Detailed response to the comments of referee 1

We want to thank referee 1 for his accurate and helpful review of our manuscript. In this author comment, we list how each of the remarks provided by the referee was addressed.

1. <u>Writing style</u>

a. We agree with you that terminology issues are important. In this paper, we tried to use short terms whose definitions are given at the beginning of the paper. On top of that, we tried as much as possible to use standard terminology, inspired by *Bloschl et al. (2013)*. However, our paper presents a new approach which is sometimes out of the scope of the existent literature. Regarding the term 'prescription', we do not define it as '*apriori* parameter estimation' which, in the context of distributed modelling, usually refers to giving spatialized parameter values based on spatialized catchment characteristics values. Here, 'prescription' is a broader term which encompasses '*apriori* parameter estimation' but also the assignment of a value based on the user experience.

➔ We propose to modify the manuscript as follows (Introduction):

p2, l.10 "Prescription consists in assigning parameter values based on the literature, the user experience on the hydrological model or empirical relationships between model parameters and catchment characteristics"

b. Before submission, the paper was corrected by a native English person.

c. We agree that the structure of our article is a little bit different from conventional papers since the methodology of combining several regionalization methods necessarily implies to describe parameter spatial patterns which can be seen as a kind of results. Therefore, we cannot consider a method/results structure. Instead, we propose a 'parameter spatial patterns' section followed by a 'performance analysis' section. Regarding the order of the figures, we first tried to respect the order of appearance in the text (see "Figure_order.pdf") but we found it very difficult to compare the different parameter spatial patterns. Thus, we chose to make one figure for the Loire basin (Fig. 4) and one figure for the Durance basin (Fig. 5). However, if you believe that it is better to present the figures in their order of appearance in the text as in "Figure_order.pdf", we are quite open to do it.

2. <u>Methods</u>

a. We used an awkward wording. Section 4.2.1. consists in fixing insensitive parameters to some values based on the user experience about the hydrological model. On top of that, we modified the introduction as discussed above not to suggest that prescription is only '*apriori* parameter estimation'.

➔ We propose to modify the manuscript as follows (section 4.2.1):

p9, l.8 "Consequently, we propose to prescribe the insensitive parameters uniformly at a value derived from literature or user experience, namely five parameters for the Loire catchment and two for the Durance catchment".

b. Constraint is a little applied method first introduced by Yadav et al. 2007 which can be considered in different ways. We can understand this method as providing a range of possible values. However, we can also understand it as a method based on estimations at ungauged sites, and this is how we considered the constraint method. Indeed, in our view, the constraint method uses streamflow estimations at ungauged sites to determine the parameter sets while the calibration method uses streamflow observations. We thus believe that name it calibration would be quite misleading. That is why we believe it is better to name it 'constraint'. But once again, if you think that this is a mistake, we are open to name it "ungauged calibration".

c. The fourth approach cannot be applied to all validation basins due to the inter-dependance of the basins. For example, consider a validation basin A upstream a calibration basin B, and on another side a validation basin C upstream a calibration basin D. If the most similar basin of A according to physio-climatic characteristics is basin D and the most similar basin of C is basin B, the parameter set of A should be the one calibrated on D and the parameter set of C should be the one calibrated on B.

A-v [D]-> B-c [B] ; C-v –[B] > D-c [D] parameter set into brackets

However, the calibration of B depends on the parameter set of A (upstream) that is to say the parameter set of D, which depends on the parameter set of C (upstream) that is to say the parameter set of B. We should therefore do several calibration passes until the parameter sets stabilize. Thus, at the scale of several dozens of catchments, this is not an option given the computation time.

