Responses to comments posted by Referee #1

We thank the first referee, Shervan Gharari, 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: 1- The use of English language is far from being perfect. It took me more than usual to read through the manuscript due to insufficient use of English language. The sentences are very long, the wording are sometimes very awkward. For example, the first sentence of the abstract is very hard to follow. It is amalgamation of information which, at the end, does not say much about the intention of this study (the intention being representative model with lower computational demand). Here I give examples from the text. Page 4, line 4, “offers full code control”; full code control is very subjective. Page 4, line 21, “. . .straightforward non-iterative forward-in-time numerical scheme”. Page 7, line 19 “majority vote”. Page 11, line 6 “for adaptive clustering to make sense”. Page 13, line7 “gold standard” instead of synthetic. Page 14, line 3, “keep things simple”. I would strongly recommend the authors to have this manuscript proofread by a native speaker.

Reply 1: 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 2: 2- The Introduction seems to be superficial. I would say the paper is about representation of the system in a model vs computational time/resource. In land [surface] modeling community there is significant body of literature devoted to the effect of grid size (computational burden vs spatial representation) and example of them can be Melsen et al., 2016 (and many more). This is the case in hydrological rainfall/runoff models as well (Liu et al., 2016 and many more). In its current form the Introduction starts with general reflection on sophisticated processed-based models; then moves to concept of co-evolution (which is not directly relevant to the message this study wants to convey) and then comes down to clustering. I would suggest to re-organize the Introduction to reflect on pervious works on computational burden vs spatial representation, clustering and its application in hydrological similarities and finally make it clear what the reader should expect from this paper.

Reply 2: We thank the referee for pointing at useful literature related to the effect of grid size in land surface modeling. We will integrate it in a revised version of the manuscript. However we do not agree that the introduction is superficial. We discuss conceptual modeling and its main shortcomings to show the merits of distributed modeling. We then discuss the computational challenges of distributed modeling and how they are mitigated by adaptive time stepping and adaptive gridding. We discuss the shortcomings of available adaptive gridding methods, namely that they require spatial adjacency, and argue that similarity of sub systems of natural systems is not necessarily limited to neighboring elements. Taken together, this shows that adaptive clustering i) can be useful, ii) it has novel aspects compared to existing methods of adaptive gridding, and iii) similarity is not only an artificial effect caused by representing natural systems in models, but it is also property of the real-world system (although the inevitable simplifications associated with representing real-world system in models can increase similarities). We therefore prefer to keep the structure of the introduction as it is.

Comment 3: 3- Section 2 can be better organized. There is still mention of the CATFLOW in this section. I would suggest the authors to shorten the text in section 2.1 and directly explain the model.

Reply 3: Thank you for this suggestion, which was also made by referee #2. 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 4: I suggest using transpiration coefficient instead of crop coefficient in Table-1 as the region is partly covered with forest.

Reply 4: The correction factors we use to adjust Penman reference ET from the reference surface (short grass) are available in Dunger (2006) not only for crops, but for a range of vegetation cover, including coniferous forest and deciduous forest. So while the term 'crop coefficient' is widely used for this kind of correction factor, it is not limited to crops (see e.g. http://www.fao.org/3/X0490E/x0490e0b.htm#crop%20coefficients), but we agree that the term is misleading. In a revised version of the manuscript, we will replace 'crop coefficient' by 'vegetation correction factor'.

Comment 5: It seems to me that the routing is only a linear reservoir (eqation-11). Can this representation simulation lag function or unit hydrograph which is often used in the models? I would say no, therefore the highest streamflow peaks are the same as precipitation peaks.

Reply 5: Thank you for raising this point. A single river element is indeed represented by a single linear reservoir. There are altogether 147 such river elements in the model, which together form a linear reservoir cascade. A linear reservoir cascade can reproduce both translation and retention effects associated with streamflow. Therefore, streamflow peaks in the model do not necessarily coincide with the precipitation peaks. To avoid misunderstandings, we will in a revised version of the manuscript mention the 147 river elements and the linear reservoir cascade.

