Review of the final manuscript: "Selecting a conceptual hydrological model using Bayes' factors computed with Replica Exchange Hamiltonian Monte Carlo and Thermodynamic Integration" by Damian N. Mingo, Remko Nijzink, Christophe Ley, and Jack S. Hale1

General comments

The manuscript has undergone a significant improvement compared to its previous version. It is now much more coherent and pleasant to read. The extensive explanation of the algorithm and numerical methods remains, but now with a much clearer presentation of the results. In particular, the results are now easier to follow and more cohesive. The authors have addressed all the feedback provided by me and Reviewer #2. Overall, I find the manuscript to be cleaner and more organized, with unnecessary clutter removed or rephrased, making it far easier to understand and engage with.

I recommend accepting the manuscript as it is. However, I am offering a few optional suggestions for potential changes, which the authors may choose to reconsider, though they are not required for acceptance.

Suggestions:

1. The Contribution section feels a bit too long, particularly the paragraph starting at line 110, which seems somewhat out of place and could potentially be moved…
2. Consider repositioning Figure 1, I believe it is generally preferred for figures to appear after they are referenced in the text.
3. In Equation 7, the meaning of 'D' is unclear—while it seems to represent a general discrepancy measure, it may be helpful to clarify this by adding a phrase such as, 'To compute the PPP, the chosen discrepancy measure, D, is calculated…'
4. It might be worth considering whether the methods section, starting from 2.3, should be placed in a separate section.
5. I would suggest moving Table 3, which currently appears on page 22, closer to its first reference on page 19, as it is relevant to the data presented in Figure 6.
6. There is some confusion regarding the use of capital 'N,' which is sometimes used to refer to the number of chains or replicas (e.g., Figure 2 and Algorithm 1) and at other times (e.g., page 21, line 445) to the number of posterior samples.
7. I noticed the absence of a concrete BF value, which is central to the paper. While it is mentioned that one can easily compute BF by just subtracting the values of the presented log marginal likelihoods (paragraph 455), I believe it would be beneficial to explicitly include a row in the results tables presenting the BF values, as these are among the most important results of the paper.

8. In section 3.3.1, the hydrographs presented in Figure 16 are analyzed, particularly the two peaks that the model(s) failed to capture, corresponding to the dates 04/05-02-1980 and 19/20-03-1980. However, it seems to me that the first peak not captured by the model's uncertainty band is actually closer to 20-02-1980, which to me, suggests a mismatch between the dates mentioned in the text and the data shown in the figure… though, it might be that I am wrong.
9. I found the last sentence in the conclusion regarding future work somewhat confusing. The sentence suggests combining REMC with NUTS to automatically turn all parameters in the HMC, but to me it seems like this idea was not previously mentioned in the text as a possibility in any form…