Response to Referee #2

We thank the anonymous referee for their valuable comments and constructive reviews. These comments certainly help us to identify the sections in the manuscript that required improvement. We highlight the key points of this revision as follows:

- We improved the discussion of the role of D16's mechanism on SST predictive skill and included a more detailed comparison of statistical significance of skill differences, adding a new figure (Fig. 9);
- We carried out an additional analysis of ocean heat transport convergence in two regions covering tropical and subtropical North Atlantic, adding two new panels in Fig. 4;
- We show a comparison of the AMOC fingerprints between assimilation experiment and hindcasts;
- We carefully indicated in each section and caption which experiment is being addressed;
- We performed a major revision of the text in the entire manuscript to make it clearer and improve reading flow.

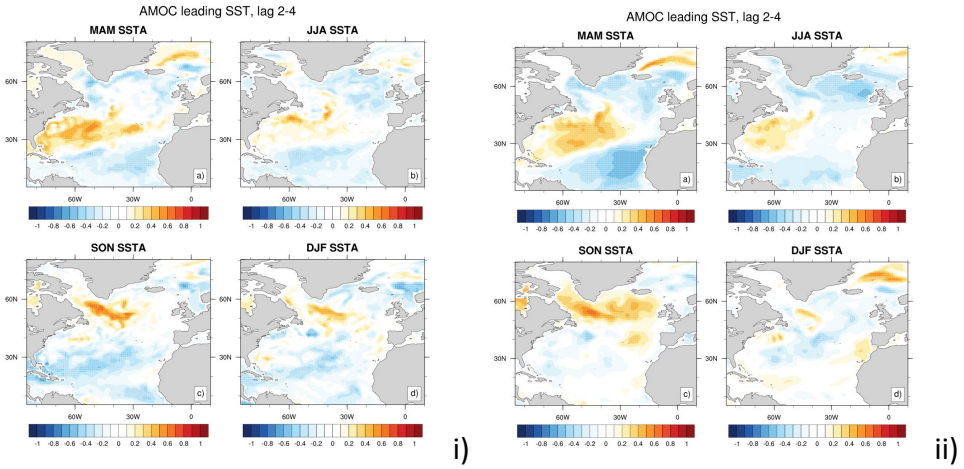
We took into consideration all suggestions made by the reviewer and a point-by-point response to each comment is reported below. Please note that the referee's comments are highlighted in **bold** font, while our answers are in regular font.

Specific Comments:

1) I believe it would help the clarity of the manuscript if in each section of the paper it would be clearly stated which experiment is being analysed, as it took me some time to figure out if it was the assimilation experiment or the seasonal predictions. This should also be indicated in the figures, for example in the caption of figure 1, 'The AMOC in MPI-ESM-MR' is confusing as it does not refer to the experiment type. It would also help to motivate in the text why the experiment being used is suitable for the analysis. Since the first part of the paper considers only the assimilation experiment and it is constraint to observations, do you expect the conclusion to hold for the seasonal forecasts? The authors could consider computing the AMOC fingerprints with the seasonal forecasts.

We thank the reviewer for raising these points. We agree that the indication of the experiment was not obvious at times, and we now carefully revised the manuscript and made sure to explicitly state which experiment was used in each session and figure caption. Regarding the reviewer's second comment, in Fig. 1 of this document we compare the AMOC fingerprints for the assimilation experiment (Fig. 5 in the manuscript) and the seasonal forecasts. Despite differences in the magnitude of correlations, we find reasonable similarity between AMOC fingerprints for the ensemble mean and assimilation, with overall

agreement in the sign of the correlations. We are therefore confident that our conclusions for sections 3.1 and 3.2 using the assimilation experiment can be extended to section 3.3, where we analyse seasonal forecasts. We add a comment on that in lines 223-225 in the revised manuscript.



i) ii)

Fig. 1. 2-4 month lagged correlations between the SSTA over the North Atlantic and the AMOC at 26◦N during 1982-2014, with the AMOC leading (a-d, as labelled) i) in the assimilation experiment (same as Fig. 5 in the manuscript) and ii) for the 2-4 month lead time ensemble mean.

