

Review 2

Please find below the answers to the comments point-by-point. For clarity, all reviewer comments are in **black** and responses in **blue**. Modifications in the manuscript are indicated with underlined text.

The manuscript by Couallier et al., presents an approach to sub-select ensemble members from an uninitialized large ensemble to predict precipitation over France with higher skill than using the full ensemble. The paper is generally well written though there are places where it can benefit from improving grammar. The manuscript falls within the scope of the journal and may be accepted for publication after clarifying major concerns.

We are grateful to the reviewer for the careful reading of our manuscript and for the constructive comments, which helped us substantially improve the clarity and rigor of the work.

Before addressing the specific comments, we would like to note that we identified an error in the computation of 8-year mean summer precipitation for both observations and uninitialized projections, resulting in a one-year lag relative to the decadal predictions. This issue has been corrected. All analyses, figures, and associated text have been updated accordingly. This correction affects the quantitative skill scores for summer precipitation predictions (former Figs. 9 and 10), although the overall conclusions of the study remain unchanged. All modifications related to this correction are clearly highlighted in the revised manuscript.

Major points:

Major point 1. The study is missing important methodological details. By simply pointing towards an unpublished study (Alkama et al.) for methodological details is not appropriate in my opinion. Also, I am not sure how the drift in initialized predictions is treated and whether it affects their selection process. Missing these details are the major shortcoming of this study. Also, how are the model and observational data treated, for example, remapping, use of anomalies or mean values etc.

We thank the reviewer for highlighting the need for additional methodological details. We address each sub-point in turn.

(i) Reference to an unpublished study: Alkama et al. has since been accepted for publication in Science Advances and will be publicly available shortly, ensuring that readers will be able to access the full methodological details. We have updated the citation accordingly in the revised manuscript. We have also expanded Section 2 to provide a self-contained summary of the key steps of the Alkama et al. method:

"We apply the post-processing method developed by Alkama et al. (in press) that improves the earlier work by Smith et al. (2020) and Marcheggiani et al. (2023). This approach includes two key steps:

1. The DCP-A ensemble mean is aggregated across lead times using an optimized averaging scheme. For each target 8-year calendar window, rather than combining members from four consecutive start dates as in Smith et al. (2020) – which introduces a systematic two-year lag between the predicted and observed periods – Alkama et al. restrict averaging to hindcast years that fall within the target calendar window. Specifically, for each calendar year in the window, the predicted values from all contributing hindcast start dates (lead years 2 to 9) are first averaged, before computing the 8-year mean. This ensures temporal synchronization between predictions and observations and increases the ACC of the boosted winter NAO from 0.58 to 0.83 (Alkama et al., in press). Here, the method is applied using lead times 2-10, rather than 2-9 in Alkama et al. (in press). Alkama et al. restricted average to lead years 2-9 for consistency with Smith et al (2020); in the present study, lead year 10 from previous start dates is also included in the averaging on each calendar year in the window of the predicted values from all contributing hindcast start dates. This slightly increases the effective ensemble size contributing to each prediction without altering the 8-year averaging window, therefore potentially further increasing the skill scores.
2. The ensemble mean forecast is variance-rescaled so that its standard deviation matches that of the observed signal, correcting for the systematic underestimation of signal amplitude associated with the signal-to-noise paradox (Scaife and Smith, 2018).

The resulting predictions are referred to as 'boosted' decadal predictions. For some indices (wWEPA, sNAO, and sMedScand), the predicted linear trend is opposite in sign to the observed trend (Fig. S1.1). In these cases, the predicted trend is adjusted to match the observed trend over the historical period, following Kharin et al. (2012); for future forecasts, this correction is applied by extending the instrumental observed trend.”

(ii) Drift correction: No drift correction was applied in the original prediction framework. This choice was motivated by previous studies showing that initialization drift primarily affects sea surface temperature fields, while its impact on large-scale sea-level-pressure-based indices is generally limited, and often not corrected (e.g., Smith et al., 2020, Alkama et al.). To evaluate the potential impact of drift on our results, we applied a lead-time-dependent bias correction to each model from the DCP-A ensemble prior to applying the aggregation method of Alkama et al. The lead-time-dependent climatology was evaluated over 1970–2016 to ensure a consistent baseline across all lead times. Figure R1 illustrates the climatology by model and lead time, revealing very small drift across lead times for SLP indices, though a larger drift is observed for the uAMV index (Fig. R1d), consistent with its dependence on sea surface temperatures. Since predicted indices are 8-year averages over lead times 2 to 10, the impact of drift is further attenuated. Figure R2 confirms that this drift correction has a negligible effect on the SLP-based index predictions.

(iii) Data treatment: Full details have been added to Section 2.2a of the revised manuscript: “To extract atmospheric indices, observations and model outputs are remapped to a common 2°x2° regular grid using bilinear interpolation prior to spatial averaging over the index definition regions. For observations and the uninitialized IPSL ensemble, 8-year rolling means are first computed from the raw seasonal values, and anomalies are then obtained by subtracting the climatological mean of these smoothed times series, estimated over the full 1966-2019 period. For DCPD boosted predictions (see Section 2.2.b), the 8-year aggregation follows the lead-time-aware method of Alkama et al. (in press); anomalies are subsequently computed relative to the same 1966-2019 climatological baseline.”

“In addition to the atmospheric indices, we include an AMV index to capture potential low-frequency oceanic influences on European climate (Sutton and Hodson, 2005). [...] We define the index used here as the area-weighted mean of annual-mean SST over the North Atlantic domain (0°–60°N, 76°W–0°E). For models with curvilinear or irregular ocean grids, weighting is performed using the model-native grid-cell area field; for models on regular latitude-longitude grids, cosine-of-latitude weighting is used.”

