Responses to comments on Manuscript HESS-2022-312 | Research article by Clayer et al.

Original title: Inertia and seasonal climate prediction as sources of skill in lake temperature, discharge and ice-off forecasting tools

Revised title: Sources of skill in lake temperature, discharge and ice-off seasonal forecasting tools

Editor's and reviewers' comments are underlined for clarity. Line numbers refer to line in the revised (clean) version of the manuscript, if not specified otherwise.

Editor:

Both reviewers have indicated that this is an important topic, but they have also raised important issues requiring considerable work. Thank you for responding to both reviewers. However, your responses remain very descriptive, and it is challenging to assess whether the changes will sufficiently address the issues raised by the reviewers. Together with the revised manuscript, please provide a detailed point-by-point responses to all reviewers' comments.

Thank you for your time and consideration. We provide below a point by point response to your comments and to each reviewer's comments.

I have finally highlighted two points that I believe need special attention when working on the revised manuscript: * clarity. Both reviewers have stressed that the clarity, languages and terminologies have not reached a publication level. See specifically the detailed comments from RC2.

We have taken special care in rephrasing many imprecise sentences and have tried our best to use the most precise terminology. We believe the manuscript has been improved significantly.

* Retention time. Given the short retention time of the selected lakes, the advective heat might be highly important (see RC1). Is it realistic to not take into account the heat change from the throughflow? See for instance: Schmid, M., & Read, J. (2022). Heat budget of lakes. In T. Mehner & K. Tockner (Eds.), Vol. 1. Encyclopedia of inland waters (pp. 467-473). https://doi.org/10.1016/B978-0-12-819166-8.00011-6 Råman Vinnå, L., Wüest, A., Zappa, M., Fink, G., & Bouffard, D. (2018). Tributaries affect the thermal response of lakes to climate change. Hydrology and Earth System Sciences, 22(1), 31-51. https://doi.org/10.5194/hess-22-31-2018

<u>We had overlooked this in the previous version of the manuscript, but indeed, we have taken inflow water temperature into account in the heat bugdets. A side analysis, for review only, shows that the assumption the reviewer 1 raised doesn't hold. See our response to his comment for additional details.</u>

Reviewer 1:

This article treats the important and so-far under developed field of seasonal forecasts (here 4 months into the future) of lake and drainage area properties including water temperature, ice-off, and river discharge. The authors combined two lake models (simulating surface and bottom temperature and ice cover, two lakes each) with four Hydrologic models (simulating discharge, one drainage area each). The method was applied at four lake-river system located in Norway, Spain, Australia, and Germany. Modeled systems include lakes spanning 19 o 60 meters depth and with a retention time from 0,2 to 1,1 years.

The coupled model setup was calibrated towards measurements (lake temperature and river discharge) and forced with reanalyzed data from ERA5, I would define this as a general circulation model (GCM). Hydrological-lake model performance was evaluates with KGE, NSE and RSM. Thereafter calibrated models was spun-up during one year and forecasted discharged and lake surface and bottom temperature during four months (one month initialization) spanning 13 years (1993-2016). Future forcing comes from 25 forecasts from the global forecasting tool SEAS5. SEAS5 was bias corrected and downscaled (grid adjustment) towards ERA5 to enable comparison.

The correctness of the forecasts (Lake_F) was evaluated trough a sensitivity analysis, comparison of Lake_F towards in-situ measurements and towards daily pseudo-observations (Lake_PO, daily output from the coupled hydrological-lake model setup forced with ERA5). The end product of this manuscript consist in an evaluation (sensitivity analysis), of forecasting correctness for each river-lake system and a evaluation of forcing parameters influence on forecasts.

The manuscript show potential but is lacking in some areas which I list hereunder.

Thank you for this thorough and helpful review. We have carefully rephrased many imprecise sentences and have streamlined the manuscript. We also have added details on the case-study sites, observation data, our modelling procedure and how we dealt with inflow water temperature. Below are our responses for each of your specific comments.

The manuscript and language has improved but as my replies show hereunder it still needs to be made clearer for the reader.

Chosen drainage areas and lakes

The authors put forward that seasonal predictions work best next to the equator and worsen with increased latitude (line 50 to 58). Yet, no system was chosen in this region, Spain being the closest. The manuscript could still benefit from an analysis of latitudinal effects for the used forecasting method to improve forecasting towards the North/South pole.

