Reply to Referee #2 Ciaran Harman:

Ciaran Harman (CV): Summary and Recommendation: "*This paper addresses a very interesting idea and is very well written and presented. I have only one major concern, but it is right at the core of the matter and addressing it may affect the interpretation of the results."*

Ralf Loritz (RL)*:* We would like to thank the second referee for the time and the effort he put into writing his review. The points he raises are relevant and addressing them will help improving our manuscript. We hope that after this discussion (as well as after we revised our manuscript) all issues he raises can be clarified.

CV: *The main issue has to do with the calculation of entropy. My question, succinctly, is whether in equation 2 refers to hillslopes or to time steps? Or sometimes one, and sometimes the other? I have scoured the manuscript and find contradictory statements that could lead to both conclusions. The conflation between these seems to me to cast doubt on the assertion in equations 4 and 8 that the entropy can be used to determine the number of functional groups necessary to model the catchment in a non-redundant way, and that results suggest the ideal number is time-varying - major conclusions of the paper.*

However, there are other points in the paper that suggest this interpretation cannot be the case (at least not all the time). What would be the meaning of y in equations 3,5, 6 7? It would have to be the set of discharges at a different timestep from x. But then all the calculations would be about the mutual information between timesteps, and that is not what this paper is about. It seems then that x and y are now different hillslopes (one Shannon entropy for each hillslope). Now p (x) is the marginal probability that a timestep has a particular discharge at in a given hillslope. Also p(x) must be normalized such that the total probability across all timesteps for each hillslope is 1 (rather than across all hillslopes for each timestep). The p associated with a particular hillslope/timestep pair will differ depending on which marginal distribution it is being considered a part of.

RL: This is an important point and indeed the manuscript is separated into two parts. The first part deals with the identification of the required spatial complexity of a hydrological model (in our study this is expressed by the number of hillslopes). Here, we use the Shannon entropy (equation 2) as well as the concepts of the minimum number of questions (maximum compressibility; equation 8) to analyze how much redundancy with respect to storage or runoff we produce with our hillslope models (redundancy here means that two or more hillslopes return, at a given time step, discharge or storage values falling into the same bin). We calculate the entropy of the discharge ensemble (= set of discharge values from all hillslopes) for each time step and show that they are partly redundant. In addition, when looking at the time series of entropies, we see that the degree of redundancy varies in time. While we agree that the magnitude of Shannon entropy of the discharge simulations at a given time step does not necessarily related to dynamic similarity in general, it is indeed the case for the special conditions of our virtual experiment: Here the hillslopes differ only with respect to topography, all further parameters and the forcing are equal. In that special case, we can establish a direct link between cause (similarity of topography) and effect (redundancy of discharge).

The main conclusion of the first part of our study is that the model setup is in general compressible, and that the degree of compressibility varies with time. In the second part we make use of the first finding, but for the sake of brevity ignore the second. This means that we simply took the mean of the time-series of Shannon entropies from part one, which expresses the mean compressibility of the model over time. Clearly, for a particular timestep this may be an over- or underestimation of the true compressibility. We discuss the shortcoming that the spatial resolution of a compressed catchment model should also vary in time in detail in section 5.2.

However, to avoid confusion and to make this point clear we will add a section at the end of our objectives where we clearly explain the two main research questions of our manuscript.

CV: *However, if this is the case, the entropy would be maximized at any given timestep if the discharge from a randomly chosen hillslope has an equal probability of occupying any particular discharge bin. This conflicts with the definition given in equation 4. Wouldn't the maximum entropy be log(M) , where M is the number of bins? Given that you have log-spaced bins starting at 0.01 and increasing by 8.5% each time up to a maximum of around 0.7 mm/hr, it looks to me like you have about 26 bins. That would give a maximum entropy of about 4.7 bits. It seems important that 4.7 bits is slightly larger than the largest the entropy values you observe in Figure 4c.*

I find myself doubtful that Equation 4 and 8 hold. Equation 4 does not seem to me to have the meaning implied, for the reasons I give above - maximum entropy occurs when all discharges are equally likely and has nothing to do with the number of hillslopes. Similarly, the entropy of a coin-toss experiment is maximized if the coin is fair, and has equal likelihood of landing on either side, and has nothing to do with the number of times I flip it.

This doubt then extends to equation 8. I find myself uncertain as to whether it is valid to interpret the entropy in figure 4c as saying anything about the time-varying number of non-redundant hillslopes required to simulate the result. If the theoretical maximum is indeed 4.7 bits, does that mean we know, even before we run the model, that we will need at most 26 non-redundant hillslopes? Is this how the uncertainty in the discharge immediately limits the required model complexity?

RL: Thanks for this point. Also the second referee made this point which highlights that we need to address this issue in a revised manuscript. Most importantly, we must make clear that this discussion of maximum entropies contains two perspectives (experiment and system).

In statistical mechanics the maximum entropy relates to the log of the maximum number of different microstates of a system with a given energy. This leaves us with the question what determines the number of microstates in our study. We think that your interpretation of the maximum number of bins and our interpretation of the maximum number of models do not at all contradict each other but are supplementary.

Let's assume we have a fair dice with six possible outcomes. The entropy of this system is linked to the possible states the "*system*" could reach and would hence be $\log_2(6)=2.58$. Now depending on our "*experiment*" (maybe we lost our glasses) we only ask the question is the value larger or smaller than 3. In this case the maximum entropy of our "*experiment*" would with two possible states be $\log_2(2)=1$.

Following this line of thought, we argue that there is a difference between the maximum entropy of the system and of the experiment. In our manuscript the maximum information our model ensemble can produce about any given output is $\log_2(105)=6.7$ bits. If our goal is to simulate a distribution with a higher entropy our model is doomed to fail. On the other hand, as you and Mr. Weijs correctly commented: in our specific experiment, given the uncertainty in our discharge observation, it is unlikely that we reach this theoretical value as long as our models produce reasonable results. Log_2(No. of possible occupied bins) reflects hence a physical viewpoint of our specific experiment because we do not expect that our discharge is larger than a certain threshold. Here picking the maximum discharge value (0.75 mm/hr) seems reasonable (leads to 54 bins and a maximum entropy of 5.7). However, one could also argue that a value of 1 mm/hr is physical reasonable in this environment which would result in a number of bins around 58. Nevertheless, in our experiment it would still not lead to 105 bins meaning that our model will always produce redundancy.

As already stated in the answer to the first referee we need to make this clear in a revised manuscript and rephrase section 3.1.2 in this respect. We think that both concepts are useful if we want to identify the needed spatial complexity of a hydrological model and we are thankful that both referees pointed out the difference between the two concepts.