Response to referees for
"Brief Communication: The Interdecadal Pacific Oscillation is not a Reliable Indicator of the Timing of an Ice-Free Arctic"
by Christopher Wyburn-Powell and Alexandra Jahn

We thank both referees and the editor for their careful and constructive reviews. Below, each comment is reproduced in italics, followed by our response. Text added or changed in the revised manuscript is shown in bold italics. All changes described below have been implemented in the revised manuscript.

RC1: 'Comment on egusphere-2026-2082', Anonymous Referee #1, 05 May 2026 reply
This short paper is an important contribution. It takes as a starting point the work of Screen and Deser, which suggested with a single model that the IPO could be a useful constraint on the timing of an ice-free Arctic. As with all such potential constraints, it is valuable to test their robustness across multiple models. This submission tests the constraint across 8 models and finds that it is not robust. I support publication after revision and/or suitable responses to the following points:

We thank the anonymous referee for their detailed comments and suggestions, which we have addressed in the revised manuscript (see below).

Line 60: I don't understand the rationale for the ensemble pooling. The text states "to incorporate multiple initial conditions", however, as far as I understand it, these are not independent runs but multiple samples from the same ensemble member. Unlike a distinct ensemble member, which is free to evolve differently to all other members, these pooled samples are not independent and are simply offset by a few years. This process seems to artificially inflate the ensemble size, without providing independent samples, and needs better justification.
We agree that the ensemble pooling could have been better motivated and explained. We have revised the methods to explain it more fully. In short, we use ensemble pooling to expand the analysis to multiple initial mean sea ice states beside the specific 4.7 million km^2 initial mean state used by Screen and Deser (2019), as our analysis showed that using a different starting mean state (e.g., 4.2, 4.3, 5.2 million km^2) changes the results even for the CESM1 data used in Screen and Deser (2019). As there is no physical reason that the response should be dependent on the exact initial mean state of 4.7 million km^2, the ensemble pooling using multiple initial mean states makes the results more robust (in particular in light of the uncertainties in the exact magnitude of the satellite-derived mean sea ice extent). We have revised the methods to refer to multiple "mean states" rather than "initial conditions," and added the following expanded description to the methods section:

L75-77: ***"To incorporate multiple mean states, for each ensemble member we select 11 trajectories at 4.2-5.2 million km² at 0.1 increments. This has the effect of repeating the Screen & Deser (2019) method for 11 mean states before averaging to a single set of 'pooled' trajectories for each LE as displayed in Figure 2."***

We address the associated concern about independence and significance testing in our response to the next comment.

Line 76: Does this increase in the percentage of LEs showing significant divergence come from the alternative IPO definition and/or the pooling of initial states? If the latter, I'm concerned that significance is inflated artificially and has limited physical meaning. Does the significance testing assume each sample is independent? They aren't (see above). Besides, regardless of the significance or not of the divergence, the more important point seems to be that the divergence is small around the time of first ice-free conditions, which is independent of the pooling (as the samples still have the same timing of ice-free conditions). Basically, I'm wondering if the pooling is necessary, even if it could be justified.

This divergence is from the pooling methodology, however, the significance test is performed on the number of ensemble members, not the repeated sampling of those ensemble members. Thus, the significance is not affected by the pooling. We have added the following sentence to the methods for clarification:

L81-82: "***Significance testing uses the number of LE ensemble members as independent samples, not the multiple samplings of the same ensemble member."***

Section 3.2: I think this section could be clearer. There are lots of kind-of "what if" statements and the key result – which I think is it that there is no consistent lagged relationship between IPO and ice-free conditions across models (in fact, in most models there is no relationship at all) – gets a bit lost. I think this point could be made more succinctly and clearly, alongside some simplification of Figure 3 (see comment below).

We agree, and we have rewritten Section 3.2 so that it leads with the key result and removes the speculative "what if" framing. The section now opens directly with the main finding:

L136-139: ***"Despite the significant impact of the IPO phase on sea ice concentration two decades from similar mean states as today, no model except CESM1 shows a significant correlation between the IPO at a two-decade lead time and the timing of consistently ice-free conditions. Across the LEs, no common lead time shows a significant correlation, and in most LEs no significant correlation exists at any lead time up to the 45 years examined (Figure 3a-h)."***

The more speculative passages that raised questions not posed in the introduction or pertinent to the conclusions have been removed, and we have simplified Figure 3 as described in our response to the next comment, which further helps to keep the focus on the main result.