However, as we can see on the figures below, there is a clear pattern that small catchments inheriting parameters from huge calibrations basins have poor performance. In the figures below, the surface ration RS is defined as :

$$RS_i = \frac{Drainage\ area_{validation\ station\ i}}{Drainage\ area_{calibration\ station\ downstream}}$$

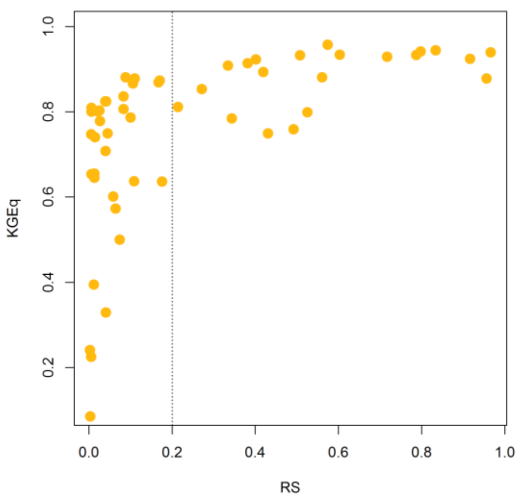


Figure 1: KGEq performance according to surface ratio - Loire@Gien

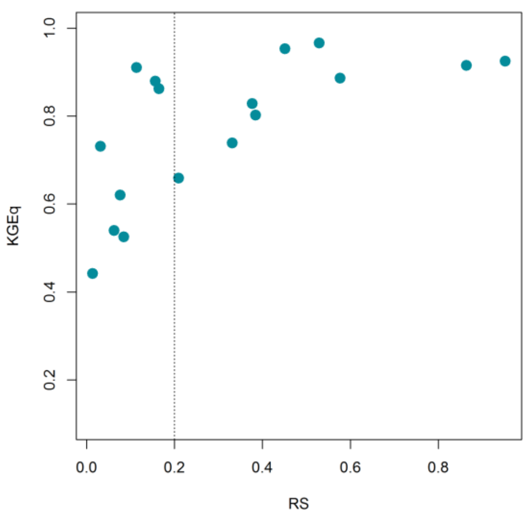


Figure 2: KGEq performance according to surface ratio - Durance@Cadarache

According to these figures, we therefore adopted the arbitrary value of RS = 20 %: the basins with a RS <20 % were submitted to the physio-climatic transposition whereas the others were still submitted to the previous upstream-downstream transposition.

➔ We propose to modify the manuscript as follows (section 4.2.4):

"Indeed, according to preliminary attempts of regionalization not detailed here, we found out poor performance for validation stations whose drainage area ratio with the downstream calibration station is low. To address this, we propose to rearrange the calibration sub-basin pattern with physio-climatic information to become a physio-climatic calibration sub-basin pattern. To do so, the idea is to no longer inherit parameters from this calibration station but inherit parameters from the most similar calibration station in terms of physio-climatic descriptors. The selection of the new donor calibration station is carried out through a Euclidian distance calculated over the principal components of the physio-climatic descriptors. As basins are nested ones, applying this method for all the validation basins is not an option since it would implies the inter-dependance of all the parameter sets. Therefore, we adopt the 20% arbitrary value of drainage area ratio meaning that all the validation basins with a drainage area ratio lower than 20% are concerned by this physio-climatic transposition while the parameters of the basins with a drainage area ratio higher than 20\% continue to be transposed as before, i.e. as described in section 4.1."

Following your suggestion, we applied the KS-Statistic as an alternative to our Enhancement Index. However, results show that it is not well suited to our problem. If we take the example of KGE seasonality over the Durance catchment, we have the figure below. The KS distance between the CDF of Exp1 and the gauged modelling (right border of the grey area) is 0,59 while the distance between Exp2 and the gauged modelling (right border of the grey area) is 0,71. Therefore, this test suggests that Exp1 provides better results than Exp2, wheareas the figure clearly shows the contrary. This is because the KS statistic calculates the vertical maximum absolute difference between the two cumulative distribution functions. Figure 4 shows that the KS test between the CDF of Exp1 and the gauged modelling in blue and the one between the CDF of Exp2 and the gauged modelling in red. So the KS test which does not represent the distance between the whole distribution functions and is not well suited to our analysis.

