Please find below the responses to Referee#1 (pages 1-9), followed by the responses to Referee#2 (pages 10-12)

Referee 1, 2nd review

We thank the anonymous referee for their valuable input and for highlighting weaknesses in the text. In response to their comments, we thoroughly revised the text to reflect more nuance on the results obtained. We also added analysis and refined some of the existing findings. We now find the manuscript to be substantially improved and thank the reviewer for helping us to get there. Major points addressed were as follows:

- We rewrote major parts of the manuscript to stress to what extent D16's mechanism influences the skill of SST and elaborated our interpretation and discussion. We particularly emphasise the seasonal and regional dependence of D16's mechanism.
- We include difference maps in the hindcast skill analysis to allow a more detailed comparison of the effect of D16's mechanism on the SST predictive skill in Fig. 8.
- Thanks to the reviewer's main concern on our surface heat fluxes analysis, we revisited our code and found a mistake, which resulted in a new Fig. 6 now in line with observations and facilitates the discussion of our results.
- We performed an additional analysis of ocean heat transport convergence and SST, which supports our discussion of the regional influence of D16's mechanism on SST predictability.

We provide below a point-by-point response to each comment. Please note that the referee's comments are highlighted in **bold** font, while our answers are in regular font.

The authors have responded to by comments and criticism, and the manuscript has been improved. The new analysis of ocean heat transport convergence is useful, and the new figure 9 summarizing prediction skill is most helpful. However, the new figure 9 also highlights what is still my major concern. Throughout the text the authors highlight the skill differences consistent with their proposed mechanism, but these differences are sometimes marginal and (based on the new figure 9) do not seem to be significant. Considering that this is the key result of the paper, this apparent lack of significant differences should be better discussed. To conclude, although the manuscript has been improved, I still believe that it is in need of some substantial revisions before it is ready for publication. Detailed comments are found below.

We thank the reviewer for this assessment. Upon revisiting the manuscript, we have made several changes (as noted above and below) and are now confident the reviewer will find the figures reflected in both presentation and discussion of our results in the text.

Major comments:

1. Differences in prediction skill: As mentioned above, it is not clear from the text and figures to what extent there is a significant increase in prediction skill by considering the state of the AMOC. On l.310 it is stated that "significant skill difference (defined by non-overlapping confidence intervals) between strong and weak AMOC phases", but in my opinion the differences in skill shown in figure 9 do not reflect how the authors present the results and conclude. Comparing with the similar figure 11 in Borchert et al. (2018), it looks like the skill differences in the present study are much smaller (less significant). Hence, I think the presentation and discussion of skill differences could still be improved.

We thank the reviewer for pointing this out. In response to this reviewer comment as well as the major comment from Referee 2, we computed skill difference maps based on AMOC phases (added to Figure 8), introduced persistence forecast as a benchmark in Figure 9, and re-considered our discussion and conclusion section as a result. We also added additional analysis on the OHT convergence issue (see major comment 2). All of these new analyses lead to a reframing of our results and the entire manuscript body, including the title.

In the Discussion and Conclusions sections specifically, we now make the point that the influence of AMOC and D16's mechanism on seasonal SST prediction skill is very nuanced (it is really only significant during summer in the subtropics), and so the value of this mechanism for predictions is very specific. We made an effort to reflect this nuance in every section of the manuscript and hope that the reviewer agrees that the results are now presented and discussed in a much more balanced fashion.

2. Ocean heat transport convergence: The authors assess OHT convergence as the difference between two latitude bands (l.109-110, l.200-201). Why not calculate the actual convergence? From the location of Box 1 and 2 it is hard to believe that zonal transports are negligible. This should at least be checked and mentioned in the text.

Unfortunately, due to data availability and issues of computational effort, calculating the actual heat budgets for the two boxes was not possible in the revision of this manuscript. However, we agree with the reviewer that more effort was needed to detail the influence of AMOC-related meridional heat transport and convergence on SST in boxes 1 and 2.