2) The main concern I have with the paper is the interpretation of the seasonal SST skill results in terms of D16's mechanism for the strong and weak AMOC phases (section 3.3). I find the discussion insufficient to conclude whether the SST skill results are consistent D16's mechanism.

Thanks for this remark, we agree. To make our discussion more satisfactory, we extended this discussion in the updated manuscript and highlighted to what extent D16's mechanism can explain SST predictive skill in the model. In particular, we added a new figure (Fig. 9) of area average skill following strong and weak AMOC phases across season for two regions in the North Atlantic (as in Fig. 4.a) that illustrates how significantly different skill estimates can be expected after strong and weak AMOC phases in both regions during summer, in line with D16's mechanism. We also improved this discussion to show more clearly how D16's mechanism may be hampered by atmospheric contributions to SST variability in the other seasons.

Comments by line:

Line 16: 'tropical seasonal SST anomalies' ->'Seasonal SST anomalies in the tropics'?

We agree and changed as requested. *'Seasonal SST anomalies (SSTAs) in the tropics have been linked to the intensity and genesis of tropical cyclones and heatwaves'.*

Line 21: Could explain briefly D16's mechanism to make the paper self sustained, having only to refer to D16 for specific details.

Agreed. To make the manuscript more self-sustained, we changed as follows: *'Here, we examine the seesaw mechanism proposed by Duchez 2016 (henceforth D16), which links the transition in strength of the Atlantic Meridional Overturning Circulation (AMOC) at 26N and North Atlantic SSTs on monthly time scales. '*

Line 21: 'links transition' -> 'links the transition'?

We modified this sentence to reflect this suggestion, see also previous comment.

Lines 29-32: These sentences could be made clearer to make it easier to follow.

We rewrote the sentences as follows, to improve clarity: *'The AMOC is estimated to transfer about 1.3 PW (10^{15} W) of heat northwards at 26N (Johns et al., 2001). This heat transport, however, shows little meridional coherence at seasonal to interannual timescales (Bingham et al., 2007; Hirschi et al., 2007). Through local convergence or divergence of ocean heat transport (OHT, e.g. Cunningham et al. (2013); Borchert et al. (2018)), AMOC fluctuations could therefore influence the seasonal to interannual predictability of SST. The SST response to AMOC results in recurring large-scale patterns, generally known as AMOC fingerprints (Zhang, 2008).'*

Lines 41-46: This paragraph could also be made clearer. The authors could explain what is meant by 'incorporating known physical mechanisms into their seasonal prediction analysis'. From the latter sentences I infer it means through the initial conditions? Perhaps the physical mechanisms could be briefly mentioned.

We agree that this paragraph could have been clearer, changing the sentences as follows: *'Recent studies have found improved hindcast skill in the North Atlantic region after considering known physical mechanisms into their seasonal prediction analysis. Mechanisms were invoked in two possible ways: by identifying and explaining times of low and high skill, including precursors of high skill, as so-called* windows of opportunity *(Borchert et al., 2018, Mariotti et al. 2020); or by establishing physical mechanisms in the hindcast ensemble by sub-selecting ensemble members that meet certain physical criteria, thus filtering atmospheric noise in the ensemble (Dobrynin et al., 2018, Neddermann et al., 2018). The present study focuses on oceanic processes that are arguably less noisy than atmospheric dynamics (Gulev et al. 2013).*

Line 47: The term 'potential predictability' could be briefly explained. Does it refer to the ability of the model to predict itself?

We refer to 'SST potential predictability' as the prediction skill of SST which quantifies the fraction of long-term variability (signal) that may be distinguished from the internally generated natural variability (noise), which is unpredictable. Therefore potential predictability of SST refers to the maximum prediction skill of SST, as a function of signal-to-noise ratio, which may improve if reducing variance of systematic error and variance of noise. To address the reviewer comment, we changed the main text as follows: *'In particular, seasonal SST potential predictability, i.e. the fraction of long-term variability that may be distinguished from the internally generated natural variability, was shown to improve for better represented ocean initial states in the tropical Pacific boreal winter (Alessandri et al., 2010), and in parts of the Atlantic (Balmaseda et al., 2013).'*

Line 49: Does 'initialised coupled simulations' refer to a set of hindcast (retrospective forecasts)? If so I would use this term to be more precise. The authors could briefly explain the evaluation technique in Borchert et al. (2018) to make the paper self sustained.

