

Response to reviewers (*Deficient ocean–atmosphere feedbacks constrain seasonal NAO prediction* by Erik W. Kolstad)

I thank the reviewers and editor for once again assessing the manuscript. Below, you'll find your comments in italicised and blue fonts and my responses in black.

Response to Reviewer #1

Thanks to the reviewer for taking into account my earlier points. I am happy they have dealt with all the major points. Providing they can add suitable caveats in both the main text and the abstract about other remote effects (e.g. ENSO) that could influence pre winter SST and the winter NAO and hence give rise to apparent but non-causal effects between pre winter SST and winter NAO then I recommend publication. This last point was also raised by reviewer 2.

Thank you for the positive assessment and the helpful suggestion. I have added caveats in the Abstract (“While remote influences...”), Introduction (“It is furthermore acknowledged that remote effects...”), and Discussion (“Another limitation...”). I also added a short subsection (2.4 Scope). Accounting for these confounding factors would by the way make for an interesting follow-up study.

Response to Reviewer #2

The author has made a thoughtful and thorough revision. However, I think some of the new analyses illustrate the issues with the overall interpretation, which are related to the issues I already raised with the previous version. Due to these issues, I do not believe the manuscript addresses two of the things it sets out to do, namely (1) clarify the causal directionality of the relationship between the NAO and SHF/baroclinicity and (2) to assess whether biases in the mediation relationship underly the limited NAO prediction skill in SEAS5. I appreciate the authors approach and think there are some interesting results, but I think the paper still needs a lot of work on the overall framing of what it aims to do and what can be concluded from the analysis.

Major comments:

1. *Problems with ratio used to infer direction of Y/Z relationship: You've introduced on line 170 a ratio of regression coefficients with different orders of predictor and predictand to assess which direction of relationship is stronger. I would urge caution about introducing such an approach without a careful discussion of regression dilution. Regression dilution biases the OLS regression coefficient towards zero when there is noise in the independent variable, which means that this ratio will be highly influenced by which variable has more noise in it.*

Thank you for this insightful comment. First, I would like to clarify that never intended to claim that the $Z \rightarrow Y$ pathway (oceanic “nudge”) dominates over the $Y \rightarrow Z$ pathway (atmospheric driving). I fully agree that the primary directionality in the North Atlantic is $Y \rightarrow Z$. My goal was to identify specific regions where the $Z \rightarrow Y$ link was not significantly overwhelmed by the reverse driving.

While I still think the ratio has potential, I see that there are potential issues related to the interpretation of it. In the revision I therefore removed references to the ratio and instead focus on α' versus α , where both coefficients are estimated for the same dependent variable Z . α captures the total $X \rightarrow Z$ effect (including routes via Y), while α' isolates the SST-forced, NAO-independent component by regressing out Y . This comparison addresses the causality ambiguity and avoids the unit inconsistency you identified in your comment below.

Regarding the dilution issue, I appreciate your guidance. I've now cited *Econometric Analysis*; Greene, 2003 for this. Because both α and α' are affected by attenuation, I devised a conservative sign-consistency test. The null hypothesis is that they have opposite signs. This is conservative because both parameters are biased towards zero, making sign swaps more likely in the bootstrap sampling process. Rejecting this null indicates that the SST-forced component is aligned with the total contemporaneous relationship despite attenuation; I explicitly do not claim that ocean forcing dominates atmospheric feedbacks. I hope you find this framing to be consistent with the literature (e.g., Patterson et al. 2024, which I now cite). I've also attempted to soften my conclusions accordingly to emphasise two-way coupling and regional heterogeneity rather than SST-driven dominance.

However, I think the problems with the inferences from this ratio (gamma/beta) run deeper than just regression dilution. Thinking about the model world (SEAS5), in equation 5 you are regressing out a significant relationship between X and Z when determining the relationship between Y and Z, but in equation 2 you are regressing out an insignificant relationship between X and Y when determining the relationship between Z and Y. However, the relationship between X and Y is significant in ERA5, and I think the model-ERA5 differences in this ratio is therefore telling you about differences in the relationship between X and Y between SEAS5 and ERA5 and not about the differences in direction of Y->Z or Z->Y relationship. Any analysis that is telling you that SHF->NAO is more important than NAO->SHF should be taken with extreme skepticism given the extensive literature about the importance of atmosphere driven SHF (one recent example out of many: Patterson et al. 2024, <https://doi.org/10.1175/JCLI-D-23-0602.1>).

I agree about the caution. Further, the comparison between ERA5 and SEAS5 was not intended to be a key point. The intention was to investigate ERA5, where as you point out the $X \rightarrow Y$ relationship is significant. In the new version of §4.5 with α and α' I have rephrased the discussion. For instance, the last paragraph there now starts with “In

summary, the directional picture is heterogeneous but broadly consistent with a dominant NAO→mediator pathway”.