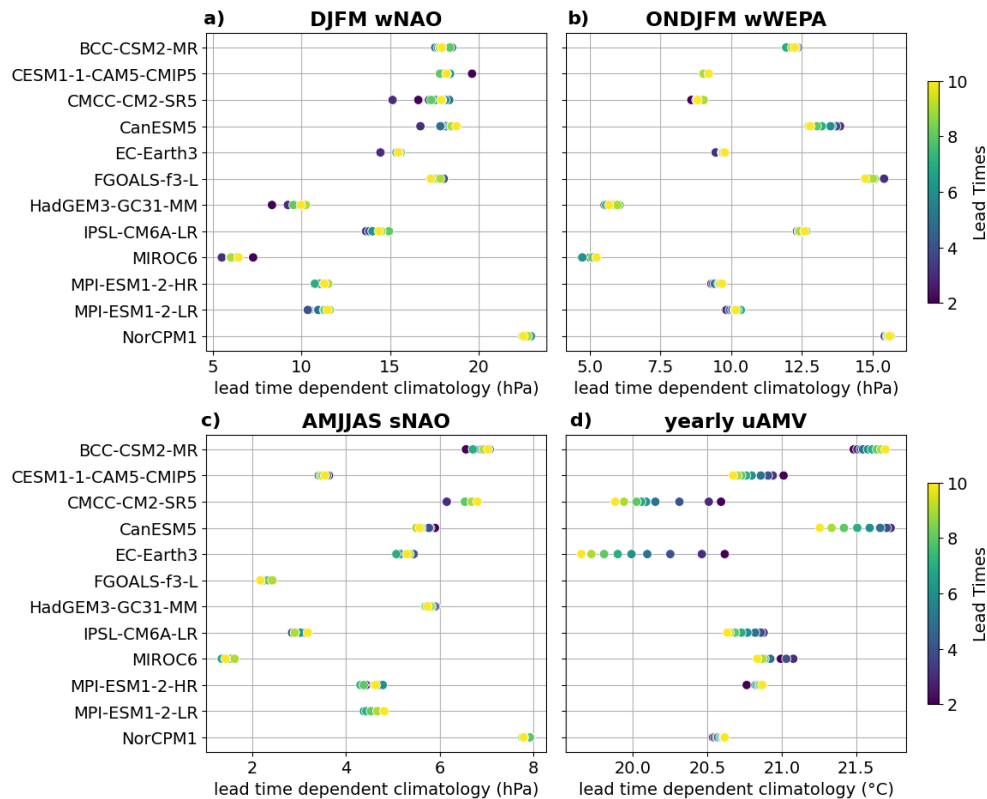


Fig. R1: Lead-time-dependent climatology of large-scale climate indices used to assess potential drift in decadal predictions. Each panel shows the climatological mean by model and

lead time for: (a) boreal winter wNAO, (b) extended winter wWEPA, (c) extended summer sNAO, and (d) yearly uAMV, computed over the 1970–2016 period. Colors indicate individual lead times (2–10). The small spread across lead times for SLP-based indices (a–c) indicates negligible drift, while the larger spread for the uAMV (d) is consistent with its dependence on sea surface temperatures.

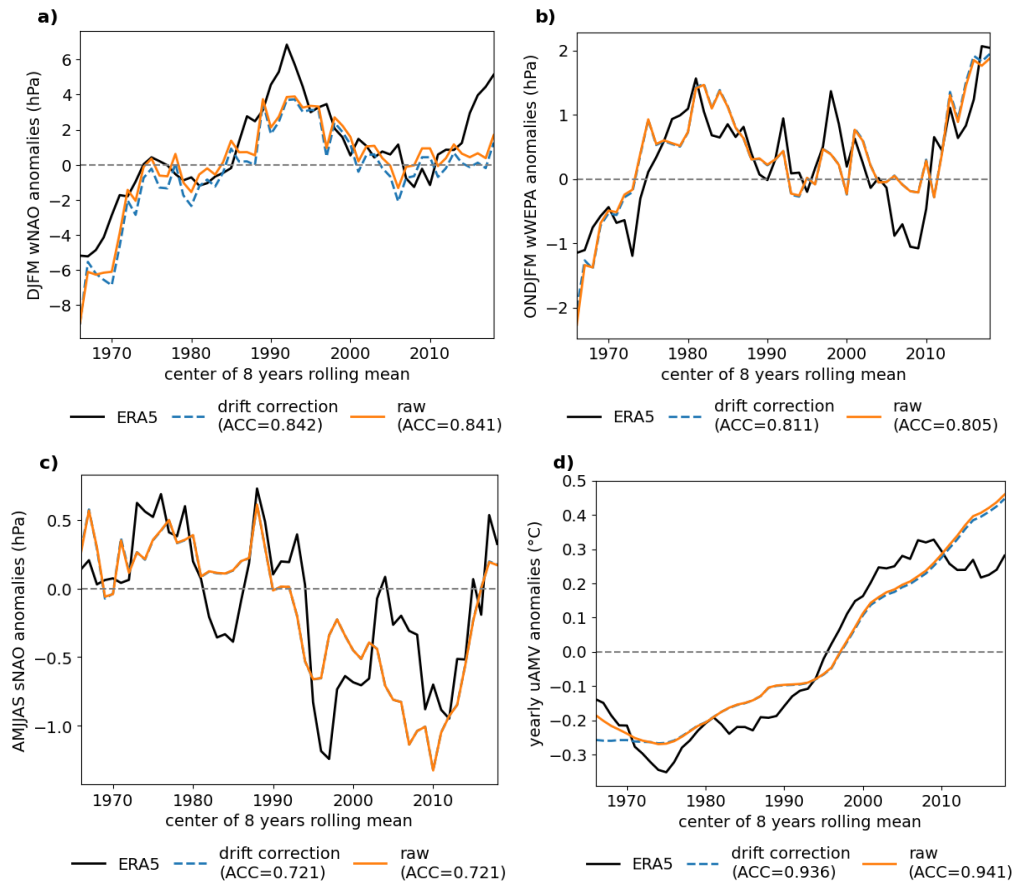


Fig. R2: Observed and predicted decadal variability of large-scale climate indices, with and without lead-time-dependent drift correction. Eight-year running means are shown for: (a) wNAO (extended winter, ONDJFM), (b) wWEPA (extended winter, DJFM), (c) sNAO (extended summer, AMJJAS), and (d) uAMV (yearly). The black line shows ERA5, the blue line shows boosted DCPD predictions with drift correction, and the orange line shows predictions without drift correction. Pearson correlation coefficients with ERA5 are reported in parentheses in each panel legend. The close agreement between the blue and orange lines confirms that lead-time-dependent drift correction has a negligible impact on the predicted indices.