Yes, 'initialised coupled simulations' refers to a set of hindcasts (retrospective forecasts). We changed the sentence to *'Analysing an ensemble of yearly initialised hindcasts with MPI-ESM-LR covering 1901-2010, Borchert et al., 2018, 2019 showed that the AMOC at 50◦ N influences the SST variability and predictability for several years, with higher skill after years of strong AMOC and vice versa. Borchert et al. (2018) perform a predictive skill analysis of SST conditioned to strong and weak OHT anomalies at 50◦N separately, showing a robust influence of the ocean on windows of opportunity for decadal subpolar North Atlantic SST predictions.'*

Line 71: This sentence should be written more precisely. 'full-field' initialisation refers to the initialisation method of the forecast system. While nudging refers to assimilation method used in the reconstruction where the initial conditions are taken from. The assimilation experiment is fully coupled right?

We agree that this description needed rephrasing and changed as follows: *'Initial conditions of the hindcasts are taken from a fully-coupled assimilation experiment with MPI-ESM-MR. In the assimilation experiment, full temperature and salinity fields in the ocean component were nudged (Dobrynin et al., 2018) towards the ORA-S4 reanalysis (Balmaseda et al., 2013). Temperature, vorticity, divergence, and surface pressure in the atmosphere component were nudged towards ERA-Interim (Dee et al., 2011), and the sea ice component was nudged to NSIDC observations (Comiso, 1995).'*

Line 98: Why is a linear trend removed?

We detrend the model data to remove, as an idealised approach, the effect of anthropogenic greenhouse gas induced global warming from the analysis and focus on the internal variability, which is in line with Duchez et. al 2016 for observations. We modified the sentence as follows: *'This data set is deseasoned by removing the 12-month climatology obtained from the monthly data, and the linear trend is removed as an idealised approach to remove the externally forced signal from the time series and focus on internal variability.'*

Line 108: 'Model verification for AMOC' -> 'Verification of the AMOC in the assimilation experiment' or 'Verification of the AMOC in the reconstruction'?

We agree and renamed the section as 'Verification of the AMOC in the assimilation experiment'.

Line 109: 'to' -> 'with', 'to' -> 'in'.

Please see below in answer to 'Lines 109-110'. As we changed the section name as suggested in 'Line 108' to highlight the experiment used, we deleted lines 109-110.

Lines 109-110: I don't think these sentences are necessary. As already mentioned I find the sentences unclear as the experiment analysed it not mentioned.

Thanks. As we changed the section name as suggested in 'Line 108' to highlight the experiment used, we deleted lines 109-110.

Table 1 is not discussed, so either comment on it or remove the table. What does 'seasonal range' refer to? Should be defined.

We added a few more comments regarding the table: *'The opposite is found for the AMOC seasonal range, which is smaller for model (2.79 Sv against 3.90 Sv, Table 1)'.* Seasonal range here refers to the range of climatological values (peak-to-peak amplitude) for each time series.

Line 112: To compute the anomalies, why is the annual mean of each year removed rather than the annual climatology?

Thanks for the remark. This sentence was misleading and needed rephrasing, since throughout this manuscript we refer to anomalies computed by removing the annual climatology (as described in lines 118-120 in section 2.2). In this sentence we were referring to the grey lines in Fig. 1.a-f, which represent the variability within each analysed year. For Fig. 1-a,c,e, each grey line shows anomalies w.r.t. the mean transport of a given year. To make this clearer we rephrased as follows: *'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.'

Lines 111-122: In this paragraph figures 1b,d and f are not referred in the text or discussed, so are they necessary? Also the figures are referred in the text with 'c.f.' when it is not necessary. The clarity of the paragraph could be improved. In lines 119-120 the 'seasonal range' is commented but no reference to table 1 is provided.

Regarding figures 1b, d and f, we added: '*There is no relevant effect of the mean state on these findings, which is why we use anomalies from now on.'* We also corrected the use of c.f. Throughout the manuscript. Lastly, we add a reference to table 1 in the sentence mentioned: '*The opposite is found for the AMOC seasonal range, which is smaller for model (2.79 Sv against 3.90 Sv, Table 1).'*

Figure 1: The labelling of the panels in the bottom right corner instead of in the top left may be slightly confusing, although a minor detail. Is AMOC-EKM (red line) necessary in the figure 1g?

