Comment on hess-2022-312
Anonymous Referee #2

Referee comment on "Inertia and seasonal climate prediction as sources of skill in lake temperature, discharge and ice-off forecasting tools" by François Clayer et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-312-RC2, 2022

Review of “Inertia and seasonal climate prediction as sources of skill in lake temperature, discharge and ice-off forecasting tool” by Clayer et al

The study assesses whether seasonal (meteorological) predictions can be used within a hydrometeorological forecasting tool to produce skillful lake temperature, discharge and ice-off forecasts. In general, the topic can be of interest for the community and gives a nice example of how seasonal (meteorological) predictions can be combined with hydrological models to produce application-relevant and user-specific outputs that provide a baseline for decision makers.

However, I struggle with the manuscript in its current form. Hydrometeorological prediction systems are per definition interdisciplinary as it combines meteorological and hydrological models. It is further complicated when the output is further used by decision makers. When reading the manuscript I got the impression that the authors do not have a strong background in hydro-meteorological forecasting systems and use terminologies that are very uncommon in this field. These imprecise formulations make the understanding of the manuscript very difficult. I suggest the authors to (further) familiarize with the literature body about hydrometeorological forecasting systems and to follow more closely the terminology used in standard literature.

In general, the language should be much more precise and terminology should be used from the community. Overall, it seems a bit like the manuscript is a composite of multiple different text/styles. I suggest unifying the manuscript to enhance its readability. Use commonly used terminologies throughout the manuscript (abbreviations, model descriptions). Delete obsolete/imprecise sentences or specifically mention the related numbers and where they can be found (e.g. 158-159: “…scores for hydrological and lake
modelling were calculated.” But there is no indication where in the manuscript the scores are provided.

Necessary clarifications:

windows of opportunity: From the manuscript I do not understand how “windows of opportunity” are defined. As this is a crucial baseline for the analysis, the authors should make an effort to properly describe how these windows are selected and what exactly the temporal resolution (or aggregation) of the forecast data is that they use to calculate the scores.

Hindcasts: Fore each forecast that is produced, the prediction system calculates as well the hindcasts for the same date in the past XX years. It is not clear to me, which hindcasts have been used explicitly by the authors as they mention hindcasts from years (1994-2016). I assume they thus use all hindcast from the forecasts produced during 2017. In addition, on the same line it is written that the hindcasts from 1993-2016 are use, what might just be a typo.

Another point is the source of predictability. Sources of predictability are from my point of view physical processes and/or connections within the atmospheric/hydrological system. E.g. a source of predictability are sea surface temperature that influence the large scale dynamical patterns, such as for example in the ENSO or NAO, or as a hydrological example, initial conditions of snow or soil moisture can be a source of predictability for river discharges thanks to the memory of the system. I struggle with the terminology used in this manuscript that assigns a seasonal forecasting system as a source of predictability. It is rather the boundary condition, provided by the seasonal forecast system as an input to the hydrological model, that can be seen as the source of predictability. This is already a problem in the title. I suggest the authors to carefully revise the manuscript and the title. The analysis rather aims at determining if seasonal predictions can be used to produce skillfull lake temperature, discharge and ice-off forecasts.

SEAS5 system does have a resolution of 1° whereas ERA5 has a resolution of 0.25 degree. The Authors mention that ERA5 data was used to bias correct the SEAS5 hindcast. Due to the mismatch in resolution this correction inherently exhibits a downscaling step. Please elaborate what the actual hindcasts are that are used to run the hydrological models and the lake models.

Furthermore, how is the bias correction implemented? Do you use a leave-one-year-out methodology? There is not enough information about this pre-processing step in the manuscript.
More specific comments

Introduction

Line 24-26: two model outputs are compared: seasonal lake hindcasts (forced with SEAS5) and pseudo observations (forced with ERA5). In the next sentence it says “the seasonal lake hindcasts was generally low but higher then SEAS5 climate hindcasts”. These sentences are confusing, what is analyzed in the SEAS5 climate hindcasts exactly. Do you mean the skill of the meteorological predictions (SEAS5 outputs) are worse then the skill of the hydrological predictions?

Line 53-55 “Hence, seasonal climate forecasts are usually not the main source of predictability outside the tropics, at least for stream flow (Greuell et al., 2019; Harrigan et al. 2018; Wood et al. 2016)”

I struggle again with this formulation: Sources of skill can be the initial conditions (e.g. snow, soil moisture) or the forcing variables (e.g. temperature forecasts that determine the skill of evapotranspiration). But the seasonal forecast itself cannot, at least from my point of view, be seen as a source of predictability. I suggest that a careful reformulation of this (and similar) sentences throughout the manuscript. In particular, the publications referred to in this sentence can be used as a starting point and the formulations and terminologies used there should be used as well in the current manuscript.

