then those obtained using a single objective function and in the third stage the results obtained when combining two single objective functions. Last, the general reach of the results is analysed with the best single objective function and the best combination by analysing (i) their temporal robustness and (ii) their sensitivity to the model structure.

#### 3.1 Assessment criteria as objective function

Figure 3 shows the radial plot with the results in calibration and in validation when using the seven indices within the objective function ( $OF_{AC}$ ). We observed that the seven indices are quite well simulated in calibration periods with SIE values between 0.95 for the simulation of the Q75 and 0.98 for the simulation of the MAM10(5). However, we observed a drop for the simulations of these seven indices in validation, with SIE values around 0.7. Using the seven assessment indices as an objective function seems to simulate indices that are not robust. These results differ from the ones obtained by Olsen *et al.* (2013). However, these authors used an objective function combining low-flow and high-flow indices, which can be more related to an objective function based on the FDC and lead to more robust results. Looking at the continuous simulations, we observed that this objective function did not provide good continuous simulations since it did not take into account the temporal goodness of fit of the hydrographs.

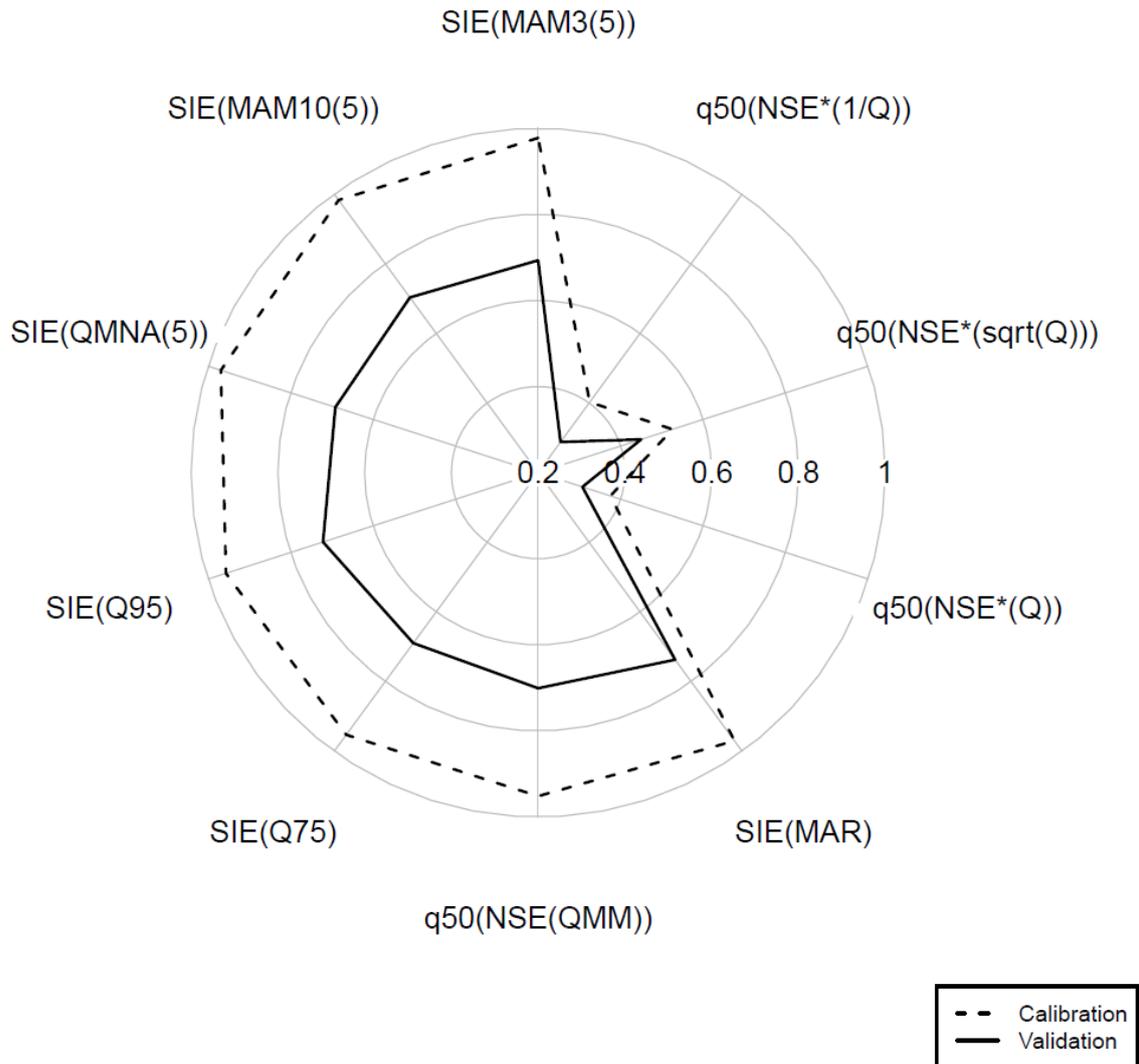


Figure 3: Radial plot showing the mean results in calibration and in validation for the two periods for the seven assessment criteria used as an objective function.

### 3.2 Single objective function

Figure 4 illustrates the impact of the choice of the objective function for one catchment, the Loir River at Durtal (7 920 km<sup>2</sup>, located by the red lines in Figure 1) in the validation period 1970–1991. Figure 4 (a) shows the simulations of seasonality with the six single objective functions for one period in validation and for this catchment. The lowest discharges, in summer, were better simulated with the KGE(1/Q) function, whereas they were underestimated with the other functions, especially with those applied to the discharge. In contrast, the highest discharges, in winter for this catchment, were better simulated with KGE(Q<sup>0.5</sup>), but they were underestimated with the functions applied to the inverse discharge and overestimated with the function applied to the discharge. Figure 4 (b) shows the

simulation of the FDC, also for one period in validation. Interestingly,  $KGE(1/Q)$  and  $KGE(1/Q_{\text{sort}})$  allowed better simulations of the part of the curve depicting the low flows, whereas the part depicting the high flows was better simulated with  $KGE(Q)$  and  $KGE(Q_{\text{sort}})$ . For the simulation of the FDC, the three objective functions based on the sorted discharge gave better results than the temporal functions. Figure 4 (c), (d) and (e) shows the simulations of, respectively, MAM3, MAM10 and QMNA. These three indices were better simulated with  $KGE(1/Q)$  and  $KGE(1/Q_{\text{sort}})$ , whereas they were underestimated by the other four objective functions. Figure 4 (f) shows the simulation of the annual runoff, AR. This index was better simulated with the two objective functions based on the discharge or the square root of the discharge, whereas those based on the inverse of the discharge showed a tendency to underestimate the AR.

