Referee 1

We thank the anonymous referee for their valuable input and for suggesting ways to improve the clarity of the manuscript. 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 the following:

- We rewrote major parts of the manuscript to improve readability and to clarify several points raised by the reviewer, in particular in section 3.3
- We substituted panel a) in Figs. 3, 6 and 7

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:

1) **Methodology. The authors provide a brief overview of the used methods, in particular causality, via referencing Runge et al., 2015, Kretschmer et al.,2016, Di Capua et al., 2020b. The authors mention the "causal inference-based tool", but it is not clear which tool is meant. The authors do not mention particular settings used in the causal discovery algorithm, such as τmin, τmax, pc_alpha, alpha_level, conditional independence test etc. Without these parameters, it is very hard to reproduce particular results.**

   We specified in the manuscript that we use causal effect network based on PCMCI algorithm as our causal inference-based tool (L28-30). We also included the information on which CEN parameters have been used to perform the analysis in the Methodology section (L166-167).

2) **While, for example, in L193 the authors indicate that their analysis focuses on 3 and 4 months lag only (guessing τmin=3, τmax =4), it is not clear from the context of the paper why contemporaneous links were not included.**

   We thank the reviewer for this remark. We rewrote this sentence for clarity as *"*We test whether the spring SST index causally drives changes in summer air temperature over central Europe and under which circumstances this holds true. Therefore, our CEN analysis focuses on τmin = 3 months and τmax = 4 months, which for simplicity we refer to as 3 and 4-month lags." (L236-239)

   We'd like to emphasise that our primary goal is to identify predictive power in forecasting, particularly focusing on discerning causal predictors in the North Atlantic, such as persistent SST features. That is, our focus is to identify causal drivers for the summer EA and air temperature which can provide information before or at

the initialisation of predictions typically in May. For this reason we are interested in quantifying links at 3-4 month lags, with contemporaneous links falling outside the scope of our study. Please see the reply to the specific comment L141-142 for more details.

3) **The authors used different variables for reanalysis and model simulation to construct their causal graphs/CENs and then compared outcomes. I address this issue in more detail in "Specific comments", but I highly recommend that the authors stay consistent in their analysis, especially in drawing proper conclusions.**

   We thank the reviewer for raising this important point. As explained in the answers to the specific comments in lines 222, 223 and Figs. 6 and 7 (see below), we made sure that the computed causal graphs consisted of the same variables and hyperparameters, when comparing model and reanalysis.

4) **Naming of variables. The authors use SST to indicate the North Atlantic extratropical surface temperature (P1L2) or SST index (P1L4) for the meridional SST gradient over the North Atlantic region. I highly recommend that the authors use the SST term to indicate SST and include a geographical indicator to their variable, e.g. NA-SST, when they refer to the North Atlantic extratropical surface temperature.**

Thanks for this, indeed it makes things clearer, to include the SST region. We went through the text and edited accordingly, making sure that the reader better understands when we are referring to SST index or NA-SST in general.

5) **The term "causal associations", apart from the title, is used twice in the manuscript. I suggest the terms "relationships/links/connections".**

   We thank the reviewer for the suggestion, we changed the term to "relationships".

Minor comments:

Abstract.

P1L6 "We only find this link to be causal, however, during the period 1958 - 2008." → the authors did not mention the analyzed period for ERA-20C reanalysis therefore "however" in L6 sounds odd. Moreover, later in the manuscript the authors provide the reasoning and explanations why there are no detected causal links for specific periods. Therefore, I suggest to keep the sentence simple "We find the causal link during 1958-2008".

Thanks for the suggestion. We agree that the “however” sounds odd, so we rewrote the sentences to emphasise which period is investigated, and in which part of it we find the link to be causal. “We apply Causal Effect Networks to evaluate the influence of spring North Atlantic extratropical surface temperatures (SST) on the summer East Atlantic Pattern (EA) seasonal predictability during the period of 1908-2008.” (...) “Notably, this causal link is only evident during the period from 1958 to 2008, and is not observed throughout the entirety of the investigated period.”

P1L9 “we find that spring SST” → do the authors mean spring North Atlantic SST? “NA-SST” would be helpful to follow the text. In P1L4 the authors introduced the SST index, so do the authors talk about SST index here as well?