Major point 2. The study seems to have ignored several pioneering works related to climate predictions using ensemble selection approaches. It is important that the **discussions** include similar works done before. I have listed a few studies who use ensemble selections for seasonal to decadal climate predictions.

We thank the reviewer for drawing our attention to this body of work on ensemble selection approaches. We have incorporated these references and contextualized our contribution relative to them in two locations in the revised manuscript.

In Section 2.2c (Methods), the following text has been added: “To transfer predictability from large-scale indices to precipitation, we adopt an index-based subsampling strategy inspired by the NAO-matching method (Smith et al., 2020; Alkama et al., [in press](#)). [The idea of selecting climate simulations based on predicted or observed criteria has been employed in various contexts: in seasonal forecasting \(Ding et al., 2018\) and on decadal-to-multi-decadal timescales \(Befort et al., 2020; Mahmood et al., 2021, 2025\).](#) Smith et al. (2020) introduced the NAO-matching method by selecting a subset of 20 members from multi-model decadal simulations at each 8-year window whose NAO values best match the NAO prediction from the whole DCP predictions ensemble mean.”

In Section 4 (Discussion and Conclusions), the following text has been added:

[“Previous work has demonstrated the potential of ensemble selection approaches for predicting climate variability \(Befort et al., 2020; Mahmood et al., 2021, 2025\). These studies focused primarily on predicting temperature variations using SST-based selection criteria. Mahmood et al. \(2025\) extended this approach by incorporating the NAO to select members that reproduce the observed NAO–temperature teleconnection during the 20 years prior to the prediction period. Our study applies a conceptually similar framework, but targets precipitation predictions and introduces an explicit step to identify the most skillful combination of large-scale predictors. Our results confirm and extend previous findings demonstrating that improved predictions of large-scale climate indices, such as the boosted NAO forecasts developed by Alkama et al. \(\[in press\]\(#\)\), can be effectively leveraged to enhance regional hydroclimate predictions \(Nicolì et al., 2025; Tsartsali et al., 2023\). While earlier studies focused primarily on winter NAO and its influence, our work broadens this scope to include both winter and summer seasons, incorporating combined index approaches such as the sNAO, together with the uAMV index for summer precipitation forecasting.”](#)

Major point 3. One of my major concern is that the authors use a rather small ensemble from which the selections are made and also only from single model. Why is the selection not performed using a multi-model ensemble from CMIP6 simulations? Using (sub-selecting) members from just one model can be a limitation when applying the same method to other regions as the author seems to claim. I do not agree that it would be easier to apply the methodology to other regions particularly when selection is performed using a single model ensemble. Also the ensemble size from which the selection is made is rather small.

We thank the reviewer for raising this important concern. We address the multi-model and ensemble size questions in turn.

Regarding the choice of a single-model ensemble: while a multi-model ensemble would in principle provide greater diversity, its use for subsampling raises non-trivial methodological challenges. Different models may exhibit substantially different amplitudes for the climate indices considered, as well as different precipitation responses to a given index state. Subsampling across a multi-model ensemble would yield predictions with varying proportions of members from each model at each start date, which complicates the interpretation of skill scores and the attribution of skill improvements. A single-model ensemble, by contrast, provides greater dynamical and statistical coherence in the predictions. Additionally, the present framework involves downscaling the subsampled uninitialized simulations to a high-resolution (8 km) grid using the CDF-t method. Applying this downscaling procedure to multiple

models simultaneously would entail substantially greater computational and data storage costs, which is why this proof of concept was carried out for a single model. We acknowledge the single-model choice as a limitation of the current study and have stated this more explicitly in the revised manuscript.

Regarding ensemble size: the subsampling consists of selecting 5 members out of 32 at each start date — a ratio of approximately 1:6. We acknowledge that both the size of the selected sub-ensemble and the size of the full ensemble from which it is drawn may influence prediction skill. Regarding the sub-ensemble size, Supplementary S8 from the first submission presents a sensitivity analysis in which the number of selected members varies from 2 to 8, providing insight into this aspect of the method. Regarding the full ensemble size of 32 members, we acknowledge that this sensitivity was not explicitly investigated and represents a limitation of the current study. Larger uninitialized ensembles from CMIP6 would provide a richer pool from which to subsample, potentially improving prediction skill and reducing sampling uncertainty. The evaluation of this method with larger or multi-model uninitialized ensembles represents an avenue for future research, as noted in the revised Discussion section:

"In the present study, subsampling is performed by selecting 5 members out of 32 from a single-model ensemble (IPSL-CM6A-LR), which limits both the statistical diversity of the selected sub-ensemble and the generalizability of the results across models. Exploring larger single-model ensembles or multi-model CMIP6 ensembles in future work could help assess the robustness of the subsampling approach and potentially improve prediction skill. [...] Future work should explore adaptive seasonal definitions, integration of additional indices, optimal weighting strategies and exploration of alternative or larger uninitialized model ensembles to further enhance predictive skill."

