Response to the referee #1 of the research article egusphere-2024-2962: Skilful Seasonal Streamflow Forecasting Using a Fully Coupled Global Climate Model

Gabriel Fernando Narváez-Campo, Constantin Ardilouze

February 26, 2025

We thank the referee for the feedback. We will prepare a revised manuscript addressing the comments. We have organized this reply document as follows:

- The referee comments are in black.
- Our responses are in blue.
- The additions/modifications proposed for the manuscript revision are in red.
- Figures prepared for the reply have the prefix "R", such as Figure R1.

Revisions

1. The author evaluated summer and winter streamflow forecast one month in advance, I wonder whether two or even three lead months are considered (not 3-month mean)? Are the results consistent with the conclusions in the manuscript, or are the results of the online coupled forecast system used superior to the ESP, as the authors state in the introduction, "while modified versions of ESP can improve streamflow predictions for shorter lead times, their skill decreases faster over time compared to NWPB systems", maybe for longer lead times, the effectiveness of online coupled forecast system in predicting seasonal streamflow improves more.

   We understand the comment regarding the lead time skill dependence and thank the referee for the opportunity to clarify these key points in detail in the following and the revised version of the manuscript. We demonstrated that the discharge 3-month mean excluding the initial month (see Fig. R1) is better predicted by the coupled forecast system (e.g., see Figures 7d and 7h of the manuscript). However, the effect of lead time has not been explicitly discussed. To address this, we have performed a monthly lead time analysis to verify that our findings regarding the 3-month mean are consistent for a 1-month analysis, taking lead time dependence into account.

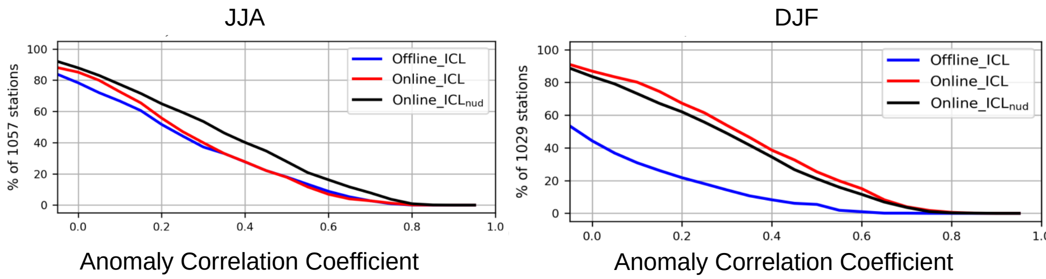

Figure R1. Global performance of forecasting systems for summer JJA and winter DJF seasons (discharge 3-month mean). Cumulative frequency distribution of anomaly correlation coefficient of hindcasts, for 1993-2017. (Same Figures 7h and 7d of the paper).

Figure S4 shows the cumulative frequency distributions of ACC for each of the four months forecasted. From the figure, we can remark on the following points.

- As expected, the performance decreases with lead time in all the forecast system configurations.
- In boreal summer JJA, the hindcast with improved land initialisation Online_ICL$_{nud}$ outperforms Offline_ICL benchmark at all lead time months. Conversely, Online_ICL is better than Offline_ICL for lead times over one month, as indicated by the slightly higher number of stations with positive but low correlations (about ACC<0.3).
- In boreal winter DJF, Online_ICL and Online_ICL$_{nud}$ performance is similar but always better than the Offline_ICL. Online_ICL is slightly better than Online_ICL$_{nud}$ in January (for 0.2<ACC<0.6) and February predictions (0.1<ACC<0.3).

The remarks are consistent with the conclusions derived from the 3-month discharge mean presented in the manuscript.

