Reply to Referee #1 Daniel Wright:

Daniel Wright ***(DW): Summary and Recommendation:*** *The authors present a framework for dynamically adapting the level of spatial detail re-solved within a physics-based rainfall-runoff model depending on the spatial variability in precipitation. I found this the be one of the most interesting manuscripts that I've ever reviewed, and commend the authors on this innovative work. Nonetheless, there are some issues that should be addressed before the manuscript is suitable for publication in HESS, and that could help maximize the impact of the work.*

Ralf Loritz (RL): We would like to thank Daniel Wright for his positive comments and the time he invested to review our Manuscript (MS). The revised MS will follow the reviewer's recommendations and include among other things a re-structured section 3 (model introduction) as well as a more extensive discussion about its connection to the land surface modeling community. Furthermore, will we carefully check the references the reviewer recommended and see whether they help us to improve our argumentation.

Major comments:

1. DW: *I believe the discussion could be strengthened by deeper consideration of how this approach would "scale up" to larger watersheds or regions. Part of my reason for encouraging this is that the land surface modeling (LSM) community is at least as concerned as the rainfall-runoff community about model computational demands of long-term/ensemble simulations, and are seeking ways of representing fine-scale (e.g. hillslope and below) over continental-to-global domains. In fact, land surface modeling was the focus of the well-known Wood et al. (2011) hyperresolution modeling opinion piece. In addition, there has been relevant progress in LSM development that the authors should cite. I will mention these below. But in terms of scaling up, the key aspects seem to be acknowledgement that heterogeneity of model parameters will increase with modeled area, while the rainfall spatial coverage will, on average, decrease.*

RL: Thank you for raising this point. We mentioned in our MS in the method section page 14 line 28 to 31: "*This entails, however, also that if we extend our research area to a catchment that is divided, for instance, into two geological settings that function hydrologically differently (regarding their filter properties) we would always need to run at least two structural different models where each of these models represents one of two geological settings.*"

We agree with the reviewer that this section is rather short and will provide a more extensive discussion in a revisited MS (see also the following points). We will also carefully read the proposed references by Woods et al. (2011).

2. DW: *I believe the discussion could also be strengthened by some discussion of how well this approach might fit with specific types of spatial discretizations. It fits quite naturally with hillslope-based models. The fit is less clear with gridded or TIN-based models-or at least with high-resolution gridded models in which individual model grids must "communicate" with each other to transmit water via overland or subsurface flow to channels.*

RL: This is an important point. Indeed our approach is limited to hydrological models that are based on a division of the landscape into partly independent spatial units (similar to the work of *Chaney et al. 2016*). However, at least in theory, there is no limit on how complex the interaction between these independent sub-units are as long as there is redundancy/similarity when different model elements "*communicate"* with each other. However, the question at what point of model complexity we would still save computational times by reducing redundancy depends on a series of factors (e.g. model, resolution, no. of processes and state variables). We will discuss this in a revisited MS.

DW: *It seems that the computational advantages of the approach might be limited in that case. In addition, models such as GSSHA in which overbank river flow can return to the land surface would have some limits here too. These issues are worth discussing because such models constitute important current directions in physics-based model development.*

RL: From our perspective, it makes much sense to divide a landscape into different building blocks such as hillslopes, sub-basins, etc. (*e.g. Zehe et al., 2014*). This is the case as current physically-based models are still constrained to small areas if they are set up on an appropriate grid size. We see hence no way around dividing a landscape into some kind of independent sub-units and either run models in parallel or/and group similar model elements (dynamically or time-invariant) if we want to work on larger scales. That said, we also believe that it would be rather difficult to implement a spatial adaptive modeling approach in a current model like Delft2d or GSSHA.

We wrote in our MS (Pg. 16 line 10): "*While we use CATFLOW as a model here, the proposed approach is not restricted to this model and can be used in any hydrological model that distributes a catchment into independent spatial units.*". To underpin this point we will discuss the limitation of our approach in more detail (please also see the discussion with the second referee and the third). Again we thank DW for this valuable comment.

3. DW: *While there may be other relevant LSM developments, the one that I am aware of is Hydroblocks (Chaney et al. 2016). While I recommend reading that paper, the basic approach is similar to this manuscript's in that spatial units are grouped into hydrologically similar clusters to reduce the computational demand.*

RL: Thank you very much for pointing us to the study of *Chaney et al. (2016)*. We will examine it carefully.

DW: *The difference is that in Hydroblocks, these clusters are not dynamically reassigned according to time-varying characteristics (unless the developers have recently added that capability). So in fact, your approach appears to be superior in some respects. Specifically, within Hydroblocks, since there is no dynamic reassignment, you can never have a cluster that extends beyond the spatial extent of a single precipitation grid cell, which means that their approach loses computational efficiency with higher-resolution precipitation datasets. Your approach thus seems to hold more promise in terms of flexibility to advances in precipitation inputs*.

RL: Interesting comment. We would like to highlight that we only showed that our approach is theoretical and practically feasible. It remains an open question if we could actually save more computational times in comparison to HydroBlocks or similar time-invariant approaches. The question is open for discussion as our approach is also more complicated. We will discuss this in a revised discussion section.

