Responses to comments posted by Referee #2

We thank the second referee for reviewing our article and providing feedback. The referee comments identify some unclear issues and help to improve the presentation of our research. In the following, we answer to all of the comments one by one. The Referee comments are in blue. Please note that our enumeration differs from that of the referee, because we split several comments into individual points. The original order of the referee comments however remains unchanged.

Comment 1: The main objective of this paper was to propose a new way to analyze hydrological systems using adaptive clustering. The study is interesting by dynamically identifying and clustered similar model elements. Representatives per cluster inferred the dynamics. Although the application of the proposed framework looks promising, additional investigations and explanations are necessary before this paper can be published in HESS. In the next sections, I outline my major comments and suggestions that should allow the authors to improve their manuscript. Thus, my recommendation is to be accepted after the following points are clarified:

Reply 1: We are glad that the referee finds the topic of our study interesting.

Comment 2: 1. The dynamic, comparability, and similarity were emphasized in the proposed framework for adaptive clustering. However, the definitions of dynamic, comparability, and similarity are ambiguous. For example, the dynamic in hydrological simulation includes temporal dynamic and spatial dynamics. In particular, time scales for the application of hydrological models in the temporal dynamics are critical.

Reply 2: Agreed. In a revised version of the manuscript, we will provide brief definitions of relevant terms when they are first mentioned.

Comment 3: 2. The main result of high redundancy in geology and climate information is obvious. The key reason is the intricately linear or nonlinear correlation among the subsystem. The technique of adaptive clustering extracts the nonlinear correlation but not nonlinear information. Further analysis of the nonlinear relationship between subsystems is suggested.

Reply 3: We are sorry but we do not understand this comment.

Comment 4: 3. Only one basin, Attert basin, was studied. Moreover, Figure 2 and the relevant analysis did not elaborate on the geology and climate conditions, which are vital for the clustering of subsystems and hydrological simulation. The number of study areas is suggested to increase. The generalization of the framework proposed in this study is recommended to be demonstrated.

Reply 4: Spatial distribution of geology and rain gauges is indeed important for clustering. In Figure 2, geology is already indicated by the background colors (see legend 'Geology'). Rain gauge locations are also shown as yellow dots. In order to make clear their respective area of influence as calculated by the Nearest Neighbor method, we will add to the plot the related separation lines. As mentioned in the text (p8 lines 7-12), all other climatological data are taken from a single station. In addition, Table 7 shows which data are used for each sub catchment.

This paper is meant to introduce and to provide a proof-of-concept of the adaptive clustering method, and we think this can be done at the example of a single, typical hydrological model. We agree with the referee that testing the method with other models and in other catchments is clearly desirable (see our conclusion on page 29, lines 1-9), and in fact we are currently doing so. However we suggest that this is beyond the scope of this paper. In a revised version of the manuscript, we will make better clear in the abstract and in the summary and conclusions the proof-of-concept character of the paper.

Comment 5: 4. One critical issue is that the definitions of some technical terms were vague, such as aggregated characteristics and dynamics of such systems, co-evolution, catchment-uniform, and multi-criteria estimation. Please explain them in detail or replace them with easy-to-understand terms to enhance the readability of the study.

Reply 5: Will do. Please see our reply to comment 2.

Comment 6: 5. Unfortunately, the grammatical errors, confusing sentences, redundant vocabulary, and an erratic writing style hinder the message that the authors want to convey, and in some cases, render some statements ambiguous or even mistaken. I recommend that the authors encourage further to undergo a resubmission process.

Reply 6: We will re-read the paper and re-write with shorter sentences and streamlined vocabulary where necessary. Also, in the production phase, the manuscript will undergo further copy-editing.

Comment 7: Data and methods section and results section are confusing, vague wording. I suggest elaborating on the description of the adaptive clustering. Otherwise, it is hard to understand how does analyze dynamical similarity.

Reply 7: Referee #1 also mentioned that section 2.2 (adaptive clustering) is hard to understand. We will re-write this section and the related Figures to improve comprehensibility.

Comment 8: The main steps in 2.2.1 section are suggested to describe in points.

Reply 8: Please see our reply to comment 7.

Comment 9: 6. The resolution of Figure 2 is low. The information cannot be extracted.

Reply 9: Agreed. In a revised version of the manuscript, we will provide all figures in higher resolution. For a final version, all figures will be provided in separate high-resolution files as required by HESS.

Comment 10: The explanation for Figure 3 and Figure 4 is difficult to understand.

Reply 10: We will improve the explanations. Please see our reply to comment 7.