Unfortunately, this is out of the scope of this study. We only have looked at four case-studies outside the tropics to investigate any opportunities and have "ready to go" workflow when seasonal meteorological predictions improve significantly. See L. 126-129

My comment did not entail an extended survey adding more systems, but rather if latitudinal effects could be distinguished in the present setup from the lakes and rivers you used. Furthermore skill is related to each individual lake in the manuscript. I could not find details regarding the impact of each river model (SimplyQ, mHM, GR..) nor lake models (GOTM, GLM) on the forecasting results. This need clarification in the manuscript, at least in the form of a discussion in order for the reader to correctly interpret your results.

Additionally, the river-lake systems chosen contain lakes with very short retention time, i.e. big impact of rivers on water constituents, including temperature. The model method used include the effect of changing lake volume, but not the effect of heat being transferred into the lakes by upstream drainage area (input temperature I could not find). Thereby it is reasonable to assume that the lake models (through calibration) had a better connection between surface and deep waters than is the in-situ case.

Could this show up in your analysis of forcing parameter importance ("Tracing of forecasting skill" section 3,4, Fig. 4)? This needs to be addressed/analyzed since you link forcing to lake processes, which in fact could be caused by upstream heat fluxes in the drainage area and not in the lakes themselves.
Admittedly, this was not clear in the manuscript. Nonetheless, we did take water temperature of the inflows into account. This is now announced early in the methods section, see L. 117-118 and described in detail at L. 207-216.
As a side-analysis, for review only, we have looked into the correlation coefficients between lake surface and bottom temperature for observations and pseudo-observations (modelled temperature with ERA5 data as forcing data). This analysis revealed that the correlation for pseudo-observations was not necessarily higher than for observations discrediting the reviewer's assumption that "lake models (through calibration) had a better connection between surface and deep waters than is the in-situ case" and giving further confidence that the heat lake budget is robust. In fact, the correlation coefficients for both observations and pseudo-observations ranged between 0.22 and 0.96 depending on the season and case-study. In given seasons, the observations even showed higher connections between surface and bottom water than pseudo-observations.
Now, we don't believe that adding inflow temperature in our analysis in section 3.4 would add any significant insight since inflow temperature is largely based on air temperature which is already accounted for. In order to avoid any redundancy in our analysis and following principles of parsimony. Note that we have added an assessment of the various lake heat fluxes in the supplementary information and refer to it at L. 330.
My concern was that river temperature looked to be excluded from the simulations, now it is apparent that it was included so I am fine with the authors additions to the manuscript. I looked at the annual heat fluxes in the S.I., the through flow was between 5 and 12% of the total heat flux and at 3 out of 4 lakes it came at 4^{th} place (e.i. in order largest to smallest: short-, long-wave, latent, throughflow and sensible heat fluxes).

Data

This manuscript use ERA5 reanalysis as a stand in for in-situ measurements. Why is this, due to large spatial extent of drainage areas? If possible, show how this influence your modelling locally, or refer to documents where the reader can find this comparison between ERA5 and in-situ measurements, in best case for the regions being analyzed.
We forced our models with ERA5 meteorological data to ensure that our workflows were comparable between each case-study and future transferability of these workflows. We clarify this now at L. 126-129. In addition, weather observations covering the whole range of variables needed to force our models were not available over the whole period from 1994 to 2016. Yes, this has likely influenced our modelling locally. Note however that, for ground truthing, we include a forecast verification step compared to a reference forecast based on observations which is not often included in forecasting studies (Table 7).
This is good, but why do you have data gaps in verification statistics in table 7? I.e. you have $ROCSS_{original}$ which require Lake_PO and Lake_F values but do not show verification statistics for Lake_PO.

Clarity

The manuscript could benefit greatly from an index defining the many acronyms used, as well as improved description of tables and figures. Ex Table 4 and 5 is hard to understand.
Thank you. We have added an index at the end of the manuscript. L. 577.
Good, but have a look so that you haven't missed any acronyms, I found R^2.

