RESPONSE TO REVIEWS OF WCD PAPER (2023):

“Predictable Decadal Forcing of the North Atlantic Jet Stream by Sub-Polar North Atlantic Sea Surface Temperatures”

RESPONSE TO REVIEWER #1

General response

We thank the reviewer for their helpful comments. Before replying to specific points, we want to draw the reviewers attention to some key changes made in response to Reviewer 2:

- The CMIP multimodel analysis has been removed entirely. A footnote is included in Section 5.1 commenting on the result briefly for the interested reader.

- The analysis related to the signal-to-noise paradox (and related speculation on turbulent heatfluxes) has been removed, due to the issues the authors raised with the editor and reviewers in the most recent review period. This also means the question of the role played by how well atmosphere-ocean coupling is simulated is not considered.

- The quantitative estimates of the stochastic forcing using two different methods has been removed.

In terms of structural changes, we have also now moved the analysis on timescales (“Decadal forcing as the accumulation of seasonal timescale forcing”) to Section 5, which has been renamed to “Pathways and timescales”.

These changes were done to make the paper more concise and flow more naturally, as requested by Reviewer 2.

We now respond to specific comments.

Reviewer #1: *I see the point raised by the other reviewer and the editor regarding the usefulness of section 6 on potential drivers of the SPNA surface temperatures: it is perhaps not strictly necessary given the main topic in this paper, namely the predictability of the jet speed arising from the SNPA. The paper is still somewhat long and topics in section 6 might as well be the object of a separate, more detailed* *study on its own, thus its exclusion might improve the readability of the paper. At the same time, it provides (some) insights into more fundamental drivers of jet predictability. All in all, I leave it up to you whether to include it or not.*

We thank the reviewer for their feedback on Section 6. We agree that the results of Section 6 could form the starting point of a separate study, and perhaps in an ideal world that's what ought to happen. Unfortunately, the authors do not currently have any plans of doing such a follow-up study any time soon. Since the role of aerosols/AMOC comes up very often when discussing these results with colleagues, we have opted to leave Section 6 in anyway. Thanks for your understanding.

Reviewer #1: *Line 233: You could specify the actual value in brackets (C = -0.35)*

We have rephrased these lines to now say: "The jet latitude correlations are never significant with respect to a similar null hypothesis, including the relatively large negative correlation emerging for 30-year jet latitude variability in DPLE ($C=-0.35$)."

Reviewer #1: *Line 288: There is an extra 'is' after 'suggests'*

The guilty line has vanished as a result of edits in response to reviewer 2.

Reviewer #1: *Line 305: I guess you are referring to Figures 3 and 4? Can you refer to them explicitly?*

We now do.

Reviewer #1: *Line 415: Apologies for my pedantry, can you include units for R and f?*

We accept both the request and the apology: units are now included.

Reviewer #1: *Line 465: 'Understand' is repeated twice, could you change the first into something like 'Identifying drivers of the variability in SSTs…'?*

We made the suggested change.

Reviewer #1: *Line 481: I would rather have '100-75W' than '260-285E'*

We've changed it to 100-75W.

Reviewer #1: *Line 700: Typo in 'North Atlantic' ('Noth')*

We fixed the typo, thanks.

Reviewer #1: *Figure 12: Hard to tell apart the ‘thin’ and ‘thick’ boxes, perhaps make one of the two thinner/thicker? Just a suggestion, it is clear from main text which is which, of course.*

We played around with that but it was a bit tricky to balance getting the NA box thick enough to easily tell it apart vs making it so thick it looked silly. In the end we decided that the contextual clues from the text should suffice, and we are glad the reviewer found this to be true. We’ve thus left the boxes as is.