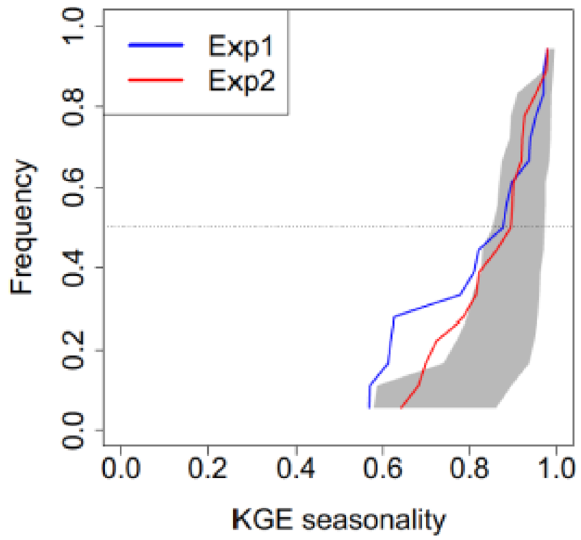


Figure 3: Cumulative distribution functions of performance for the runoff seansonality over the Durance catchment

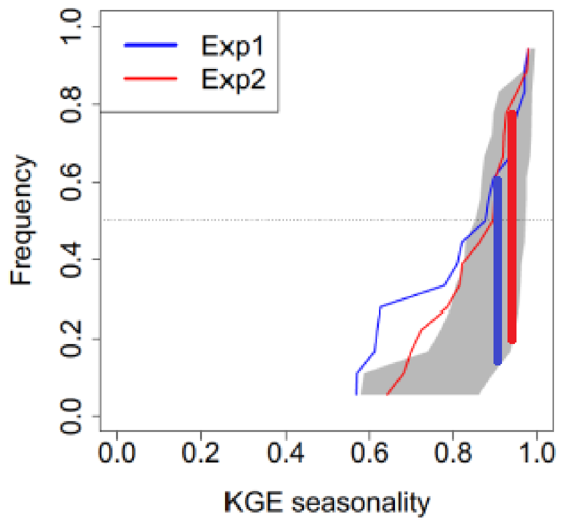


Figure 4: Cumulative distribution functions of performance for the runoff seansonality over the Durance catchment with the KS test

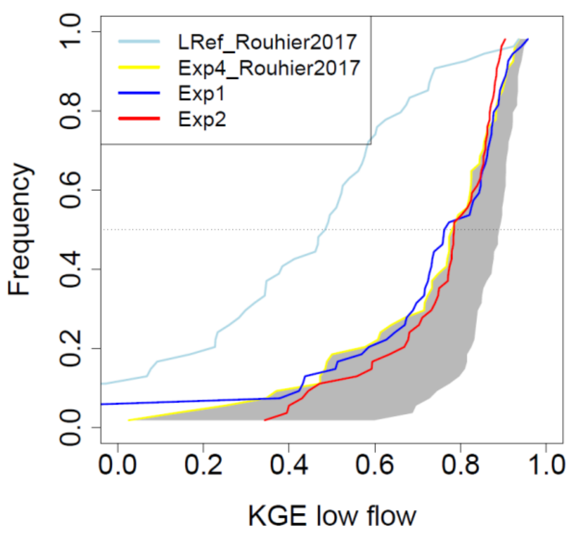
3. Introduction

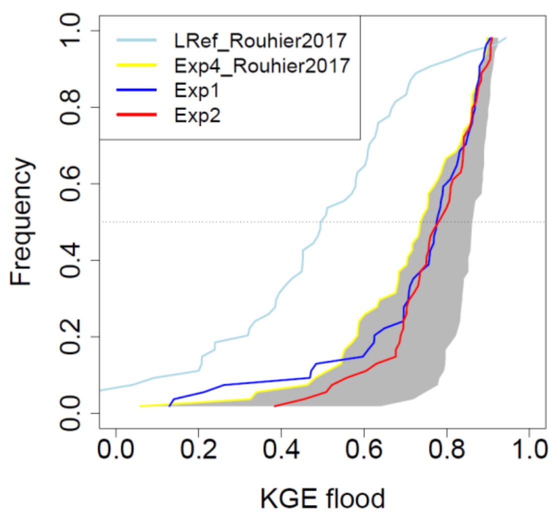
Thanks for your suggestions. We added the reference of Gotzinger and Bardossy (2007), which was effectively missing.