Regarding transferability to other regions: we agree with the reviewer that the choice of uninitialized model may affect the applicability of the method to other regions. We have revised the corresponding claim in the manuscript to reflect this nuance more accurately:

"This system leverages large-scale atmospheric and oceanic teleconnections relevant to precipitation variability over metropolitan France. Since the methodology is not specific to this region, it can in principle be applied to other regions of the globe, if appropriate large-scale predictors and suitable uninitialized ensembles are available."

Major point 4. The results shown in Fig. 6/7 does not support authors claims about reduced uncertainty when observation is almost always outside the spread of the selected ensemble. That would imply that the predictions from the sub-selected ensembles are not reliable, no? The reduction in ensemble spread is not really useful when it does not include observations inside.

We thank the reviewer for this important comment on reliability. We wish to clarify both the scope of our claims and the interpretation of Figs. 6 and 7.

Figures 6 and 7 display the predictions of the large-scale climate indices (e.g., the wNAO, wWEPA, sNAO, and uAMV), not the precipitation predictions. For the index predictions, we do not claim reduced uncertainty in the manuscript. The figures illustrate the temporal agreement between the predicted and observed index time series. Crucially, the red shading shown in

Figures 6 and 7 represents the spread of the *selected sub-ensemble*, which was not chosen to optimally predict the index itself, but rather to identify members whose large-scale circulation state is dynamically consistent with the predicted index phase — and whose associated precipitation response is therefore expected to be more skillful. The sub-ensemble spread in index space may thus appear narrow or imperfectly centered on the observed index, yet still yield improved precipitation predictions, which is the actual target of the selection procedure.

The claim regarding reduced uncertainty therefore applies specifically to the precipitation predictions, as shown in Fig. 11. In that figure, the subsampled ensemble spread is reduced relative to the full uninitialized ensemble, and observations fall predominantly within the selected ensemble spread, suggesting improved reliability at the regional precipitation scale.

We have revised the manuscript (section 3.1) to ensure that the distinction between index-level spread and precipitation-level uncertainty is stated unambiguously:

“Subsampling the five closest members to the boosted hindcasts in terms of SLP indices from the uninitialized IPSL ensemble results in predictions closely matching the boosted forecasts. Selecting the five uninitialized IPSL-CM6A-LR members whose SLP indices most closely match the boosted hindcasts results, by construction, in a sub-ensemble with a substantially narrower spread than the full ensemble. As a consequence, observations frequently fall outside this spread, which should not be interpreted as a failure of reliability. The sub-ensemble was not designed to optimally predict the index itself, but to identify members whose large-scale circulation state is dynamically consistent with the predicted index phase. The skill target of this selection procedure is regional precipitation, whose prediction reliability will be assessed later on in the manuscript.”

Major point 5. It could be useful to compare the skill predictions from the sub-selected ensemble relative to the skill of actual prediction systems (DCPP-A). This could highlight the benefits of using the current approach in predicting the future climate.

We thank the reviewer for this constructive suggestion. We agree that a comparison with DCPP-A precipitation prediction skill is important to contextualize the added value of our approach. We have therefore computed the precipitation prediction skill of the DCPP-A ensemble mean against ERA5 observed precipitation and included the results directly in the revised manuscript. The following description has been added to the Methods section (Section 2.2e, “*Comparison with raw DCPP ensemble mean*”):

“The forecast skill is also evaluated against the raw DCPP-A ensemble mean. The DCPP ensemble includes 11 models (and not 12 because precipitation data for CESM1-1-CAM5-CMIP5 are unavailable on ESGF). Drift is removed from DCPP-A predictions by computing precipitation anomalies relative to a lead-time-dependent climatology for each model. Eight-year means are computed by averaging lead times 2 to 9. The multi-model ensemble mean is then formed by averaging the individual model ensemble means.

The comparison between the two prediction systems is not straightforward due to their different spatial resolutions. DCPP-A predictions are available at a coarse native resolution of approximately 2°, while our subsampling-based predictions are statistically downscaled at 8

km resolution. To ensure inter-model comparability, we first remap all DCPD-A outputs to a common 2° resolution grid. Remapping these predictions to a finer scale would not be physically meaningful given the models' native resolution.

Concerning the subsampling-based predictions, they have been downscaled and bias-corrected against the SAFRAN observational dataset (~8 km, ~0.08° resolution). Remapping the downscaled subsampling predictions to a 2° is not straightforward. Indeed, since these predictions are only available over metropolitan France, regridding to a 2° would result in incomplete coverage of several coarse grid cells at the domain boundaries, introducing artefacts in the skill computation.

For consistency, ACC from both prediction systems (raw DCPD and subsampled predictions) are compared using ERA5 dataset. DCPD-A predictions are compared against ERA5 remapped to 2°, while subsampling predictions are compared against ERA5 at 0.25° resolution. Note that the subsampling predictions are spatially interpolated onto the 0.25° ERA5 grid for this comparison, no additional bias correction is applied at this stage.”

The following results and figure have been added at the end of section 3.3:

“Finally, in winter, wNAO-based subsampling predictions substantially outperform the raw DCPD-A ensemble mean (Fig.12.a-d), with significant ACCs over the majority of metropolitan France, whereas DCPD-A skill is largely confined to the French Atlantic coast. In summer, skill improvements from sNAO+uAMV subsampling are more localized, with gains over Brittany and northern and north-eastern France. In southern France, however, the subsampling skill is lower than that of DCPD-A (Fig.12e-f). This is because the summer DCPD-A skill in this region is largely attributable to a linear trend, likely of external origin (Fig.12.g). We also note that the skill of the subsampling predictions evaluated against ERA5 at 0.25° differs slightly from the evaluated against SAFRAN (Figs.8,9,12), with higher correlations in winter and lower correlations in summer when ERA5 is used as the reference.