To make it clearer, we edited L4 as “We find in the ERA-20C reanalysis that a meridional NA-SST gradient in spring (SST index) causally influences the summer EA (...)” and L9 as “In addition to the summer EA, we find that the spring SST index has an estimated causal effect of about (...)”. Besides, we edited L1 as “We apply Causal Effect Networks to evaluate the influence of spring North Atlantic extratropical surface temperatures (NA-SST)”.

P1L14 What is the meaning of “moderately” here?

Our intention was to express that the range of causal link strength values (beta-values) reproduced by resampling the 30-member ensemble encompassed the observed causal link value (Fig. 7c). We removed the word “moderately” from the sentence to not cause misunderstanding.

P1L13 “We find that while MPI-ESM-MR…”→ both pre-industrial and historical simulations?

Yes, so we rewrote the sentence to make it clearer as “We find that while both the pre-industrial and historical simulations using MPI-ESM-MR are mostly unable to reproduce the causal link between spring SST and the summer EA among the different datasets, the 30-member initialised ensemble can reproduce a causal link between spring SST and summer 2-metre air temperatures over a region west of the British Isles.” (L13-16)

Introduction.
P2L26 What is this causal inference-based tool?

We rewrote the sentence to include which tool we used: “Here, we apply the Peter and Clark momentary conditional independence (PCMCI) causal discovery algorithm to evaluate the

influence of North Atlantic extratropical surface temperatures (SST) on the predictability of EA at seasonal timescales." (L28-30)

P2L40 SST is already introduced in P2L27. I refer the authors to my comment to improve the notation of the North Atlantic extratropical surface temperate instead of simply using SST.

We took care of including the reviewer's suggestion in L27 and introduced North Atlantic sea surface temperatures as NA-SST. So we rewrote this sentence as "While there is no consensus on the physical processes driving the EA, spring NA-SST (...)".

P3L55 Runge et al., 2015 does not use term "Causal Effect Networks". It was Kretchmer et al., who was one of the first authors to use this term.

Indeed, thanks for bringing this up. We removed Runge et al. 2015 accordingly.

P3L58 regarding overcoming spurious correlations: see also Runge et al., 2014; Runge, Bathiany, et al., 2019.

Thanks for the suggestion, we included both papers in the citation.

P3L59 please add more examples for the application of causal discovery for other teleconnections. For example, Atlantic-Pacific teleconnections: Karmouche et al., 2023; Arctic-midlatitude teleconnections: Kretschmer et al., 2020, Siew et al., 2020; Galytska et al., 2023, marine cold-air outbreaks: Polkova et al., 2021, Walker circulation: Runge, Bathiany, et al., 2019 and others.

Thanks for the recommendation. We rewrote the sentence to accommodate some of the studies as "CEN overcomes spurious correlations due to autocorrelation, indirect effects, or common drivers (Runge et al., 2014, 2019). It has been successfully used to complement hypothesis testing for other tropical and mid-latitude teleconnections in the Atlantic-Pacific region (e.g. Karmouche et al. (2023) and Indian Ocean (e.g. Di Capua et al. (2020a)), as well as in the Arctic region (e.g. Siew et al. (2020); Kretschmer et al. (2020))." (L60-64)

Methodology.
P3L73 Here SST stands just for Sea Surface Temperature. That is confusing with the North Atlantic SST introduced before.

Indeed, so we rewrote the sentence as "The physical variables analysed are NA-SST, SLP and air temperature at 2 metre height (T2m)." (L78-79)

L106 EOF is already introduced. The calculation of EA index is already mentioned in the Introduction.

We rewrote the sentence as "As a first step, we define a reference EA index as the second principal component (PC) of the EOF of JA anomalies of SLP over the Euro-Atlantic sector (...)". (L111-113)

L114. The reference to Fig. 1b comes before Fig. 1a. What is the meaning of the colors in Fig. 1b? Please fix this. Generally, I suggest that the authors insert the Figure and/or Table, which would summarize/show the indices used in their studies and explain them in Sect. 2.2. In that case panels cf from Fig. 1 would solely address the discussions from Sect. 3.1

These were good suggestions, thanks. Below we answer to each specific point separately:

- The first reference to Fig.1a now appears in L116, whereas the first reference to Fig. 1b appears in L119. We thus keep the panels in Fig.1 in the same order.
- We chose colours to illustrate the boxes in Fig.1b to facilitate the visualisation for the reader. In the text (L114-L118) we refer to these colours when explaining which regions were taken into account to calculate the indices. To make it clearer, we now include labels in the boxes;
- As suggested, we provided a table (Table 1) summarising the indices used in the paper.