Comment 6: The are kr and kb identifiable/related? it seems to a redundancy in the two processes/parameters.

Reply 6: Both parameters are indeed retention constants of linear reservoirs, however they strongly differ in magnitude. Kr is the retention constant of a single river element. As all river elements in the model are of about the same length (1 km), we used the same kr for all river elements. Its value of 1.1 hours was found by maximizing the agreement between observed and simulated streamflow in river stretches where up- and downstream gauges were available (this way we could see the effect of translation and retention in the river stretch).

Kb is the retention constant of the base flow reservoir. We determined kb values, separately for each geology (see p.9, lines 4-28), by maximizing agreement of simulated and observed streamflow during times of summer low flow. Kb values are in the range of 500 hours (for Schist) and 20000 hours (for Sandstone), i.e. at least two orders of magnitude higher than kr. We are therefore confident that there is little redundancy between these processes/parameters.

Comment 7: It is unclear how the computational units are set up, based on superposition of all the geospatial data or at the sub-basin level (later in Table-7 it becomes more apparent that the setup is at sub-basin level). Which data is directly used in setting up the model? I would suggest providing a table or bullet points for that; now it is very scattered around.

Reply 7: Indeed the computational units are the sub-basins (termed sub catchments in the manuscript). The setup of the sub catchments, and how land use and geology are assigned to each of them is described on p.7, line 15 – p.8, line 14. We agree with the referee that it would be helpful to show Table 7 (currently in section 2.3.2 'Adaptive clustering') already in section 2.1.2 ('SHM Attert'). The reason for placing it in section 2.3.2 is that it shows the static optimal clustering of sub catchments, which is an important benchmark for adaptive clustering. As we will move most of section 2.1 to the Appendix (please see our reply to comment 3), we decided to leave the table in section 2.3.2 also in a revised version of the manuscript, but will refer to it in the 'SHM Attert' section. Also, to make clearer which rain gauge data were used for each sub catchment, we will add to Figure 2 lines separating the area of influence of each rain gauge (please see our reply to referee 2, comment 4).

Comment 8: I suggest removing equation 12, to a separate Section with entropy as measure of performance and uncertainty. Why there is NS values report in this Section?

Reply 8: The main topic of this manuscript is to introduce and show a proof-of-concept of adaptive clustering. SHM Attert in this context is a means to an end rather than a central topic. We therefore decided to discuss all aspects of SHM set up, calibration and validation in short form and in a single section, and to keep it separate from the discussion of entropy related to system analysis and evaluation of adaptive clustering. We therefore prefer to keep the discussion of calibration/validation results and entropy separate, all the more because we will move most of section 2.1 (The SHM hydrological model) to the Appendix (see our reply to comment 3).

Comment 9: is the satellite based evaporation a result of more sophisticated model (such as a land model)?

Reply 9: Yes, the ET estimates are produced by forcing a SVAT model (a simplified version of the ECMWF TESSEL SVAT scheme) by Land-SAF radiation products (DSSF, DSLF and AL) and ECMWF meteorology. A detailed description is given in Trigo et al. (2011), section 3.1, and on the LSA-SAF pages (https://landsaf.ipma.pt/en/products/evapotranspiration-energy-flxs/met/).

Comment 10: Section 2.2 is the heart of this manuscript. I would say it should be presented separately in a Section before the model is presented as this approach is model independent.

Reply 10: We agree. We will move most of section 2.1 to the Appendix to give more visibility to the main topic of the manuscript (please see our reply to comment 3).

Comment 11: Don't give sub-section "main steps" title (2.2.1).

Reply 11: Agreed. In a revised version of the manuscript, we will use a different title.

Comment 12: why section 2.3.1 is called "hydrological system analysis"; it is about entropy as a measure of uncertainty (as NS is a measure of performance).