Figure 3: It's unclear what value panels I-P add, and they are not directly referred to in the text. If I understand correctly, the correlation shown in panels I-P can already been inferred from the blue lines in panels A-H. Panels A-H include a second metric for ice-free, which is not used (or introduced) before. Is this necessary? What is the added value of the red line? In such a short note, I'm in favour of parsimony: keeping things as simple as possible.

We agree. In the revised manuscript we have removed panels I-P, and we have also removed the red lines (the first ice-free metric) from panels A-H, as they were not introduced elsewhere

and distracted from the main findings. Figure 3 now shows only the eight correlation panels (A-H) with the single consistently-ice-free metric, and the caption has been updated accordingly.

Conclusions: The lack of robustness of the constraint across models is an important result, but there is scope for deeper consideration of the causes of this lack of robustness, its implications, and potential ways forward. The IPO constraint can only ever reduce the model uncertainty due to internal variability. Its utility is dependent on the timeframe for an ice-free Arctic, which also depends on the simulated forced response. In this regard, it might not be appropriate to consider all models as equal. There is no effort to critique if any of models have (un)realistic forced responses, for example, by looking at their historical trends or sea-ice sensitivity (sea-ice loss per degree of global warming). At the very least, this requires some thought and discussion. A more sophisticated approach might be to first constrain the model spread in the forced response, before considering if internal variability can additionally constrain the spread around this forced response.
In using the 8 available large ensembles (LEs), we recognize that many of the LEs have biases relative to observations in terms of the forced response, internal variability, or in the case of MIROC6 a weak Pacific-Arctic teleconnection. However, none of these models are extreme outliers for CMIP5 or CMIP6. To the authors' knowledge, if we were to select some of these LEs to use only the more 'realistic' models on certain metrics, we would not see an increase in the consensus of the IPO as a good/poor indicator for the timing of an ice-free Arctic. To address this comment directly, we have made the following changes:

1) We have added realism metrics to Table 1, so that readers can assess the forced-response and IPO realism of each LE. We report the dSIA/dGMST forced response and historical mean September sea ice area, as calculated in the CMIP6 sea ice overview by Notz and SIMIP-Community (2020). Additionally, we include an IPO skill metric as assessed by Qin et al. (2024), which is referred to in the introduction to explain the variation in realism across models:
L32-34: ***"Finally, because CMIP5 and CMIP6 models inconsistently reproduce the IPO's observed spatial pattern, amplitude, and phase evolution (Qin et al., 2024), detection of relationships between the IPO phase and Arctic sea ice is further complicated."***

(2) We have added text to the methods motivating the inclusion of all eight LEs despite their varying realism:
L42-54: ***"Five of the six CMIP6 models used in this analysis were shown by Jahfer et al. (2024) to have strong Pacific-Arctic teleconnections: ACCESS-ESM1-5, CanESM5, CESM2-LENS, EC-Earth3, MPI-ESM1-2-LR; only MIROC6 was shown to have a weak (asymmetric) Pacific-Arctic teleconnection (CMIP5 models were not assessed in Jahfer et al. (2024)). IPO simulation skill has been assessed for all models used here except for EC-Earth3 (Qin et al., 2024), with model skill ranging from 1st (CESM2-LENS) to 45th (ACCESS-ESM1-5) place from 52 models assessed (Table 1). In terms of their sea ice simulation, the eight LEs vary across common sea ice realism metrics such as forced response and September mean sea ice area (Table 1). For mean historical September sea ice area, 5 of the 8 models are within the uncertainty range from observational estimates, with three (CESM2-LENS, MPI-ESM1-2-LR, and CanESM5) below the observational***

estimate (Table 1). The sea ice area loss per degree of global warming is below the observed estimate in all models except CESM2-LENS, which agrees with the observed estimate (Table 1). Based on all of these model metrics relevant for this study, no model subset emerges as clearly superior for the purpose of detecting an influence of the IPO on the timing of an ice-free Arctic. Thus, we analyze all 8 ensembles in the following."