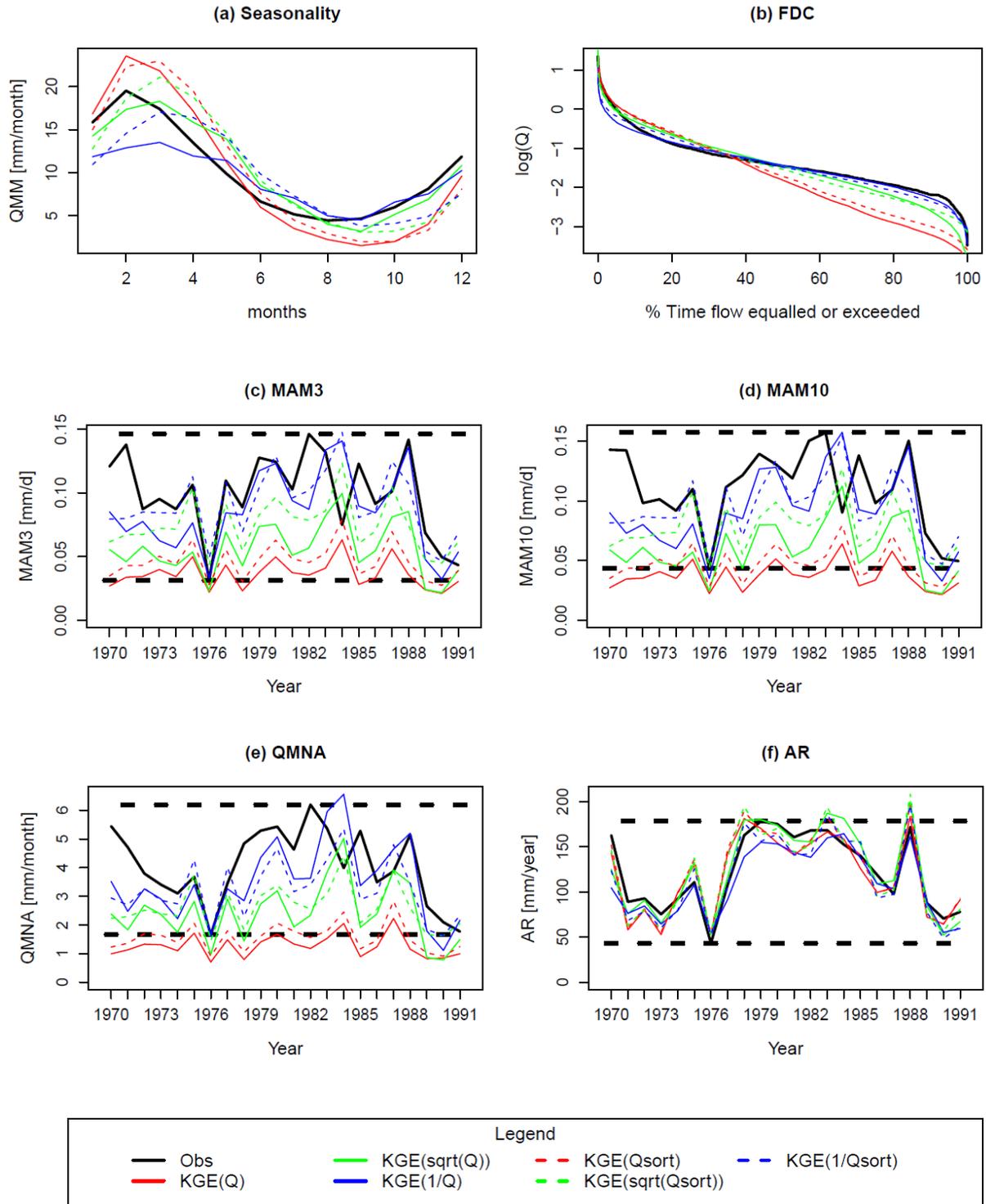


Figure 4: Representation of seasonality (a), the flow duration curve (b), the annual MAM3 (c), the annual MAM10 (d), the annual QMNA (e) and the annual runoff (f) observed (in black) and simulated with the six single objective functions (in colours) in the validation period 1970–1991 for the Loir River at Durtal.

Figure 5 shows the radial plot with the results in validation for the simulations with the six single objective functions. As for the example of the Loir River at Durtal, the two functions applied to the inverse discharge,  $KGE(1/Q)$  and  $KGE(1/Q_{\text{sort}})$ , provided better simulations of the four low-flow indices with SIE values around 0.9, whereas the two functions applied to the discharge,  $KGE(Q)$  and  $KGE(Q_{\text{sort}})$ , gave relatively poor results with SIE values between 0.55 and 0.65. The opposite is observed for the simulation of mean annual runoff and flow seasonality. The two functions applied to the square-root of the discharge,  $KGE(Q^{0.5})$  and  $KGE(Q_{\text{sort}}^{0.5})$ , provided good simulations of mean annual runoff and seasonality and reasonable simulations of the low-flow indices. These results confirmed that  $KGE(Q)$  and  $KGE(Q_{\text{sort}})$  put more weight on high flows,  $KGE(1/Q)$  and  $KGE(1/Q_{\text{sort}})$  on low flows, whereas  $KGE(Q^{0.5})$  and  $KGE(Q_{\text{sort}}^{0.5})$  weighted neither type of flow. Looking at the continuous simulations, we observe the same trend. The objective functions with no transformation or the square-root transformation gave good results for the high and intermediate flows, whereas the inverse transformation gave better results for the low flows. As for the Loir River at Durtal study, only a few differences were observed between objective functions in the temporal and frequency domains. However, these differences increased for the continuous simulations since the three single objective functions in the frequency domain did not take into account the temporality of the hydrographs. Since the first goal was the simulation of low-flow indices, the functions on the inverse discharge seemed to provide better results, but the simulations of mean annual runoff and seasonality were overly reduced. Therefore, the function applied to the square-roots seemed to provide the best results. Since there is clearly a trade-off between low-flow indices, mean annual discharge and flow seasonality, the next section tests various combinations of the objective functions tested.

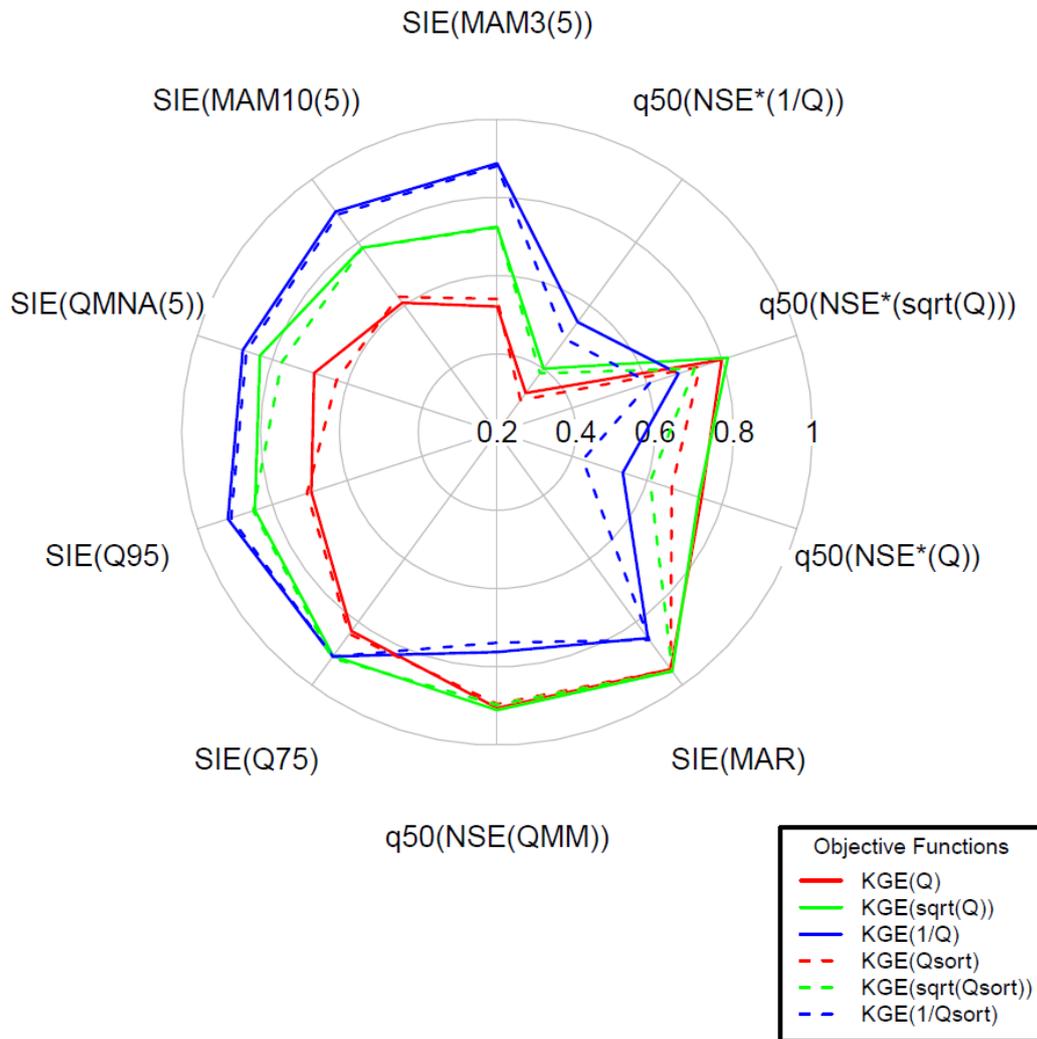


Figure 5: Radial plot showing the mean results in validation for the two periods for the six single objective functions.

### 3.3 Combination of two single objective functions

Figure 6 shows the simulation of seasonality, FDC, MAM3, MAM10, QMNA and AR in the validation period 1970–1991 for the Loir River at Durtal, for the parameters calibrated with  $KGE(Q^{0.5})$ , the combination of  $KGE(Q)$  and  $KGE(Q_{sort})$ , and the combination of  $KGE(Q)$  and  $KGE(1/Q)$ . In all the sub-figures except the last one, the combination between  $KGE(Q)$  and  $KGE(1/Q)$  provided good simulations of all these indices, whereas the other combination did not improve the results. For the simulation of AR, we observed a tendency to overestimate this index with the combination of  $KGE(Q)$  and  $KGE(1/Q)$ .