Reply 12: Actually the section is about the question of how to analyze hydrological systems in terms of the time-varying similarity of its sub-elements (see p. 17, lines 2-4), and we suggest entropy as a suitable measure to do so. However, we agree that the section title does not perfectly reflect that. In a revised version of the manuscript, we will instead use 'Entropy as a measure of hydrological similarity'.

Comment 13: Section 2.3.2 is again called adaptive clustering, similar to section 2.2. and again, in Section 2.3.2 the authors are referring to CATFLOW and MIKE SHE, etc.

Reply 13: Respectfully, we do not understand the concerns of the referee here.

Comment 14: 4- Section 3 is also rather hard to follow as well. The primary message of the manuscript is about tradeoff between spatial representation and computational resources (time). It is best to start from Figure-6 and 7 and then move to Figure-5 for example.

Reply 14: We agree that the primary message of the paper is about how adaptive clustering can reduce computational efforts of distributed modeling while maintaining, by and large, modelling quality. However, adaptive clustering is built on the observation that i) hydrological similarity among sub systems exists, and ii) that this similarity is time-variant. If i) would not be true, there would be no potential for clustering; if ii) would not be true, time-invariant clustering would do the job. We therefore think it is important to first show to the reader results with respect to i) and ii), which we do in section 3.1, before moving on to the main topic in section 3.2. However, the referee comment reveals that we can do better to show the connection between these topics. In a revised version of the manuscript, we will add a related explanation at the beginning of section 3.2.

Comment 15: Page 20 line 9, why it is "striking" that the entropies are lower than the uniform? I would always expect so. It is also expected that the entropy is lower for the recession and higher for rising discharge. This is kind of similar to the heteroscedasticity assumption on the error as well (more diffused with higher discharges). If only observation is used with varying error assumption, higher streamflow will have higher entropy and lower streamflow will have lower entropy.

Reply 15: We agree that it is no surprise to see entropies below the entropy of the corresponding uniform distribution. 'Striking' here refers to the fact that entropies are well below the uniform entropy, and often close to zero. In our opinion this is indeed noteworthy, and it shows the high potential for adaptive clustering. In a revised version of the manuscript, we will make this point clear. We also agree with the referee that hydrologists have known since long that the degree of similarity between sub systems varies with the hydrological situation. The point we want to make here is that i) entropy of normalized, binned distributions of states and fluxes expresses this in a conveniently dimensionless way, and ii) that we make use of this knowledge.

Comment 16: I don't really see the information the first and second paragraph of Section 3.1 provide.

Reply 16: These sections show several important points: i) hydrological similarity among sub systems exists, and ii) this similarity is time-variant. iii) the temporal and spatial patterns of similarity are in accordance with hydrological expertise, which increases confidence that our way of expressing similarity in terms of entropies is reasonable. In a revised version of the manuscript, we will summarize these points at the section end.

Comment 17: The third paragraphs, starting with "In Fig.6", is also kind of obvious. Are these what we have trained the model to do? The result is like a self-fulfilling prophecy (I will elaborate on this later).

Reply 17: Please see or reply to comment 16.

Comment 18: 5- The final Section is lacking proper discussion on what we have read in this manuscript. What can be the take home message for a modeler who want to model this basin in the future. A bullet point conclusion of this study is also appreciated.

Reply 18: Respectfully we disagree. The point of the paper is not about how to model the Attert basin in the future, it is about the potential of adaptive clustering (as the referee correctly states in comment 14). We will however check in both the abstract and the conclusions if we can convey this message better.

Comment 19: 6- For me personally, moving from the world of conceptual models to land models, I would like to question the motivation of this study. Although saving time is valuable but having method that needs model re-run or updating for more complex model is terribly cumbersome. This is the reason why the authors have chosen to use SHM rather than CATFLOW for example.