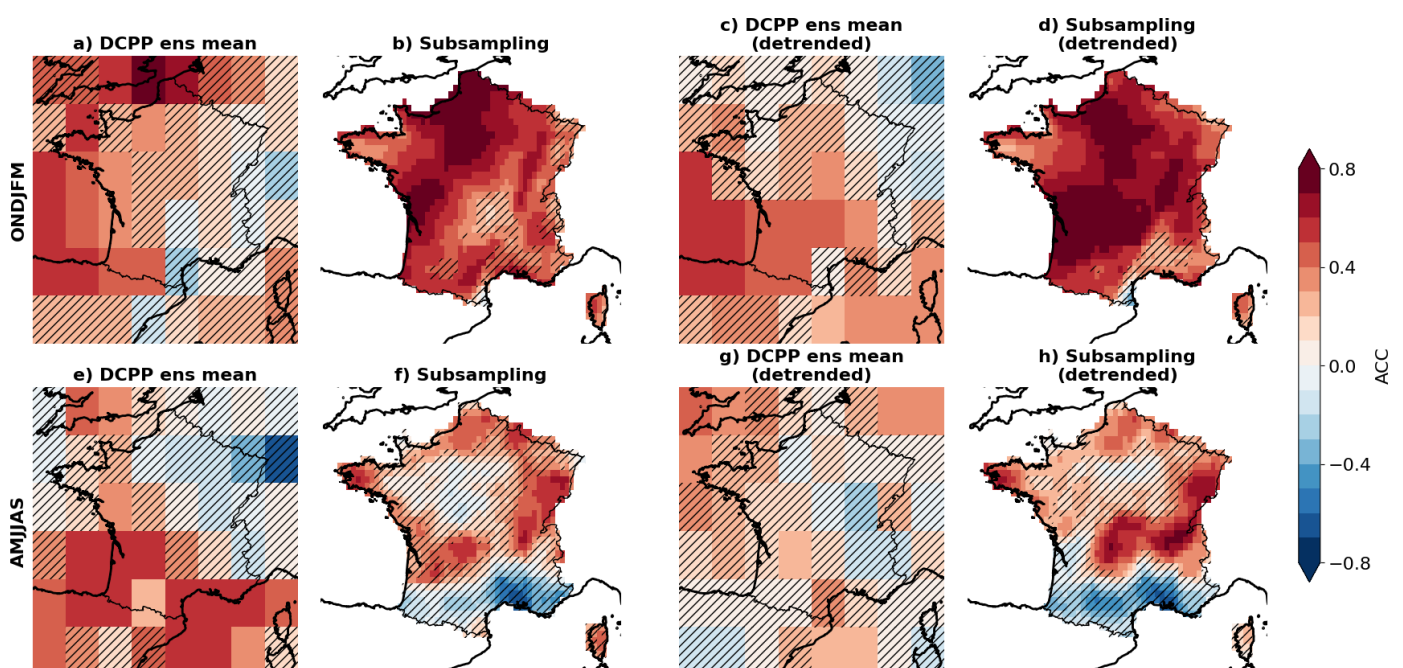


Fig 12: Precipitation prediction skill for the DCP-A ensemble mean and the subsampling method, for extended winter (ONDJFM, a–d) and extended summer (AMJJAS, e–h), evaluated against ERA5 over 1966–2019. Panels (a, e): ACC between the DCP-A multi-model ensemble mean and ERA5 remapped to 2°. Panels (b, f): ACC between subsampling predictions (wNAO-based for winter; sNAO+uAMV-based for summer) and ERA5 at 0.25°. Panels (c, g) and (d, h): same as (a, e) and (b, f), respectively, but computed after removal of the linear trend from both predictions and observations, isolating skill arising from internal climate variability rather than the externally forced signal. Note that panels (a, c, e, g) are evaluated on a 2° grid while panels (b, d, f, h) are evaluated on a 0.25° grid, reflecting the respective native or target resolutions of the two systems; a direct gridpoint-by-gridpoint skill comparison is therefore not computed. Hatched areas indicate regions where skill scores are not statistically significant at the 95% confidence level, assessed using a 1,000-sample block bootstrap.”

Regarding the drift correction, Fig.R3 shows that the standard deviation of the lead-time-dependent precipitation climatology across lead times 2-9 is close to zero for all DCP-A models over Europe in winter, indicating negligible drift in the precipitation climatology across lead times. Consistent with this, we verified that applying lead-time dependent drift correction has no material effect on DCP-A winter precipitation skill scores (not shown).