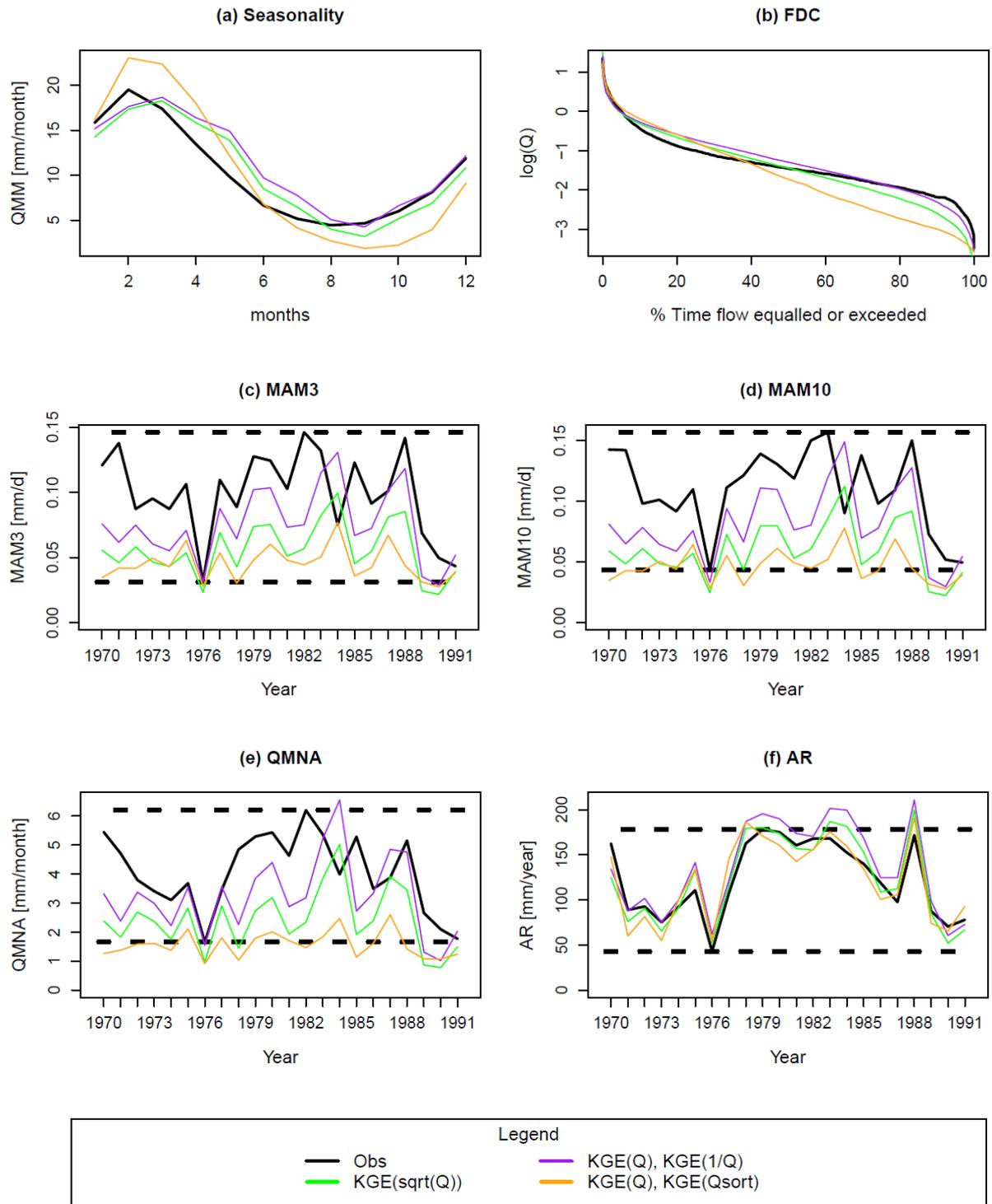


Figure 6: Representation of seasonality (a), the flow duration curve (b), the annual MAM3 (c), the annual MAM10 (d), the annual QMNA (e) and the annual runoff (f) observed (in black) and simulated with the three objective functions, KGE(sqrt(Q)) and two combinations (in colours) in the validation period 1970–1991 for the Loir River at Durtal.

All the possible combinations between two single objective functions from the pool of objective functions assessed in the preceding section were tested and the results for the seven assessment criteria are shown in Figure 7. The complementarity of the objective functions

taken into account demonstrates good efficiency for the several assessment criteria. In this context, mixing an objective function applied to the inverse of the discharge with an objective function applied to the discharge provided the best results, as shown by the purple curve for the simulation with  $KGE(Q)$  and  $KGE(1/Q)$ . Conversely, mixing two objective functions on similar discharge transformation provided the worst results overall, such as the orange curve on the graphs, which draws the results for the simulation with the combination of  $KGE(Q)$  and  $KGE(Q_{sort})$ . The same results were observed for the continuous simulations.

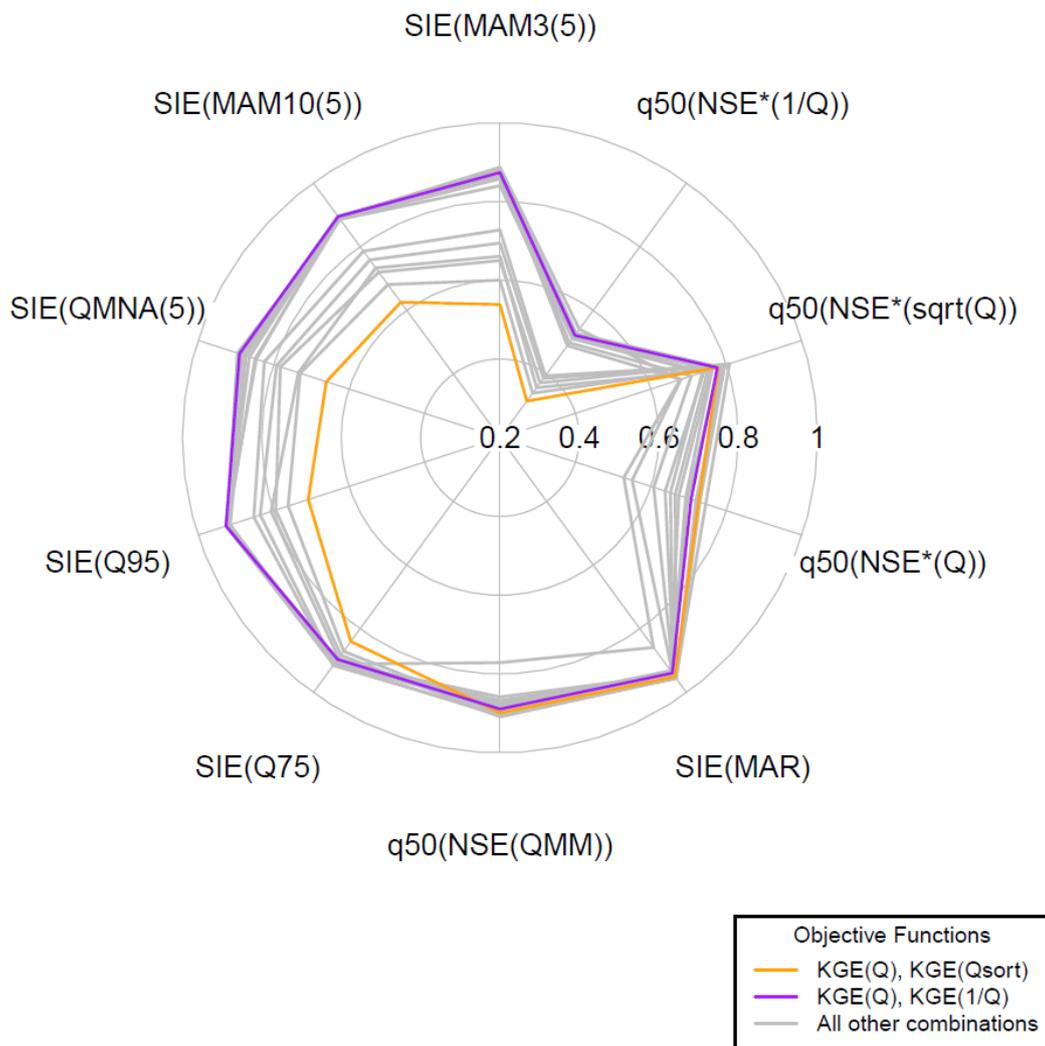


Figure 7: Radial plot showing the mean results in validation for the two periods for all the combinations of two single objective functions.