RESPONSE TO REVIEWER #2

Reviewer #2: *The manuscript presents investigations of different topics / tools: atmosphere-ocean coupling, seasonal and decadal predictability, CMIP-like model bias, AMOC, stratospheric aerosol… In all these topics, the authors focus on the Atlantic jet speed and its link with subpolar Atlantic SST. But the links between all these analyses remain not clear. This is also linked to major point three, as a lot of discussions and speculations are needed to somehow link all the analyses. I strongly recommend that the authors be more focused and concise. I listed quite a lot recommendations in the previous review to help if needed.*

We accept the reviewers recommendation to be more concise. The following major cuts have now been made:

- The CMIP multimodel analysis has been removed entirely. A footnote is included in Section 5.1 commenting on the result briefly for the interested reader.

- The analysis related to the signal-to-noise paradox (and related speculation on turbulent heatfluxes) has been removed, due to the issues the authors raised with the editor and reviewers in the recent review period. This also means the question of the role played by how well atmosphere-ocean coupling is simulated is not considered.

- The quantitative estimates of the stochastic forcing using two different methods has been removed.

In terms of structural changes, we have also now moved the analysis on timescales ("Decadal forcing as the accumulation of seasonal timescale forcing") to Section 5, which has been renamed to "Pathways and timescales". A number of other minor cuts have been carried out to trim the length of the paper further.

In addition to the cuts already made in the previous round of revisions, this has had the effect of reducing the paper from 32 pages and 16 Figures (original submission) to 27 pages and 13 Figures (present submission); note that the Figure count has only not dropped by more because (a) in response to the reviewer's comments on Figs 3 /4 we made changes to analysis which prompted a new standalone figure (new Figure 5) and (b) the reviewer's comment about whether the timescales could be assessed further prompted us to include a new figure addressing this (new Figure 11).

We hope that the paper flows more concisely and naturally now. In essence the structure now goes as follows:

1. Show that predictability is associated with the jet speed.
2. Show that the SPNA is the only obvious source of a jet speed signal.
3. Pathways and timescales:
   * Argue that the link can be understood as a result of changes to the

meridional temperature gradient.
* Timescales: Argue that the link is taking place on seasonal timescales with no multiannual lags needed.
4. Briefly consider if the jet speed perspective sheds further light on the relative role of the AMOC and aerosols in generating decadal atmosphere-ocean covariability.
5. Discussion/Conclusions

We accept that the paper could be made even more concise by removing Section 6 (about the AMOC and aerosols). However, our experience is that the analysis we carry out there answers questions that very often arise when communicating this work to other interested colleagues. We also do not think the results of Section 6 by themselves would be worthy of publication in a separate article, and the authors have no present intention of seriously expanding on this part of the work. Finally, the result showing that AMOC oscillations can generate predictable jet speed variability adds further weight to the argument that there is a causal influence from the SPNA.

Since the other reviewer and editor have indicated they are happy for Section 6 to remain in the paper, we hope that reviewer will be willing to accept us having left it in.

Reviewer #2: *The statistical analyses are much improved. But I still have major issues with some of the results presented.*

- Figure 3 & 4 and text line 274-289 : the authors present in the figure with first 'a visual inspection'. Then the reader learns that the authors used on purpose in the figure a statistical test not adapted (not accounting for autocorrelation of SST). This is because the authors assume that the statistical level of significance found in changes that are significant in the four datasets would then lead to a correct test. This is quite confusing.

Why not just provide a figure with correct statistical testing? And then say that qualitatively, the SPNA SST/jet speed is significant in most of the dataset with the same sign (negative) in all four dataset.

- The authors sometimes discuss 5% and 10% significance in each of the four datasets and discuss all the small differences obtained. Sometimes such discussion is absent. I believe that the authors can illustrate the same conclusion while keeping the same level of statistical significance (10% or 5%) in the whole paper… this would make the paper more concise.

About Figures 3 and 4, we have implemented the reviewer's suggestion. Stipling now denotes significance at p=5% with respect to a null hypothesis that is now consistent with what is done elsewhere. On interannual timescales, the jet speed is modelled as white noise, and SSTs at gridpoints modelled using the phase shuffle method. On decadal timescales, the phase shuffle method is applied to both the jet speed and SSTs.