L117 The authors explain how they calculated an SLP index, however the introductions lacks the explanation of the impact and significance of the pressure over this region on East Atlantic Pattern and North Atlantic SST.

We thank the reviewer for the suggestion, we rewrote the sentence as "To comprehensively investigate the influence of spring NA-SST on summer SLP variability, we incorporate the SLP index introduced by Ossó et al. (2018) alongside the EA index. This approach aims to address the broader significance of pressure dynamics in the region, particularly in relation to the physical mechanism proposed by Ossó et al. (2020)." (L121-123)

L121 "wherever useful"→ where exactly?

We rewrote the sentence as "We use a two-tailed Student's t-test to calculate the statistical significance of point-wise correlations maps." (L131-132)

L127 Cite Spirtes et al., 2000 for PC part of the algorithm.

Done.

L134 “circles” → nodes?

Done.

L137 The authors should also explain the meaning of the color of the nodes.

We rewrote the sentence as “We visualise the output of PCMCI in a CEN, i.e. a causal graph where nodes represent the investigated variables, arrows indicate the direction of the causal links, and colours denote the strength of these links.” (L155-156)

L141-142 I assume the authors used Tigramite for their research. I find it important that the authors stay transparent on the software that they used as well as the settings that were applied, τmax, τmin, pc_alpha, alpha_level. I am wondering if the authors already tested PCMCI+ algorithm for their study, which is able to capture causal contemporaneous connections. Why authors did not include contemporaneous connections?

We thank the reviewer for this important question. Our primary aim in this study was to identify precursor signals in spring, specifically before the initialisation of prediction systems typically in May, that could forecast the summer EA. Thus, we concentrated on uncovering interseasonal causal links, i.e. those between spring and summer. This approach led us to exclude contemporaneous links from our analysis, as our focus was on understanding how extratropical North Atlantic SSTs in spring influence the EA teleconnection over a span of three to four months. For this reason, performing the analysis with PCMCI+, despite potential interest from a methodological point of view, would not add any physical insight to the analysed mechanism.

We included more information on the used settings in Sect. 2.3. We added “CEN analysis is a causal discovery tool which implements the so-called Peter and Clark momentary conditional independence algorithm (PCMCI, Runge et al. 2019, Spirtes et al. 2000). We specifically use the PCMCI version 4.2 from the Python package Tigramite (https://github.com/jakobrunge/tigramite).” (L136-138)

At the end of Sect. 2.3, we added “Lastly, the PCMCI parameters are chosen as follows: pc alpha = 0.2, Alpha level to print results = 0.1, τmin = 3 months and 155 τmax = 4 months, independence test = parcorr, significance=’analytic’, masking type ‘y’.” (L165-167)

Results

P6L146-147 If the authors follow my suggestion and introduce a separate Figure for Sect. 2.2 and summarize the used indices, I suggest to move this sentence to Sect. 2.2.

We included a table (Table 1) containing an overview of the indices used in the analysis, including abbreviations, long names and the geographical regions used for the calculations.

P6L147-149 Fig. 1 c and d do not show "below average temperature/precipitation", they show correlation between EA-index and temperature/precipitation, which is associated with below average temperatures/precipitation over this region.

We rewrote the sentence as "A typical surface climate imprint of the summer EA in positive phase correlates with below-average temperatures in continental Europe (Fig.1c) and below-average precipitation in the British Isles and northwestern Europe (Fig.1d)." (L192-194)

P6L149-150 This sentence should be moved to Sect. 2.2. The simple usage of term "SST index" is very confusing.

We thank the reviewer for bringing this up. Indeed, we already have a similar sentence in Sect. 2.2 (L112). Our aim here was to recap the idea, but we agree it can be confusing. Therefore, we decided to exclude the sentence.