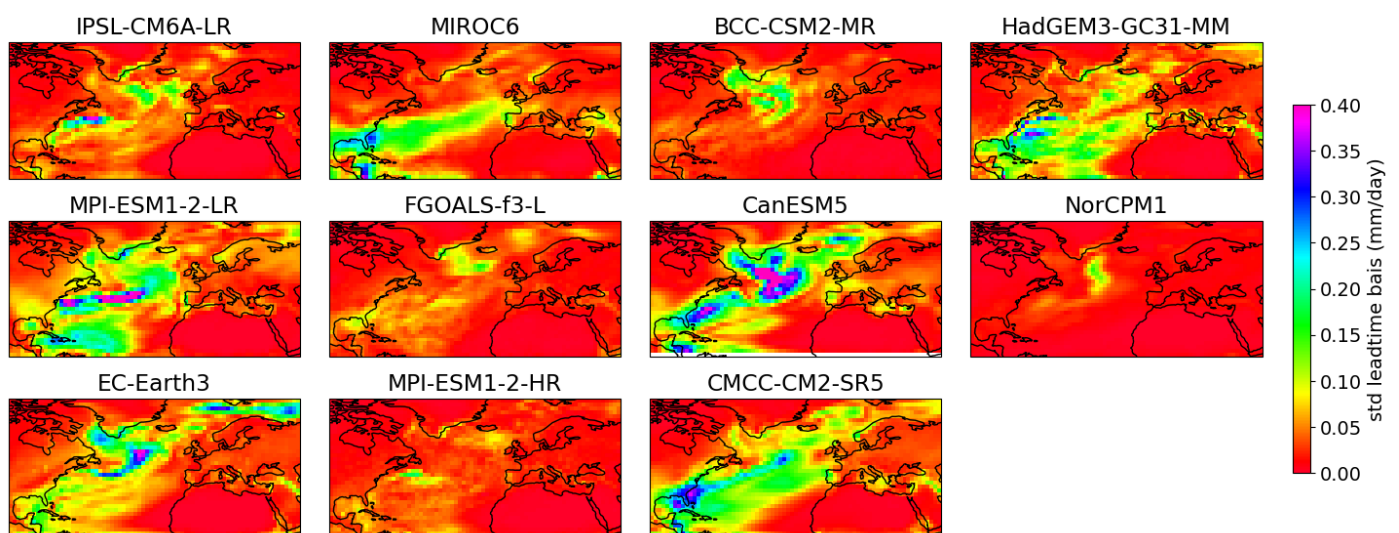


Fig.R3: Standard deviation of the lead-time-dependent precipitation climatology across lead times 2–10, for each model in the DCP-A ensemble, evaluated over the North Atlantic for the extended winter season (ONDJFM) and the 1970–2016 period. Values close to zero indicate negligible drift in DCP-A precipitation predictions across lead times, confirming that drift correction has no material effect on winter precipitation skill.

Finally, the following text has been added to Section 4:

"The resulting precipitation predictions over France demonstrate skill improvements relative to both the uninitialized simulations and the raw DCP-A ensemble mean. Comparison with DCP-A reveals large improvements in winter, where significant skill extends across the majority of metropolitan France, and more localized improvements in summer."

Other points:

Line 44: I am not sure if the wording is clear: “Seasonal forecasts provide initialized climate information” what do you mean by “provide initialized climate information”?

We thank the reviewer for this comment. The sentence has been revised to improve clarity:

"Decadal climate predictions (DCP) are designed to fill the gap between seasonal forecasts and long-term climate projections (Meehl et al., 2009; Solaraju-Murali et al., 2022). Seasonal forecasts are initialized from the observed climate state and provide predictions up to approximately 12 months ahead, with skill potentially arising from those initial conditions. Long-term climate projections, in contrast, simulate future climate under external forcing scenarios and do not attempt to predict internal variability at specific dates."

Line 63: "thereby enhancing downstream precipitation forecasts",... It is not clear what does "enhancing forecasts" mean here?

We agree with the reviewer that this phrasing was ambiguous. The sentence has been revised as follows:

"Building on the well-established relationship between the NAO and European precipitation, Smith et al. (2020) proposed an innovative approach: rather than directly predicting regional rainfall, the first step is to skillfully predict the NAO itself. This enables the identification or subsampling of model ensemble members whose atmospheric states are dynamically consistent with the predicted NAO phase, thereby producing more skillful decadal precipitation predictions over Europe."

Line 69: A recently study (<https://iopscience.iop.org/article/10.1088/1748-9326/adde75>), used NAO and temperature relationship to sub-select the ensemble members and make skill predictions over Eurasia. Maybe author could comment on this.

We thank the reviewer for pointing out this relevant recent study. As detailed in our response to Major Point 2, this reference has been incorporated in both the Methods section (Section 2.2c) and the Discussion (Section 4).

Line 91: "extracted"?

The sentence has been revised for clarity: "The precipitation observational data used in this study come from the SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige) analysis..."

Line 97: "through homogenization technique"...did you mean “thorough”?

We thank the reviewer for identifying this typographical error. The sentence now reads: "...and is widely used for hydrological and climate studies due to its temporal consistency, thorough homogenization technique, and fine spatial detail (~8 km × 8 km resolution)."

Line 99: AMV was already unfolded in the introduction.

We thank the reviewer for this observation. The acronym AMV has been removed from this location to avoid redundancy.

Line 107: Double check the abbreviation “Decadal Climate Prediction Project Phase A”, is it phase A.

We thank the reviewer for this correction. The text now reads: "Decadal Climate Prediction Project, [Component A](#)".

Line 108: please double check the start dates for DCPA, at least some of the predictions systems are initialized from 1960 and onward. Also, I am not sure about predictions from CESM1, does the initialization extend to 2014 for this model?

We thank the reviewer for drawing attention to this inaccuracy. We had incorrectly stated the first start dates, confusing them with the first available winter prediction. Regarding CESM1, according to the ESGF portal, start dates are available from 1960 to 2017 in DCPA. The corrected sentence now reads: "The ensemble includes decadal hindcasts from 12 models (see Table A1), with a total of 188 members initialized annually from [1960](#) to 2014 and run for one decade."

Line 235, it is not clear to me what exactly this means. In section 2.2b, there are several things being discussed. It would be better to describe here what had been done.

We agree that the cross-reference was insufficiently informative. The sentence has been revised to: "To ensure consistent autocorrelation properties, we apply [to the selected members the aggregation of post-processing method described in Section 2.2b — specifically, the 8-year running mean aggregation with the lead-time weighting scheme.](#)"

Eq. 2, please describe, what does alpha and beta mean here and how they are derived?

We thank the reviewer for identifying this oversight. Upon reflection, we acknowledge that the free parameters alpha and beta were not explored and do not add analytical value as presented. They have therefore been removed from the loss function. Furthermore, we identified that the original equation was incorrectly written: the loss function should use standardized errors, since the indices are not on the same scale. Standardization was already applied in the code but was not reflected in the written formula. The corrected loss function, now included in the revised manuscript, reads:

$$D_i(t) = \sqrt{(d1_i(t)/std(\{d1_i(t)\}_{i=1..32,\forall t}))^2 + (d2_i(t)/std(\{d2_i(t)\}_{i=1..32,\forall t}))^2}, \forall i \in [1,32]$$

(2)

where $d_i(t) = u_i(t) - f(t)$ for each index, and standard deviations are computed over the entire historical period to provide a more stable estimate.