Additionally, I believe there is a mistake in the equation for the ratio (line 170), since beta has units of Y/Z, whereas gamma has units of Z/Y, such that one or the other is inverted before the ratio is taken. This is true even if all quantities are normalized.

I'm not fully convinced this is a problem after standardising. Nevertheless, I have now, as mentioned above, moved away from the γ/β ratio. This is not a problem when comparing α and α' .

2. Remaining interpretive issue affecting most conclusions: Figure 7 is an excellent demonstration of an interpretive issue that I believe runs throughout this analysis. As I've mentioned in my comments on the previous version, if there is a coincidental correlation between the November SST pattern and the DJF NAO, this will show up as a large mediated effect, because both of the mediated fields are strongly correlated with the NAO (with the causality primarily NAO → mediator, at least for SHF). I do see your point in the last paragraph of 4.5 that the NAO skill is a measure that doesn't depend on the SST or mediator fields, and I therefore agree that this analysis is showing a relationship between the NAO skill and total effect of X on Y (as was already shown in Fig. 2). However, bootstrapped samples that have a large skill/total effect will necessarily have a large alpha, because the mediators are strongly correlated with the NAO. And since beta is always large (the mediators are strongly correlated with the NAO), this will show up as a large mediated effect.

From this, I think one can conclude that there will be a large apparent mediated effect when there is coincidentally a large SST-NAO relationship, even if the mediators play no causal role whatsoever. While you could point to the small direct effect in something like Eq. 2, all this really shows is that the contemporaneous mediators are more correlated with the DJF NAO than the November SSTs, such that they are given more weight in the multiple linear regression. However, it is still valid to interpret the mediators as being caused by the NAO, since there are also problems with the analysis used to infer the directionality of this relationship.

I want to acknowledge that Section 5 (lines 459-462, 469-473) does discuss these potential issues with some nuance, however, I think this raises doubt as to the importance of the mediation pathways introduced, which in turn partially undermines the premise of the study. I think there is still value in the analysis presented, but quite some work is needed on how it is framed and interpreted.

Thank you for raising these important points. I believe there is a small but important conflation between “skill” and the internal “total effect” in the statement that “bootstrapped samples that have a large skill/total effect will necessarily have a large alpha, because the mediators are strongly correlated with the NAO”. In the manuscript

these are distinct quantities. The total effect (the internal correlation between the November SST index and the DJF NAO index) is computed entirely within SEAS5. By contrast, “skill” is an external validation metric: the correlation between the SEAS5 DJF NAO index and the ERA5 DJF NAO index for the same bootstrap sample. It is correct that a bootstrap sample with a large internal total effect will often exhibit a larger α (and hence a larger internal mediated effect) because the mediators co-vary contemporaneously with Y (which contributes to a realistic β pattern in SEAS5). However, this implication does not extend to the external skill metric. A sample can display a strong internal $X \rightarrow Y$ relationship yet still have low skill if its DJF NAO is out of phase with ERA5 or differs in amplitude. Conversely, high external skill can arise in samples that, for unrelated reasons, capture other sources of predictability (e.g., stratosphere–troposphere coupling, tropical forcing, or internally generated atmospheric variability) that have little relation to the November SST pathway examined in the paper. For these reasons, a large internal mediated effect does not necessitate high external skill, and high external skill does not require a large internal mediated effect.

This distinction underpins the interpretation of Figure 7. In that figure, ERA5 enters only through the external skill calculation via the ERA5 DJF NAO time series. The mediated effect is derived solely from SEAS5’s internal covariances among X , Z , and Y within each bootstrap sample. The mapped correlation in Figure 7 therefore relates an external validation metric to an internal SEAS5 pathway; it is not an algebraic consequence of mediator–NAO covariance within the model. The salient feature is that the association emerges in the regions where ERA5 indicates robust mediation (see the white contours in Figure 7), while the (mean) mediation in SEAS5 is not necessarily robust in the same areas (Figures 4f and 5f). Within SEAS5, those resampled subsets that happen to realise a more ERA5-like mediation structure in these dynamically sensitive regions are also the subsets that better match the ERA5 NAO. This is not presented as evidence that the mediators causally drive the NAO; rather, it indicates that reproducing the observed structural pathway is associated with improved forecast performance, even though the model’s mean mediation in those regions may be muted. To avoid any residual ambiguity, I have now clarified already in §4.2 that “skill” denotes external agreement with ERA5 and is distinct from the internal total effect or the internal mediated effect. In practice this meant moving some of the text from the final paragraph of §4.6 to §4.2.

Minor comments:

- *Just a note that if you detected slight disapproval of the term “indirect effect”, it was only because this is commonly used for aerosol effects on climate. However, the context is sufficiently distinct that I didn’t find it too confusing. Nevertheless, it’s nice to use “mediated effect” to avoid this terminology overlap.*

Noted!

- Line 26: I suggest “regional” instead of “local”, since the latter implies a more grid-cell-level relationship (at least to me)

Changed.