Line 70 “...temperature predictions and forecasts” from a meteorological point of view forecasts and predictions are basically synonyms. Thus, I suggest not using both terms to make the manuscript more readable.

Line 72:

“...it doesn’t take seasonal climate forecast ensembles i.e. climate data products specifically designed for seasonal forecasting...” this formulation is again confusing for me. The output of seasonal forecasts are seasonal forecasts and not a climate data product. I suggest to avoid such sentences to make the manuscript more clear.

Line 76 & 77: The authors mention “When forecasting river flow, for example, predictability can originate from two main sources: (i) catchment water stores of initial soil moisture, groundwater, and snowpack, which are directly linked to the water residence time; and (ii) seasonal climate prediction (Greuell et al., 2019). “I agree that there a two
main sources (1) the initial conditions and (2) the boundary conditions, i.e. the relevant variables from the driving meteorological forecasts. Again the formulation that “seasonal climate predictions” are a source of predictability is misleading and is to my knowledge not used in literature, as it is a very vague formulation. This makes me feel, without being rude, that the authors should invest more time to familiarize with the commonly used terminology in hydrometeorological forecasting and try to carefully revise the manuscript.

Line 80-84: "When dealing with standing water bodies, antecedent conditions are also likely to provide significant predictability, given that the water storage in lakes and reservoirs is large compared to river channels, providing higher inertia. Water residence time is thus expected to exert a strong influence on water flow predictability. Water temperature, on the other hand, is influenced by multiple meteorological variables, e.g., wind, and radiation, in addition to water stores which can affect the source of its predictability."

I do not understand what the authors try to argue, what is meant by “water flow predictability”, do you mean discharge of the rivers? Lake level heights? Can you make this sentence clearer?

Line 90. I suggest to quickly introduce what ice-off is.

Line 90-91: You mention that you quantified the forecasting skill of each meteorological variable. This is later provided in Table 4. However, this table is hard to understand and does not provide quantified values of the fair RPSS and the ROCSS. Per definition a skill score ranges from one (perfect forecast) to zero (= no skill of the forecast) and negative values indicating that the forecast is less skillful then the reference. Here the authors just indicate if a given skill score is significant for the given variable and tercile. Why don’t you provide the actual skill scores and the numbers? And what is your definition of a significant skill? Is everything above 0 a significant skill? Furthermore, the abbreviations of the variables are non-intuitive and seem to be directly from a model output. I suggest using more common/readable abbreviations (eg. rsds = rad or solar radiation). In addition, the variable rlds is only given in this table but nowhere in the manuscript described, what is the meaning of this variable?

Methods:

Line 101: Introduce the definition of ice-off already in the introduction section, this helps the reader to understand from the beginning what you aim at.

Table1: Maybe it helps to add a map, where these reservoirs are located. Are they in complex terrain? In arid or humid climate regions?
Line 109: Climate data

I suggest changing the name of this paragraph to e.g. meteorological input data. Seasonal forecast data is not really a climate dataset, it is in a forecast dataset.

Line 111: what do you mean with "relatively homogeneous spatial and temporal coverage"? Both datasets, the reanalysis and the forecast data are datasets produced by global earth system models and thus provide a global coverage and are continuous in time as defined by the model integration steps.

Line 114-115: Please add a bit more detail about the bias correction. Did you use a leave-one-year-out approach? What quantile mapping approach did you use? How do you deal with values above the 99th percentile? Do you use an additive or a multiplicative method? How do you account for intervariable dependencies? Although there are some more information in the supplement of the indexed publication, I think some more information within this manuscript will help the reader to understand what you have done in this study.

Line 116: Here you use the term “impact models” and “impact variables”. Are these the water quality models and variables? In the rest of the manuscript the terms “catchment models”, “hydrologic models”, “lake models”... It makes it hard to follow, when you jump between the different models. I suggest unifying these terms and use one definition throughout the manuscript.

Line 125

Here you give more information about the EQM (referred to my comment for line 114-115). I suggest referring to these lines already at line 114. Nevertheless, it is crucial how the bias correction of the seasonal forecasts is done, I suggest adding some more information about how it is actually done. In addition, did you look at the performance of the bias correction? How do skill scores change before/after the correction? In forecast verification, a specific score only shows part of the story, it could be fruitful to look at additional scores and measures, such as e.g. reliability diagrams to get more information about the full ensemble.

Line 127-128:

Here the variable rlds is missing which is shown in Table 4 later. In addition, the abbreviations of the variables are non-intuitiv. I suggest using standard, or more readable abbreviations (e.g. temperature = T, pressure = p or even use there full names in the
text). In addition, do you use daily values or lower temporal resolution? Is the daily air temperature a mean daily air temperature? Are these the parameters you really use from SEAS5 and ERA5? Is the air pressure the surface air pressure? You mention wind speed here, but use the u and v wind components, that’s completely fine, but then I suggest only mentioning the wind components and not the wind speed, as wind speed is a different variable then (i.e. a combination of) u and v.