Furthermore, I could not find/understand if the drainage area and lake models are coupled in time (run simultaneously), or if the drainage area models where run in advance to provide discharge for the lake models.
This is now clarified L. 114-118
Good

The language

Certain words in the manuscript cause some confusion. Bellow I have stated some that might need to change

Skill – is associated with people. A fast car (a tool) has no skill it has performance, the driver on the over hand has skill. That said, I know skill is used more commonly to describe models (tools) in meteorology than hydrology. So I suggest that you define what you mean by skill if you want to keep this formulation.
Skill is now defined twice in the introduction (L. 53 and 88-89) and repeated in the results (L. 347)
Good

Climate & climate prediction – studies involving effect of climate focus on longer time periods (>30 years) than what is the focus in this study (<1.5 years). Both SEAE5 and ERA5 comes from global GCM models, which could be used for climate studies. But in the context of this manuscript I do not think this is the right phrase describing the models you used.
Agreed, there was some confusion between climate and meteorological variables. This is now clarified throughout the manuscript. We rather refer to "seasonal meteorological forecasts/predictions" and mention explicitly the variables, when possible.
Good

Hindcasts – is usually used in the setting of running models with data from past events, close to reanalyze with the aim to improve said models. Here this word is used in combination with SEAE5 forecast simulations. The authors have adjusted these to ERA5 (real data proxy) but the intention is still to use SEAE5 as forecasting forcing. Therefor consider other alternatives in the manuscript, or define this word in the context of your manuscript.
Hindcast is now properly defined in the introduction, see L. 93
Adjust this sentence to read clearer, otherwise acceptable

Water quality – for drinking water and the biosphere, temperature is considered an important water quality parameter. Here we do not look at lakes and rivers in this sense, water quality one would assume here to entail dissolved constituents (nutrients, oxygen…). To avoid misunderstanding, consider using something else.
Agreed, we have replaced water quality by relevant alternatives throughout the manuscript, e.g., water temperature, lake.
Good

Line 19 : "as previously presented". Avoid need for reference in abstract.
We have removed this part from the sentence.
Good

Line 67 : Consider adding the following reference https://doi.org/10.1016/j.watres.2020.115529
Thank you!
Good

Line 72 to 74: partly untrue, air2water can run perfectly with seasonal forcing as you do here (only air temperature as forcing), and ice-off is currently available indirectly.
The sentence has been updated. L. 71-72
This air2water model is constrained to stop at 0 °C, i.e. ice-on and ice-off is indirectly modelled while other ice processes such as ice thickness is omitted. Adjust text accordingly.

Section 2,1,1 : The reader do not know where the lakes and drainage area (rivers) under investigation is situated. Add a map showing the global location and regional extent of each drainage area-lake system (rivers and lakes). Additionally add these system details (names, stations, etc.) where appropriate, ex. Table S1.

We have added a map (new Fig. 1) and detailed catchment maps are given by Jackson-Blake et al. 2022 as well as more background information on the case-studies. We now refer the readers to this study. See L. 110-111.

Good, but I read through the following paper and could not find the detailed catchment maps referred to in Figure 1. Add spatial catchment extent of upstream rivers that was modelled. Furthermore, using a figure already published requires permission from original creator, now since the author list below is similar to the list in this manuscript I leave it up to the editor to decide if a written permission to use the adjusted figure is required or not.

Jackson-Blake, L. A., Clayer, F., de Eyto, E., French, A. S., Frías, M. D., Mercado-Bettín, D., Moore, T., Puértolas, L., Poole, R., Rinke, K., Shikhani, M., van der Linden, L., & Marcé, R. (2022). Opportunities for seasonal forecasting to support water management outside the tropics. Hydrology and Earth System Sciences, 26(5), 1389–1406. https://doi.org/10.5194/hess-26-1389-2022

Line 111 or 112 : add reference: SEAS5: the new ECMWF seasonal forecast system. Stephanie J. Johnson, (2019), https://doi.org/10.5194/gmd-12-1087-2019

Done

Good

Line122 to 123 : „Climate data where downloaded….". What do you mean in this sentence, ERA5 and/or SEAS5?

This sentence has been removed and we now provided much more details on ERA5 and SEAS5 data pre-processing steps, as requested by the other reviewer. See L. 139-150

Good

Line 139 ")" missing

Thank you, Corrected.

Good

Line 156 to 157 : Add details (equations and ex. RMSE) of this linear regression between in- versus outflow.

This is now described in more details L. 195-206 and in the supplementary material.

Good

Figure 2. consider showing mean of SEAS5 predictions and ERA5 at the same time (i.e. continue black lines into transition and target season).

We already show the mean of SEAS5 predictions, adding a line would overload the figure (now Fig. 3). Besides, the main concept of this figure is to illustrate the input data used to force the model. ERA5 data was used only over the warm-up period.