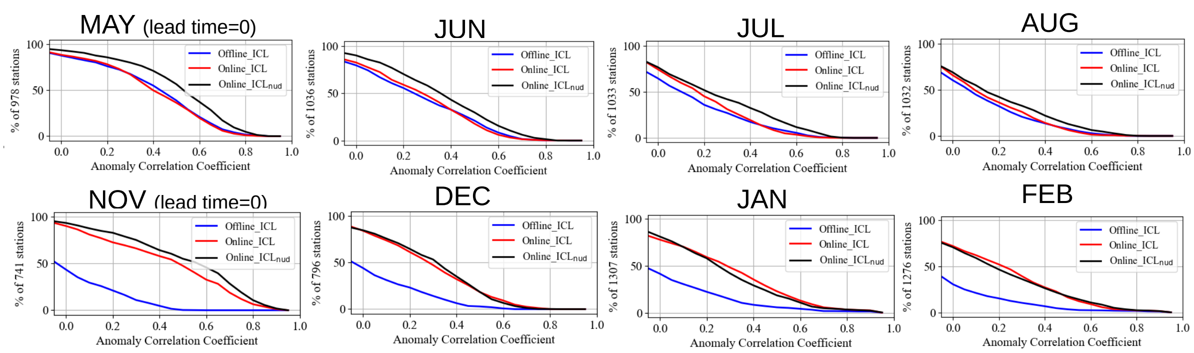

Figure S4. Global performance of forecasting systems at different lead times for summer JJA and winter DJF boreal seasons. The anomaly correlation coefficient for 1993-2017.

To strengthen our conclusions derived from the 3-month discharge mean, we will briefly analyse the monthly lead times in the manuscript, analogous to what is presented here, with additional figures located in the supplementary material.

2. Figure 1 demonstrates comparison of model configurations, but it is not sufficiently intuitive to understand and what the numbers in the figure represent is not explained. Also, "to generate the benchmark hindcast Offline_ICL, the land-river model ISBA-CTRIP is forced by ERA5 historical climate (Figure 1) so that each year produces one of the 25 forecast members" (Line 143), why does each year produce one of the 25 forecast members? The author mentioned 25 members several times, what specifically does members refer to?

   We thank the referee for the comments on improving the article's readability. To include the considerations in section *2.2.2 Forecast experiments*, we propose enhancing Figure 1 (and caption) to include the configuration details more clearly while rewriting the referred paragraph to enhance the readability.

   To generate the benchmark hindcast Offline_ICL, the land-river model ISBA-CTRIP is forced by ERA5 historical climate (Figure 1) so that each year produces one of the 25 atmospheric forecast members. We use leave-three-years-out cross-validation (L3OCV) to select the forcing. In L3OCV, the year of the climate forcing cannot match the hindcast year nor the preceding year and the two following years to avoid artificially inflating the skill due to large-scale climate–streamflow dependence with influences lasting from seasons to years like the North Atlantic Oscillation (Dunstone et al., 2016). For example, to apply the L3OCV selection method to the hindcast of 1993, forcing of years 1991 and 1996-2019 ensures 25 members. For the hindcast of 2000, forcing from 1991 to 1998 and 2003 to 2019 is used. Unlike in the current hindcasting for validation, in operational forecast systems based on the ESP Offline approach, future climate information is unavailable; thus, only past climate information can be employed.

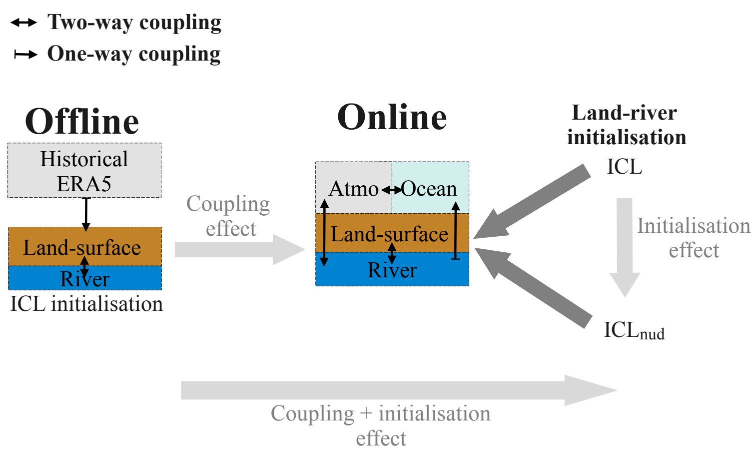

   Figure 1. Schematic of offline and online forecast system configurations and corresponding land-river initialisations. ICL: initial condition from the historical run with the online system; ICL_{nud}: initial conditions from a historical run with soil moisture relaxation to fields reconstructed from the offline land simulation SMR. As illustrated by the grey-filled arrows, the design of the experiment allows the evaluation of the coupling effect, the initialisation effect or both.