We include the AMOC-EKM since we use this quantity in other sections of this manuscript. Panel labelling is kept as is as we believe the current version does not significantly impact readability of the figures.

Line 127: 'model' -> 'assimilation experiment'

We changed as suggested. *'We find that a dipole pattern represents the influence of AMOC on Atlantic SST variability in the assimilation experiment up to 7 months in the subtropical and tropical regions, similar to D16.'*

Lines 125-142: The paragraph structure and the clarity of this section should be improved.

We modified the structure of paragraphs and rewrote the section.

Lines 126-130 the authors describe the lagged correlation patterns between the AMOC and the SSTA as being the same for all lags. However, to me it seems that the pattern correlation evolves from lag 0 to month 7, as later described by the authors, which seems contradictory. I believe this should be clarified.

To make it clear that the correlation patterns change with time, we added: *' This dominant SSTA correlation pattern evolves over time.'*

Line 145: Could reference Fig. 2 in this sentence.

We modified as follows: *'We now analyse the impact of AMOC on SSTs over 36 years, to assess the consistency of the previous results (Fig.2) over a longer period.'*

Line 152: 'not shown', could show in the supplementary material.

We included one animation as supplementary material to illustrate this time dependency.

Line 155: The first sentence seems unnecessary.

Thanks for the suggestion. We removed the sentence and now refer to Fig.4: *'To further explore the sensitivity of AMOC fingerprints to the length of the observational window used, we show the AMOC-SST relationship averaged over two regions comprising the dipole lobes (Fig.4).'*

Lines 159-164: This paragraph seems a repetition of the previous paragraphs.

We rewrote the paragraph as follows, to clarify how this analysis adds to previously presented findings: *'Elaborating on findings based on spatial fields (Figs.2,3), we here show spatially aggregated SST variability (Fig.4). As before, AMOC fingerprints over the RAPID period are stronger than over the full time series. We find high anticorrelations up to 5-month lag over Box 1, ranging from -0.57 at 5-month to maximum magnitude of -0.69 at 2-month lag. In contrast, 180 when the full time series is considered, values drop to the order of -0.4. This results in a significant skill difference between the RAPID and the full period for lags 1-5 months, evaluated by non-overlapping uncertainty envelopes for the two correlation estimates (Fig.4c). Similarly, we find high correlation values above 0.6 up to 2-month lag over the RAPID period for Box 2, which drop to 0.24 at 5-month lag. The magnitude of correlations for Box 2 over the full time series reaches a maximum of 0.25. Correlation estimates for box 2 are significantly different between the two periods for lags 0-4 months (Fig.4c). Weakened AMOC fingerprints particularly during the 90s are likely responsible for the decline of the correlations computed for the full time series.'*

Lines 155-159: Related to a previous comment, a limitation with the definition of box 2 (Fig. 4) is that the positive correlation at later lags shifts towards the Eastern North Atlantic which is not included in the box. This indicates that the lag correlation patterns are not constant.

We agree, the lag correlation patterns do move towards east and thus out of Box 2. This is to us a nice illustration of how the impact of the fingerprint differs between the RAPID and

the full period. We therefore prefer to keep this definition. In comparing RAPID and 30-year period, we add the following remark to highlight the differences between the correlation patterns: '*This results in a significant skill difference between the RAPID and the full period for lags 1-5 months, evaluated by non-overlapping uncertainty envelopes for the two correlation estimates (Fig.4c).'*

Line 166: delete 'here'.

We changed the sentence as follows: *'To further investigate the variability of the SST dipole pattern, we analyse the role of SST seasonality.'*

Lines 166-175: It is not clear what period is being used, I suppose that the 30-year Period?

We added the period analysed as follows: *'Using the assimilation experiment for the period of 1982-2014, we perform correlations of the AMOC anomalies at a given month with the mean seasonal SSTA 2-4 months ahead (Fig.5).'*

Line 174: Why atmospheric drivers? Is this the only option? In fact, this seems inconsistent with the results shown in the next section (Fig. 7m).