We are however still keen to include a more 'objective' filtering method than just visual inspection, which could be accused of being non-rigorous and places some extra burden on the reader. We have therefore added a new Figure 5 which simply highlights gridpoints for which the correlation has the same sign across all 8 panels of Figures 3 and 4, with no restrictions made about significance of the correlations; the old subplots (e) have been removed from Figs 3 and 4 as a result. We hope the reviewer will agree this is a simple and transparent method which achieves the same ends, namely to clearly highlight the SPNA. We then remark, as the reviewer suggests, that the correlations in the SPNA are in general significant.

Note that some of the discussion here has been altered or rearranged as a result of this change. The relevant paragraphs are highlighted in red.

About the 5 vs 10% significance: currently the only place a 10% significance threshold is discussed is in Section 3.1, where we discuss the significance of the forecast skill. This is motivated by the discussion at the end of this section, where we compare our results against those of Smith et al. and Athanasiadis et al. They both report significance with p=5% and we do not, and we consider it proper to discuss this at least briefly given typical conventions and assumptions in the field concerning p-values. The alternatives do not seem ideal to us: if we simply pick a 10% threshold everywhere then many readers will naturally wonder why we don't pick the more common 5% threshold, and it may appear as if we are concealing something. On the other hand, simply stating that we don't achieve 5% significance with no further comment also does not seem ideal.

Comparing our results to Smith/Athanasiadis et al requires that we provide some additional information beyond just an assertion of significance, such as an explicit p-value estimate or explicit significance thresholds: once this information is given it becomes obvious that the 'discrepancy' between our results and those of Smith/Athanasiadis is probably just noise. Here we have opted to provide explicit thresholds because of our general belief that thresholds are more easily interpreted than p-values.

We have thus left this one mention of the 10% threshold as is: the rest of the paper discusses only 5% thresholds. We hope this will be acceptable to the reviewer.

Reviewer #2: T*he discussion p24-29 (5 pages) should be reduced… I suggest reducing at least section 7.2 , 7.3 and 7.5.*

Section 7.5 has been removed, along with the general analysis of air-sea coupling. Section 7.2 has been shortened by removing the figure related to stochastic forcing: we now just refer to quantitative estimates that are not shown. Section 7.3 has been trimmed slightly. We hope that these changes have left the discussion a more reasonable size.

Reviewer #2: *Regarding section 5.3 and Figure 11. The atmospheric forcing of the SST is simulated in ERA20C, as the observed large scale atmospheric conditions are consistent with the observed SST. In ASF20C, the atmospheric large scale conditions are only weakly related to the SST, as the atmosphere does not influence* *the SST (which is prescribed), and as the SST only has a weak influence on the atmosphere. This implies that one expects a large correlation between the atmosphere and the ocean in ERA20C when SST is lagging the atmosphere or at lag 0. As the atmosphere is somehow persistent (autocorrelation remains significant lag 10 to 15-days) the dominant atmospheric forcing of the SST is expected to be visible when the SST leads by 0 to 10-15 days. Therefore, I disagree with the conclusions in line 443-446. This is not a bias of ASF20, but only reflects that the SST and atmosphere are mainly uncorrelated in ASF20C in the absence of assimilation. Then, the link between this analysis and the signal to noise paradox or the link SPNA-jet speed is not clear.*

We agree: as communicated to the reviewers via the editor, there were several problems with this analysis. While the authors think that the datasets we considered could be used to analyse the role of air-sea coupling biases in generating the paradox, this is now more properly left to future work. We have deleted all relevant sections, including this one.