Table S1, missing references for the prediction systems.

We thank the reviewer for this observation. The missing references for each prediction system have been added to Table S1 in the revised manuscript:

Model	Ensemble size	Reference
BCC-CSM2-MR	8	(Wu et al., 2019)
CESM1-1-CAM5-CMIP5	40	(Yeager et al., 2018)
CMCC-CM2-SR5	20	(Cherchi et al., 2019)
CanESM5	20	(Swart et al., 2019)
EC-Earth3	15	(Döscher et al., 2022)
FGOALS-f3-L	9	(He et al., 2020)
HadGEM3-GC31-MM	10	(Sellar et al., 2020)
IPSL-CM6A-LR	10	(Swingedouw et al., 2026)
MIROC6	10	(Tatebe et al., 2019)
MPI-ESM1-2-HR	10	(Müller et al., 2018)
MPI-ESM1-2-LR	16	(Mauritsen et al., 2019)
NorCPM1	20	(Bethke et al., 2021)

Line 470: “French climate”?

The sentence has been revised for precision: "This system leverages large-scale atmospheric and oceanic teleconnections relevant to precipitation variability over metropolitan France. Since the methodology is not specific to this region, it can in principle be applied to other regions of the globe, if appropriate large-scale predictors and suitable uninitialized ensembles are available."

Lines: 500-506, there has been studies using sub-sampling and predicting climate on seasonal to decadal timescales, which surprisingly authors have completely ignored. A few of those are listed below:

This point is fully addressed in our response to Major Point 2 above, where the relevant references (Befort et al., 2020; Mahmood et al., 2021, 2025; Ding et al., 2018) have been incorporated into both the Methods and Discussion sections of the revised manuscript.

Line 513: That’s also my concern, why not use uninitialized simulations from different CMIP6 models?

This concern is fully addressed in our response to Major Point 3 above. We acknowledge the limitation of using a single uninitialized model and have incorporated a clearer statement of this limitation, along with a forward-looking statement regarding the evaluation of the method with alternative or multi-model CMIP6 ensembles, in the Discussion section of the revised manuscript.

Relevant references:

Befort, D. J., O'Reilly, C. H., and Weisheimer, A.: Constraining Projections Using Decadal Predictions, *Geophysical Research Letters*, 47, e2020GL087900, <https://doi.org/10.1029/2020GL087900>, 2020.

Ding, H., Newman, M., Alexander, M. A., and Wittenberg, A. T.: Skillful Climate Forecasts of the Tropical Indo-Pacific Ocean Using Model-Analogs, *Journal of Climate*, 31, 5437–5459, <https://doi.org/10.1175/JCLI-D-17-0661.1>, 2018.

Mahmood, R., Donat, M. G., Ortega, P., Doblas-Reyes, F. J., and Ruprich-Robert, Y.: Constraining Decadal Variability Yields Skillful Projections of Near-Term Climate Change, *Geophysical Research Letters*, 48, e2021GL094915, <https://doi.org/10.1029/2021GL094915>, 2021.

Mahmood, R., Yang, S., and Donat, M. G.: Constraining the NAO-temperature teleconnection in CMIP6 simulations enables skillful multi-annual predictions of Eurasian winter climate, *Environ. Res. Lett.*, 20, 074016, <https://doi.org/10.1088/1748-9326/adde75>, 2025.

Alkama, R., Swingedouw, D., Mignot, J., Gastineau, G., and Ogée, J.: Clearing the noise to predict the rhythm of the North Atlantic climate, *Submitted to Science Advances.*, in press.

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., Samuelsen, A., Langehaug, H., Svendsen, L., Chiu, P.-G., Passos, L., Bentsen, M., Guo, C., Gupta, A., Tjiputra, J., Kirkevåg, A., Olivié, D., Seland, Ø., Solsvik Vågane, J., Fan, Y., and Eldevik, T.: NorCPM1 and its contribution to CMIP6 DCPP, *Geoscientific Model Development*, 14, 7073–7116, <https://doi.org/10.5194/gmd-14-7073-2021>, 2021.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., Masina, S., Scoccimarro, E., Materia, S., Bellucci, A., and Navarra, A.: Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model, *Journal of Advances in Modeling Earth Systems*, 11, 185–209, <https://doi.org/10.1029/2018MS001369>, 2019.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J.-P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégoz, M., Miller, P. A., Moreno-Chamarro, E., Nieradzic, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wårlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q.: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, *Geoscientific Model Development*, 15, 2973–3020, <https://doi.org/10.5194/gmd-15-2973-2022>, 2022.

He, B., Yu, Y.-Q., Bao, Q., Lin, P., Liu, H., Li, J., Lei, W., Liu, Y., WU, G., CHEN, K., GUO, Y., Zhao, S., Zhang, X., Song, M., and Xie, J.: CAS FGOALS-f3-L model dataset descriptions for CMIP6 DECK experiments, *Atmospheric and Oceanic Science Letters*, 13, 1–7, <https://doi.org/10.1080/16742834.2020.1778419>, 2020.

Khariin, V. V., Boer, G. J., Merryfield, W. J., Scinocca, J. F., and Lee, W.-S.: Statistical adjustment of decadal predictions in a changing climate, *Geophysical Research Letters*, 39, <https://doi.org/10.1029/2012GL052647>, 2012.