Line 130-131: Please describe the data you used and where it was measured? Is the station directly at the lake/reservoir inlet? Is it an official observational station? Who is responsible for the measurements (public agency? Scientific group?)? why are there so many data gaps? How trustful are these observations? What method is used to estimate the discharge? What temperature do you use (daily mean, daily max, daily min)? There is no need to include all this information, but at least a reference about how trustworthy the measurements are would be necessary.

Line 142: Here you use the variable abbreviations, as mentioned before, I suggest using more intuitive abbreviations or writing the full names of the variables, to improve the readability of the manuscript.

Line 143-145:

The authors mention that all hydrological models have been validated against observations. Do you have any publication you can refer to? What is the performance in terms of NSE for the calibration? Were the models calibrated for the same time period/locations? Was the same observational dataset used? Please give some more information.

Line 152: Maybe it is worth to introduce these performance measures in a separate chapter, e.g. together with the skill scores.

158-159: “Most common statistical goodness-of-fit parameters, e.g., Kling-Gupta efficiency (KGE), NSE and RMSE, for hydrological and lake modeling were calculated.”

What is the results in terms of these scores? Specifically mention the numbers, otherwise this sentence is obsolete. In addition, I struggle with the formulation “goodness-of-fit paramters”, KGE, NSE, RMSE are performance criteria or scores to evaluate the model performance. I suggest revising this formulation and use more commonly used formulations from the field of forecast verification.

Line 167-168: What do you mean with “ total predictions skill” and what observations
have been used to assess the performance? Do you here refer to the hydrological part I assume? I suggest a more careful formulation to better discriminated between the meteorological and the hydrological parts.

Line 173-174: “Over the initialization month, the 25 members of SEAS5 progressively diverge from ERA5 to their respective SEAS5 member.”

How is this transition done? In the text you say initialization month, whereas in Figure 2 you mention transition month, I suggest unifying the terminology to avoid confusion.

Line 174-175:

“Model outputs for the final 3 months, i.e., the target season, were selected and used to calculate the probabilistic forecasts of seasonal summary statistics.”

This sentence illustrates what I mean with imprecise language. “The model outputs for the final three months” are already the probabilistic forecasts and for these forecasts, the scores are calculated. I suggest carefully revising this (and similar) formulations. In addition: Does it mean that you only use 3 month lead times? The seasonal forecasts provide predictions up to 13 moths, why do you use only 3 months lead time? I think this should be mentioned in the very beginning of the manuscript (and already in the abstract), as it is important to know what lead time horizon you are focusing on.

In addition, it is not clear to me how you do your verification. Is it based on daily values? Weekly averages? Monthly averages? The full 3 month period? Do you do a lead time dependent verification as well?

Line 191:

“Both skill scores are expressed as relative to a reference forecast, i.e., climatology.”

It would be worth to more carefully explain the concept of skill scores. A Skill Score is a comparison of two scores, one calculate for the forecast of interest and one for a reference forecast. Thus again this sentence is somehow imprecise and should be revised. Hence, for calculating the skill scores, what climatology did you use? I assume you used the climatology based on the pseudo observation experiments, is that correct? If so, please mention this in the manuscript.
I do not understand the concept of “windows of opportunity”. Can you better explain how these windows of opportunity were selected? And why only these windows were taken into account for the evaluation?

Table2 and full paragraph:

I struggle with understanding your procedure. Is the “Evaluation data” in table 2 the data used to compute the reference forecast for the skill scores? If so, please change the name from evaluation data to reference forecast data or similar. For evaluation of a forecast, you anyway need explicit data at the same temporal resolution as your forecasts (being real or pseudo observations). If you use a climatology (based on real or pseudo observations) as your reference forecast, how is the climatology constructed? How long is the timeseries that you take into account and how many data gaps do you allow for constructing your climatology based on real observations. If you mention that there are large gaps, it might be difficult to construct a reliable climatology.

Line 205-207: these two sentences are a repetition. I suggest reformulating it.

Line 200-227: I struggle a lot to follow the explanation. First of all what is ROCSSs/ ROCSSw/ ROCSSw+t

Refer to table 3 early in this paragraph, then it becomes already a bit clearer!

Line 209-230: Again the formulation makes it very difficult to follow and the abbreviations further reduce the comprehensibility. I suggest to reformulate this paragraph and emphasize on better understanding.

Maybe it would help to choose simpler or more descriptive titles for the paragraphs. E.g. 2.2.2 could be sensitivity analysis to initial conditions and meteorological forcing (input periods); and 2.2.3 Sensitivity to individual input variables.