Reviewer #2: *Regarding section 4.3 and figure 6: I do understand the analysis but I do not understand why the authors did it. The results look obvious. The reader already knows with Figs. 3 and 4 that the correlation is significant at both interannual and decadal time scales. If the covariability at decadal time scale dominates, it will affect both the interannual (includes both interannual and decadal variability) and decadal time scales. This can explain the results of Fig. 6 and this is not really in agreement with the conclusion L352-353 : ‘’the decadal time scale correlations [...] can be completely explained by the interannual […] link”. Maybe one can investigate the coherence spectrum to understand the time scale behind the correlation?*

As discussed in the paper, there have been outstanding questions around the timescales involved in AMV-NAO links. Most fundamentally, several studies have observed the apparent multi-annual lags between the two, but it has remained unclear exactly how to interpret these lags. This gives clear motivation for considering the question ourselves. We are simply making the point here that multi-annual processes are not required for the teleconnection: the seasonal timescale interaction between the SSTs and atmosphere completely suffice to explain the teleconnection.

We do not agree with the reviewer that this is obvious, since other papers (like Kwon et al.) explicitly suggest an idea by which there is a lag of multiple years (which they interpret to be the time it takes for the SST pattern to evolve to one more optimally situated to perturb the jet).

Our analysis certainly doesn’t completely answer all questions about timescales, it just proves sufficiency. We consider it beyond the scope of the paper to delve more deeply into the coherence spectrum, especially since we have been generally asked by the reviewers and editor to shorten the paper. However, we have added a new figure which both supports our assertion about timescales and the discussion on how our results compare to the AMV-NAO links: the new Figure 11 shows yearly lag-correlations between both the SPNA and JetSpeed and AMV and JetSpeed. The large lags in the AMV-NAO links are also seen here, but when using the SPNA these vanish, with the correlations peaking at lag 0. For increasingly negative lags (SPNA leading by more and more years) the correlation diminishes to zero. This is consistent with the ‘instantaneous’ impact of the (decadally varying) SPNA and the argument that the long lags in the AMV are an artefact of not having isolated the key region (discussed in Section 7.3)

We hope this is acceptable. See also the response to the next comment for related comments.

Reviewer #2: *L89-90: ‘’with no need to invoke mechanisms spanning multiple seasons as in several other studies (Peings and Magnusdottir, 2014b; Kwon et al., 2020)”. Indeed, all these studies invoke mechanisms leading to the evolution of the SST over several years, involving the ocean circulation. However, the atmospheric response to SST is argued to be fast (few weeks or month) in the two studies mentioned. I do not think that the mechanism for the atmospheric response to ocean surface anomalies disagrees with the findings shown. I therefore suggest that this sentence should be revised. The same remark applies to L250-252.*

Yes, upon reexamination we agree that our comments on these papers were misleading. The main point we wish to make is that their results, and results more broadly concerning the multi-year lags seen when correlating the AMV and the NAO (or similar) create ambiguity about the timescales involved in generating a strong atmospheric response. We have rephrased both parts to rather just say that the multi-year lags discussed in those studies do not seem essential for generating a strong predictable atmospheric response.

We have also rephrased Section 4.3 somewhat for the same reason. Finally, we added the new Figure 11 (see above response).

Reviewer #2: *L169-173: please mention if the NAO indices are standardized.*

The removal of a seasonal cycle already implies the mean is zero. We have clarified now that the timeseries is not standardised further.

Reviewer #2: *L199-202 : I wonder why DPLE is used in this study : the authors show the results but decide to only discuss ASF20C and CSF20C in case DPLE do not agree with the other data.*

We consider it crucial to show consistency of the results with DPLE, because ASF20C/CSF20C are not genuine decadal forecasts. If we did not show DPLE, it would be entirely unclear if the assertion that decadal jet predictability can be obtained from the SPNA holds in an actual decadal forecast. Similarly, it wouldn’t be possible to rule out that the lack of jet latitude predictability is a feature particular to seasonal hindcasts which might not hold in an actual decadal forecast. Knowledge about which SSTs DPLE can and can’t predict is also crucial: the fact that the SPNA is one of the few regions it can predict well on decadal timescales adds much weight to the importance of the SPNA.