(3) We have added a new paragraph in the conclusions to consider the causes of the lack of robustness and the way forward, emphasizing that generalized realism metrics such as forced response do not, on their own, identify which model best captures the IPO-Arctic response:
L171-178: ***"Lacking model consensus, can we determine which model is the most realistic in capturing the processes involved? Unfortunately, the currently available realism metrics (see Table 1) cannot clearly determine which model, if any, is most realistic in capturing all processes involved. Qin et al. (2024) place CESM1 and CESM2-LENS in the upper quartile for IPO spatial realism (EC-Earth3 was not assessed), and CanESM5, CanESM2, and MIROC6 in the upper half (see Table 1). Jahfer et al. (2024) classify all CMIP6 models except MIROC6 as having a strong Pacific–Arctic teleconnection (CESM1 and CanESM2, the two CMIP5 models, were not assessed). Thus, despite good relevant metrics for several of the models, we see large differences between the spatial and temporal IPO–sea ice responses, making it unlikely that a single common physical mechanism underlies them. We therefore find insufficient evidence for a single, realistic IPO–ice-free relationship."***

Other minor points:

Typo line 18: "showed the IPO has been shown"
Corrected. The sentence now reads as follows (L18-19): ***"Screen and Deser (2019) showed that the IPO phase at mean sea ice extent states similar to today's affects the time of a consistently ice-free Arctic by 9 years, using the Community Earth System Model version 1 (CESM1)."***

Figure 1: might be helpful to list the number of members on the titles
The eight LEs are considered equally valid as they have 30 or greater ensemble members, hence we consider displaying the number of ensemble members would highlight these differences which do not impact the results as significance is tested relative to the number of ensemble members. We note that the full list of LEs with ensemble members are listed in Table 1, and we prefer to retain that as the single reference for member numbers.

Line 134: "at today's lead times" confusing and probably not needed. The sentence makes (more) sense without this bit.
We agree and have removed the phrase, rewording the sentence to reference the mean-state range used in the pooled methodology:

L168-170: ***"Thus, we conclude that there is currently no consensus of the utility of the IPO for determining when consistently ice-free conditions will occur, when assessed from September mean sea ice extents between 4.2-5.2 million km².***"

Citation: https://doi.org/10.5194/egusphere-2026-2082-RC1

RC2: 'Comment on egusphere-2026-2082', Anonymous Referee #2, 05 Jun 2026 reply

This Brief Communication presents a relevant contribution to the study of Arctic sea-ice climate variability. In particular, the authors investigate whether the relationship between the current phase of the Interdecadal Pacific Oscillation (IPO) and the timing of an ice-free Arctic Ocean, previously identified in CESM1 projections by Screen and Deser (2019), is robust across other models. They extend the analysis to multiple large ensembles from CMIP5 and CMIP6 and find no consistent agreement across models that the initial IPO phase affects the timing of an ice-free Arctic Ocean. Nevertheless, in the majority of large ensembles, they do find evidence that the initial IPO phase significantly affects the sea-ice extent trajectory, although at different lead times across models.

The manuscript is well written and presents an interesting analysis that contributes to a better understanding of Arctic sea-ice variability. However, a few points should be clarified before publication.

We thank the anonymous referee for their positive assessment and their constructive and detailed comments. We have clarified all of the points raised, as described below.

Lines 54–56: Screen and Deser (2019) use an 11-year running mean for the IPO. Please clarify whether the same approach is used in the present study.

We have added clarification that we isolate low-frequency variability in the IPO by applying a zero-phase, 11-point Lanczos filter window via forward-backward filtering (scipy.signal.filtfilt), which follows Phillips et al. (2014). While this is not exactly the same method as Screen & Deser (2019), the member categorization in Figure 1 aligns precisely with their analysis for CESM1.

L61-64: ***"The IPO is calculated using the Henley et al. (2015) tri-polar index of annual sea surface temperature anomalies, with low-frequency variability isolated by applying a zero-phase, 11-point Lanczos filter window via forward-backward filtering (scipy.signal.filtfilt) following the method in the Climate Variability Diagnostics Package (Phillips et al., 2014).***"

Pooling experiment methodology: Please clarify the methodology used for the pooled experiments.