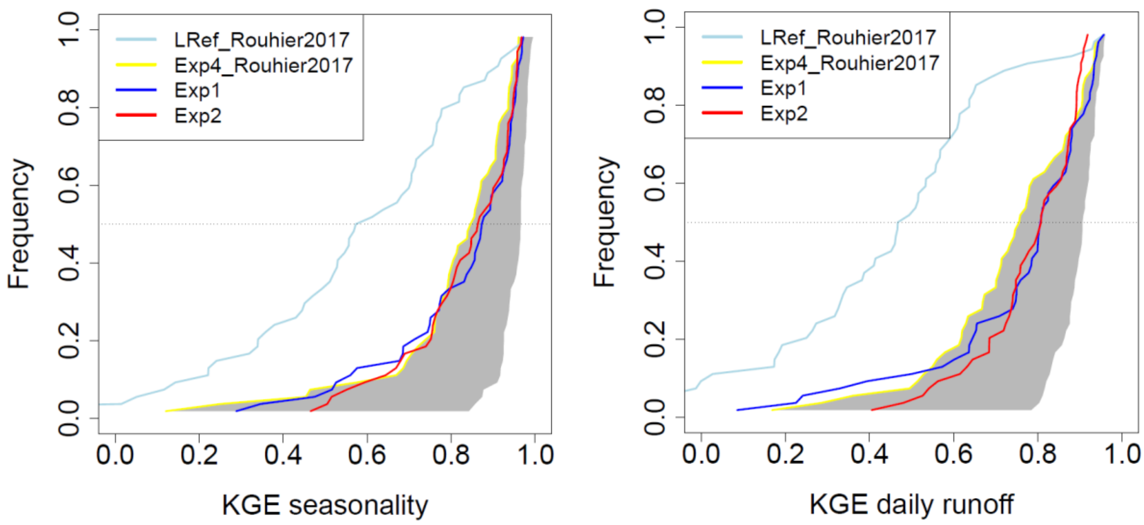
4. Results

The results section was renamed “Performance analysis” to be more explicit. We also renamed the title of section 4 ““Construction of parameter spatial patterns”.
Exp1 corresponds to Exp6 of Rouhier et al. (2017). It is the best solution of Rouhier et al. (2017). Obviously, it is not the best for all signatures and the whole distribution. For example, as you mentioned, this experiment is not the best for low flow (see Rouhier et al. (2017)). However, it is clearly the best compromise of Rouhier et al. (2017) with the best overall performance.
➔ We propose to modify the manuscript as follows (section 5):
” It is referred to as *Exp1* and corresponds to the regionalisation method with the best overall performance of those discussed in Rouhier et al. (2017).”

Regarding the comparison with the various experiments in Rouhier et al. (2017), we do not want to add to many experiments in the present paper as it would make reading difficult. However, you can find on the 4 figures below the comparison with two experiments of the previous paper : LRef (uniform climatic inputs + uniform parameters) and Exp4 (spatialized climatic inputs + uniform parameters), for the Loire basin.
Be careful, the grey area of the present article is different from the one of the previous article.







5. Discussion

As we explained before, we believe that it is not a good idea to add the experiments of Rouhier et al. (2017). On top of that, the purpose of the present article is not to compare climatic inputs spatialization and parameter sets spatialization. Nevertheless, we could add a supplementary material with the 4 figures above and three enhancement indexes:

- Enhancement index about climatic inputs spatialization:

$$EI_c = \frac{area\ KGE(LRef) - area\ KGE(Exp4)}{area\ KGE(LRef) - area\ KGE\ (gauged)}$$

where the gauged experiment corresponds to the right border of the grey area.

- Enhancement index about parameter sets spatialization:

$$EI_p = \frac{area\ KGE(Exp4) - area\ KGE(Exp2)}{area\ KGE(LRef) - area\ KGE\ (gauged)}$$

		EIc	EIp
Loire	Daily runoff	64%	10%
	Seasonality	62%	7%
	Flood	61%	15%
	Low flow	73%	6%
Durance	Daily runoff	67%	11%
	Seasonality	69%	5%
	Flood	64%	16%
	Low flow	69%	11%

According to these results, climatic inputs spatialization provides an enhancement of about 65% while parameter sets spatialization brings an enhancement of about 10%, in ungauged context.

Let us know if you consider that this supplementary material may be relevant.

➔ We propose to modify the manuscript as follows (section 6):

“As for Gotzinger and Bardossy (2007), our paper thus shows that combining several regionalisation methods make it possible to benefit from the advantages of each to provide better performances in ungauged basins.”