Furthermore, after the various cuts and revisions, there are no longer any parts of the paper where we end up explicitly disregarding DPLE in favour of ASF/CSF20C. We have thus removed the line in the Methods mentioning that we will favour ASF/CSF over DPLE if there are disagreements.

Reviewer #2: *Table 1: why not just show a symbol (*) or (**) if the correlation shown in panels of Figs. 1 and 2 is significant at the 10% or 5% level, respectively?*

Because we believe it is more transparent to just provide actual numbers, and that these are easier to interpret. The question of whether or not there is actual forecast skill is obviously critical, so we want to be very transparent about the comparison of our significance tests vs those done in Smith et al and Athanasiadis et al.

Reviewer #2: *Figure 3 and 4: perhaps a better presentation would be to show all the 1y correlations in one Figure and all the 10y correlation in another figure.*

Good idea: we have made this change, and reworded the text accordingly.

Reviewer #2: *L327-331 : this discussion is perhaps not needed.*

Perhaps, yes. However, it is our general experience that the point raised by Shepherd (to not get obsessed about p-values and forget to incorporate prior knowledge) is not one universally acknowledged in our field. Indeed, Shepherd's paper came out only in 2021. We would therefore like to err on the side of reinforcing the point for the benefit of readers less familiar with it. We hope the reviewer is understanding of this choice.

Reviewer #2: *Figure 6 and L348 : ''observed". Do the authors mean the actual correlation obtained in each dataset? This is not calculated from observations, right?*

Yes, that's what we mean. We reworded to say "actual" rather than observed. A similar change is made in the caption of Figure 6 (which is now Figure 10 after the move of this subsection to Section 5).

Reviewer #2: *P18-19: again, it is obvious that the geostrophic balance applies here. Indeed, the tropospheric temperature below 30°N is only weakly linked to the jet speed (Fig. 7a). This could be related to the weak horizontal gradient of temperature in the tropics. Therefore, it seems obvious that the jet speed variability is mostly related to the extratropical temperature. To explain the mechanism linking the SPNA*

SST to the jet, one expects to see more analyses using lags or diagnostics of the storm tracks and eddies.

It is obvious that standard geostrophic balance will hold to very good approximation on decadal timescales. However, we do not agree that it is obvious that the meridional temperature derivative across the jet core can be completely approximated using only the difference in air temperature between the jet core and the SPNA region. We also do not think it is obvious that one can reconstruct almost all the relevant decadal variability in this air-temperature-difference from the SPNA region alone (i.e. assuming that temperatures in the jet core are constant in time).

In particular, the fact that tropospheric temperatures south of the jet core are only weakly linked to the jet speed does not seem obvious in advance. The reviewer provides a plausibility argument (weaker gradients in the tropics) but if we hadn't confirmed this quantitatively (e.g. with our Figure 8b showing CMIP multimodel spread, which is what we presume the reviewer meant to refer to) then we do not think this plausibility argument is strong enough to make our quantitative reconstruction of the jet speed variability totally obvious in advance. After all, earlier studies (like the one by Ceppi et al. we discussed in the previous draft) explicitly link the speed of the eddy-driven jet in the southern hemisphere to the Hadley Cell and hence tropical heating. The fact that the tropics don't seem important for the North Atlantic jet speed does not seem to have been noted before: Ceppi et al just say they don't find a similar link to the Hadley Cell in the North Atlantic with no further comment. In fact, the fact that Ceppi et al. obviously looked for a link with the tropics in the North Atlantic but didn't find one is why we explicitly discussed our CMIP multimodel results in the original draft of our paper.

We have therefore left this quantitative computation in the current revision. We have highlighted further the assumptions we make and how these go clearly beyond basic geostrophic balance.

We emphasise that of course we are not making any profound statements here and do not for example shed any more light on the exact fast timescale dry and moist processes which bring the jet into geostrophic balance following a heating anomaly. We are simply arguing that if the SPNA is heating the atmosphere above it then geostrophic balance forces the jet speed to respond the way we observe. We have added a line emphasising this in the discussion of this result.