3. In chapter 3.2, the author shows the performance of the atmospheric seasonal forecast is presented in Figures 5 and 6, in particular, precipitation and near-surface temperature. Please highlight in the figures where the author mentioned in the paragraph. In Figure 5, the ACC of global precipitation is overall lower in Online_ICLnud than in Online_ICL in summer, especially in South America and Australia, also Online_ICLnud has more blue areas than Online_ICL in winter that means more negative ACC of precipitation and near-surface temperature. Can the author give some explanations?

   From a global view, we showed that nudging the soil moisture (SM) towards the reconstructed fields has a positive impact on the streamflow May initialisation (Figure 3i) and, therefore, on the associated JJA forecast (Figure 7d). Meanwhile, the effect in boreal winter DJF is negative (Figure 4i and 7h). This effect is consistent with the reduction in the anomaly correlation coefficient (ACC) for both precipitation and near-surface temperature of Figures 5 and 6. The atmospheric impact in summer is lower, as runoff production tends to be driven more by the initial water storage in the basin at the time of forecast initialisation rather than by atmospheric conditions.

The dependence of the atmospheric forecast performance on different land initialisations reveals the critical role of land-atmosphere coupling, which can positively or negatively impact the atmospheric forecast skill. We believe that the negative impact of ICLnud initialisation on the precipitation and temperature seasonal forecasts is linked to the SM nudging, which is expected to improve the variability of soil water content (hence the positive impact on the forecast of river streamflows) but can induce adverse effects on land-atmosphere coupling simulated by the model. For example, the initial soil moisture conditions brought by the offline nudging technique may shift the coupled system away from its equilibrium state. When the forecast integration starts, the nudging constraint is switched off, and the model adjusts to its equilibrium, producing potentially spurious heat and water fluxes (including for the snowpack, if any) at the land-atmosphere interface. Ultimately, this could alter the atmospheric circulation and degrade the temperature and precipitation forecast skill. This makes it difficult to provide further explanations for the regional reductions in ACC noted by the referee (for example, in the northeast of South America or Australia) without performing a deeper dedicated analysis to reveal any causation effect (e.g., Runge et al., 2019).

Hence, Following your suggestion, we propose highlighting the main concerned regions through cyan and red boxes in Figures while modifying the paragraph of the figure discussion to include an explanation of the Online_$\mathrm{ICL_{nud}}$ degradation, as follows.

A global view does not reveal marked changes in terms of ACC for the atmospheric predictions. However, from a continental to regional view, differences are noticeable. In boreal summer (Figure 5), enhanced initialisation $\mathrm{ICL_{nud}}$ tends to increase precipitation correlation in the middle region of South America, including the Paraná River basin and southern Amazon basin (red box), with degradation in the northeast of Brazil, Australia, and some areas of North America and Asia on the north of $40°N$ (cyan boxes). Notably, Europe experiences improved precipitation predictions. Temperature predictions are less sensitive to the land initialisation in summer, but degradation is concentrated in higher latitudes (north of $40°N$ and south of $20°S$). In winter, regions with reduced performance for both precipitation and temperature predictions are primarily found in North Africa, Europe, and Asia (Figure 6).

We have found that the $\mathrm{ICL_{nud}}$ initialisation can harm the accuracy of precipitation and temperature seasonal forecasts in some regions of the globe. This is due to soil moisture nudging, a technique intended to improve the variability of soil water content and the forecast of river streamflows. However, it can also lead to adverse effects on the land-atmosphere coupling simulated by the model. The initial soil moisture conditions introduced by the offline nudging technique may shift the coupled system away from its equilibrium state. When the forecast integration begins, the nudging constraint is deactivated, and the model adjusts to its equilibrium, potentially generating spurious heat and water fluxes at the land-atmosphere interface. This could ultimately alter the simulated atmospheric circulation and reduce the accuracy of the temperature and precipitation forecasts.