6. Minor comments

1. See 1.a.

2. Changed for ‘represented by’

3. Quite a number of citations are given in the paragraphs below (page 2 - line 5 to page 3 - line 14)

4. See 2.b.

5. Silvestro et al. (2015) don’t calibrate one parameter with land surface temperature but the parameter sets.

6. Changed to ‘spatial validation’ .

7. There is already a verb in the sentence: ‘comprises’.

8. Corrected

9. Added

10. It is not possible to provide equations for each objective function since for some of them it is not only an equation but an algorithm. However, we can illustrate this in a supplementary material with the figures below :

>> Daily Runoff

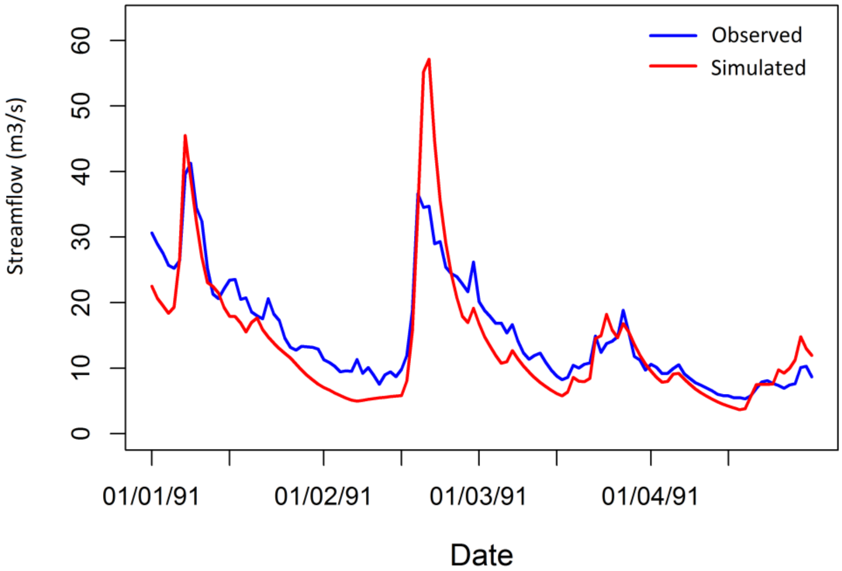


Figure 5: Daily runoff

>> Seasonality

The hydrological signature on which the criterion is a time series of 365 values, the first value being the average of the discharges of every January 1st.

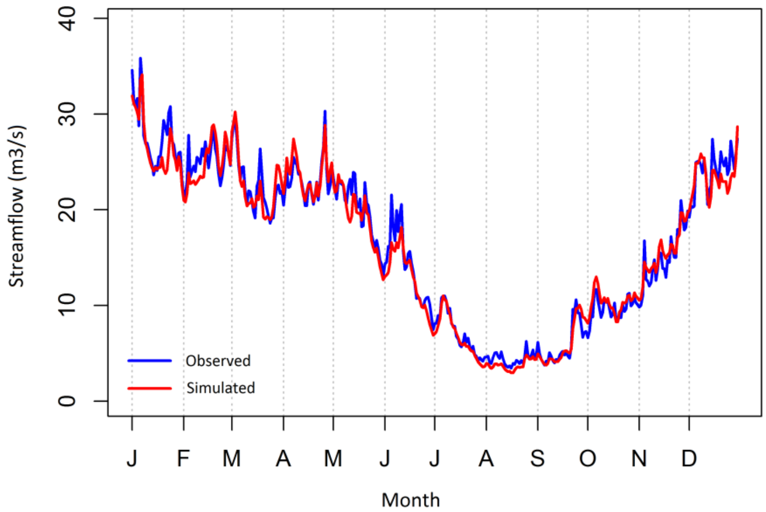


Figure 6: long-term mean daily streamflow

>> Flow duration curve

KGE is applied to every monthly flow duration curve. Then, the 12 KGE values are aggregated with a weighted mean for which the weights depend of the monthly means, as

$$KGE_{flood} = \sum_{i=1}^{12} \frac{\overline{Q_i}}{\sum_{j=1}^{12} \overline{Q_j}} KGE_i$$

