Dear Reviewer,

Thank you for this constructive review. Please find our first responses below (**in bold**).

Best regards,

Wouter Berghuijs (on behalf of all authors)

This is a very interesting paper on the long-term memory in precipitation, evaporation, streamflow and storage. The analyses are clearly described. The writing is clear and the figures are all very informative. I have no major comments and highly recommend publication of the manuscript in HESS.

Thank you.

My main comment is related to the structure of the paper. The model part is mentioned at the end of the introduction but is not part of the methods and almost came as a surprise to me. I would move some of the parts of the model section into the methods (e.g., the analyses and the description of the three model structures) and then divide the results into those related to the data analyses and those related to the model results. That way, a) the model section has a better division of methods and results/discussion, and b) the model part is part of the methods and doesn't appear to almost come as an afterthought. The model results are very interesting but I think that they could use a bit more discussion, e.g., it would be useful to highlight which lumped bucket type models have the tested structures, to highlight that this type of analysis may help to determine what type of model structure one needs to use, and that it means that any model structure testing with short datasets needs to be done with care!

In the revised manuscript, we will extend the modeling part in the methods section. In addition, we will extend the discussion on our modeling and its results.

My other comment relates to the datasets used. I fully agree with the choice of the datasets for ET and storage but it would be useful if there was at least some critical reflection of the datasets. Afterall, there is some "modelling" already involved in getting the "data". Thus, as with any data, there are some uncertainties in the data. There is currently no discussion on how this may influence the outcomes.

In the revised manuscript, we will provide critical reflections on the datasets and our use of these data.

Minor suggestions –really just suggestions:

- L98: Explain why you left out these arctic areas, rather than only stating that you left them out.

In the revised manuscript, we will clarify this.

- L126, L188: Considering all uncertainties, I would not include the decimal.

In the revised manuscript, we will not include this.

- L173: Is this 67% of the 79% or 67% of all the pixels? This could be worded more clearly.

In the revised manuscript, we will clarify that this is 67% of all pixels.

- L176, L181: Replace 'Spearman rank coefficient' by a symbol, but ideally not rho (see comment below).

In the revised manuscript, we will use r_s for the Spearman rank correlation

- L195, L199: Add symbol after 'mean' for greater clarity.

In the revised manuscript, we will add this.

- L209: I don't think that the header is very fitting to the contents of the section. In fact, I think that you can just leave the header out and include the text as a continuation of the previous section.

This header covers the entire 3.2 section, including empirical analyses (3.2.1) and model experiments (3.2.2), both addressing the physical causes of memory. Therefore, we do not think that removing this header is beneficial

- L226: It would be helpful for the readers if you gave your thoughts on why the larger catchments have a stronger long-term memory. Is it the presence or importance of larger (alluvial) aquifers? The fact that there are likely more (large) lakes for larger catchments? Or that larger catchments are overall flatter?

In the revised manuscript, we will discuss the potential causes of stronger memory in larger catchments.

- L236: I found it a bit confusing that rho is use for both the memory (r_y) and the Spearman rank correlation (r). Consider using r_s for the Spearman rank correlation instead.

In the revised manuscript, we will use r_s for the Spearman rank correlation

- L239-240: Are these very low correlations statistically significant?

In the revised manuscript, we will report these (significant) *p*-values