

Responses to RC 3 posted by Referee #1

We thank the first referee, Shervan Gharari, for the interest in our study and the replies to our questions. In the following, we answer to all of the comments one by one. The Referee comments are in blue.

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.

Re-Reply 2: What the authors mentioned here are clearer than the introduction at its current format. I would say keeping the interlocution as it is, is a disservice to the manuscript. There is only one paragraph about clustering in the introduction. For a reader, the concept of co-evolution is somehow presented as the main topic, while I think the concept of similarities of the sub-system behavior (grouped response units, GRU, hydrological response units, HRU;) should be more elaborated. The examples the authors mentioned, “north facing. . .”, actually fall very well in the concept of GRU and HRU and pave the way for better presentation in the manuscript rather than co- evolution. I agree that co-evolution is the process of creating

similarities/rules but in my point of view it should not be the main point of discussion here. I leave this to the authors and editor to decide.

Re-Re-Reply 2: We agree that the concept of GRU/HRUs, namely time-invariant grouping of model elements without the requirement of spatial adjacency is not discussed in the introduction, but relevant to the topic. We will add a related discussion to the introduction in a revised version of the manuscript. Nevertheless, we prefer keeping the part about co-evolution, as it explains why the similarities that we exploit by GRU/HRU or adaptive clustering methods occur.

Comment 6: The are k_r and k_b 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. k_r 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 k_r 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). k_b is the retention constant of the base flow reservoir. We determined k_b 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. k_b 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 k_r . We are therefore confident that there is little redundancy between these processes/parameters.

Re-Reply 6: sorry I mean k_i , for the fast reservoir, instead of k_b . Indeed, the slow reservoir does not even need to be routed through river network due to its long reaction time. k_i can be said to be at the scale of hillslope temporal response which should be more or less in scale of hours. Interested to know more about k_i and how it is interpreted from the field data and separated from k_r .

Re-Re-Reply 6: The time scale of river routing is indeed negligible compared to that of the base flow, so from a practical point there is not much of a difference whether base flow is routed or not. But as it actually enters the river before the catchment outlet (otherwise the rivers would fall dry in summer, which they did not even in the very dry summer 2015), we routed it. k_r we found, as mentioned in our initial reply, by maximizing the agreement between observed and simulated streamflow in river stretches where up- and downstream gauges were available. The values of k_i were determined by calibration. As mentioned in the manuscript, for the entire model setup and choice of reasonable parameter ranges, we relied on the detailed investigations and findings of Fenicia et al. (2014, 2016).

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/>).

Re-Reply 9: – So good that you gave a low weight to its NS value.

Re-Re-Reply 9: Indeed this ETP product is not a rock-hard observed truth, but nevertheless it was a useful benchmark to evaluate our model in terms of daily and seasonal ETP patterns both qualitatively and quantitatively.

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.

Re-Reply 13: I meant both Sections have the same title. Also again in the middle of the manuscript it is referred to CATFLOW and other models (page 18, lines 3-4).

Re-Re-Reply 13: We will change the titles in a revised version of the manuscript. But we still do not understand why we should not refer to examples of physically based hydrological models in this context.

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.

Re-Reply 15: I would say it is not still sticking. In my point of view any model set up (even worst ones) can easily show very good behaving entropy as they are mostly affected by forcing and memory of the forcing rather than parameters.

Re-Re-Reply 15: We would like to reply along two aspects the referee mentions: The first is about forcing as a control of similarity, the other is about a high degree of similarity observed in many models.

Forcing: We agree that forcing is an important control of sub catchment dynamical similarity, but it is not the only one. Just one example: Two hillslopes may behave identical during snowmelt conditions if they are exposed to radiation in the same manner. But they may function quite differently in summer if one of them is vegetated and the other is not.

Similarity: We agree that most distributed models show a high degree of similarity among the states and fluxes of its sub elements, which is exactly why methods exploiting this similarity such as adaptive gridding, GRUs or HRUs are so useful. However, by far not all models make use of these methods. We therefore think it is noteworthy to discuss similarity, and especially the time-variant character of similarity, to show the potential computational savings.

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.

Re-Reply 19: Yes, testing the method for the more sophisticated models is desirable of course. I just wanted to draw the attention of the authors that to the fact that running a more complex model means more technicalities. Given those technicalities, and time/resources to fix them, it is not really clear if the final gain will be higher. The technicalities can be how to efficiently read/write/update this adoptive clustering; how to efficiently do a warm start for a model; how to pass this over various processors if needed; do the mentioned models' capabilities allow such an approach? Etc.

I give an analogy of the sensitivity analysis of land models. Land models may fail (crash) for some given parameter sets therefore may not result in output values (objective functions) which are essential given the struct of parameter sampling method. This may cause issues for sensitivity methods which should be thought through. I would say adding one or two sentences on those technicalities/obstacles at the end might be useful for the reader.

Re-Re-Reply 19: We agree with the referee that it remains to be proven that the benefits of adaptive clustering will not be eaten up by its overhead in more sophisticated applications. And in fact it would be highly interesting to discuss with the referee the potential of adaptive clustering in the land surface models he uses. We are carefully optimistic about the potential benefits of adaptive clustering, as established methods to save computation time, like adaptive gridding and adaptive time stepping, also come with a considerably overhead and are nevertheless very effective. Nevertheless, in a revised version of the manuscript, we will add to the conclusions a short discussion about potential challenges when applying the method to more sophisticated models.

Yours sincerely,

Uwe Ehret, on behalf of all co-authors

References

Fenicia, F., Kavetski, D., Savenije, H. H. G., and Pfister, L.: From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, *Water Resources Research*, 52, 954-989, 10.1002/2015wr017398, 2016.

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., and Freer, J.: Catchment properties, function, and conceptual model representation: is there a correspondence?, *Hydrol. Process.*, 28, 2451-2467, 10.1002/hyp.9726, 2014.