Review of the article:

The sub-annual calibration of hydrological models considering climatic intra-annual variations

By **Binru Zhao, Huichao Dai, Dawei Han and Guiwen Rong**
Submitted for possible publication in *Hydrology and Earth System Sciences.*

1 GENERAL COMMENTS

The manuscript presents the results of different calibration procedures that are based on climatic similarities between sub-periods and on one rainfall-runoff model, methods applied on a unique catchment in UK. The issue of the rainfall-runoff model parameter dependence to the climatic period considered for the calibration is very interesting, especially in the context of the quantification of the climate change impacts on hydrology. Thus, the paper subject is highly relevant and the tested methodology is interesting and original, but the paper is lacking significant information about the applied methodology, the studied catchment and is lacking elements on how this methodology could be applied in an operational context. Moreover, the consideration of only one catchment is a strong limitation of this paper and is not enough discussed in the conclusion. Some of the paper figures are useless; the other ones are poorly presented in the paper and in their caption. These comments are detailed in the first part of this review and specific comments are given in the second part.

1.1 Studied catchment

Considering only one catchment for such study is a strong limitation for the generalization of the obtained results. Why not considering other catchments and applying the same methodology on an ensemble of different catchments?
The paper lacks some justification on the choice of this particular catchment regarding the objectives of the study. What are the particularities of this catchment in terms of hydro-climatic variability (both inter and intra-annual)? Moreover, information on the quality of the studied times series is lacking. The potential temporal variability of the measurement quality is highly important in such studies. For example, the poor hydrometric quality of the flow time series on several particular years could significantly affect the performance of the rainfall-runoff model calibration on this sub-period and thus misleading the result interpretation.
Finally, the presentation of the catchment regime and of the temporal variability of the hydro-climatic series (flow, temperature and precipitation) is lacking.

1.2 Calibration methodology

The presentation of the developed methodology is lacking some important information and the applied methodology presents some limitations that need be discussed.
Considering only one calibration and evaluation criterion in a study based on only one catchment is somehow disappointing. Why only looking at the Nash and Sutcliffe (1970) Efficiency (NSE) criterion? I think that considering the Kling and Gupta Efficiency score (KGE, Gupta *et al.*, 2009) and analyzing its different sub-criterions will be interesting for studying the benefits of the different calibration procedures in terms of flow mean bias, variance bias and temporal correlation…
The choice of the calibration and validation periods is important in this type of study. Why the selection of this particular periods for calibration (1960-2000) and validation (2001-2011)? Why only using 10 years for validation and why not considering different validation periods?
Is it not clear to me why you did not choose an index considering both precipitation and temperature variables for grouping periods, such as the aridity index, cited in the introduction section and used by Brigode *et al.* (2013)?
In addition, I think that performing a calibration on a "randomly grouping" for each time steps would be an interesting reference to compare with climatic grouping.

In the subsection 3.4 (line 182 to 184), you stated that the "sub-periods in the validation period are matched into the most similar cluster of all clusters in the calibration period". This point needs to be discussed. What about potential differences between clusters of the calibration period and validation period? What about potential new clusters? This should be addresses in the results section by comparing the characteristics of the calibration and validation sub-periods.

Finally, it is unclear how the model parameters are obtained for each calibration process. I think that you should explain how you perform a continuous rainfall-runoff simulation over a given period and how you calibrate the model only over several timesteps and sub-period.

1.3   Seasonal bias of the model?

I think that an analyze of the seasonal performances of the model should be added before applying the different calibration strategies, as an analyze of the performance on the different sub-periods considered. For example, the calculation of NSE for each season and each month would be interesting. Thus, potential seasonal biases in the rainfall-runoff model calibration could be identified and discussed.

1.4   Use of "only" one hydrological model

Could you please discuss the fact that you only considered one rainfall-runoff model in this study? What would be the conclusion if you applied the same calibration methodology with one other hydrological rainfall-runoff?

1.5   Operational use of this methodology?

Could you please discuss the potential uses of your developed methodology in applied studies? How this method could be applied for the quantification of the climate change impacts of catchment hydrology? For each catchments?

2   SPECIFIC COMMENTS

- **Line 27:** could you detail what you mean by "satisfactory performances"?
- **Line 31:** could you detail what you mean by "stationary"? Such word has to be clearly defined in this context of climate change.
- **Line 32:** could you detail what you mean by "catchment conditions": climatic, land use, hydrological conditions?
- **Line 32 to 35:** this sentence is very unclear. I think that you should be more precise on what you mean by "change of catchment", "climate change" and "catchment conditions"…
- **Line 36 to 38:** please give more details on what is a "calibration error" and if validation performances have been quantified in this study, and on which catchments the methodology has been applied.
- **Line 38 to 42:** again, on how many catchments, where (and thus in which climate) this test has been conducted? How many years of calibration were available? Are these results obtained in calibration or in validation on an independent sub-period?
- **Line 44:** "worth" compared to what? The report of the conclusion of this paper is unclear although it seems particularly interesting considering the aim of the submitted paper.
- **Line 48:** what is "different climatic" conditions?
- **Line 51:** Merz *et al.* (2011) worked on catchments in Austria and not in Australia.
- **Line 53:** could you clarify that the difference between calibration and validation periods are in terms of climate?
- **Line 55**: could you clarify what is, in this context, the aridity index and how it is calculated?
- **Line 55:** what is a "sub-period group" in this context?
- **Line 56:** again, could you explain what you called "performances" here? In terms of what score?