We approach this issue with a twofold analysis. First, we compare correlation between OHT convergence and SST variations in both boxes to an equivalent analysis using SST in the latitudinal bands covered by the OHT convergence analysis instead of the boxes (Fig. R1). If correlations are similar between the SST boxes and the corresponding latitudinal bands (which are independent of zonal transports), that would be an indication that zonal transports have a small impact on SST variations in those boxes. We find for both boxes a decrease of correlation when using the latitude bands instead of the boxes. This indicates a non-negligible influence of zonal transports, as suspected by the reviewer. That being said, the correlations remain significant (95% level) when using box 2 or the corresponding latitudinal band, which indicates that meridional heat transport and convergence contribute

significantly to SST variations in the subtropics, despite a relevant contribution from zonal transports of heat. Box 1 SST exhibits insignificant impact of heat convergence as a whole.

Second, an analysis of spatial SST correlation patterns to meridional OHT convergence (Fig. R2) supports the previous assessment. This analysis was done to find the spatial imprint of meridional heat convergence on SST. Across all seasons (but most prominently MAM and JJA, see manuscript), meridional convergence of heat in the box 2 latitudinal band explains SST variations in the box, which indicates a relevant impact of OHT convergence there, whereas OHT convergence in latitudes of box 1 does not explain a significant portion of SST variability there (in line with the assessment shown in Fig. R1). It is noteworthy that compared to the corresponding figure using AMOC instead of OHT convergence (paper Fig. 5), correlations are reduced for the box 2 latitudinal band. This indicates (as before) that zonal transports contribute to a non-negligible extent to SST there. Since correlations in Fig. R2 (right column) are mostly significant in box 2, however, we conclude that the impact of meridional heat convergence on SST in box 2 is significant.

We made an effort to reflect these new results in the manuscript, rephrasing many parts (maybe most prominently in lines 210-212), and to discuss our findings in a more nuanced fashion. We find this greatly improved the manuscript and therefore thank the reviewer for this suggestion.