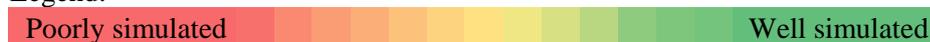
Table 5 summarises the values of the seven assessment criteria and the sum of these seven criteria for each objective function. As shown in Figure 7, combining with an objective function applied to the inverse discharge gave good results for the simulation of low-flow indices. The simulations of seasonality and mean annual runoff were improved with these

combinations, compared to the results when calibrating only with  $KGE(1/Q)$  or  $KGE(1/Q_{\text{sort}})$ . Combining with the objective functions applied to the discharge or the square root of the discharge gave better results for the simulation of seasonality and mean annual runoff. Looking at the sum of the values of the seven assessment criteria, three objective functions gave the best performance: combining  $KGE(1/Q)$  with  $KGE(Q^{0.5})$ ,  $KGE(Q)$  or  $KGE(Q_{\text{sort}}^{0.5})$ . Therefore, we chose the objective function that combines  $KGE(Q)$ , an objective function with no transformation that puts more weight on high flows with  $KGE(1/Q)$ , an objective function that puts more weight on low flows. This function has already been used in a study to calibrate the parameter set of the rainfall–runoff model GR6J conducted by Nicolle *et al.* (2014).

Table 5: Simulations of the seven assessment criteria in validation for the two periods for all objective functions.

KGE on	SIE on MAM3(5)	SIE on MAM10(5)	SIE on QMNA(5)	SIE on Q95	SIE on Q75	$q_{50}$ of NSE(QMM)	SIE on MAR	Total
$Q^{0.5}, 1/Q$	0.877	0.895	0.886	0.924	0.896	0.894	0.937	6.309
<b>Q, 1/Q</b>	0.873	0.894	0.890	0.926	0.896	0.889	0.938	6.305
$1/Q, Q_{\text{sort}}^{0.5}$	0.865	0.893	0.882	0.926	0.909	0.884	0.945	6.305
$Q^{0.5}, 1/Q_{\text{sort}}$	0.879	0.893	0.874	0.920	0.901	0.880	0.931	6.279
$1/Q, Q_{\text{sort}}$	0.840	0.885	0.895	0.924	0.912	0.873	0.944	6.272
$Q, 1/Q_{\text{sort}}$	0.877	0.892	0.864	0.922	0.899	0.877	0.936	6.267
$Q_{\text{sort}}^{0.5}, 1/Q_{\text{sort}}$	0.876	0.894	0.845	0.925	0.905	0.867	0.940	6.252
$Q_{\text{sort}}, 1/Q_{\text{sort}}$	0.857	0.886	0.849	0.921	0.904	0.858	0.943	6.219
$1/Q, 1/Q_{\text{sort}}$	0.887	0.892	0.873	0.915	0.910	0.771	0.858	6.106
$1/Q$	0.887	0.896	0.878	0.918	0.908	0.762	0.850	6.099
$1/Q_{\text{sort}}$	0.880	0.887	0.868	0.909	0.909	0.738	0.856	6.047
$Q^{0.5}, Q_{\text{sort}}^{0.5}$	0.727	0.785	0.824	0.852	0.915	0.902	0.955	5.961
<b><math>Q^{0.5}</math></b>	0.725	0.782	0.832	0.846	0.907	0.910	0.955	5.958
$Q_{\text{sort}}^{0.5}$	0.723	0.782	0.775	0.850	0.913	0.897	0.955	5.894
$Q, Q_{\text{sort}}^{0.5}$	0.694	0.758	0.792	0.835	0.910	0.906	0.955	5.850
$Q, Q^{0.5}$	0.650	0.719	0.782	0.794	0.882	0.909	0.954	5.690
$Q_{\text{sort}}, Q_{\text{sort}}^{0.5}$	0.661	0.733	0.734	0.806	0.892	0.893	0.954	5.672
$Q^{0.5}, Q_{\text{sort}}$	0.600	0.681	0.729	0.761	0.870	0.900	0.954	5.494
$Q, Q_{\text{sort}}$	0.538	0.625	0.660	0.708	0.839	0.899	0.950	5.218
$Q$	0.521	0.609	0.688	0.695	0.828	0.905	0.949	5.194
$Q_{\text{sort}}$	0.540	0.627	0.628	0.707	0.837	0.895	0.949	5.184

Legend:



### **3.4 Comparison between $KGE(Q^{0.5})$ and the combination of $KGE(Q)$ with $KGE(1/Q)$**

Figure 8 compares the distribution of the relative errors or the NSE when the parameter sets are calibrated with  $KGE(Q^{0.5})$  or the combination of  $KGE(Q)$  with  $KGE(1/Q)$ . Figure 8 (a), (b), (c), (d) and (e) shows the results for the relative errors of the five low-flow indices, respectively, MAM3(5), MAM10(5), QMNA(5), Q95 and Q75. For some catchments, the relative error values are very high, up to factors 10 to 20. These catchments are intermittent streams with observed low-flow indices very close to zero. When the parameters are calibrated with  $KGE(Q^{0.5})$ , these indices are overestimated for many basins. This trend decreases with indices that characterised less severe low flows, such as Q75. Figure 8 (f) compares the relative errors on the MAR. No huge trend is observed, but the results are underestimated for some basins with the combination of  $KGE(Q)$  and  $KGE(1/Q)$ . Figure 8 (g) compares the simulations of seasonality. For some basins, seasonality is better simulated when the parameter sets are calibrated with  $KGE(Q^{0.5})$ .