P6L153 "correlation reaches significant values" → what is the definition of significance here? Did the authors use significance test here?

We thank the reviewer for this question. Yes, we calculated the significance of the correlation using a two-tailed Student's t-test. We included the p-values in the manuscript. In addition to that, upon rechecking results we came across an error. The correct correlation coefficient for the period between 1958 and 2008 is 0.43, and not 0.51. The sentence was edited as "A Pearson correlation analysis reveals a time-dependent relationship between the SST index in spring and the EA in summer (Fig. 1e). Over a span of 101 years (1908-2008), this relationship appears weak ($r = 0.22$, $p < 0.05$). However, examining the most recent 51 years (1958-2008) shows a doubling of correlation values ($r = 0.43$, $p < 0.05$). Furthermore, focusing on the latest 30 years (the period analysed in Ossó et al. 2018) results in correlation values increasing even further to 0.60 ($p < 0.05$) ." (L195-199)

Figure 2. Caption. I suggest: "Distinct spatial characteristics of the spring SST influence on the summer circulation over the 20th century (for ERA-20C) for early (1908-1957, left), late (1958-2008, middle) and full periods (1908-2008, right column). …" …"Box in Fig.2i illustrates…". To label y-axes the authors use "Influence of AM SST index", which implies the direction of the impact, however by this point it was not yet discussed. I suggest terms "relationship/connection" instead. Another minor suggestion: if the authors draw coastlines in gray, then SLP contours in panels d-f could be plotted in black and become more visible. Further, since Fig. 5 shows only correlations (models vs reanalysis), I suggest

that the authors similarly restructure the panels in Fig. 2, e.g. first two rows showing correlations, the third raw showing regression.

We thank the reviewer for the helpful suggestion, we rewrote the label and edited the figure panel as the reviewer suggested. Please see revised Fig. 5 in the main text.

L166 Is it about panels a and b?

We rewrote the sentence as "Surrounding this high correlation region, the sign of correlations is opposite between early (Fig2a) and late (Fig2b) periods." (L210-211)

L166-167 Not shown in this paper? Or which Figure/panels?

Not shown, so we rewrote the sentence as "We find similar results using March-April (MA) NA-SST means, only in weaker strength (not shown)."

L171 But not in panel d?

Exactly, not in "panel d" of the previous version (now panel g), i.e. not in the early period. To increase clarity we rewrote the sentence as "In the late period (Fig.2h), these anticyclonic conditions coincide specifically with the position of the EA centre of action, whereas this association is absent in the early period. (L216-217)

L173-174 In which geographical region?
We rewrote the sentence as "We find significant correlations between the AM SST index and JA T2m, showing a similar pattern of significant positive correlations west of the British Isles, as in Fig.1c corresponding to JA EA - T2m." (L218-220)

P9L191 "correlation" → anticorrelation

Modified.

L193 I assume $\tau_{min} = 3$, $\tau_{max} = 4$.

We rewrote the sentence as "Therefore, our CEN analysis focuses on $\tau_{min} = 3$ months and $\tau_{max} = 4$ months, which for simplicity we refer to as 3 and 4-month lags."

L200-201. In regard to the full periods, the authors already stated that the spring SST -summer EA relationship is nonstationary for the full period, which would lead to the non-realistic causal graph/CEN (also see Runge 2018).

Indeed, using the full period with Tigramite could lead to a non-realistic causal graph due to nonstationarity. To emphasise this point, we revised the sentence to: "While PCMCI cannot handle the nonstationarity identified in the full period (e.g. Runge et al. 2018), we find no significant causal links in the early period." This adjustment highlights our awareness of the nonstationarity issue and underscores the need for creating sample subsets (the early and late periods) to overcome the problem.

L208 "actor" → variable. The authors did not define the term "actor", but used term "variable" throughout the manuscript.

For consistency, we rewrote the sentence with "variable", as suggested.

L208 excluded → "not included".

Modified as suggested.

Figure 3. Given that panel a consists of two subfigures that aim to show the same result, I strongly recommend that the authors combine them into one figure. For example, in plot_graph function, the authors can use node_pos.