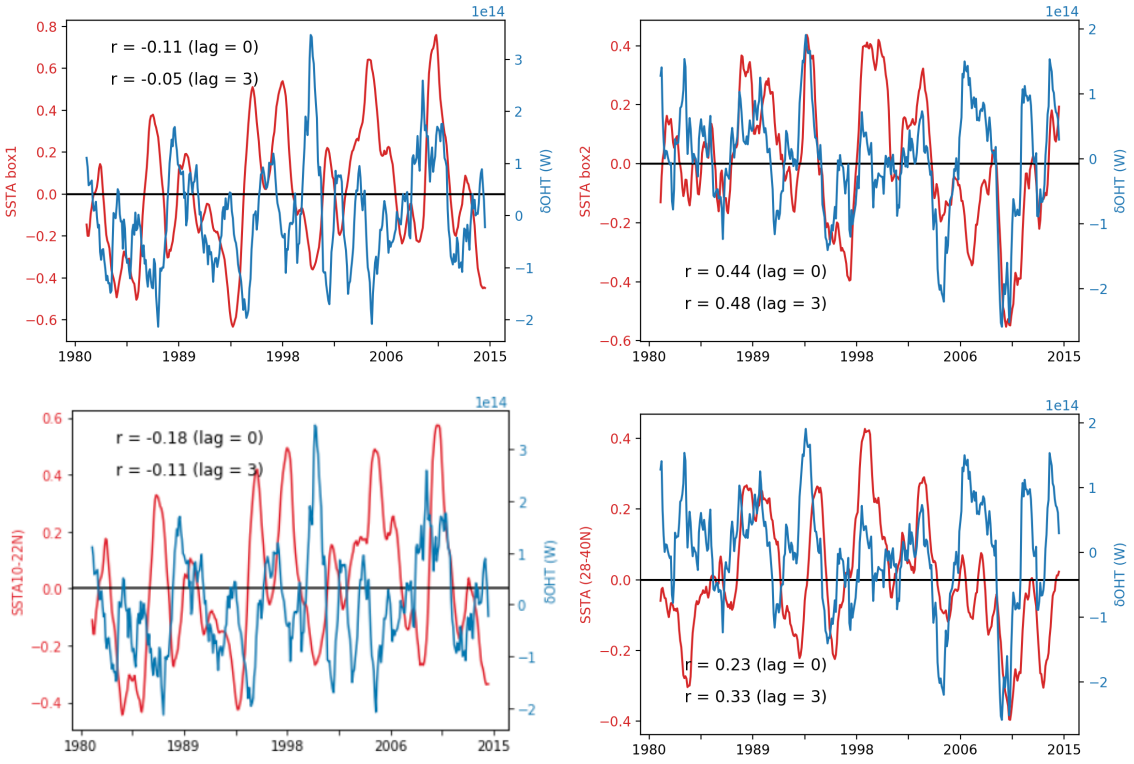

Fig R1. Time series of averaged SSTA and delta OHT. In the top row, SST is averaged over Box1 and Box2 as defined in Fig4a in the manuscript, while on the bottom row the average takes the latitudinal bands corresponding to each box, as for delta OHT.