- **Line 61:** what is a "30-day data sets" in this context?
- **Line 64:** could you clarify what is an "hydrological similarity"?
- **Line 65:** do you refers to Toth and Brath (2007) instead of Toth (2009)?
- **Line 71:** could you clarify what are the difference between the "serial" and the "parallel" calibrations in this context?
- **Figure 1**: this figure needs to be strongly improved, with the addition of:
  - a general map of the UK,
  - a scale bar,
  - the elevation of the catchment,
  - the position of the rivers and of the gauging station.
- **Line 94:** the Figure 2 needs to be more deeply presented, with explanation on the period considered and on the obtained results, for example.
- **Line 110:** could you define what the word "flexibility" means in this context? Also, the "," after flexibility needs to be deleted.
- **Figure 3:** this figure seems to be useless. I think that a complete diagram of the rainfall-runoff model with the different parameters would be more useful.
- **Table 1:** please add parameter units.
- **Line 119:** please consider to change the title of this subsection into "recognition of… with climatic similarities" since you choose your sub-periods based only on climatic variables. I think that this change has to be made all over the paper.
- **Line 124:** please consider changing "hydrological" into "climatic".
- **Line 129 to 131:** please consider to merge these two sentences and rephrase them, since they are unclear to me.
- **Line 138:** could you clarify what is the "periodic rainfall" variable?
- **Line 154:** please cite the "previous studies" you mentioned.
- **Line 160 to 161:** thus, why considering this validity index (cf. section 1.2 of this review) ?
- **Line 164:** again, I think that you should clarify and define first in the introduction what you mean by "stationary" or you should avoid this word.
- **Line 164 to line 169:** this paragraph lacks some clear explanation on how model parameters are obtained (cf. section 1.2 of this review).
- **Line 176 to 179:** this paragraph is unclear: why using Latin Hypercube Sampling? What is the *nlminb* function?
- **Line 182:** no, the similarities of sub-periods are only "climatic" and not "hydroclimatic" in your approach.
- **Line 185:** please rewrite this unclear sentence.
- **Line 194:** could you give some explanation on the obtained results presented in the Table 2?
- **Figure 4:** please correct the figure legend by writing "calibration". You should explicitly state in the figure caption that these results are obtained with the monthly time scale.
- **Line 196:** I think that the results obtained with the other timesteps are interesting and may be somehow presented in the paper. Please consider to add these results in the paper.
- **Line 196 to 199:** please rewrite this unclear sentence.
- **Line 207:** please change "hydrological similarities" into "climatic similarities".
- **Line 209:** please change "hydrological similarities" into "climatic similarities".
- **Figure 5:** I do not understand how the figure 5 presents the difference between two classifications. For me it only shows the results of one classification. Moreover, why only presenting this 5-year period? This has to be addressed in the paper. Finally, why the number of groups is different considering different time steps?
- **Line 209 to 215:** this paragraph is unclear. It seems to me that the authors are in the end analyzing the rainfall regime through the calibrations results, while a basic analysis of the observed regimes (cf. section 1.1 of this review) would a priori give the same information.
- **Figure 6:** please indicate the quantiles used for the construction of the boxplots. What are the points outside of the boxplots? Please give the scale of the box widths, which is proportional

to the size of the group. Please also state explicitly in the caption legend that you are presenting the results for the monthly time step only.

- **Line 223 to 224:** could you define what is a outlier in this context and why you consider that better performance are obtained for classification with "fewer outliers in clusters" ?
- **Line 225 to 227:** It seems to me that the CBG is, by construction, better able to capture the flow seasonal pattern since it exists a clear seasonal pattern for the temperature of the studied catchment, while there is no clear seasonal pattern for precipitation. The climatic regimes of the catchment needs to be plotted and presented before (cf. sections 1.1 and 1.3 of this review). Please consider this observation for the analysis of the Figure 7. Finally, why not showing the same figure for the other time steps, for which the results could be less obvious?
- **Line 231:** change "hydroclimatic" to "climatic".
- **Line 233 to 235:** this sentence needs to be clarified and strongly improved in terms of explanation quality (cf. section 1.1 of this review). Is there any indication of climatic change on the studied catchment or is there "only" a seasonal bias in the rainfall-runoff model performances?
- **Line 243 to 249:** are you sure that this "new" classification method obtained with the "temperature-dominated FCM algorithm" is not the same classification that the calendar-based one?
- **Line 251 to 258:** this paragraph needs to be improved or deleted. Please state what is NDVI, where and how you define this index. Why did you analyze the correlation over the 2001-2011 period? Why only a sub-period is plotted on the Figure 10?
- **Figure 11**: why only this sub-period (2005-2008) is plotted and why only the CBG calibration is considered? This figure is useless in this form, since it is difficult to compare the calibration strategies.
- **Line 273**: change "hydroclimatic" into climatic.
- **Line 274:** define or delete the "stationary" word.

3   REFERENCES

Brigode, P., Oudin, L., Perrin, C., 2013. Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change? Journal of Hydrology 476, 410–425. doi:10.1016/j.jhydrol.2012.11.012

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377, 80–91. doi:10.1016/j.jhydrol.2009.08.003

Merz, R., Parajka, J., Blöschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water Resour. Res. 47, W02531. doi:10.1029/2010WR009505

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