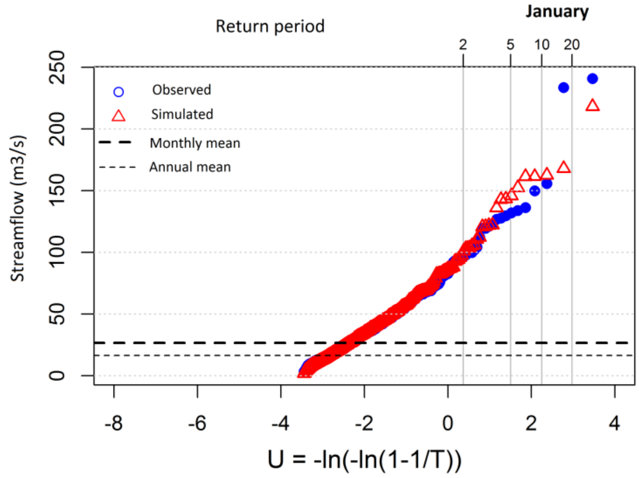


Figure 7: Flow duration curve of January

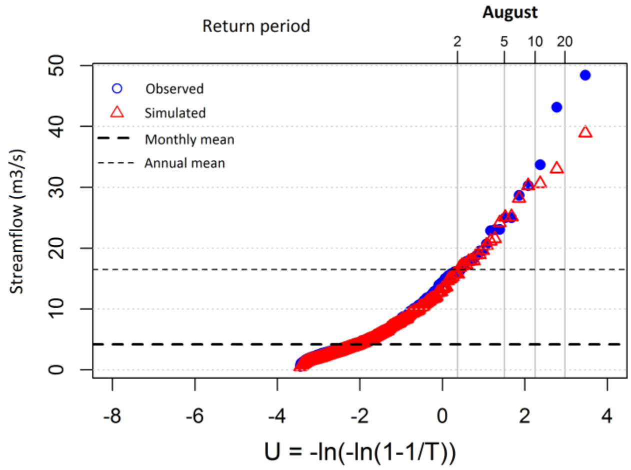


Figure 8: Flow duration curve of August

>> Flow recessions

The selection of the recessions of the daily streamflow time series result of: (i) the smoothing of the time series with a 7-day window, (ii) the selection of the decreasing periods, (iii) the removal of the first 7 days of each decreasing period.

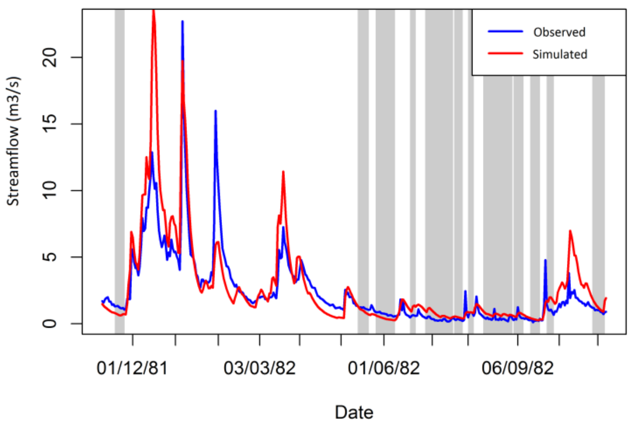


Figure 9: flow recessions are identified in grey

11. We did not find satisfying alternatives for these titles. Hence, we propose to keep the titles of sections 4.1 and 4.2 which explain our approach.

12. The sensitivity analysis was done for several stations and for all the four objective functions. Thus, it gave the sensitivity for all the parameters at a given station for a given objective function. It is too ambitious to describe in details this sensitivity analysis here and to provide sensitivity indices. However, a paper might be written about this sensitivity analysis that we cannot sum up here.

    ➔ We propose to modify the manuscript as follows (section 4.2.1):
    "A sensitivity analysis of the MORDOR-TS model has been conducted over hundreds of French catchments according to the approaches for each of the four objective functions according to the approaches of Sobol (1993), Homma and Saltelli (1996) and Liu and Owen (2006)."