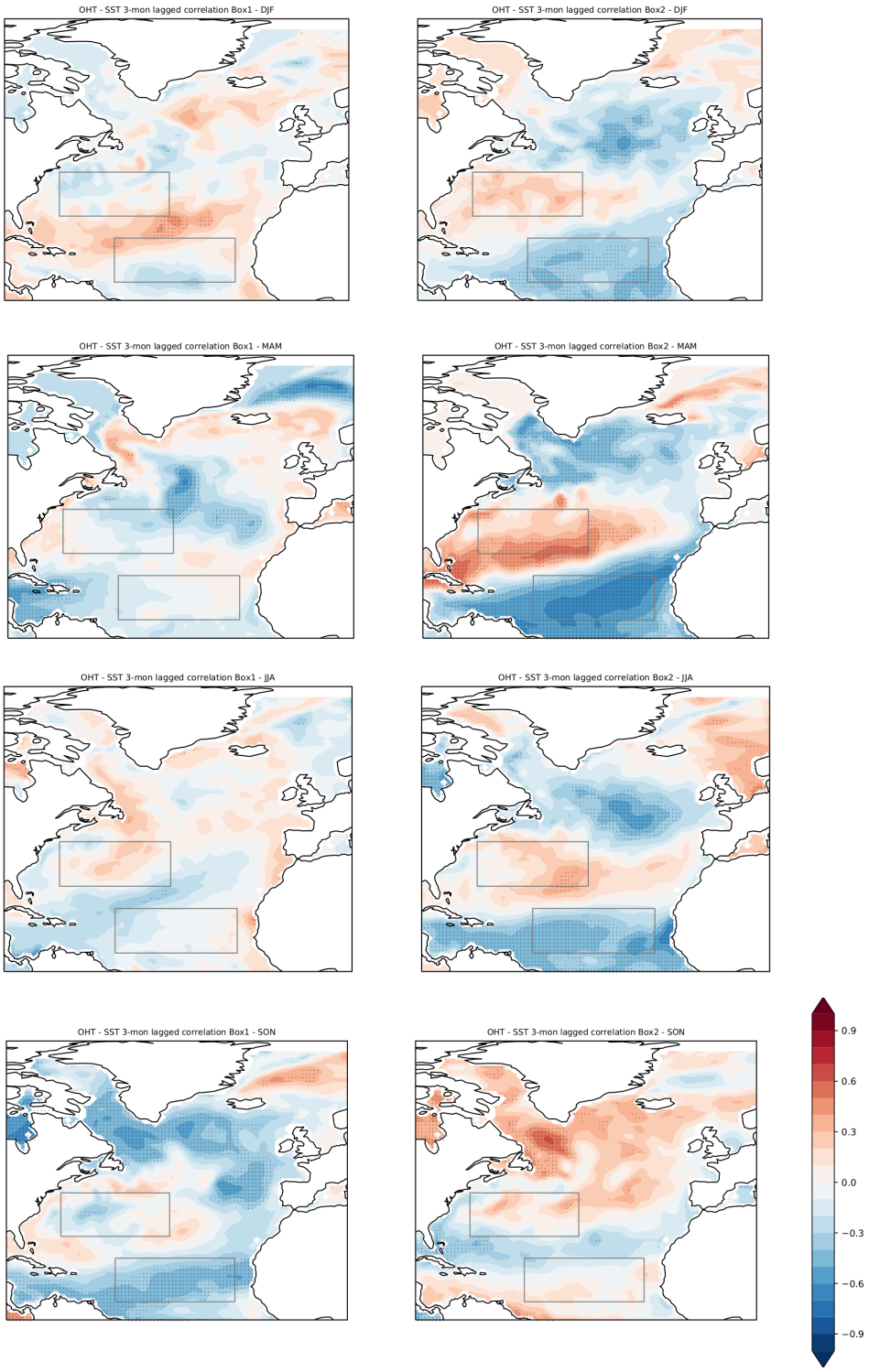

Fig R2. Correlation at 3-month lag between seasonal SST and delta OHT (with the respective season as indicated by subtitles). Delta OHT is calculated with the latitude bands respective to Box 1 (southern box) for plots on the left, and Box 2 (northern box) on the right, as in Fig.4 in the manuscript.

3. Surface heat fluxes: The authors find negative correlations between SST and surface heat fluxes, i.e., the ocean forcing the atmosphere (l.243). However, as stated in the

manuscript (l.228), it is known from observations that seasonal SST anomalies are strongly linked to atmospheric forcing. Why would the model be different, and would this in any way influence your interpretation of your results?

In pursuing this comment, we unearthed a coding error in our original scripts. Fixing that error made the resulting ASF-SST correlation plot look much more like what would be expected from observations. So the model is in general agreement with observations. Since the regions of low ASF influence continue to overlap with those where we find significant skill influence of AMOC, this change does not impact our central findings. We rewrote the entire ASF section to accommodate our "new" results, and discuss the influence of ASF more carefully in the discussion section now (e.g. lines 237-250).

Minor comments:

l.3: It is not clear what "a seesaw-like mechanism" means.

We made an effort to briefly explain the mechanism in the abstract.

l.24: You should consider explaining the SST anomalies associated with the tripole.

Done.

l.37-41: Suggest to move these sentences up to l.34 (before "[Here] We evaluate...")

Done.

l.44: delete "as"

Corrected.

l.45-46: "by sub-selecting ensemble members that meet certain physical criteria, thus filtering atmospheric noise in the ensemble" -> I think this argument needs to better explained.

This was identified as not pertinent to the story, so we deleted this bit.

l.47-48: I don't see the logic between the first and second part of this sentence ("...which is why it focuses").

See above.

l.133-135: The use of smoothing is still confusing. In their reply, the authors state that "applying the low-pass filter only for plotting time series (e.g. Fig. 1, 4) , but not for any analysis of seasonal means". This is not easy to understand from the text. Are the grey lines in Fig.1 used/necessary? You could rather show some form of error bars.

Here, the smoothing is used to highlight the seasonal cycle. This is now pointed out in the text: *'To show the spread of the annual climatology, grey lines in Fig. 1.a, c, e represent anomalies w.r.t. the mean transport of a given year calculated for the full time series (1979-2014), and smoothed with a 3-month running average to highlight the seasonal cycle.'*

l.153: "up to 7 months" – ahead?

Corrected.

l.156-157: I don't think you need a new paragraph here.

Agreed and fixed.

l.160-161: I'm still not convinced that any pronounced displacement of correlation is seen along the northern boundary of the STG. Could you help the reader somehow by e.g., adding the mean barotropic streamfunction to one of the panels?

Since this was a side note that is not particularly pertinent to the story and apparently trips readers, we rewrote: *'The magnitude of the correlation (anticorrelation) drops to maximum of 0.4 (minimum of -0.5) with increase in lag. With increasing time lag (5-7 months specifically), the subtropical lobe of positive correlation shows a displacement towards the east.'*

l.177: "Hence,..." Not sure this sentence is consistent with the previous two, but rather agrees with what is stated on l. 173.

Agreed, we rephrased.

l.186: No need for a new paragraph here

Changed accordingly.

l.203-205: The authors state that in the tropical lobe (Box 1) the correlation is significant (negative), whereas in the subtropical lobe (box 2) the correlation is weakly positive.

