Hybrid forecasting: using statistics and machine learning to integrate predictions from dynamical models

Response to Reviewer 2
Anonymous Referee #2, 09 Nov 2022

Summary

This paper reviews - indeed it defines - the burgeoning field of hybrid dynamical-statistical hydrometeorological forecasting. The paper is timely and I believe it to be of wide interest to readers of HESS (and very likely beyond). I generally like to balance positive and negative feedback in reviews, but it was very difficult for me to find any suggestions to improve in this paper. It is skillfully organised, placing a very wide range of studies in sensible categories and highlighting specific themes with more detailed discussions of some papers. I didn't think there were really any major gaps in the literature and ideas they presented. The paper is also brilliantly written, with concise, lucid sentences making it an easy read - I believe even for non-experts. In short, in my view this review does everything a review should do: summarises the literature comprehensively, shapes the literature sensible themes, makes an argument - in this case the paper is essentially arguing for the recognition of hybrid forecasting as a distinct field (or at least a subfield within hydrometeorological forecasting) - and makes clear recommendations on the future direction of hybrid forecasting. I congratulate the authors on a remarkable review paper, one that I believe deserves to be widely cited.
Reply. We are most grateful to the Reviewer for this kind assessment of our work! The Reviewer's comments are copy-pasted below verbatim in black font, and our replies are in blue font. We label the comments in the following manner: "R2.C1" indicates Reviewer 2, Comment 1.

Specific comments

R2.C1: L33 "We do not provide a prescriptive definition of hybrid forecasting as it exists along a continuum, which may include a wide range of modeling and 'big data' type Earth Observation (EO) datasets" Fair enough - a sensible choice.
We are glad the Reviewer agrees with this choice!

R2.C2: L156 "ML models are also employed during the dynamical climate model simulations to correct model biases" I suspect the use of 'ML' to describe Bayesian techniques like Schepen and bias-correction methods like Meyer may be a bit unusual to many. Suggest the broader term 'statistical models' or 'data driven models' (consistent with the definition given in the introduction) to encompass all these.
We have updated this to "data-driven models".

R2.C3: L156 "The use of ML..." same issue with this paragraph - I would say that neither Bennett et al. nor McInerney et al. really qualify as ML - they are error models, which I think in general usage don't get lumped in with ML. These distinctions may well be arbitrary, but I'd suggest if the authors want to broaden the common use of ML to include a wide range statistical models that this be defined up front somewhere (in the way the authors have done with 'data-driven').
Fair enough - we have updated this paragraph to "data-driven models" also.

R2.C4: L453 "4 Key challenges and opportunities of hybrid forecasting" I guess I would add to the topics covered in this section the effective use of probabilistic forecasts in decision making. One of the major efforts in hybrid forecasting systems has been to achieve reliable predictive distributions; but it's not yet clear that this effort will necessarily result in better decisions. It's likely that automated decision systems/optimisation will be the means to take advantage of reliability in ensemble distributions. In my view this still requires considerable research - existing methods of optimsation do not necessary take advantage of this property. But I also understand that this may be outside the scope of what the authors wish to address - the paper is really comprehensive in the areas it does choose to address, so they may feel they cannot do this area justice (even if they agree that it is worth discussing). I will leave it to the authors to decide whether this is worth including in their paper.

We entirely agree with the Reviewer that the development of probabilistic forecasts and their subsequent uptake in decision making (and potential for improving decisions) is an important topic to address. In revising our manuscript, we will review the existing literature and see if it is feasible to include this point in the section on Challenges and Opportunities.