- Line 60: Consider adding a bit more transition here from the previous sentence to help the reader understand the overall flow of the introduction

Good idea. I’ve modified the paragraph starting with “A large body of work”, where I now first cite the Patterson study and emphasise that the atmospheric impact tends to dominate. The flow works better now, so thanks!

- Line 81: I am always a bit skeptical of statements such as this about its previous use in a field. I recently became aware of a land-atmosphere interactions paper (Khanal et al. 2024) that uses partial correlations in a similar way, although they do not spend as much time writing out the formalism. Still, it seems that this sub-field is somewhat aware of this type of approach. I don’t have a specific recommendation here, but consider watering down this statement.

I appreciate this observation. I have followed the suggestion to water down the claim of novelty and included two recent references.

- Line 90: I was anticipating O’Reilly et al. 2024 (<https://doi.org/10.1038/s41612-023-00335-0>) to be an example here (but I don’t have a strong opinion about whether you should include it or not)

That’s a very nice example. I have included it in the new version. Thanks!

- Line 106: Italics on “informed” implies it is defining a term, when that doesn’t really seem like it is the case here. I’m also not entirely convinced by this word choice

You’re right. I deleted the word.

- Line 112: Is the word “observed” necessary?

The intention was to distinguish it from sensitivity experiments, but I now think that whole sentence can be deleted (so I did).

- Line 116: Avoid “it is obvious” without an explanation. I’m assuming you mean because of the time offset (among other reasons)?

Indeed I did, and I’ve specified this now (and deleted the “obvious” part).

- Line 123: “absurd” is a word I generally don’t recommend in scientific writing

Fair enough. I’ve rephrased now.

- Line 148-149: Isn’t the last sentence only true if the other mediator were correlated with Z, and if so, is it maybe worth adding “if they are correlated with Z” at the end?

No, it’s entirely possible (and likely) that there exist other pathways that don’t involve any of the mediators investigated in the paper. However, as the last sentence was superfluous and could be misunderstood, I deleted it.

- Line 158-161: Too much going on in this multi-part sentence. Please break it up for readability.

Done.

- Line 163: Since the mediators have not been introduced yet, please add “that we consider” or similar somewhere in this sentence

I’ve rephrased this now.

- Line 185: Presumably you mean “ordinary least squares” (this is an important distinction to make, related to my major comment about regression dilution)

Yes. Added.

- Line 212: “qualitatively similar” would be better than “qualitatively identical”, because “identical” is more of a quantitative statement

Replaced “identical” with “similar”.

- Line 272: “still weaken” is confusing. I think you mean “weaken further”

It actually says “skill”, not “still”.

- Line 273: “The sample parameter used in the remainder of the analysis” doesn’t make it very clear what this is, i.e., that it is the member-level regression coefficient. Please also quote the range across ensemble members for this and for the correlation coefficient

I have rephrased this now for clarity and specified for each metric that they are not significant. I’m not sure what you mean by “the range”; presumably a confidence interval? I think it suffices to write that the values are not significant.

• Line 315-317: This sentence is confusingly worried. Also, it's necessary to call this a weak negative SHF bias, because it is hardly visible compared to the surrounding positive biases. This will also help readers understand what anomalies you are referring to.

Rephrased.

• Line 357: I am getting tripped up by "it is useful" here. Specifically, it is not clear whether "it" refers to the hypothesized $X \rightarrow Z \rightarrow Y$ relationship (I don't think so, because this wouldn't make sense to me) or whether "it is

I think the last part of this comment didn't make it through. I have rephrased the sentence to remove the vague "it" and explicitly state that the goal of this specific step is to isolate the component of the $X \rightarrow Z$ link that is statistically independent of the concurrent NAO variability.

• Figure 6a,d: Are these DJF SST and SLP? I assume so, because I don't think November SST and SLP would make sense here, but the caption and text don't really state this as far as I can see.

I agree that this should be made clear and have done this.

• Line 379: Typo. There should be no X here.

Thanks for spotting this. I have rewritten this paragraph now based on your major comment.

• Line 385: It would be helpful to clarify that this sentence only applies to the suppression in the Gulf Stream region and not to the SHF relationship as a whole (e.g., the positive mediation in the subpolar region that dominates the full mediated effect)

I deleted this whole sentence, as it just repeated what comes just before.

• Line 396-398: I think this summary text understates the importance of NAO- \rightarrow mediator feedbacks that are evident in the differences between equivalent panels of Fig. 6 and Fig. 4/5.

I agree. The last paragraph in that section has been rewritten (see response to major comment).

• Line 399: Your analysis does not demonstrate "the model's essentially one-way representation of these pathways". See major comment 1.

Agreed. Deleted.

- *Line 496: I don't know what is generally recommended on this, but it seems strange to me to acknowledge a tool that itself does not acknowledge its sources.*

The journal instructions required me to acknowledge the use of such tools.