However, in Fig.5 it is stated the mean correlation for Box 1 is -0.33 and +0.46 for Box 2. So I'm not sure I understand the authors interpretation here. Also, as the relation between AMOC and SST vis OHT convergence/divergence is central in the mechanism of D16, I think these correlations (or lack thereof) should be discussed in more detail (l.209-210 just states "other factors", could be elaborated).

Thanks! This was an error based on mis-interpretation of earlier plots. The reviewer is completely correct. We rephrased the paragraph substantially to accommodate their concerns.

l.216: "an assessment of an attribution" – unclear

Rephrased to be clearer.

l.217 (and elsewhere): the authors refer to other drivers than AMOC as "non-oceanic". However, this is not justified and there are also other sources of oceanic variability.

Yes, agreed. We rephrased to be unambiguous now.

l.238-239: Some repetition, consider rewriting.

We considered the reviewer's suggestion, but decided to keep the repetition for the sake of making the point of atmospheric heat flux contribution to SST and the associated sign completely clear. That being said, rewriting in that paragraph unrelated to this particular reviewer comment might have alleviated some of the redundancy (lines 238-240).

l.273-275: Some repetition in the description of skill. Consider rewriting.

Rewritten to avoid redundancy.

l.288-289: How does this statement resonate with Fig.9? Except for MAM in Box 1, it seems like the confidence intervals are overlapping for "strong AMOC" and "all years".

Both figure 8 and figure 9 changed substantially in the process of revising the manuscript. Simultaneously, we rewrote the section on AMOC influence on SST predictions substantially. We hope that the reviewer will now find the text to be consistent with all figures.

l.299: "skill differences we find agree to some extent with D16's physical mechanism" – the authors should better explain what skill differences they refer to and more specifically how this relates to D16's mechanism.

We added some information on our skill difference pattern expectation in lines 285-287. The sentence that this comment references does not strictly exist anymore, but we made an effort to rephrase this section to be clearer, more intuitive, and logical.

l.308: Again, how do relate this to what is shown in Fig.9? In Box 1 there is a skill decrease for "strong AMOC" versus "all years" (although overlapping).

We realized in preparing this revision that we never formulated our expectation of skill increase/decrease connected to the D16 mechanism. In fact, a skill increase in box1 would be expected after weak AMOC due to the accumulation of heat. We made an effort to make this clearer in the current version of the manuscript. Still, the results did not support all claims made in the last version of the manuscript, so we now discuss this in a more nuanced fashion (lines 298-306).

l.310: "only during summer a significant skill difference" – Confidence intervals seem to be overlapping for JJA. Ans what about MAM in Box 1?

This was corrected in the rewrite.

l.353: Better skill for MAM in box 1?

Indeed, the reviewer is right. The skill for MAM SST does not improve in Box 1 after weak AMOC according to the D16 mechanism, and we now corrected the text as follows: *'After weak AMOC phases, we find high tropical SST hindcast skill during DJF and JJA (among which the JJA improvement in line with the D16 mechanism is significant, cf. Fig. 9), in particular over the central hurricane main development region.'* Since the skill improvement in spring does not comply with D16's mechanism (Box 2 has higher skill after weak AMOC and vice versa), we add the following lines for the sake of completeness: '*We also find enhanced prediction skill in that region during MAM, but after strong AMOC phases. Since this skill increase in spring does not fit the D16 mechanism, it is unlikely to originate from the examined mode of AMOC fluctuations, making room for different mechanisms to be explored and discussed in the future.'*

l.399-400: Again, from Fig.9 skill improvement is only seen in JJA (but not significant?).

In the process of refining the language in the manuscript to reflect the nuances of our findings, this was rewritten such that it (hopefully) reflects all caveats (lines 395-402).

l.402-403: This I assume refers to only the small patch also referred to earlier in the text. The area-averaged skill is still only 0.4 (Fig. 9).

See previous comment.

l.407-408: I know I'm repeating myself, but the skill improvement the authors refer to is marginal for DJF and not significant for JJA.

Thank you indeed for repeating these comments, as they pointed us to the urgency to discuss these results in a more nuanced fashion. That being said, we are going to return the favour and repeat ourselves when we respond to this comment: See previous comment.

Figure 4: It's very hard to see the significance lines.