Nevertheless, we give some details below :

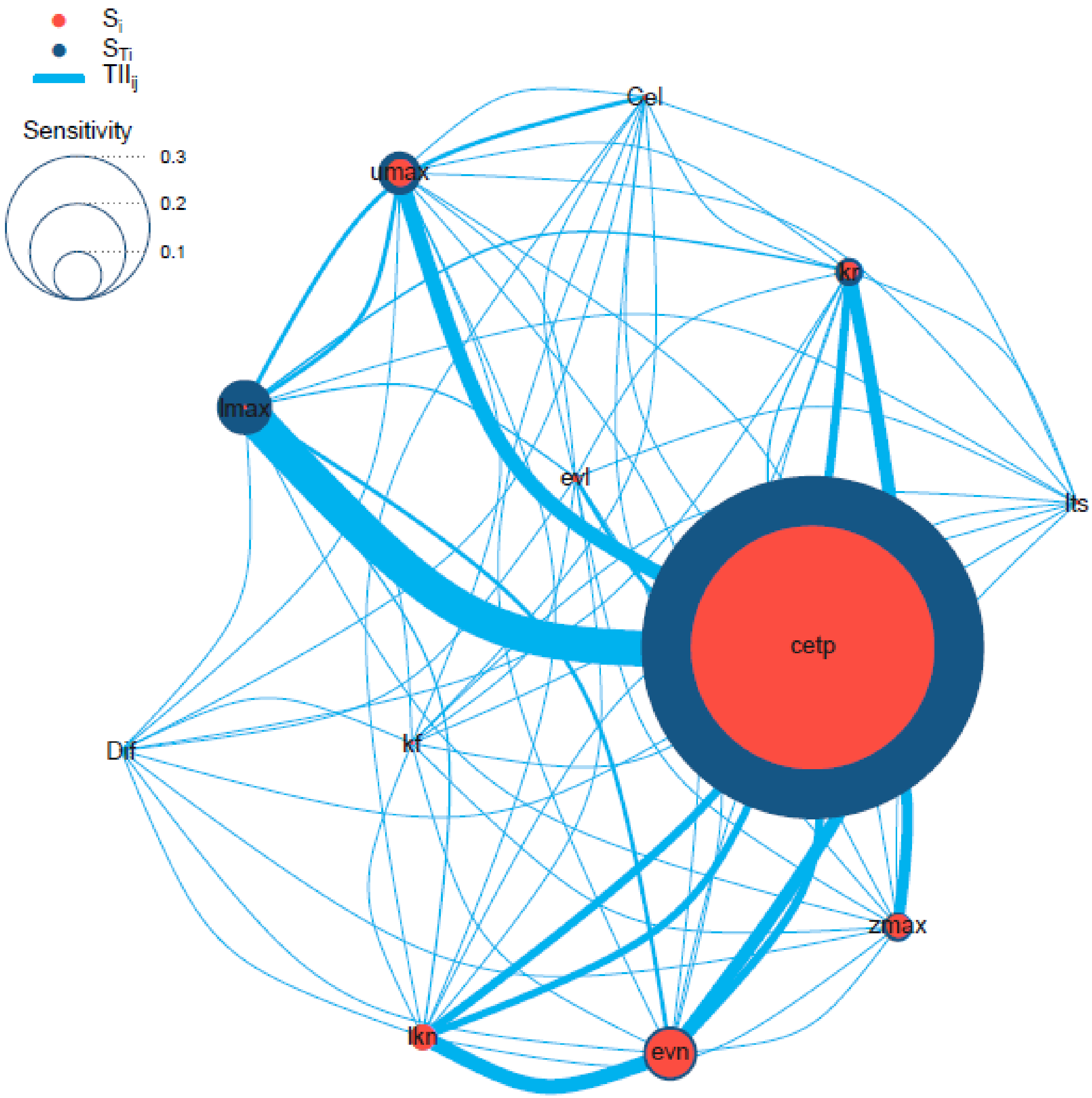


Figure 10: Fanova graphs - KGE daily runoff for Loire@Gien

This figure shows the results obtained for the discharge station of Gien in our 12-parameter configuration as regards KGE daily runoff. This graphical representation is inspired by the FANOVA graphs of Muehlenstaedt et al. (2012). The radius of the red disc reprensents the value of the first order sensitivity Si of the parameter (Sobol, 1993), while the radius of the blue disc gives the total sensitivity STi of the parameter (Homma and Saltelli, 1996), i.e. first-order sensitivity plus its sensitivity in interaction with the other parameters. The red disc is superimposed on the blue disc since the first-order sensitivity is always lower than the total sensitivity. The larger the red disc, the greater the first-order sensitivity. The larger the differences between the blue and the red discs, the greater the interactions with the other parameters. The distribution of these interactions is represented by the blue lines between the parameters. The greater the thickness of the line, the greater the interaction TIIij between the two parameters (Liu and Owen, 2006). Figure 10 therefore informs us that five parameters are not sensitive at all: the snow parameters (kf and lts), the parameter generating the delayed flows (evl), the diffusivity (Dif) and the celerity (Cel). This outcome is confirmed by the sensitivities distributions over the 106 discharge stations of the Loire catchment shown in Figure 11. The same outcome is obtained for the other three signatures : KGE daily regime, KGE flood and KGE low flow.

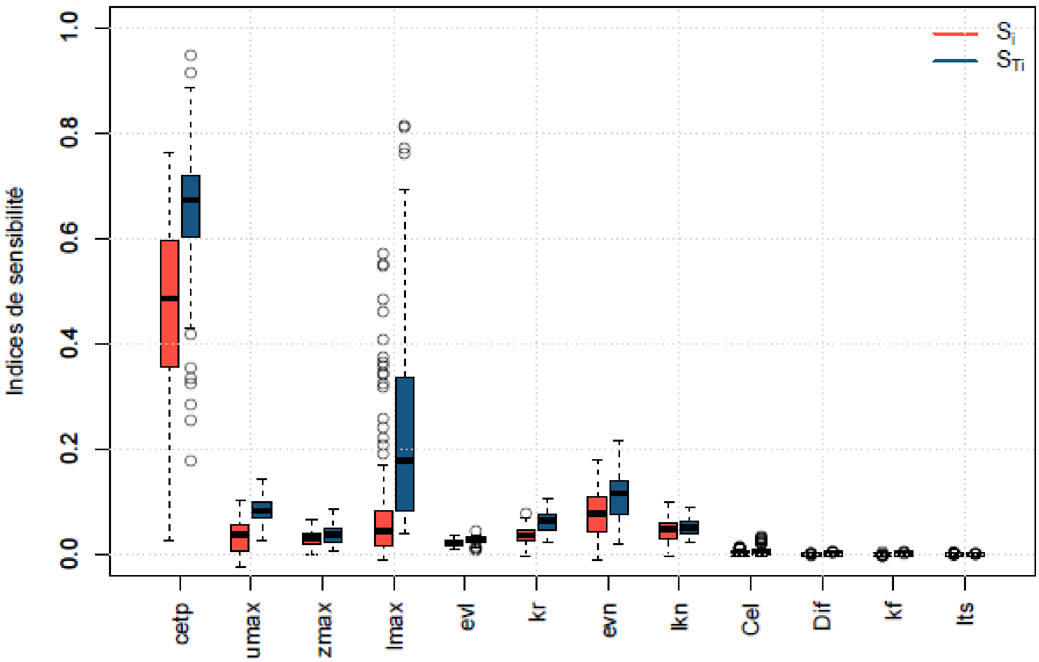


Figure 11: Boxplot of parameter sensitivity over the 106 streamflow stations of the Loire catchment as regards the KGE daily runoff

References

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H.: Runoff Prediction in Ungauged Basins. Synthesis across Processes,Places and Scales, Cambridge University Press, Cambridge, 2013

Homma, T., Saltelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering & System Safety 52 (1), 1–17.

Liu, R., Owen, A. B., 2006. Estimating mean dimensionality of analysis of variance decompositions. Journal of the American Statistical Association 101 (474), 712–721.

Muehlenstaedt, T., Roustant, O., Carraro, L., Kuhnt, S., 2012. Data-driven kriging models based on fanova-decomposition. Statistics and Computing 22 (3), 723–738.

Sobol, I. M., 1993. Sensitivity estimates for nonlinear mathematical models. Mathematical modelling and computational experiments 1 (4), 407–414.