Upon reconsideration, we decided to modify the panel and present the causal graph without the illustration. Considering the reviewer`s point in L222, we included instead a causal graph containing the same set of variables analysed in Fig.6a,c.

Figure 5. Since panels g-h are the same as Fig. 2a-c, the authors can either remove these panels completely, or show the differences instead.

Following the reviewer's suggestion, we removed panels g-h from Fig.5, since these are the same as in Fig. 2a-c. We modified the caption to accommodate this change as "Spatial characteristics of the SST-SLP relationship over the 20th century in MPI-ESM-MR. Correlation maps show point-wise correlation coefficients for the April-May SST index and July-August SLP means considering early (1908-1957; a,d,g), late (1958-2008; b,e,h) and full periods (1908-2008; c,f,i), respectively. Top row shows results for the MPI-ESM-MR historical simulation and bottom row for MPI-ESM-MR 30-member ensemble. The reader may refer to Figs.2a-c for a comparison with ERA-20C."

P9 L222 I do not recommend to comparing causal graphs/CEN between reanalysis and model simulations while using different sets of variables. The assumption of causal sufficiency says that Measured variables include all of the common causes. For the reanalysis, the authors motivated the usage of 2m Temperature. It is important to explain, why authors did not include this variable for model simulations. For example, the

correlation of AM SST index and 2m Temperature is not that strong in the model. But this might lead to the other discussion, e.g, low correlation could be a consequence of different state(s) of AM SST index in the model vs reanalysis. Currently the results from Fig3a and Fig.6a,c are not directly comparable.

We thank the reviewer for raising this very important point. To overcome this issue, we included in Fig.3b a causal graph for ERA-20C using the same set of variables analysed for MPI-ESM-MR (i.e. not including 2m temperature). By doing so, the reader is able to directly compare the causal graphs between model and observations.

P9 L222-223 Here the authors indicate that for the model simulations the time lag was 3 and 4 months, however Fig. 6a also shows two causal lagged links with 2 months delay. Please, clarify this.

We kindly thank the reviewer for pointing this out. We unfortunately had used the wrong figure by mistake, since over the course of our analysis we also tested the effect for tau_min = 2. The correct causal graph containing the same set of variables but focusing on tau_min = 3 and tau_max = 4 is now in Fig. 6a. Below you can find a comparison between the causal graphs. The principal link (SST -> EA, lag=3) has a marginal difference between graphs, with beta = 0.04, as opposed to beta = 0.03 in the first version of the manuscript. Besides that, other qualitative differences are:

- As expected, the autocorrelation of SST_index decreases when tau_min=3, indicating that SST persistence is higher when tau_min=2
- A weak (beta=-0.02) negative link between SST_index -> SLP_index (tau=3) appears
- The EA -> SST_index link is found for lags 4 and 3.

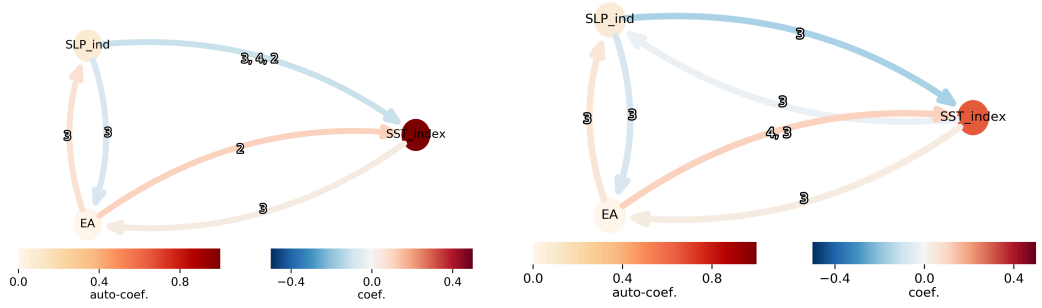

Fig. 1 *left*: Fig.6a in the first version of the manuscript, showing tau_min=2; *right*: Fig.6a in the revised manuscript, for tau_min = 3.

To convey these changes we rewrote the text (please see below in the reply to the specific comment concerning L227-230).

P10 L224 "suggesting an atmospheric forcing into ocean" → please, also specify exact links. This will improve readability and comparison with the results from Fig. 6. Same for L225.