We increased the colour intensity of significance lines.

Figure 9: As in Borchert et al (2018), you should also add the persistence skill here as a comparison.

We added a persistence baseline for comparison as requested.

Referee 2, 2nd review

We thank the anonymous referee for their valuable comments and suggestions. In response to both reviews, we have thoroughly revised the text to reflect more nuance on the results obtained. We also added analysis and refined some of the existing findings. We now find the manuscript to be substantially improved and thank the reviewer for helping us to get there. We highlight the major improvements in response to this reviewer's comments as follows:

- We compute and include correlation difference maps in Fig. 8, highlighting the nuance of what we are trying to achieve in this document.
- We rewrote major portions of the manuscript to make it clearer, convey subtleties more clearly, and improve the reading flow.

We took into consideration all suggestions made by the reviewer and we provide below a point-by-point response to each comment. Please note that the referee's comments are highlighted in **bold** font, while our answers are in regular font.

Major comments:
The manuscript has greatly improved since the first version. My main concern with the paper is section 3.3.1. This section contains some of the main conclusions of this paper regarding the improvements in skill, however this mainly relies on visual comparison of the maps and it is not very robust. Perhaps computing maps of correlation differences including the statistical significance, as done in seasonal and decadal forecast studies, would help the clarity and would be a more robust way of determining skill improvements, providing stronger conclusions.

We thank the reviewer for their assessment and the suggestion. We agree that including differences in skill does improve the interpretation of Fig. 8, and we added a new figure showing the difference maps and the regions where we expect to see an influence of the seesaw mechanism.

Comments by line:
Line 144: AMC-EKM -> AMOC-EKM

Changed.

Line 153: Second 'subtropical' should be 'tropical'?

Yes, changed to tropical.

Lines 189-196: Check this paragraph as I believe that several sentences reference wrong panels of figure 4.

Indeed there were wrong references to Fig. 4. We fixed them accordingly.

Lines 203-205: In figure 4e the correlation is -0.33, while in figure 4f the correlation stated is 0.46 (although the latitudinal bands are inconsistent with the figure caption). Are these values correct? If so, how come the latter is referred to as 'weak' even though it's magnitude is greater than the former?

Thanks. Indeed there was a mistake in the latitudinal bands caption, which we fixed. One correlation value was also incorrect (0.48 and not 0.46), and the reviewer is right in pointing out the inconsistency in the text. We rewrote the text as follows: '*This analysis shows significant negative correlation of AMOC with OHT convergence in the latitudes of the tropical lobe (Fig. 4e), showing that AMOC-related outflow of heat represents a relevant driver of heat convergence changes in the area. Further, AMOC is strongly positively correlated to OHT convergence in the latitudes of the subtropical lobe of the AMOC fingerprint, indicating a substantial impact of AMOC-related heat transport on oceanic heat convergence there.'*

Figure 4 caption: The latitude bands in the caption for f) are different to what is indicated in figure.

Thanks, we fixed the mistake.

Line 258: 'less' is repeated.

Fixed.

Lines 288-289: Concluding that there is an improvement in ACC skill over the subtropics in the strong AMOC cases with respect to the entire time period by visually comparing the maps does not seem very evident to me, perhaps only in JJA. As indicated in my previous review comments, the magnitude of the skill greater but the regions of positive skill are smaller. Furthermore, in Fig. 9, you have shown that JJA has greater skill in strong AMOC event, but this is not the case for DJF nor MAM.

We thank the reviewer for pointing this out. As explained above, the manuscript underwent a major rewrite, pointing out the nuance of D16's mechanism for seasonal SST predictions. In the process, we added skill difference plots (see response to major comment) and discussed the deficiencies of our mechanism (e.g. its inability to explain skill changes during

spring, autumn and winter) in more detail. We hope that the edits made to the manuscript now convey our results in a more balanced fashion and thus satisfy the reviewer.

Lines 309-311: Figure 9 is a nice addition to the paper. It would help the discussion if this figure would include if the correlations and the differences in correlations are statistically significant. It would provide more robust results.

We added this information to Figure 9 by highlighting significant skill in dots and insignificant skill using crosses. We believe this really improved the message of the figure!