We have expanded the description of the pooling procedure in the methods section (see also our response to RC1, Line 60). The revised text reads:

L77-83: ***"For analyses where we 'pool' initial sea ice states in Section 3.1, we classify the IPO phase based on persistence above/below zero for 10 years. This persistence criterion is chosen both to capture the typical IPO phase transition timescale and to be consistent with the current persistently negative phase (Henley et al., 2015). Significance testing for the pooled IPO states is tested by whether the IPO phase means are different from the ensemble mean at the 95% level. Significance testing uses the number of LE ensemble members as independent samples, not the multiple samplings of the same ensemble member."***

In Figure 2, all trajectories appear to be initialized at 4.7 million km². Does the pooling procedure replace each ensemble member with an average over shifted versions of itself, or are the shifted trajectories treated as separate samples?
The pooling replaces each ensemble member with an average over the 11 mean-state trajectories for that member. Averaging the 11 initial mean-state trajectories is what produces the common 4.7 million km² starting point in Figure 2. The shifted trajectories are therefore not treated as separate independent samples and significance is tested on the number of ensemble members. For brevity, in the results section (L109) we refer back to the methods section where we have expanded the description as quoted for the previous point.

Figure 2: In panels A, there are small grey boxes near the left-hand side of the plots. Please indicate in the caption what these represent, or remove them if they are not needed. Please also clarify what the hatched regions indicate in panels B–E.
We have clarified both in the caption. The grey boxes indicate the trajectory-year averaging period used for the spatial composites in panels B-E, and the hatching in panels B-E indicates regions that do not meet the 0.05 significance threshold. The caption has been revised as follows:
"Grey boxes at the top of the subplots indicate the temporal period for which the spatial plots B-E are averaged … Hatching indicates regions of sea ice anomaly which do not meet the 0.05 significance threshold."

Section 3.2 methodology: Please clarify how the analysis in Section 3.2 is performed. Are the same trajectories as in Section 3.1 used? If so, at a given lead time before the ensemble-mean ice-free year, do different members start from different initial sea-ice extents?
Thank you for pointing out the omission of this section in the methodology. This analysis uses lead times relative to the ensemble-mean ice-free year, so at a given lead time individual members are generally at different sea-ice extents, which is now clear from the added text. We have included the following sentences in the methods section:
L83-86: ***"For the correlation analysis (Section 3.2) we compute correlations at lead times of 0 to 45 years prior to the ensemble-mean ice-free year. For each ensemble member for a given lead time, we compute the correlation between the IPO value at that lead time and the ice-free year relative to the ensemble-mean ice-free year (even if the ensemble member is already ice-free for a given lead time)."***

Line 115: Please clarify whether this sentence refers to the scatterplots in Figure 3 panels I–P.
It did refer to the former panels I-P. As those panels have been removed at the request of Referee #1, we have reworded the relevant text so that it no longer relies on them. The point is now made directly from the correlation panels (A-H).

Citation: https://doi.org/10.5194/egusphere-2026-2082-RC2

Editor Comments: Comment on egusphere-2026-2082
Dear Christopher and Alexandra
Thank you for providing author comments to both reviews.

The reviewers have already picked up most of my comments. At this stage I park detailed requests, except to work to strengthen the argument in the conclusion even further to what you indicate in the author comments.

I invite you to include information on the following in your revisions:
* Reference about the how well the IPO is captured in the LEs analyzed in your ms.
* Any relevant notes about how spatial variability of IPO in the LEs might degrade the teleconnection of IPO to Arctic sea-ice change.

Thank you, Petra

We thank the editor for her specific constructive comments. We believe that we will be able to clarify all raised points in a revised manuscript (see below).