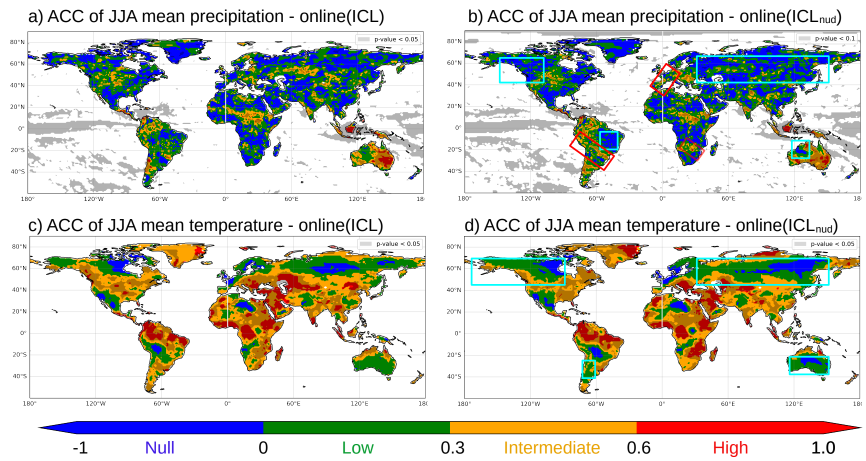

Figure 5. Comparison of Online_ICL and Online_ICL$_{\text{nud}}$ atmospheric forecasts for the anomalies correlation coefficient of the JJA 3-month mean precipitation (a and b) and temperature (c and d). Red (Cyan) boxes highlight regions with noticeable ACC increase (decrease).

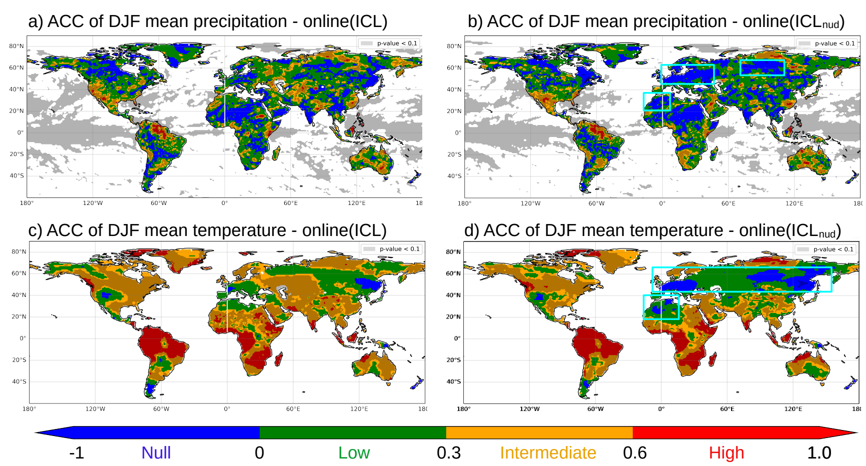

Figure 6. Comparison of Online_ICL and Online_ICL$_{\text{nud}}$ atmospheric forecasts for the anomalies correlation coefficient of the DJF 3-month mean precipitation (a and b) and temperature (c and d). Red (Cyan) boxes highlight regions with noticeable ACC increase (decrease).

4. Line 50: “Conversely, in regions dominated by rainfall, FCAs tend to significantly influence...”, what does FCAs mean? This typo has been corrected in the revised version of the manuscript. It is FSCs instead of the typo FCAs.

5. Lines 147-149: “For example, to apply the L3OCV selection method to the hindcast of 1993, only forcing from 1996 to 2020 ensures 25 members. For the hindcast of 2000, only forcing from 1992 to 1998 and 2003 to 2020 is used.” The previous article refers to the period from 1993 to 2017, please confirm the range. We appreciate the feedback about the offline configuration reproducing the ESP classical approach. We take the opportunity to correct a typo on the range and clarify the difference between the atmospheric sampling period and the simulation hindcast period. To comply with the atmospheric ensemble in the leave-three-years-out cross-validation framework, we have intentionally avoided using ERA5 atmospheric data from the hindcast year, the previous year and the two subsequent years. As

a result, we have extended the sampling period to include data from 1991 to 2019, which allows us to maintain a sample of 25 atmospheric members to force the offline configuration. Here, the conformation of the atmospheric ensemble follows the ESP approach applied in benchmarking works such as Harrigan et al. (2018) (see reply to comment 2 to see the related changes in the manuscript).

6. Line 172: The values written in the article is not the same as in Figure 2. This typo of the correlation value in the text has been corrected to fit the value in Figure 2. Before (wrong): 0.9986, now (correct): 0.9886.

