

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
<https://doi.org/10.5194/hess-2022-334-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2022-334

Anonymous Referee #1

Referee comment on "Hybrid forecasting: combining dynamical predictions with data-driven models" by Louise Slater et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2022-334-RC1>, 2022

Thank you for the opportunity to review Slater et al. "Hybrid forecasting: using statistics and machine learning to integrate predictions from dynamical models". Overall, I find this to be a timely and informative review. However, I do have a variety of comments, detailed below. I recommend at least a minor revision, if not a major revision.

My biggest concern in reading this paper is the number of different models and approaches etc. that are discussed. The paper is full of acronyms (so Table 2 is certainly helpful) such that I routinely found myself lost in the details and trying to remember the bigger picture or category that the details were supporting. If I'm someone coming to this review trying to figure out where to start with hybrid modeling, I think I would really struggle. How would I begin? Would I choose a model/paper from Table 1? How would I discriminate or know how to choose among the myriad of options? If the authors can provide some answers or guidance to these types of questions, I think it would be very helpful. Also, if there is any way to more clearly emphasize the main points even among all the details.

Terminology is really important in this paper. Can you please provide some definitions of the differences between physics-based vs. conceptual models?

One question I had was whether any hybrid schemes are currently operational. But, this is partially answered in line 93. Also wanted to see what the authors think it would take to make these models operational, which is partially addressed in the conclusion. Any further details that can be provided on this topic would be greatly appreciate (i.e., are there ANY examples of operational hybrid schemes? And if so, can they serve as pilot projects? i.e., what can we learn from their implementation that might help hybrid schemes become more widely used?).

Lines 100 and on list many hybrid models... but not all the references are in Table 1 as

well. Any reason? (e.g., Miller et al., 2021)

Section 2.4 seems to have a different focus than what is indicated on line 122.

The grammar of the sentence spanning lines 122-124 isn't quite correct. Same for the sentence spanning lines 273-274.

Lines 243: seems like a concluding statement (summarizing the overall point of the paragraph) is needed here.

Line 249: the reference to Madadgar et al., 2016 – where was this study applied?

Lines 264-266: Is this sentence a description of "mode-matching"? And if so, can that be made clear. If not, please provide a brief idea of what mode-matching is.

Line 409: by "national" does that mean the United States?

Line 440: what does "surface water" mean?

Lines 454-461: this paragraph, especially the last sentence, seems to imply there are no limitations to hybrid models.

Lines 491-509: are these paragraphs in the correct place? The information presented within seems to go in Section 2.1 on pre- and post-processing.

Lines 598-599: this is a really important point that I'm glad was made (i.e., the marginal improvement might be not worth the effort). It seems to me that dealing with this issue is critical to making hybrid schemes more widely accepted. Is there any way we can determine a priori the marginal improvement (without having to build both models in parallel and then compare)? For example, the Mai et al. (2022) study in line 616 – would be good to comment if the demonstrated superiority was enough to justify the extra effort.

Table 1: (a) Are any of these operational? (b) Any rationale for inclusion/exclusion of

studies in this table? (c) Can you add another column that describes how the statistical and dynamical models are combined? (d) Regarding column headings, in the text, "data-driven" seems to be the most generic term (lines 25-26) but here the column header is "statistical" model (and elsewhere, "empirical" is used). Again, the importance of terminology in this paper. (e) Would this table become slightly easier to digest if it was first sorted by predictand type (i.e., streamflow vs. reservoir, etc) and then horizon? I'm not sure, but I think that predictand is a larger category (and what I would first be interested in), then horizon.

Some acronyms that are not defined anywhere: RCP8.5, FV3GFS (this is just the name of the atmospheric model?), PREVAH (also a model name?)

Table 3: (a) Shouldn't "coupled" be included here also, since it is discussed in the text. (b) I find it interesting that Lee et al. (2002) is a primary reference for two of the options (serial and parallel) – given that it is now 20 years ago. Is that because it was such a foundational paper? Either way, can a more recent reference also be provided? As a corollary comment: It would be nice to have a discussion in the text of when these approaches were first tried (what was the foundational paper) on hydroclimate variables.

Figure 1: A few comments/questions on this graphic: (a) Please explain if the coloration of the boxes has any meaning. (b) Aren't large-scale predictors etc. also inputs to the hybrid forecasting scheme (not just dynamical predictors) – in other words, the straightforward left-to-right is not actually quite so straightforward? (c) Bottom middle: shouldn't it be "hydroclimate model" rather than "hydrological model" to be more general?

Figure 2: So, you obtain one value each for JJA, then take the max? Could be clarified in the caption text.