To improve readability, we rewrote the sentence as "While no causal links are found in the historical simulations, we find opposite causal links than those in ERA-20C for the pi-Control simulation, suggesting an atmospheric forcing from EA into the extratropical North Atlantic (e.g. $\beta_{EA \rightarrow SST} \approx 0.22$), but no detected causal influence from the ocean on the atmosphere (Fig.6c)."

P10 L227-230 I would like to highlight that differences between reanalysis and model simulations were expected, since the authors already showed that the timeseries (Fig.4) as well as spatial patterns differ quite a lot between different data sources. I would appreciate if the authors addressed this issue here (as well as in the Discussion).

We thank the reviewer for this comment. As rightly pointed out, we anticipated such disparities given the distinct characteristics of the datasets involved. Our intention while showing these results is twofold. First, causal discovery adds the information of whether other causal paths amongst SST, EA and SLP indices appear in the different model sets for time lags of 3 to 4 months. Second, this analysis provides a stepwise analysis of MPI-ESM-MR`s capability and limitations in capturing the observed spring SST - summer EA link. As illustrated in Figure 4, the ensemble members of MR-30 exhibit high variability, albeit with correlations mostly failing to reach significance. In light of this variability, our approach aims to first utilise the complete set of ensemble members to construct a comprehensive understanding of the MR-30`s potential to reproduce the SST - EA link. In other words, this first approach aims at testing the use of the ensemble mean. The weakly positive SST - EA link at the 4-month lag found in MR-30 (beta = 0.04) differs from the link found in ERA-20C (beta = 0.22), highlighting that relying solely on the ensemble mean might not fully capture the capabilities of MR-30. We modified the manuscript in order to make our intention clearer and to stress that the weaker strength of the causal link found in MR-30 could be anticipated by our previous analysis. We report these changes in lines L268-285:

"The observed disparities between the model and observations, as highlighted in the spatial correlations and time series analyses depicted in Figs.4-5, prompt further investigation into the causal relationships within MR-30. To address this, we proceed to assess whether the model reproduces any of the observed causal links or presents alternative causal pathways. We construct three different CEN sets to evaluate, respectively, pi-Control, historical and initialised simulations with MR-30. The variables analysed in the CEN sets are SST, EA and SLP indices and the time lag of interest is spring - summer (3 and 4 months lag). While no causal links are found in the historical simulations, we find opposite causal links than those in ERA-20C for the pi-Control simulation, suggesting an atmospheric forcing from EA into the

extratropical North Atlantic (e.g. $\beta_{EA \rightarrow SST} \approx 0.22$), but no detected causal influence from the ocean on the atmosphere (Fig.6c).

Moving on to the initialised simulations, we leverage the entire 30-member ensemble of MR-30 to construct a comprehensive CEN spanning the full period (1908-2008), resulting in each constructed time series comprising 3030 years. We find that MR-30 is able to reproduce a weakly positive SST index - EA link (i.e. $\beta_{SST \rightarrow EA|SLP} = 0.04$) at 3-month lag (Fig.6a), but not at 4-month lag as detected in ERA-20C during the late period, and in much weaker strength (i.e. $\beta_{SST \rightarrow EA|SLP} = 0.22$, ERA20C). Moreover, we find a weak negative causal link from SST index to SLP index in the model (i.e. $\beta_{SST \rightarrow SLP|EA} = -0.02$), as opposed to observations (i.e. $\beta_{SST \rightarrow SLP|EA} = 0.21$, Fig.3b). This finding aligns with Fig.5d-f, which shows that the area of positive correlations in MR-30 is displaced southwestwards with respect to ERA-20C. No causal links from SST index to EA or SLP indices are found when analysing only the late period (1958-2008). Next, we therefore investigate the causal link sensitivity to the sample size and focus on 45-year long timeseries covering the late period, allowing a direct comparison with the sensitivity analysis performed in ERA-20C (Fig.6b-d)."

L232 There is no Fig. 6e

Corrected.

Figure 6 a and c. Node "EAP"→ EA.

Modified as suggested.

Figure 7a. It is not clear why the authors did not keep SLP-ind in the causal graph? Is there a reason why the colormap for the β -coef. is changed?