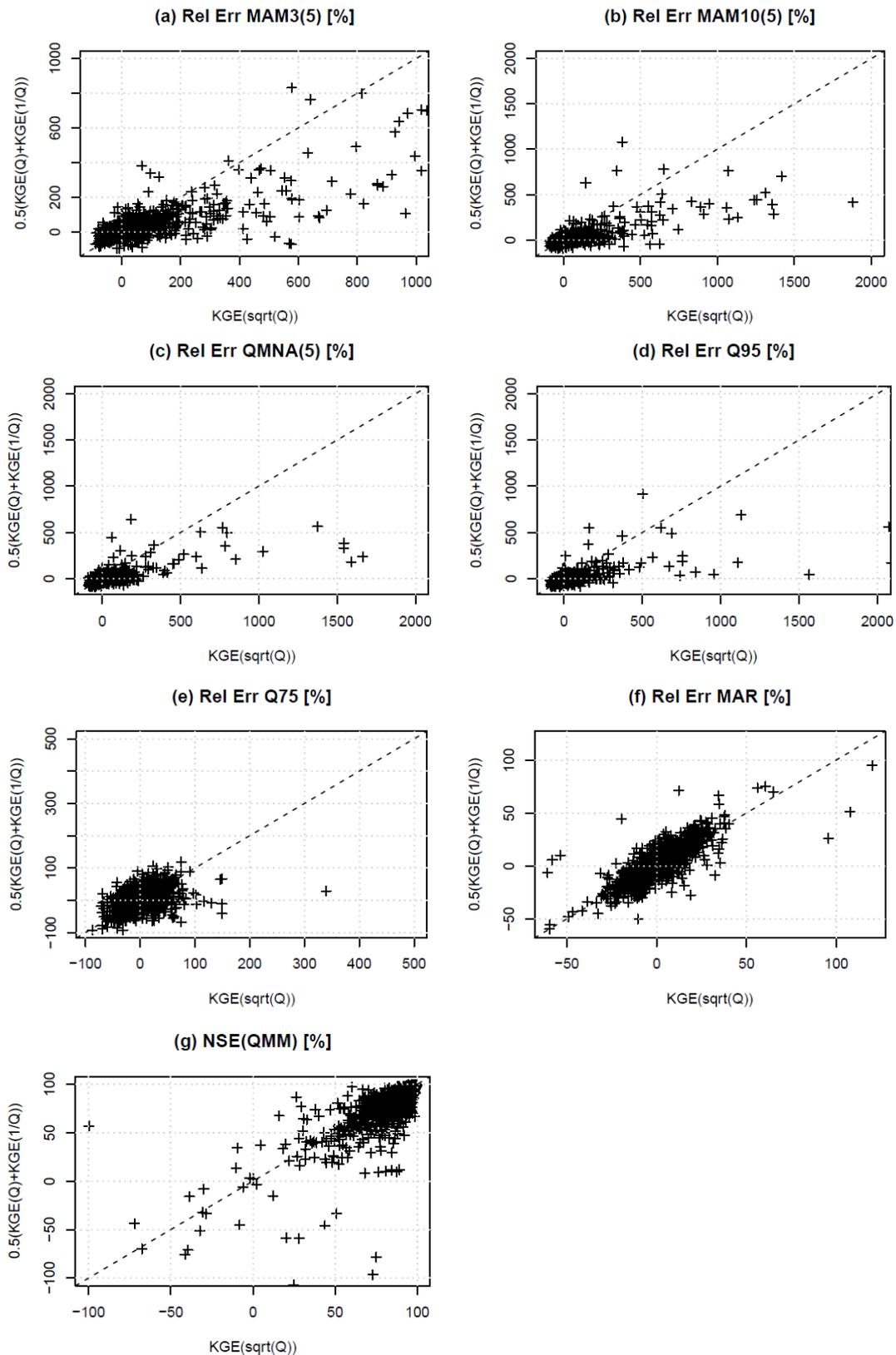


Figure 8: Scatter plot of relative errors on MAM3(5) (a), MAM10(5) (b), QMNA(5) (c), Q95 (d), Q75 (e) and MAR (f), and the NSE(QMM) (g) obtained in validation by  $KGE(\sqrt{Q})$  and the combination of  $KGE(Q)$  with  $KGE(1/Q)$ .

Figure 9 compares the parameters that are calibrated by each objective function. The values of the parameters X1 (capacity of production store) and X2 (groundwater exchange coefficient) were higher when calibrated with the objective function combining KGE(Q) with KGE(1/Q) than when calibrated by KGE(Q<sup>0.5</sup>), resulting in greater losses from evapotranspiration and lower underground losses. This trend was not observed for the other two parameters. Whatever the two objective functions, we observed that some values of the parameter X4 are outliers. These points represented catchments in the Seine–Normandy region, for which BFI values are high, between 0.96 and 0.99. Their hydrographs are very influenced by groundwater runoff and are very difficult to simulate with GR4J.

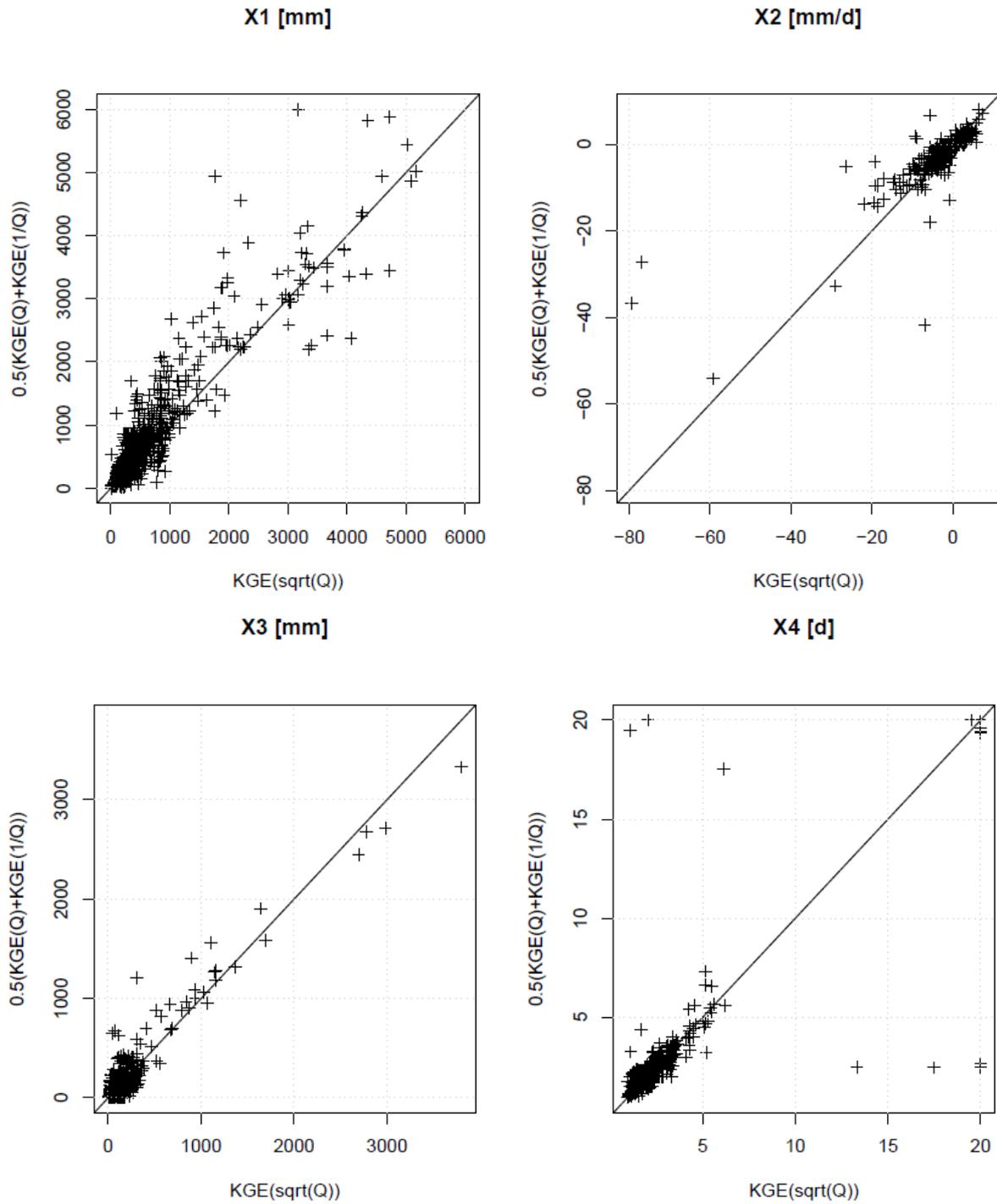


Figure 9: Comparison of the GR4J parameters calibrated by the objective function  $KGE(\sqrt{Q})$  and the parameters calibrated with the objective function combining  $KGE(Q)$  with  $KGE(1/Q)$ .

### **3.5 Weights in the combination of KGE(Q) with KGE(1/Q)**

The results in section 3.3 indicate that the combination of KGE(Q) with KGE(1/Q) improved the calibration of the parameter set to estimate low-flow indices, seasonality and the mean annual runoff. This combination was chosen as the mean between the two single objective functions, putting the same weight, 50%, on the two functions. We evaluated the impact of the weights on the results. Therefore, different weights were tested, from 10% to 90%. Figure 10 shows the radial plot results for the different weights in validation. Interestingly, the functions that gave substantial weight to KGE(Q), 90% or 80%, did not provide good simulations of the low-flow indices. For example, when the weight was 90%, the SIE on the MAM3(5) was 0.7, whereas with the other functions, it was around 0.9. When putting 90% on KGE(1/Q), the quality of the seasonality simulations and the mean annual runoff decreased slightly. For all the other weights, the quality of the simulations was nearly the same.