Comment 11: 7. In the data and methods section, excessive writing space is used for introducing the SHM model and its structure, which are not vital in this study. The structure of this article is suggested to adjust and enhance readability.

Reply 11: Thank you for this suggestion, which was also made by referee #1. During manuscript preparation, it was also a matter of discussion among the authors whether the SHM model description should be at such a prominent place, as it just serves as a testbed for the demonstration of adaptive clustering. Nevertheless, as SHM has not been described elsewhere so far, and as knowledge of the model structure and parameterization is important to understand its behavior in terms of dynamical similarity, we think the model should still be presented in detail. We will move most of section 2.1 (The SHM hydrological model) to the Appendix. In the main text, we will give a very brief introduction to the model and will refer to the Appendix.

Comment 12: In the results and discussion section, the principal results and conclusions are suggested to summarize briefly.

Reply 12: Respectfully we maintain that the principal results and conclusions are given, in short form, in section 4 (summary and conclusions).

Comment 13: 8. How to estimate the weights in Eq. 12? I think weight has a significant influence on the streamflow simulation in different phases, which is essential for the applicability of the proposed framework. Moreover, the Nash-Sutcliffe efficiency only prefers the simulation accuracy of high flow.

Reply 13: Assigning weights to different components of a multivariate calibration is a difficult task, and there is no single-best, objective solution for it. Rather, the choice of these weights expresses the user's subjective ranking of the components in terms of importance, and/or trustworthiness. In our case, the weights express our subjective ranking of the discharge, soil moisture, and evapotranspiration data available for calibration and validation with respect to data quality (see p. 9, line 16). In a revised version of the manuscript, we will explain this point in more detail.

NSE: We agree with the referee that NSE as a measure of model efficiency has well-known limitations (see, e.g. Schaefli and Gupta, 2007). Nevertheless, it is still one of the most widely used performance measures in catchment hydrology, and therefore we considered it adequate for our goal of building a robust, state-of-the-art hydrological model. However, we did not only consider NSE during calibration, we also looked at mass balance errors for all water balance components (not shown in the manuscript for brevity: For discharge at the catchment outlet, it was 2.2%), and we did sequential calibration: Base flow parameters were calibrated during times of summer low flow only, interflow parameters were calibrated during times of rainfall-runoff events. We thus reduced the risk of overfitting to flood peaks, as mentioned by the referee.

Comment 14: 9. How to identify and estimate the representatives which have a strong influence on the performance of adaptive clustering.

Reply 14: For each cluster, first the number of representatives is determined by multiplying the number of catchments in the cluster with parameter 'perc_reps' (set to 10% in our study, see Table 6), which is then rounded towards an integer number. E.g. 51 catchments * 0.1 = 5.1 → 5 representatives. These 5 representatives are then randomly picked from the 51 catchments (see p. 14, lines 17-22). Random picking is a simple method, which leaves plenty of room for improvement, and we will test alternatives in the future. In a revised version of the manuscript, we will mention this is in the last section of section 4 (summary and conclusions).

Comment 15: 10. Importantly, the mechanism for the improvement of model performance was not discussed. For example, What operations lead to improvements in model performance (also involving high flow, middle flow, or low flow)?

Reply 15: We are not entirely sure we understand this question. We assume it is about how we selected the SHM Attert model structure and how we iteratively adjusted the model parameters during calibration. If this interpretation is wrong please let us know.

The model structural choice was guided by the findings of Fenicia et al. (2014) and Fenicia et al. (2016), in fact our model structure strongly resembles the models reported in them. Likewise, choosing geology as the main control of parameter variations within the catchment is based on these papers (see p. 4, lines 13-15, and p. 8, lines 1-2). Our calibration procedure is described on p. 9, lines 4-28. As explained in reply 12, our calibration strategy was sequential: Base flow parameters were calibrated during times of summer low flow only, interflow parameters were calibrated during times of rainfall-runoff events.

Yours sincerely,

Uwe Ehret, on behalf of all co-authors

References

Fenicia, F., D. Kavetski, H. H. G. Savenije, and L. Pfister (2016), From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, Water Resources Research, 52(2), 954-989.

Fenicia, F., D. Kavetski, H. H. G. Savenije, M. P. Clark, G. Schoups, L. Pfister, and J. Freer (2014), Catchment properties, function, and conceptual model representation: is there a correspondence?, Hydrol. Process., 28(4), 2451-2467.

Schaefli, B., and H. V. Gupta (2007), Do Nash values have value?, Hydrol. Process., 21(15), 2075-2080.