We thank the reviewer for pointing this out. Upon reevaluation we decided to show in Fig.7a a causal graph containing SST index, SLP index and T2m_CE. The reason for that is that in Fig.7b we show a causal map for 3-month lag, and we want to provide a causal graph that is immediately comparable with it. This also answers your second question, why we changed the colormap for the causal graph in this panel: to make the comparison with Fig.7b easier.

Discussion
P15LL288 For which time period?

We rewrote the sentence as "Using ERA-20C, our CEN analysis confirms that the spring SST index proposed in Ossó et al. (2018) causally influences the variability of summer SLP in the Euro-Atlantic region with a 3-4 months delay during the late period (1958-2008)." (L343-344)

Conclusions

P16 L323 "we find that"… but in L303 the authors state that the nonstationarity has been previously reported by other studies.

An important finding of the present study is to find that the link between extratropical NA-SST (SST index) and summer EA is nonstationary. Rieke et al. 2021 finds that the link between tropical North Atlantic SSTs and summer EA (previously reported by Wullf et al. 2017) is nonstationary. To our knowledge, no other study has looked into the stationarity of extratropical NA-SST in particular.

L326 Please, specify if this conclusion is for the reanalysis. Same for L330.

We rewrote the sentences as "We find that this relationship in ERA-20C is only causal over the late period." and "In addition to summer EA, we find in ERA-20C that the spring SST index causally influences summer T2m (...)" (L383-385)

L328 "that an external physical mechanism not included in our analysis" → the usage of PCMCI implies that the user makes an assumption that all variables are included to represent a specific mechanism. If user does assume that all variables are included in the causal graph/CEN, then LPCMCI should be rather used.

Our assumption is that all variables relevant to the causality of the analysed physical mechanisms are indeed included in the analysis. Therefore, we chose to use PCMCI instead of LPCMCI. However, we do feel the need to point out, for transparency, that there could be other relevant variables that we cannot identify in the related literature, that may affect the results. This is true in virtually the entire body of studies that has applied PCMCI since 2014. Here, we want to stress that we analyse the causal relationship among the selected time series. Finally, although LPCMCI may results in the explicit information that (one) or more actors are missing, it would however no help us in identifying those variables in the physical set of fields.

L333. Please, also summarize why the model does not reproduce the links from the reanalysis.

We acknowledge the reviewer's observation; however, our analysis did not delve into the specific reasons for the model's inability to capture the causal links. A subsequent study would be necessary to address this comprehensively. Various factors could contribute to the low performance of MPI-ESM, including biases in the positioning of the jet stream (Pithan et al., 2016; Beverley et al., 2019), inaccurate representation of coupled ocean-atmosphere feedbacks (e.g. Ossó et al. 2020), and ensemble overdispersion stemming from initialisation

issues (Ho et al., 2013). To highlight the need for future work investigating the possible causes, we added in L397-399: “Exploring the causes behind the model’s deficiencies in this aspect—such as whether they stem from inadequacies in representing crucial coupled ocean-atmosphere feedbacks or other processes—will be a significant focus for future research. ”

L334 “weakly”→ weak

Modified as suggested.

L341 “limited performance in reproducing”→ I suggest to highlight the reasons while giving such a statement.

Please see the reply regarding L333 above.

References

Beverley, J. et al (2019). “The northern hemisphere circumglobal teleconnection in a seasonal forecast model and its relationship to European summer forecast skill.” In: Climate dynamics 52.5-6, pp. 3759–3771.

Ho, C. et al (2013). “Examining reliability of seasonal to decadal sea surface temperature forecasts: The role of ensemble dispersion.” In: Geophysical Research Letters 40.21, pp. 5770–5775

Pithan, F. et al (2016). “Climate model biases in jet streams, blocking and storm tracks resulting from missing orographic drag.” In: Geophysical Research Letters 43.13, pp. 7231–7240

Ossó, A., Sutton, R., Shaffrey, L. and Dong, B., 2020. Development, amplification, and decay of Atlantic/European summer weather patterns linked to spring North Atlantic sea surface temperatures. Journal of Climate, 33(14), pp.5939-5951.