I appreciate the authors' efforts in revising the manuscript and responding to the previous round of review. However, I think there are still some critical logical issues that remain unresolved:

First, in the synthetic case (Fig. 2), the central inference seems to be: conceptual model same as the "true model" or over-parameterized ⇒ low entropy; model different or under-parameterized ⇒ high entropy. But this does not mean that models with high entropy are "wrong," nor that low entropy means "correct." Entropy is not a sufficient condition for the adequacy of a model structure. Over-parameterized models can have lower parameter entropies. Hydrologic models with more process representations may require a larger subset of parameters to be dynamic to accommodate their complex structure for good performance, whereas simpler models may need only a few dynamic parameters. Does this imply that the latter has a better physical representation?

In addition, if only streamflow data from the "true model" is used to constrain the training (pretending this is the only information we know), we might think that Model 3 provides a better representation of hydrological processes since it has two-layer buckets (soil moisture and baseflow buckets) and gives an almost perfect streamflow prediction. However, when more information about the model structure or other variables is available, we see that this model is actually most different from the "true model." This simple example demonstrates why additional constraints are needed to diagnose whether a model reflects "bad" physics. More rigorous experiments, e.g. adding additional constraints to the training (as suggested in the previous round of review), are required to demonstrate how entropy can be meaningfully connected with the physical representation of a conceptual model. The authors argue that "we're focusing on this relationship between model complexity and difficulty in prediction." Can I then understand that entropy is related to model complexity, not to the correctness of the physical representation? However, what we need is not an easier model but a more physically correct model.

Second, the entropy of LSTM states might not reflect parameter entropy. As noted in the previous review, we should not make all parameters dynamic. For the same model, the selection of dynamic parameters will change the entropy. In addition, the parameter ranges differ between models in the synthetic case, which might make comparisons and diagnosis problematic. A hybrid model with flexible parameter ranges can still simulate well even with the wrong structure, since the governing equations of buckets the simple conceptual rainfall–runoff models are similar to each other (summarized to Eq. 1).

Third, the hybrid model using all static parameters could have similar performance to the one with some dynamic parameters in predictions for ungauged basins. I am not sure why the authors claim: "The potential overwriting of physics constraints happens during the

training phase, and hence it is natural and logical to analyze it in the respective basins that these data are available for." One can still calculate the entropy of the LSTM and parameters. In this case, can we say the model using all static parameters has a better physical representation than the one with dynamic parameters? However, they use the same physical model. How the NN part of the hybrid model is designed also matters. Again, equating low entropy with "adequate physics" and high entropy with "physics being ignored" oversimplifies the problem.

I think these issues, along with those raised in the previous round of review, need to be addressed in the current work; otherwise, the results could mislead readers. Unfortunately, the authors defer them to future work, which prevents me from supporting the publication of this manuscript in its current form.