7. I suggest the colour bar is divided by 0, which makes it possible to visualize the changes in the indicator more clearly in Figs.3-4, and whether the horizontal coordinates of the last column of Figs.3-4 are displayed incorrectly. We understand your suggestion. However, in preliminary versions of figures 3-4 (see second column), we found that dividing the colour bar exactly at zero made it challenging to identify the more significant changes in the skill. Concerning the third column axis, we corrected the horizontal labels in the revised manuscript's last column of Figures 3-4 (see the following reproduction of the reviewed figure for May initialisation).

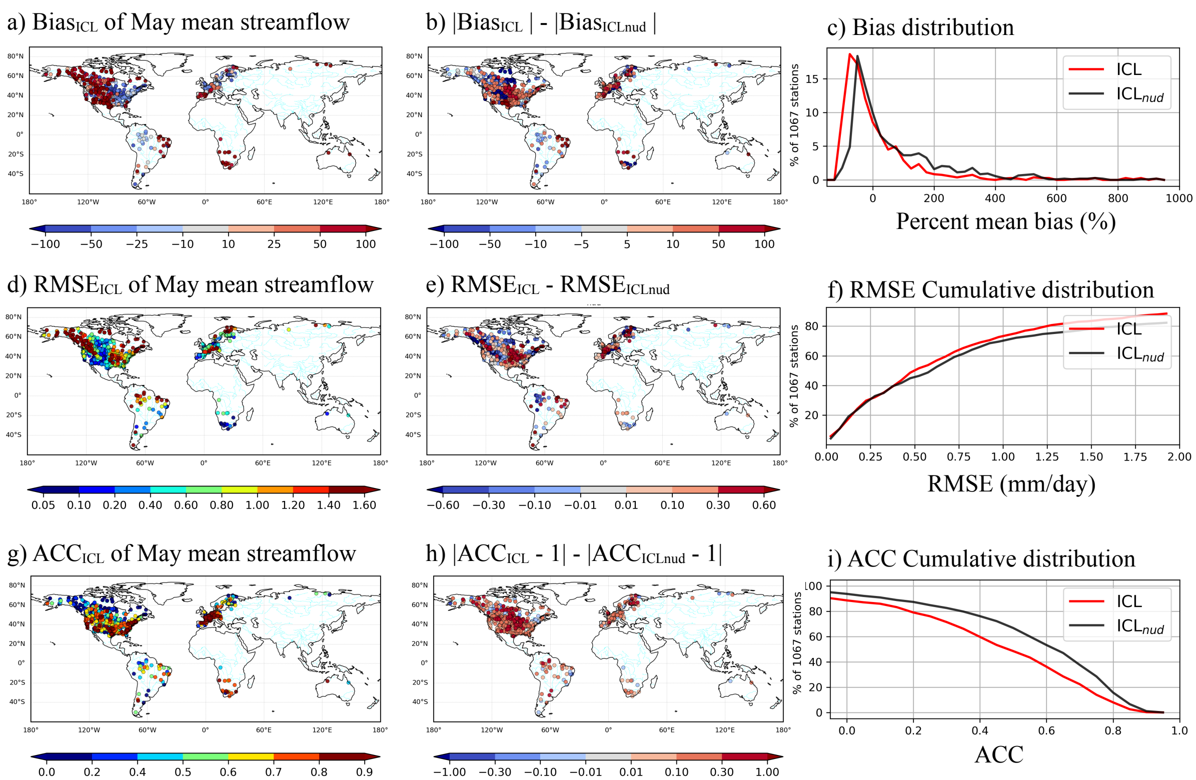

Figure 3. Comparison between May streamflow mean of initialisation run against the observed one over 1993-2017. Left column: ICL bias (a), root mean square error (mm/d) (d), and anomaly correlation (g). Middle column: difference with the $\mathrm{ICL_{nud}}$ enhanced land initialisation bias (b), root mean square error (mm/d) (e), and anomaly correlation (h). Right column: distribution of bias for each experiment (c), accumulated distributions of the root mean square (f), and anomaly correlation (i).

References

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., Andrews, M., and Knight, J.: Skilful predictions of the winter North Atlantic Oscillation one year ahead, Nature Geoscience, 9, 809–814, https://doi.org/10.1038/ngeo2824, 2016.

Harrigan, S., Prudhomme, C., Parry, S., Smith, K., and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, Hydrology and Earth System Sciences, 22, 2023–2039, https://doi.org/10.5194/hess-22-2023-2018, 2018.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., et al.: Inferring causation from time series in Earth system sciences, Nature communications, 10, 2553, 2019.