Good

Line 185 : RPSS looks to be missing from table 2 and table 3.

FRPSS was not considered for formal forecast verification because it doesn't allow to distinguish forecast performance for a given tercile. With these low performance lake forecasts that we have reported, this FRPSS didn't seem very useful to us. However, we still include it in Table 6 for reference.

Good

Line 235 to 238 : something is missing here, hysteresis should make linear relationship between ex. air temperature and water temperature rather bad. Describe how good these linear fits were (in appendix). And/or show with figure and improve explanation.

Pearson partial correlation coefficients (PPCC) are calculated from seasonal means, and not daily values, which likely yielded much cleaner correlation than expected from daily values. This point is made clear now see L. 321.
Good

Line 243 the reader are not familiar with the contributions of local heat fluxes at the chosen locations. Before disregarding for example cloud cover from the analysis, show the reader in numbers (or preferably figure as appendix) for each lake the seasonal heat budget contributions. I.e. uptake and emission of infrared longwave radiation, evaporation + condensation, sensible heat flux and uptake of surface downward solar radiation. Throughflow you only have the outflow (at some lakes?) since inflow temperature is missing.
This background information is now described L. 207-211 and we refer to the supplementary information L. 330 for further details.
Good

Line 250 : RMSE not consistent with RMSE/sd in Table S2. What is RMSE/sd? Use the same in text as in Table S2.
This is now clarified, these first performance measures are for the whole year (now in Table 4) while Table S2 shows measures by season. RMSE/sd is now defined in the caption of Table S2.
Good

Table 5. move description of asterisk under table and improve the site representation. Now you can not see what belongs to which system. And define the season duration.
Done.
Good

Figure 3 and 4 : missing Germany and Australia, add or explain.
Note that we precise that the four probabilistic sensitivity analyses, S-SA, W-SA, W+M0-SA and OAT-SA were only performed at the sites in Spain and Norway because of the significant resources needed to execute these hindcast experiments (see L. 296-298). This point is now repeated in the figure captions (now Fig. 4 and 5).
Good

Figure 4 : Something do not add up in your analysis. Top row for Spain – Bottom temperature, and Norway - Surface temperature appear to be to large compared to the individual season values taken together. I.e. if the impact is small most seasons, I do not see how it could be much larger on an annual basis.
We agree with you that it can appear quite surprising that the sensitivity calculated on an annual basis is much larger that over each single season. However, Norway's climate is subjected to strong seasonality which can be well captured over annual scales, but still have low correlation for a given season. To make this point clearer, we added a sentence in the figure caption (now Fig. 5). See L. 451-454.
I don't follow you reasoning here. One would expect that surface temperature (ST) in summer is sensitive to for example short wave radiation (SWR), but during a complete year this sensitivity should be lessened by decreased importance of SWR for ST. If the seasonal signal intrude into multiple time frames (summer, autumn etc.) I would expect that the dependency is carried over from time window to time window and thereby show up for each season in your analysis.
Furthermore from the new figure (Fig. 5) it is now more clear what the size of the circles represent. Yet I could not find the definition for R^2, is it the coefficient of determination? If so, then there is an error in the figure and on line 319 since a high coefficient of determination (towards 1) show good correlation, thereby $1-R^2$ would give small and not large circles for more influential input variables. Is this so?

Figure 5 : Why so many data gaps? Consider showing seasons where significance is worse (higher) but clearly state which significance level you trust.
This figure has been updated to show all correlation coefficients, even those that are not significant, and we now precise the significance level we trust in the figure caption (now Fig. 6) at L. 458-459.
Good, but use different color range for PPCC, try for example white color around 0.

Line 468 : add author contributions. Who did what?
Done, see L. 626-629
Good

Additional points
It need to be more clear what is included in the lower, middle and upper terciles used throughout the manuscript. As an example it is now hard to understand Table 5, which show that your predictions can be better for lower and upper terciles compared to middle. I understand it as terciles composing a diversion of the model output so that extremes end up in the lower and upper terciles. With this understanding it now looks like from Table 5 that you are better at predicting extremes rather than the majority of events occurring in your system.

Add the following from Supplement to the main manuscript
You can find all the codes and data files related to this manuscript at: https://github.com/NIVANorge/seasonal_forecasting_watexr

Figure 4: Extend y-range of plots, now boxplots are cut-off by the x-axis

Line 239: Incorrect title number labeling