Line 231: “..each Lake_PO variable” do you mean with each output variable of the Lake_PO experiment? State that explicitly, it will be much easier to understand!

Line 240-244:
Is this now based on your analysis? Or are these theoretical considerations? This is not entirely clear to me. In addition, the formulation is a bit misleading, does it mean that the variables that are retained are the variables that are used from seasonal meteorological predictions to run the hydrological models?

Results

Line 251: you refer to table S2 in the supplement. I suggest adding part of the table in the main manuscript and refer to “additional information” in the supplement.

Line 262: Here you mention fair RPSS, which accounts for a limited number of ensemble size. However, this is not introduced in the methodology section, there you should at least mention that the fair (or debiased) RPSS is used and what it accounts for.

Again, what do you mean with “significant fair RPSS”? When do you assume an RPSS to be significant? How do you test the significance of a skill score?

Line 269-270:

“Only 0 to 10% of the SEAS5 climate hindcasts are skillful, on average”.

I struggle understanding this result. The RPSS (or any score) is usually determine based on a large sample forecasts (or hindcasts). Of course, for an individual forecast this might be poor but the full picture of forecast performance can only be revealed when the Scores for many issued forecasts are investigated.

I think here you should add a figure with the results where the reader can see how the hindcast performance actually is. From the text and the table alone, it is rather difficult to follow your argumentation. In addition, what is the definition of a skillful hindcast in your context? Is every hindcast with a ROCSS>0.5 skillful or do you use other thresholds? It would be crucial to mention the numbers at least once in the results section as well.

Line 278-280: How many seasons are discarded due to missing observations? Please indicate the exact numbers such that the reader knows how many samples (hindcasts) are actually used for the analysis. Or at least refer to the table where the numbers are listed.
You mention that the ROCSS does not capture the same as the “goodness of fit statistics” do. This is by definition the case as they do not look at the same properties of a forecast. Therefore, for forecast verification multiple scores should be taken into account to properly assess the forecast performance. Can you rephrase the sentence to make it clearer what you actually mean?

Table 4

Again general comments: abbreviations are non intuitive. FRPSS is not introduced before. I do not get the message of this table. I would prefer to see the skill scores (e.g. as boxplots) over different seasons for the variables. It is not clear to me what temporal aggregation is the baseline of this analysis.

Table 5

Here you use abbreviations for the seasons (SP, WI, AU, SU), although it can be inferred I suggest to avoid introducing additional abbreviations here. Again, how do you determine the significance of the ROCSS?

Fig. 31 should be Fig. 3

Can you elaborate how the ROCSS is determined, what exact values are taken into account? Do you use daily values to calculate the scores or weekly/seasonal aggregated values? This is still unclear after reading the manuscript.

Fig 3: it is confusing that in the plot description and the text you mention ROCSSs etc. but on the x axis S-Sa W-Sa etc are displayed. I suggest to unify all and make the plot more readable.

Maybe because I do not understand what the windows of opportunity are, I do not get the message here. I suggest to more explicitly formulate what the impact is of changing the initial conditions and the forecast input. The result indicates that it is not worth using
seasonal forecasts at all, which is hard to believe. Please elaborate such that the message gets clearer.

Line 323: again the title does not reveal to me what will be considered in this paragraph. I suggest to use a more appropriate title for this paragraph.

Figure 4: It is hard to follow what is shown here. What is the relative sensitivity. Maybe it helps if you refer to the exact paragraph number in the label of the figure. In addition, is there a reason why you use a color coding and a size coding? I suggest either using color or size, otherwise it seems that multiple aspects are coded.

Line 347-348: "Hence, a significant fraction of predictability is originating from the SEAS5 dataset although the largest source remains ERA5 data over the warm-up". This sentence illustrates what I think makes the manuscript complicated to follow. The reader himself must make the connection what this means. If I am correct, it shows that the initial conditions are more important than the driving meteorological predictions. Is this correct? It would make the manuscript much more readable if you directly refer to formulations what it actually means in addition to just give such “cryptic” explanations.

366-368: ... sources of seasonal water quality skill... You mean water quality forecast skill, water quality itself does not really have skill, does it?

Line 394-395:

Literature on streamflow hindcasts broadly shows that beyond the transition month, climatology-driven hindcasts are typically 395 more skillful than hindcasts driven by seasonal climate predictions (Arnal et al., 2018; Bazile et al., 2017; Greuell et al., 2019).

This is a misleading interpretation, when you say “climatology driven hindcasts”. In all three publications a well established ESP (ensemble streamflow prediction) approach is used. This can be seen as a climatology driven hindcast, but for a scientific publication I would expect to have a clearer formulation. In addition, in these papers there is no transition month mentioned, a concept I do not understand. All papers mention the first lead time month. What exactly is the transition month in your analysis?