4. DW: *More clear description of what each model does and does not do is needed in Section 3. Specifically, I found it confusing the way that the models are briefly introduced at the beginning of the section, and then discussed further in various subsections. I also find it strange that you have text that is not assigned to specific subsections. It isn't clear why section 3.2.1 is needed...convention is that you don't include subsections unless you have at least 2 or 3 (i.e. 3.2.2, 3.2.3). This section structuring needs rethinking. […] Also, a table that compares the key features and differences of all the models could be effective. I think one think that would really help is to not use "model a", "model b", etc. but some brief descriptive names that actually help the reader understand and recall the differences. More important, I really couldn't figure out from the descriptions what the differences between some models were. I also don't understand the motivation for using a different rainfall dataset for the reference model and model a; this seems unnecessary.*

RL: We stated in our MS on page 11 line 26-29: "*We added model a to test if the performance difference between the reference model and our distributed model b is merely a result of quantitative differences between the different precipitation products measured either by a single ground station or by a weather radar.*"

However, we understand the comment of Daniel Wright and we will restructure section 3 entirely and remove the short introduction of the different models. We will also follow your advice and add a table with the key features of each model.

5. DW: *Zhu et al. (2018) and Peleg et al. (2017) both highlight how distributed rainfall structure is really important in determining flood frequency across a range of scales. Though I normally refrain from suggesting that authors cite my own work, in this case it seems appropriate to highlight these studies, since they do show that for extreme events, rainfall space-time structure is extremely important in determining hydrologic response even at very small scales (see Peleg et al. in particular), and that this importance varies with rainfall magnitude and basin size.*

RL: We have carefully read both publications and they fit nicely into our revisited discussion. Thank you for pointing us towards these two references.

DW: *Along with this, I disagree with the statement on pg. 27: "it seems that catchment size might not be the best indicator to decide if" a distributed model is needed. It probably is the best single indicator, but is still insufficient. I draw a somewhat different conclusion from your work: that a distributed approach is always needed to reap the full benefit of spatially distributed rainfall (at least in locations in which convective rainfall can occur), and that provides motivation for continued developments such as this into ways of handling this need in computationally-efficient ways.*

RL: What we wanted to convey here is that it is the combination of the drainage area and the average size of a typical rainstorm, which is important and not the drainage area alone. For instance, if you wanted to predict the runoff formation in the Colpach catchment in the winter season a spatially aggregated model driven by a single precipitation time series might be sufficient as our results show. This means also that you could invest your limited time and improve for instance the groundwater representation in your model instead of setting up a distributed model. However, if you wanted to make predictions in the summer months our results highlight that you need some sort of a distributed model to be able to capture the spatial variability of the rainfall. This means that only because the Colpach is 20 km^2 we cannot decide if we need a spatially distributed model as the catchment size does not explain how variable its meteorological forcing is. *Nicotina et al. (2008)* argued along these lines and stated that the "*total residence time of a water parcel is often controlled by the travel time within hillslopes, we find that when typical hillslope size is smaller than the characteristic size of rainfall structures (say, a correlation length of rainfall intensity), the rainfall pattern effectively samples all possible residence times and the response of the catchment does not depend on the specific rainfall pattern.*" and the second referee pointed us towards the study of *Ogden and Julien (1993)*. The second reviewer also nicely summarized their key finding: *"only for rainfall with durations shorter than the concentration time of a catchment does the spatial distribution of the rainfall matters, for longer rainfall events only the temporal distribution matters."*. Following these two studies and our own results we would argue that our first research question in our MS: "How important are spatial patterns of precipitation for the runoff generation at the catchment scale?" can only be answered if we combine information about the catchment size (e.g. average hillslope length, concentration time) with information about the meteorological forcing (e.g. intensity, correlation length, velocity). In a revisited MS we will rephrase this paragraph and explain in more detail what we meant by this statement.

DW: *Likewise, I disagree with the statement on pg. 30 line 18-19: compressing rainfall into a single time series isn't so important as the ability to only use as much computational power as is truly needed to solve the problem at hand.*

RL: Again an interesting point you raise here. In our specific setting, compression of precipitation and saving computational power are the same. By compressing the precipitation field to a single time series we also compress our model, minimize redundant calculations, which again means that we save computational power. So, we would argue that we first need to understand (test) how far we can compress our rainfall field without losing predictive performance before we can save computational times in a meaningful manner. The data-based / machine learning community most likely would disagree ☺

6. DW: *Some discussion of implications for calibration would be interesting. Is it necessary to calibrate using a fully distributed model? This would limit the usefulness of this approach in some respects such as automated calibration procedures.*

RL: Typically, one run of the reference model (a single CATLOW hillslope) for a simulation period of one year and hourly printout times takes about 2 - 3 hrs. Assuming that you run your code on a workstation with 32 cores you can run about 400 model setups in 24 hrs. As structurally similar areas are represented by the same model in our approach, testing different model parameters sets should be feasible even in larger areas if the structural properties are not too variable/complex. We will discuss this in a revisited MS.

7. DW: *There are a number of minor grammatical issues that nonetheless cause some distraction from the overall high quality of the manuscript. I will point out some of these below, but it could be worthwhile to have a native English speaker perform a careful proofreading.*

RL: We will carefully proofread the MS once more and would like to highlight that there will be another professional language check by Copernicus if the MS is accepted for publication.

References:

Nicótina, L., Alessi Celegon, E., Rinaldo, A. and Marani, M.: On the impact of rainfall patterns on the hydrologic response, Water Resour. Res., 44(12), 1–14, doi:10.1029/2007WR006654, 2008.

Ogden, F. L. and Julien, P. Y.: Runoff sensitivity to temporal and spatial rainfall variability at runoff plane and small basin scales, Water Resour. Res., 29(8), 2589–2597, doi:10.1029/93WR00924, 1993.

Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., van Schaik, L., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V. and Kleidon, A.: HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, Hydrol. Earth Syst. Sci., 18(11), 4635–4655, doi:10.5194/hess-18-4635-2014, 2014.