By atmospheric drivers we mean air-sea heat fluxes, and not Ekman transport. We understand that this should be rephrased for clarity. It is true that AMOC and AMOC-EKM show similar correlation patterns in Fig. 7m, which probably means that EKM is not the cause. However, the ASF-SST correlations for the subtropical region in figure 6f are positive and therefore indicate an influence of the atmosphere in driving SST variability. Which explains why we raise atmospheric drivers as a possible reason for the different AMOC-SST pattern. We rephrased as follows: '*In contrast, we find that autumn and winter seasons lack a characteristic dipole pattern (Fig. 5.c,d), showing instead only a narrow region of negative correlations over the subtropics of the order of -0.2 (-0.1) for winter (autumn). The absence of a dipole pattern in autumn and winter may suggest the influence of atmospheric drivers that could potentially supersede the AMOC fingerprints during these seasons.'*

Line 177: 'is response' -> 'is the response' or 'responds'

We changed as requested: *'At the seasonal timescale, much of the SST variability in the North Atlantic is the response to atmospheric forcing (Deser et al., 2010)'.*

Lines 176-213: The clarity of this section could be improved and perhaps made shorter. I believe this section could also be improved by commenting on the implications of the Results.

Thanks. We took the reviewer's suggestion and improved the text for this section. In particular, we added more comments on the implications of the results: *'We compute correlations between cumulative ASF anomalies and SSTA for 2 and 4 months (where ASF leads) for each seasonal mean SSTA (Fig.6), thus highlighting regions and seasons where the atmosphere strongly contributes to SST variability. ASFs are defined as positive into the ocean, i.e. positive correlations with SST are interpreted as SST response to atmospheric heat fluxes, and vice versa. Consequently, significant positive correlations between cumulative ASF and SSTA indicate significant atmospheric contribution to SST changes that, should they overlap with AMOC fingerprints identified in Figure 5, indicate a role for the atmosphere in these regions that are potentially unpredictable. As such, this analysis forms an important step towards the assessment of seasonal SST predictions.'*

Line 218: 'The hindcasts' -> 'The hindcasts initialised in'.

We changed as requested: *'The hindcasts initialised in FEB, MAY, AUG and NOV yield 2-4 months lead time SST targets MAM, JJA, SON and DJF, respectively.'*

Line 222: I assume that AUG and MAY refer to JJA and SON in fig. 8, if so this makes the text confusing, it would help to refer to the season like in fig. 8.

We rephrased the sentence as follows: *'SON and JJA SSTs show the lowest ACCs over the subtropics, with the former showing particularly low values below 0.1 in large areas.'*

Line 229: 'The role of AMOC' -> 'The role of the AMOC'.

We changed the section title as requested.

Line 231: What does 'separately' mean here? Is there any difference with the previous Section?

Yes, there is a difference to the previous paragraphs. In lines 215-227 we describe the ACCs for each start date considering the whole time series (Fig.8a, d, g, j). In section 3.3.1 we calculate the ACCs separately only for years of strong (positive anomalies) and weak (negative anomalies) of the AMOC at 26°N one month before initialisation. We rephrased this bit to reduce confusion.

Line 233: Does this mean that the start dates are divided into two sets, strong and weak AMOC?

Yes. Please refer to the response above. We also tried to make this clear in the updated manuscript.

Line 233 'Atlantic Meridional Variability (AMV)' -> 'Atlantic Multidecadal Variability (AMV)'?

Thanks. Please note that while rewriting, we decided to remove this sentence from the manuscript to make the discussion more focused on the AMOC strength as the chosen criterion.

Line 238: 'sub-' -> 'sub-tropical'

We changed to 'subtropical': *'D16's physical mechanism suggests that via convergence (divergence) of OHT in the subtropics (tropics), a strong AMOC at 26N drives the subtropical and tropical SST variability at a maximum of 2-5 months lead time.'*

Line 242: 'we find higher hindcast skill for DJF. . .' this statement is not so clear to me. The magnitude of the skill in some regions seems higher, but the area of skill seems smaller.

We agree with the reviewer that this bit deserves a more nuanced presentation. We therefore changed the phrasing to reflect details.