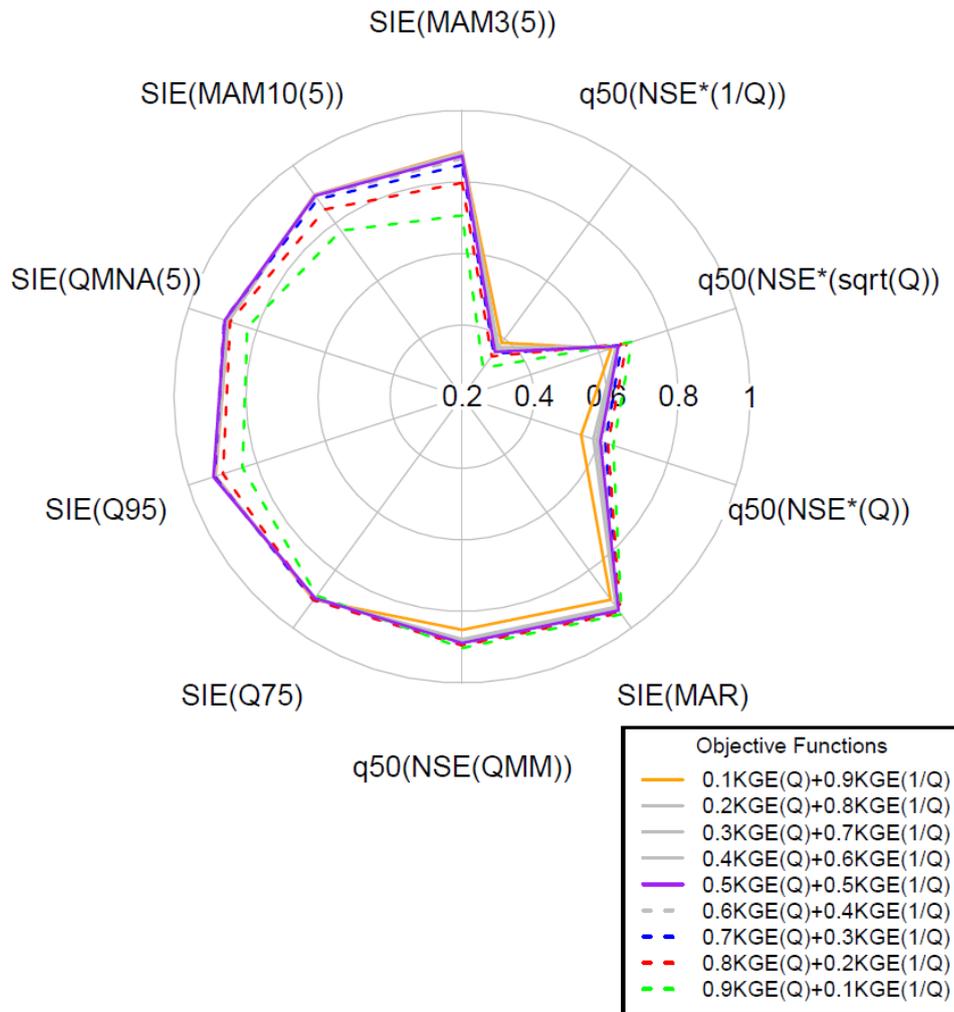


Figure 10: Evaluation of the weights in the combination of KGE(Q) and KGE(1/Q). Radial plot showing the results in validation for the two periods.

Based on the  $OF_{AC}$  calculated on the seven assessment criteria, we determined the optimal weights for each catchment. For 461 catchments, there was not one optimal weight but all weights between 20% and 70% for KGE(Q) gave exactly the same calibrated parameter values and therefore the same simulations. For the other 230 catchments, we were able to find one optimal weight between 10% and 90%, depending on the catchment. We compared these optimal weights with different low-flow indices, the BFI, the hydrological regimes of each catchment but we were not able to provide a priori descriptors to determine these weights. These optimal weights could also be related to poor quality in low-flow data leading to parameter values that compensate these uncertainties. However, the data quality for the selected catchments were deemed good by the water agency managers during low flows, intermediate flows and high flows and it is not easy to find indicators related to the quality of

low-flow data. Therefore, we recommend using the mean and thus putting the same weight on  $KGE(Q)$  and  $KGE(1/Q)$ .

### **3.6 Temporal robustness**

We first evaluated the temporal robustness of the model parameters calibrated by  $KGE(Q^{0.5})$  and by the combination of  $KGE(Q)$  with  $KGE(1/Q)$ . Figure 11 shows the results with the simulation in calibration, for the wet and dry periods, and in validation for these two periods. The loss in performance between calibration and validation for all seven assessment criteria was a little greater with the combined objective functions – 0.01 for MAR and seasonality and 0.07 for the low-flow indices – than with  $KGE(Q^{0.5})$  – between 0.01 and 0.05. However, MAM3(5), MAM10(5), QMNA(5) and Q95 were better simulated in validation with the combined objective function.

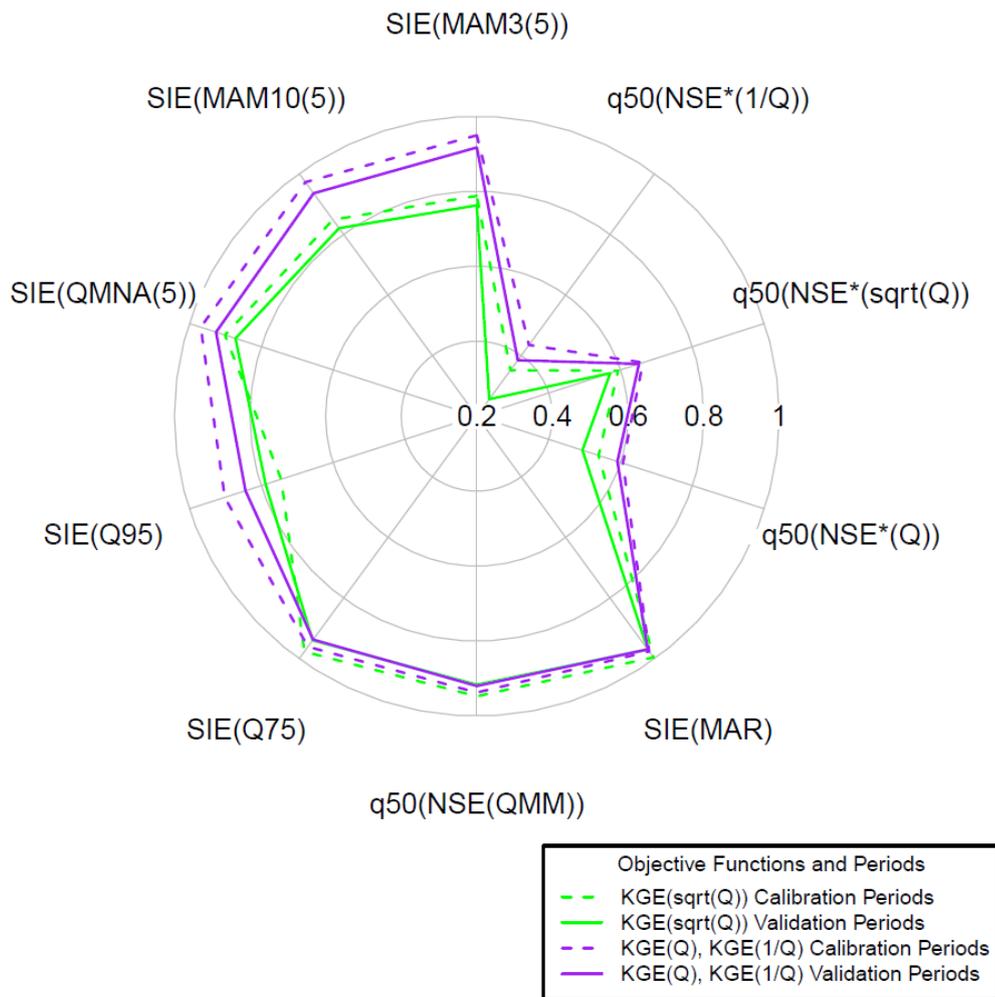


Figure 11: Assessment of temporal robustness. Radial plot showing the results in calibration for the two periods (dry and wet periods) and in validation for the two periods for the two objective functions,  $KGE(\sqrt{Q})$  and the combination of  $KGE(Q)$  and  $KGE(1/Q)$ .

Figure 12 shows the distribution of the relative errors of the MAR simulations with each objective function and for each period. We observed that when calibrating on the wet period, the MAR were overestimated on the dry validation period. The opposite was observed when calibrating on the dry period and validated on the wet period. The same trend was observed for the other assessment criteria. This lack of robustness was more important for mountainous catchments where the climate variability was higher. However, this lack of robustness is relatively independent of the objective function.