Marcheggiani, A., Robson, J., Monerie, P.-A., Bracegirdle, T. J., and Smith, D.: Decadal Predictability of the North Atlantic Eddy-Driven Jet in Winter, *Geophysical Research Letters*, 50, e2022GL102071, <https://doi.org/10.1029/2022GL102071>, 2023.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., Jimenéz-de-la-Cuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Möbis, B., Müller, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D., Nyawira, S.-S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., von Storch, J.-S., Tian, F., Voigt, A., Vrese, P., Wieners, K.-H., Wilkenskjaeld, S., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO₂, *Journal of Advances in Modeling Earth Systems*, 11, 998–1038, <https://doi.org/10.1029/2018MS001400>, 2019.

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M. A., Greene, A. M., Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D., and Stockdale, T.: Decadal Prediction: Can It Be Skillful?, *Bull. Amer. Meteor. Soc.*, 90, 1467–1486, <https://doi.org/10.1175/2009BAMS2778.1>, 2009.

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornblueh, L., Li, H., Modali, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., Tian, F., and Marotzke, J.: A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR), *Journal of Advances in Modeling Earth Systems*, 10, 1383–1413, <https://doi.org/10.1029/2017MS001217>, 2018.

Nicoli, D., Gualdi, S., and Athanasiadis, P. J.: Decadal predictions outperform climate projections in forecasting Mediterranean wintertime precipitation, *Environ. Res. Lett.*, 20, 034034, <https://doi.org/10.1088/1748-9326/adb59e>, 2025.

Scaife, A. A. and Smith, D.: A signal-to-noise paradox in climate science, *npj Clim Atmos Sci*, 1, 1–8, <https://doi.org/10.1038/s41612-018-0038-4>, 2018.

Sellar, A. A., Walton, J., Jones, C. G., Wood, R., Abraham, N. L., Andrejczuk, M., Andrews, M. B., Andrews, T., Archibald, A. T., de Mora, L., Dyson, H., Elkington, M., Ellis, R., Florek, P., Good, P., Gohar, L., Haddad, S., Hardiman, S. C., Hogan, E., Iwi, A., Jones, C. D., Johnson, B., Kelley, D. I., Kettleborough, J., Knight, J. R., Köhler, M. O., Kuhlbrodt, T., Liddicoat, S., Linova-Pavlova, I., Mizieliński, M. S., Morgenstern, O., Mulcahy, J., Neisinger, E., O'Connor, F. M., Petrie, R., Ridley, J., Rioual, J.-C., Roberts, M., Robertson, E., Rumbold, S., Seddon, J., Shepherd, H., Shim, S., Stephens, A., Teixeira, J. C., Tang, Y., Williams, J., Wiltshire, A., and Griffiths, P. T.: Implementation of U.K. Earth System Models for CMIP6, *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001946, <https://doi.org/10.1029/2019MS001946>, 2020.

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L. F., Caron, L.-P., Counillon, F., Danabasoglu, G., Delworth, T., Doblus-Reyes, F. J., Dunstone, N. J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V., Kimoto, M., Merryfield, W. J., Mignot, J., Mochizuki, T., Modali, K., Monerie, P.-A., Müller, W. A., Nicolí, D., Ortega, P., Pankatz, K., Pohlmann, H., Robson, J., Ruggieri, P., Sospedra-Alfonso, R., Swingedouw, D., Wang, Y., Wild, S., Yeager, S., Yang, X., and Zhang, L.: North Atlantic climate far more predictable than models imply, *Nature*, 583, 796–800, <https://doi.org/10.1038/s41586-020-2525-0>, 2020.

Solaraju-Murali, B., Bojovic, D., Gonzalez-Reviriego, N., Nicodemou, A., Terrado, M., Caron, L.-P., and Doblus-Reyes, F. J.: How decadal predictions entered the climate services arena: an example from the agriculture sector, *Climate Services*, 27, 100303, <https://doi.org/10.1016/j.cliser.2022.100303>, 2022.

Sutton, R. T. and Hodson, D. L. R.: Atlantic Ocean Forcing of North American and European Summer Climate, *Science*, 309, 115–118, <https://doi.org/10.1126/science.1109496>, 2005.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), *Geoscientific Model Development*, 12, 4823–4873, <https://doi.org/10.5194/gmd-12-4823-2019>, 2019.

Swingedouw, D., Reiter, R., Mignot, J., Gastineau, G., Torres, O., and Alkama, R.: Evaluation of two versions of the IPSL-EPOC decadal prediction systems, *Front. Clim.*, 8, <https://doi.org/10.3389/fclim.2026.1803256>, 2026.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'ishi, R., Yamazaki, D., Suzuki, T., Kurogi, M., Kataoka, T., Watanabe, M., and Kimoto, M.: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6, *Geoscientific Model Development*, 12, 2727–2765, <https://doi.org/10.5194/gmd-12-2727-2019>, 2019.

Tsartsali, E. E., Athanasiadis, P. J., Materia, S., Bellucci, A., Nicolí, D., and Gualdi, S.: Predicting precipitation on the decadal timescale: A prototype climate service for the hydropower sector, *Climate Services*, 32, 100422, <https://doi.org/10.1016/j.cliser.2023.100422>, 2023.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., Jie, W., Zhang, J., Liu, Y., Zhang, L., Zhang, F., Zhang, Y., Wu, F., Li, J., Chu, M., Wang, Z., Shi, X., Liu, X., Wei, M., Huang, A., Zhang, Y., and Liu, X.: The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6, *Geoscientific Model Development*, 12, 1573–1600, <https://doi.org/10.5194/gmd-12-1573-2019>, 2019.

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., Karspeck, A. R., Lindsay, K., Long, M. C., Teng, H., and Lovenduski, N. S.: Predicting Near-Term Changes in the Earth System: A Large Ensemble of Initialized Decadal Prediction Simulations Using the Community Earth System Model, <https://doi.org/10.1175/BAMS-D-17-0098.1>, 2018.