Reply 19: The referee correctly states one main motivation for this study: saving computation time. This is already a more than sufficient reason, as the referee will surely agree, as he also works on concepts to make high-resolution land surface modeling more efficient (Gharari et al, 2020). In addition, the concept of adaptive clustering yields valuable insights in the time-and space patterns of similarity among sub systems, which, we daresay, is a useful contribution to hydrology research. We have chosen SHM for the proof-of-concept as any hydrologist can easily connect to it, and hence we can show the effects of adaptive clustering in an environment familiar to most hydrologists. We agree with the referee that implementing adaptive clustering in more advanced models will be more demanding, but also the gains will be higher (see p. 29, lines 5-9). So it will be well worth the try.

Comment 20: I also didn't really understand the model set up, are we looking at comparison of models with clustering with a synthetic case? If that is the case, then the comparison of what is the best representation in a specific time of the year is dependent on the most elaborative set up and its simulations (basically the most computationally expensive model should be set up and simulated).

Reply 20: Yes, the benchmark for evaluating the model performance with adaptive clustering is the full-resolution model run (i.e. without adaptive clustering), which is our ground truth (see p.16, lines 20-22). As we have shown that the model correctly reproduces various observations (p. 10, lines 1-6), we are confident that this virtual reality approach is valid.

Comment 21: Also, one might say there are easier time stepping approaches for lower computational costs. For example, for the high flow the model temporal resolution can be maximum (1 day) and during recessions it can be up to couple of days. These type of approach to reduce computational time is much easier to implement.

Reply 21: The model time stepping is 1 hour (not 1 day) and was kept the same throughout all simulations (please see our reply to comment 23). We agree with the referee that adaptive time stepping methods can dramatically reduce computation times, without substantial losses of simulation quality. We mention this on p. 2, line 20. The point is that adaptive clustering can be used in addition to adaptive time stepping. In that sense, acknowledging the merits of adaptive time stepping does not render adaptive clustering redundant.

Comment 22: Moreover, if computational time is considered, I would say Matlab is not the best choice. Land models implemented in C or Fortran can handle hourly simulation (with much more variables, IO traffic) much faster.

Reply 22: We agree that Matlab is not the best choice for high-performance computing. But this is not the point of our manuscript. We compare the performance of models applying adaptive clustering compared to benchmark models operated in full resolution, and we are confident that the general conclusions we can draw from such comparison can be transferred from one programming language to another.

Comment 23: BTW, I also missed the modeling time stepping; please clarify.

Reply 23: The model time stepping is 1 hour (see p. 8, line 14).

Comment 24: I would suggest the authors to clarify their methodology of the clustering. Please make more time and maybe better visualization to convey the message here. As I said earlier, I would suggest the author to allocate a full Section for the clustering method they have proposed.

Reply 23: Referee #2 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.

Yours sincerely,

Uwe Ehret, on behalf of all co-authors

References

Dunger, V.: Entwicklung und Anwendung des Modells BOWAHALD zur Quantifizierung des Wasserhaushalts oberflächengesicherter Deponien und Halden. Habilitationsschrift an der Fakultät für Geowissenschaften, Geotechnik und Bergbau der TU Bergakademie Freiberg, DOI http://dx.doi.org/10.23689/fidgeo-668, 2006.

Gharari, S., M. P. Clark, N. Mizukami, W. J. M. Knoben, J. S. Wong, and A. Pietroniro (2020), Flexible vector-based spatial configurations in land models, Hydrol. Earth Syst. Sci. Discuss., 2020, 1-40.

Trigo, I. F., Dacamara, C. C., Viterbo, P., Roujean, J.-L., Olesen, F., Barroso, C., Camacho-de-Coca, F., Carrer, D., Freitas, S. C., García-Haro, J., Geiger, B., Gellens-Meulenberghs, F., Ghilain, N., Meliá, J., Pessanha, L., Siljamo, N., and Arboleda, A.: The satellite application facility for land surface analysis, International Journal of Remote Sensing, 32, 2725-2744, 10.1080/01431161003743199, 2011.