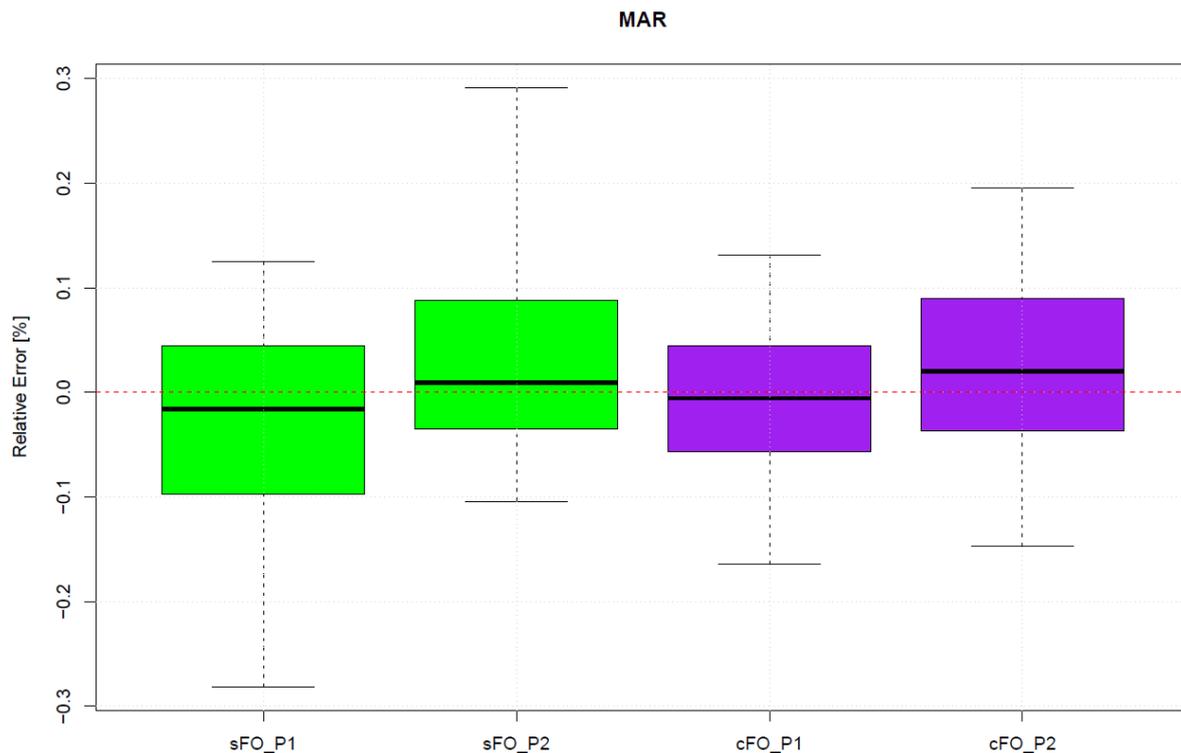


Figure 12: Box plot of distribution of MAR obtained in validation (P1: calibration on dry period, P2: calibration on wet period) with the two objective functions, KGE(sqrt(Q)) (called sFO) and the combination of KGE(Q) and KGE(1/Q) (called cFO) over the entire catchment set (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.05 and 0.95 percentiles).

### 3.7 Comparison with the six parameter model TOPMO

Figure 13 shows the radial plot results for the two objective functions,  $KGE(Q^{0.5})$  and the combination between  $KGE(Q)$  and  $KGE(1/Q)$  for the two models, GR4J and TOPMO, in validation. We observed that the low-flow indices were better simulated when calibrating the models with the combination of  $KGE(Q)$  and  $KGE(1/Q)$  than with  $KGE(Q^{0.5})$  and the calibration with these two objective functions gave the same results for the simulations of the seasonality and MAR. Looking at the continuous simulations, we observed that for high and intermediate flows, the medians of the distributions of  $NSE^*(Q)$  and  $NSE^*(Q^{0.5})$  were the same whatever the objective function. However, for low-flow simulations, the simulations were better with the objective function combining  $KGE(Q)$  and  $KGE(1/Q)$ , especially for TOPMO. Whatever the objective functions, the low-flow indices were as good simulated by GR4J as by TOPMO. The MAR was slightly better simulated by GR4J. The hydrographs for the mountainous catchments were indeed not as well simulated with this version of TOPMO.

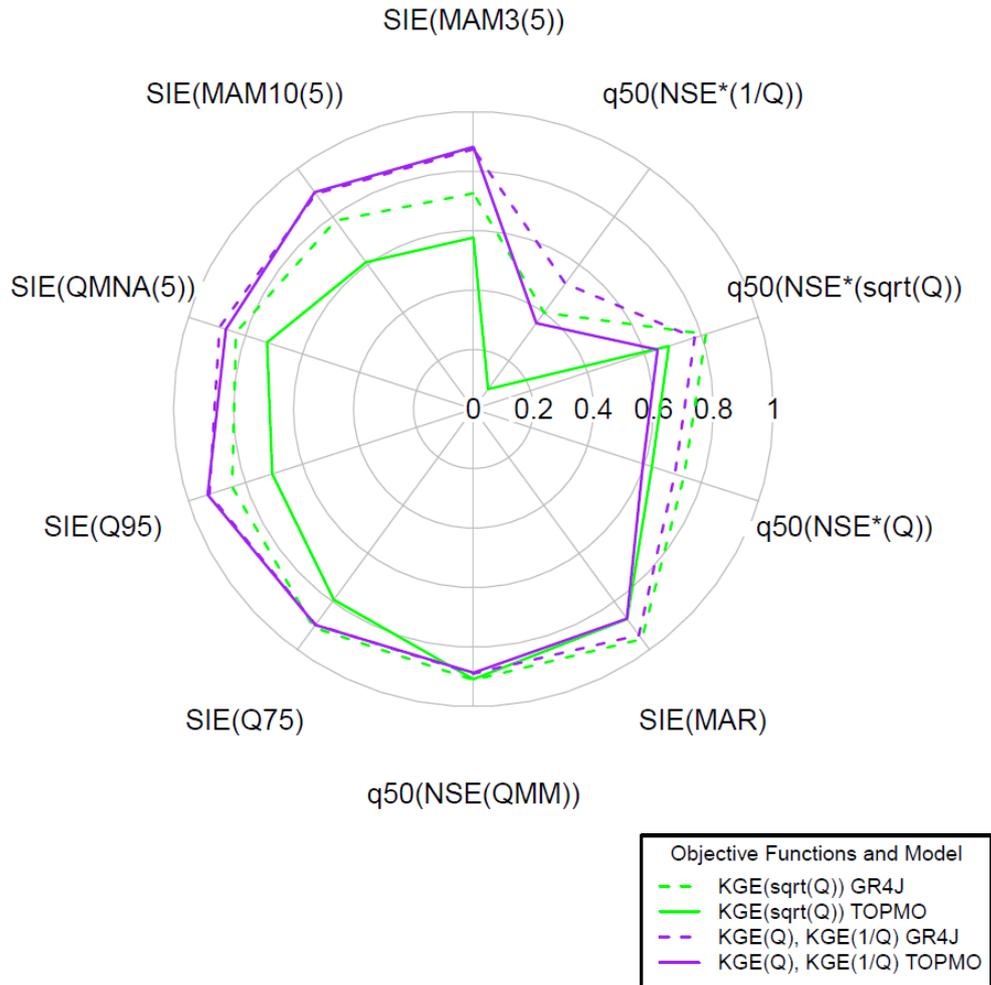


Figure 13: Radial plot showing the results in validation for the two objective functions,  $KGE(\sqrt{Q})$  and the combination of  $KGE(Q)$  and  $KGE(1/Q)$  and the two models, GR4J and TOPMO.

### 3.8 Conclusions

This study aimed at identifying the best objective function to simulate low-flow indices in order to calibrate the parameter set of a rainfall-runoff model to estimate these statistics. First, the seven assessment criteria were used as the objective function. The results of this test suggest that this objective function was not robust, giving quite good results in calibration but relatively poor results in validation. Then, six single objective functions and 15 combinations of these single objective functions were tested on a large sample of 691 catchments of various hydro-meteorological regimes throughout France. The results of our tests suggest that a combination as the mean of two single objective functions,  $KGE(Q)$  and  $KGE(1/Q)$ , improved the simulation of these seven assessment criteria. The quality of the simulations for very different calibration and validation periods did not decrease excessively and was dependent

on the climate variability of the catchments rather than the choice of the objective function. When testing the best single objective function and the best combination with an alternative rainfall–runoff model, the same conclusions were observed, suggesting that the results do not depend on the rainfall–runoff model used. Therefore, we recommend using the mean of  $KGE(Q)$  and  $KGE(1/Q)$  as an objective function to simulate low–flow indices with continuous conceptual rainfall–runoff models.

The objective function calculated on the seven assessment criteria showed results that were not robust while the combined objective functions were. A perspective of our study would be to combine the objective function calculated on the assessment criteria and on the temporal goodness-of-fit to keep the dynamic of the hydrographs.

While the results presented in this paper seem to be relatively general regarding to the model used, it should be noted that both tested rainfall–runoff models were not developed specifically for low–flow simulations and thus their structure and parametrization might not be the best suited for low–flow simulations. Olsen *et al.* (2013) showed that modifying the structure of a typical hydrological model to fit the low–flow simulations might better improve low–flow simulation than modifying the objective function.