Lines 237-258: These paragraphs constitute some of the main results of the paper, but they seem insufficient to me and hard to follow. The interpretation of D16's physical mechanism in terms of skill does not seem very clear and I struggle to follow the interpretation of the results. Also, based on the AMOC seasonal fingerprints I would expect that D16's mechanism could perhaps explain the SST skill in spring and maybe in summer, but not autumn or winter.

Thanks for alerting us to these issues; these paragraphs are indeed essential in conveying the key findings of the manuscript. We tried to be clearer by rewriting the paragraph as well as adding the new figure 9 and its explanation in the manuscript. In essence, the reviewer is right in expecting high skill during MAM and JJA from the variability analysis (Fig. 5). The findings, however, are actually even more nuanced than that: during SON and MAM, significant atmospheric contributions to SST variability may obstruct skillful predictions as well (Figs. 6 and 7). So the only season where skillful predictions can be expected to arise from the AMOC seesaw mechanism is JJA. This is shown in figure 9, where we illustrate the significance of the skill differences between strong and weak AMOC, and also reflected in the updated manuscript text.

Line 241: The statement: 'For strong AMOC phases, we find higher hindcast skill for DJF, JJA and MAM SSTAs over the subtropics in comparison to ACCs considering the full period', I am not certain about this statement. Comparing the skill maps it seems that the

magnitude of the skill is greater in some regions but the areas with positive skill smaller. Perhaps computing maps of the difference of correlation (including the significance) could help the interpretation. Could the SST skill maps be affected by the smaller sample size when comparing the entire period and positive and negative AMOC composites?

To better illustrate the difference in ACCs for strong and weak AMOC states, we add Fig. 9 in the manuscript, showing the difference of spatially aggregated skill estimates; please consider our response to your previous comment as well. The reviewer's concerns about the time period length are valid. In response to their suggestion, we rewrote this section refraining from quantitative statements about skill differences between strong/weak AMOC and the full period, and added a discussion of the skill differences between strong and weak AMOC (which are of similar time series length) based on the new figure 9.

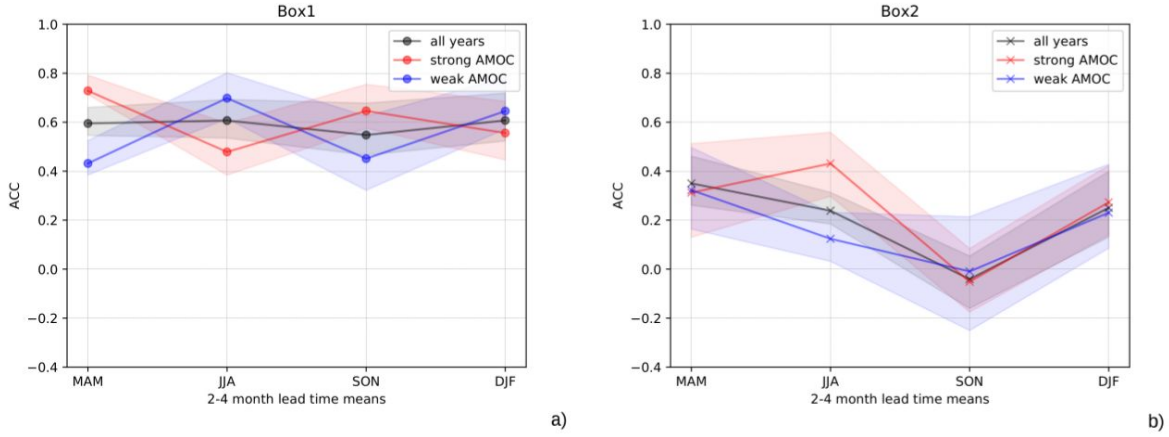


Fig. 9. SST ACCs against ERA-Interim at 2-4 months lead time averaged over the regions shown in Fig. 4a. a) Box 1 (10.5◦ - 22.5◦N, 22◦ - 55◦W), and b) Box 2 (28.5◦ - 40.5◦N, 40◦ - 70◦W. Black lines represent the ACCs considering the full time series (1982-2014), red lines for strong, and blue lines for weak AMOC phases. The shaded areas indicate the interquartile ranges.

Line 296: 'a more active ocean'? what does this refer to?

We meant 'more active' to compare oceanic and atmospheric contributions in driving SST variability. Please note that when revising the manuscript we came to delete this sentence.