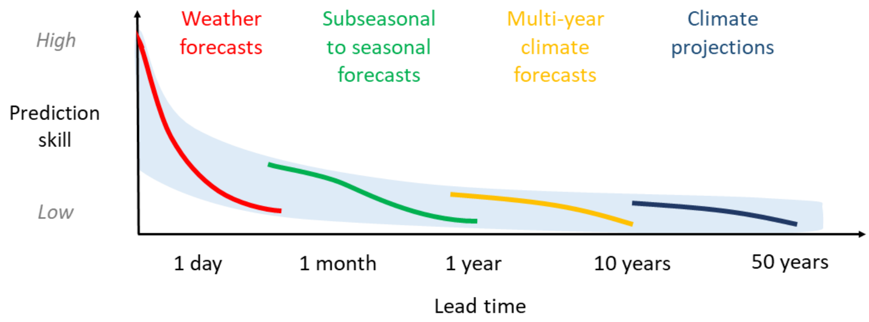
R2.C5: L456 "ML models include the requirement for large datasets (previously discussed)" This review presents the availability of large datasets for ML as a strength of ML - which it of course is - but it presents few of the difficulties associated with using these datasets for prediction, for example some of the 'curse(s) of dimensionality' described by Altman & Krzywinski (2018). ML models are still subject to some of these issues - though I realise canvassing these is not the main aim of the paper. Whether these matters are best discussed in this paper is a subjective judgment: I am happy to defer to the authors on this point.
This is a nice suggestion. We will attempt to include a couple of sentences on the difficulties associated with the use of large datasets for hybrid prediction, based on this reference. There may also be difficulties applying some techniques for subseasonal-to-seasonal (S2S) prediction that are not an issue for shorter range (and longer) forecasting because the S2S sample sizes can be so much smaller (often nearer 100 versus thousands for shorter ranges).

R2.C6: L465 "data-driven models were once thought to be unable to accurately predict values outside the range of the training" I'm not sure this is really true (or if it is, I haven't been exposed to it) - would be good to provide a reference in support of this statement. There is a long history of statistical extrapolation - not least in extreme value theory or design engineering - for exactly these purposes.
It is interesting that there seem to be different opinions on the question of data extrapolation by data-driven models. We will carefully review the literature on this point and revise the sentence accordingly (either delete it or provide references).

R2.C7: L487 "Explainability is sometimes useful to help develop trust in model predictions" this is a very interesting point - in my experience forecasting agencies frequently engage in this kind of story-telling, both for internal and external communications, so this is probably an important box to tick for the widespread adoption of hybrid forecasting systems. I'm not suggesting any change here, but I guess I also feel this kind of narrative building can be antithetical to the effective use of (usually carefully constructed) probability distributions that come out of hybrid forecasting systems.
This is an interesting point for discussion too. We will attempt to update the text to reflect the different perspectives on the use of storytelling/narratives versus the use of probabilistic forecasts, based on the literature. One reason for providing explainability of a prediction model is that when predictions evolve from forecast to forecast for a given variable (e.g. spring runoff), stakeholders want to know why (i.e. what has changed). This explainability can be important, but there are indeed cons, and we will discuss these in the revised manuscript.

R2.C8: L536 "For low flows skill may currently extend up to 20 days, but this is mostly due to the quality of the information on initial conditions and the memory effect of catchment storage" this statement may be true specifically for the study by Fundel et al. 2013, but it is phrased more generally. It is quite possible to get forecast skill of streamflow well beyond twenty days - even with simple ESP methods - (depending on catchment, time of year, etc.) so I think the authors should avoid a statement that posits a general limit on the prediction of streamflow of 20 days. Please reword this so that it is clear that this finding was specific to Fundel et al.
We have reworded this sentence so it is clear the finding is specific to Fundel et al., and that skill can be obtained beyond 20 days in other cases. Thank you.

R2.C9: Fig 4: As you've used 'prediction' generically in the vertical axis label ('Prediction skill') - implying (correctly in my view) that all the models in this plot produce predictions - I suggest changing the label "Subseasonal to seasonal predictions" to the more specific "Subseasonal to seasonal forecasts" and the label "Climate predictions" to "Multi-year climate forecasts".
These are excellent points, thank you very much! We have revised the figure accordingly, as shown below.



Typos etc.

L50 "While conceptual hydrological models..." suggest a paragraph break before 'While'
Done.

L71 Suggest paragraph break before 'Historically...'
Done.

L83 "to understand to which" typo - delete second 'to'
The sentence has been rewritten for clarity.

References

Altman N, Krzywinski M. 2018. The curse(s) of dimensionality. Nature Methods 15: 399-400. DOI: 10.1038/s41592-018-0019-x.
Thank you for the reference, which has been added to the revised manuscript.

Thank you for the helpful review!