The choice of an objective function for low–flow index simulations is strictly restricted to gauged catchments. For ungauged catchments, hydrological models are commonly used to derive streamflow indices at ungauged sites by transferring model parameters from gauged to ungauged catchments. Thus the choice of the objective function used for the gauged catchments might influence the simulation of the regionalized model on ungauged sites. However there is a high probability that the model parameter transfer from gauged to ungauged catchments will carry much more uncertainties than the choice of the objective function used in gauged catchments. A natural perspective of our study would be to test the sensitivity of the objective function in ungauged catchments.

## **Acknowledgments**

The authors wish to thank Météo–France for providing the meteorological data. We also wish to thank the two anonymous reviewers and the associate editor for their constructive comments on a previous version of this article which help improve the paper.

## References

- Beven, K. J. et Kirkby, M. J. 1979. A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin*, 24(1), 43-69.
- Booij, M. J. et Krol, M. S. 2010. Balance between calibration objectives in a conceptual hydrological model. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 55(6), 1017-1032.
- Brigode, P., Andréassian, V., Bourgin, F. et Lobligeois, F., 2014. *Reconstitution ponctuelle de chroniques hydrologiques incomplètes ou manquantes par modèle pluie-débit*. Rapport Irstea - Onema.
- Budyko, M. I., 1974. *Climate and life*.
- Catalogne, C., 2012. *Amélioration des méthodes de prédétermination des débits de référence d'étiage en sites peu ou pas jaugés*. Thesis (PhD). Irstea (Lyon), Université Joseph Fourier (Grenoble).
- Chiew, F. H. S. et McMahon, T. A. 1993. Assessing the Adequacy of Catchment Streamflow Yield Estimates. *Australian Journal of Soil Research*, 31(5), 665-680.
- Crochemore, L., Perrin, C., Andréassian, V., Ehret, U., Seibert, S. P., Grimaldi, S., Gupta, H. et Paturel, J. E. 2015. Comparing expert judgement and numerical criteria for hydrograph evaluation. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 60(3), 402-423.
- Deckers, D. L. E. H., Booij, M. J., Rientjes, T. M. et Krol, M. S. 2010. Catchment Variability and Parameter Estimation in Multi-Objective Regionalisation of a Rainfall-Runoff Model. *Water Resources Management*, 24(14), 3961-3985.
- Edijatno, Nascimento, N. D., Yang, X. L., Makhlof, Z. et Michel, C. 1999. GR3J: a daily watershed model with three free parameters. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 44(2), 263-277.
- Engeland, K. et Hisdal, H. 2009. A Comparison of Low Flow Estimates in Ungauged Catchments Using Regional Regression and the HBV-Model. *Water Resources Management*, 23(12), 2567-2586.
- Fenicia, F., Solomatine, D. P., Savenije, H. H. G. et Matgen, P. 2007. Soft combination of local models in a multi-objective framework. *Hydrology and Earth System Sciences*, 11(6), 1797-1809.
- Gupta, H. V., Kling, H., Yilmaz, K. K. et Martinez, G. F. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2), 80-91.
- Houghton-Carr, H. A. 1999. Assessment criteria for simple conceptual daily rainfall-runoff models. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 44(2), 237-261.
- Khu, S. T. et Madsen, H. 2005. Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration. *Water Resources Research*, 41(3).
- Klemeš, V. 1986. Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1), 13-24.
- Krause, P., Boyle, D. P. et Bäse, F. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89-97.
- Laaha, G. et Blöschl, G. 2006. Seasonality indices for regionalizing low flows. *Hydrological Processes*, 20(18), 3851-3878.
- Lang delus, C. 2011. Les étiages : définitions hydrologique, statistique et seuils réglementaires. *Cybergeog*.

- Le Moine, N., 2008. *Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des mod-les pluie-débit ?* Thesis (PhD). Cemagref (Antony), UPMC (Paris).
- Lobligeois, F., 2014. *Mieux connaître la distribution spatiale des pluies améliore-t-il la modélisation des crues ? Diagnostic sur 181 bassins versants français.* Thesis (PhD). Irstea (Antony), AgroParisTech.
- Madsen, H. 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources*, 26(2), 205-216.
- Magand, C., 2014. *Influence de la représentation des processus nivaux sur l'hydrologie de la Durance et sa réponse au changement climatique.* Thesis (PhD). UPMC (Paris).
- Mathevet, T., 2005. *Quels modèles pluie-débit globaux au pas de temps horaire ? Développements empiriques et comparaison de modèles sur un large échantillon de bassins versants.* Thesis (PhD). Cemagref (Antony), ENGREF (Paris).
- Mathevet, T., Michel, C., Andréassian, V. et Perrin, C. 2006. A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins. *IAHS-AISH Publication*, (307), 211-219.
- Merz, R. et Blöschl, G. 2004. Regionalisation of catchment model parameters. *Journal of Hydrology*, 287(1-4), 95-123.
- Michel, C., Perrin, C. et Andréassian, V. 2003. The exponential store: a correct formulation for rainfall-runoff modelling. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 48(1), 109-124.
- Nash, J. E. et Sutcliffe, J. V. 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282-290.
- Nicolle, P., Pushpalatha, R., Perrin, C., Francois, D., Thiéry, D., Mathevet, T., Le Lay, M., Besson, F., Soubeyroux, J. M., Viel, C., Regimbeau, F., Andréassian, V., Maugis, P., Augeard, B. et Morice, E. 2014. Benchmarking hydrological models for low-flow simulation and forecasting on French catchments. *Hydrology and Earth System Sciences*, 18(8), 2829-2857.
- Olsen, M., Trolborg, L., Henriksen, H. J., Conallin, J., Refsgaard, J. C. et Boegh, E. 2013. Evaluation of a typical hydrological model in relation to environmental flows. *Journal of Hydrology*, 507, 52-62.
- Osuch, M., Romanowicz, R. J. et Booij, M. J. 2015. The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 60(7-8), 1299-1316.
- Oudin, L., Andréassian, V., Mathevet, T., Perrin, C. et Michel, C. 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. *Water Resources Research*, 42(7).
- Oudin, L., Michel, C. et Anctil, F. 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model? *Journal of Hydrology*, 303(1-4), 275-289.
- Perrin, C., Michel, C. et Andréassian, V. 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4), 275-289.
- Price, K., Purucker, S. T., Kraemer, S. R. et Babendreier, J. E. 2012. Tradeoffs among watershed model calibration targets for parameter estimation. *Water Resources Research*, 48(10).
- Pushpalatha, R., Perrin, C., Le Moine, N. et Andréassian, V. 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology*, 420(0), 171-182.

- Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L. et Morel, S. 2008. Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France. *Journal of Applied Meteorology and Climatology*, 47(1), 92-107.
- Sauquet, E., Gottschalk, L. et Krasovskaia, I. 2008. Estimating mean monthly runoff at ungauged locations: an application to France. *Hydrology Research*, 39(5-6), 403-423.
- Smakhtin, V. U. 2001. Low flow hydrology: a review. *Journal of Hydrology*, 240(3-4), 147-186.
- Stedinger, J. R., Vogel, R. M. et Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events. In: MAIDMENT, D. ed. *Handbook of Hydrology*. McGraw-Hill, 18.11-18.66.
- Valéry, A., 2010. *Modélisation précipitations - débit sous influence nivale. Élaboration d'un module neige et évaluation sur 380 bassins versants*. Thesis (PhD). Cemagref (Antony), AgroParisTech.
- Valéry, A., Andréassian, V. et Perrin, C. 2014. 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. *Journal of Hydrology*, 517, 1176-1187.
- Vidal, J. P., Martin, E., Franchisteguy, L., Baillon, M. et Soubeyroux, J. M. 2010. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*, 30(11), 1627-1644.
- Vis, M., Knight, R., Pool, S., Wolfe, W. et Seibert, J. 2015. Model Calibration Criteria for Estimating Ecological Flow Characteristics. *Water*, 7(5), 2358-2381.
- Vogel, R. M. et Fennessey, N. M. 1994. Flow-Duration Curves. I: New Interpretation and Confidence Intervals. *Journal of Water Resources Planning and Management*, 120(4), 485-504.
- Vogel, R. M. et Fennessey, N. M. 1995. Flow duration curves II: a review of applications in water resources planning. *Water Resources Bulletin*, 31(6), 1029-1039.
- Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E. et Xu, C. Y. 2011. Calibration of hydrological models using flow-duration curves. *Hydrology and Earth System Sciences*, 15(7), 2205-2227.
- Wmo, 2008. *Manual on Low-flow Estimation and Prediction*. Operational Hydrology.