Reference about the how well the IPO is captured in the LEs analyzed in your ms
We have added a reference to the fidelity with which each LE simulates the IPO, drawing on the CMIP5/6 IPO evaluation of Qin et al. (2024). Table 1 now includes an “IPO Skill” column reporting each model’s ranking (out of 52 models assessed) for the skill score of the SST distribution regressed on the IPO. This now appears as follows in the Methods and Data section:
L46-48: “***IPO simulation skill has been assessed for all models used here except for EC-Earth3 (Qin et al., 2024), with model skill ranging from 1st (CESM2-LENS) to 45th (ACCESS-ESM1-5) place from 52 models assessed (Table 1).”***

Further, in the conclusions, we now note the range of realism of the simulated IPO and Pacific-Arctic teleconnection from Qin et al. (2024) and Jahfer et al. (2024) to determine there is no consensus by excluding some models based poor realism metrics.
L173-176: ***“Qin et al. (2024) place CESM1 and CESM2-LENS in the upper quartile for IPO spatial realism (EC-Earth3 was not assessed), and CanESM5, CanESM2, and MIROC6 in the upper half (see Table 1). Jahfer et al. (2024) classify all CMIP6 models except MIROC6***

as having a strong Pacific–Arctic teleconnection (CESM1 and CanESM2, the two CMIP5 models, were not assessed).”

Any relevant notes about how spatial variability of IPO in the LEs might degrade the teleconnection of IPO to Arctic sea-ice change

We have added a note to the introduction that inconsistency in the simulated IPO pattern itself, independent of the IPO-Arctic teleconnection, can obscure any detected IPO-Arctic relationship:
L32-34: ***“Finally, because CMIP5 and CMIP6 models inconsistently reproduce the IPO’s observed spatial pattern, amplitude, and phase evolution (Qin et al., 2024), detection of relationships between the IPO phase and Arctic sea ice is further complicated.”***

In the revised conclusions we make the broader point that generalized realism metrics do not correlate with the IPO-Arctic signal, and refining model selection on IPO-Arctic realism specifically would be needed to reach a firmer conclusion to include potential mechanisms:
L176-178: ***“Thus, despite good relevant metrics for several of the models, we see large differences between the spatial and temporal IPO–sea ice responses, making it unlikely that a single common physical mechanism underlies them. We therefore find insufficient evidence for a single, realistic IPO–ice-free relationship.”***

The revised manuscript now cites the following additional references to support the realism metrics: Qin et al. (2024), Notz and SIMIP-Community (2020), Phillips et al. (2014). Additionally we have added a discussion of DeRepentigny et al. (2022) which offers a possible explanation for why CESM2-LENS, despite being responsive to the IPO in both its SIE trajectories and ice-free timing, does not show significant sea ice concentration anomalies at 17-22 year lead times:
L115-122: ***The two exceptions, CanESM5 and CESM2-LENS, differ in nature. CanESM5 shows no significant IPO influence on sea ice in any analysis presented here. CESM2-LENS, by contrast, shows a significant influence of the IPO on both SIE trajectories and ice-free timing, making its lack of a significant 17-22 year SIC signal notable. For CESM2-LENS this lead-time window corresponds to 2009-2030 in 96% of ensemble trajectories, which overlaps with the anomalous forced sea ice evolution in CESM2 driven by its sensitivity to biomass-burning aerosol emissions (DeRepentigny et al., 2022). This transient forced signal may obscure the internal IPO-driven SIC anomaly in the composite, offering a possible explanation for the muted response in an otherwise IPO-sensitive model.”***

References:

DeRepentigny, P., Jahn, A., Holland, M. M., Kay, J. E., Fasullo, J., Lamarque, J.-F., Tilmes, S., Hannay, C., Mills, M. J., Bailey, D. A., and Barrett, A. P.: Enhanced simulated early 21st century

Arctic sea ice loss due to CMIP6 biomass burning emissions, Science Advances, 8, 2405, https://doi.org/10.1126/sciadv.abo2405, 2022.

Notz, D. and SIMIP-Community: Arctic Sea Ice in CMIP6, Geophysical Research Letters, 47, https://doi.org/10.1029/2019GL086749, 2020.

Phillips, A., Deser, C., and Fasullo, J.: A New Tool for Evaluating Modes of Variability in Climate Models, Eos, 95, 453–455, https://doi.org/10.1002/2014EO490002, 2014.

Qin, Z., Wang, T., Chen, H., and Gao, Y.: Performance of CMIP5 and CMIP6 models in reproducing the Interdecadal Pacific Oscillation and its global impacts, International Journal of Climatology, 44, 3742–3765, https://doi.org/10.1002/joc.8548, 2024.