

We thank both reviewers for their comments. Here we reply to and address all of your comments (responses are in blue, Times New Roman font). The corresponding places where changes are made in the revised manuscript (tracked changes version) are tagged accordingly. Following the guideline, the tracked changes version of the revised manuscript is appended. A 'clean' version of the revised manuscript is also uploaded.

## Reviewer 1.

### General comments:

1. Given that the authors argue that splitting ND from JF is important, I think it is important to show each of these 4 months individually in a version of figure 2 with four columns and three rows, before then showing the ND and JF composites. This additional figure would help make the point to the reader that ND are more alike, and JF are more alike, before the authors then go on to combine ND together and JF together.

We have now added further explanation in the second paragraph of Sect. 3.1 as well as a new Fig. A1. Details of the teleconnection in the four months individually have also been presented in detail by previous studies (e.g. in the cited Moron and Gouirand 2003, among others). **Tag: R1.1, R2.2**

2. There is one recent paper the authors seem to have missed that is very relevant and complementary to the present study. The size of a composite necessary for the stratospheric route in late winter to become robust was discussed in detail in Weinberger et al 2019. They concluded that the EN-weak vortex route is robust with around 15 events, though the impact of LN on the vortex was only apparent with more than 25 events (see their figure 5) as compared to climatology. They don't show the difference between EN and LN, but clearly the difference will be significant with less.

The polar cap SLP response was also found to be significant after around 15 events are taken (their figure 7) both for EN and LN. This is consistent with the results shown in the present paper.

Weinberger et al also found robust impacts on subpolar Eurasian temperatures (see their figure 7) that are robust for EN and LN individually with composite sizes similar to that used in the present paper. Weinberger et al didn't consider early winter and they analyzed historical data for a much shorter period than analyzed here, and so there are clearly new results here. But as best as I can tell the results are consistent when the late winter and the stratospheric route is considered.

Thank you for pointing out this paper and summarising the relevant parts for us. We find that this is an interesting and important aspect. Therefore, although you have not requested it explicitly, we have performed additional calculations to investigate the effect of sample sizes in our cases. The result is described in a new paragraph in Sect 3.1 and a new appendix figure. Weinberger et al.'s work is also cited now. We hope you will find the outcome interesting and agree with our revision. **Tag: R1.2**

### Minor changes

**Line 15** : I suggest adding a more general first paragraph about ENSO teleconnections. The paper jumps right in to tackling the uncertainty, and while most of the intended audience probably already knows about ENSO teleconnections, a more casual reader may not.

We have now added a general background paragraph at the beginning of the Introduction. **Tag: R1.3**

**Figure 2 and similar:** Please add to the caption and to the text the region over which the spatial correlation is computed for the Taylor diagram. Is it identical to the region shown in the bottom rows? Is area-weighting applied?

This information is now added in Sect. 2.4 and the figure captions also refer to this section. Yes, both the spatial correlation and the spatial amplitude are area-weighted. This is also mentioned and clarified in Sect 2.4. **Tag: R1.4**

**Line 195 and table 1:** please clarify how you classify years as either CP-EN, EP-EN, EP-LN, or CP-LN. It would also be appreciated if a table was added that lists which years are included in each of the four composites (or alternately mark in bold or some other designation either EP or CP events on table 1).

A new Table 2 listing the years is added and also referred to in Sect. 3.2. **Tag: R1.5**

The years/events selected follow the same ones in the cited Zhang et al. (2018). We realise that there are variations in defining CP and EP ENSO and didn't want to get into investigating the subtlety deeply. We only wanted to check briefly what the Nov-Dec composites look like under these events selected by a previous study.

**Line 306 the publication year for the Domeisen et al review paper should be changed to 2019**

Corrected. Please see References. **R1.6**

## Reviewer 2.

### General comments.

1. The authors seem to have missed a few recent and relevant studies examining the asymmetry and non-linearity in the ENSO teleconnection to the North Atlantic. It would be nice that the papers cited below are included in the new version of the manuscript. At least when discussing that these model results disagree in some points when trying to identify nonlinearities in the ENSO-North Atlantic teleconnection (e.g. lines 194-201)

Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Ineson, S., Lim, J., & Fereday, D. (2019). The Impact of Strong El Niño and La Niña Events on the North Atlantic. *Geophysical Research Letters*, 46(5), 2874–2883. <https://doi.org/10.1029/2018GL081776>

Jiménez-Esteve, B., & Domeisen, D. I. V. (2020). Nonlinearity in the tropospheric pathway of ENSO to the North Atlantic. *Weather and Climate Dynamics*, 1(1), 225–245. <https://doi.org/10.5194/wcd-1-225-2020>

Trascasa-Castro, P., Maycock, A. C., Scott Yiu, Y. Y., & Fletcher, J. K. (2019). On the Linearity of the Stratospheric and Euro-Atlantic Sector Response to ENSO. *Journal of Climate*, 32(19), 6607–6626. <https://doi.org/10.1175/JCLI-D-18-0746.1>

Weinberger, I., Garfinkel, C. I., White, I. P., & Oman, L. D. (2019). The salience of nonlinearities in the boreal winter response to ENSO: Arctic stratosphere and Europe. *Climate Dynamics*, 53(7–8), 4591–4610. <https://doi.org/10.1007/s00382-019-04805-1>

Thank you for highlighting the relevance of these papers. The last three papers are cited now in Sect. 3.2. **Tag: R2.1**

Please also see replies to your related comments R2.25, 2.26 below.

Hardiman et al. is cited in the newly added first paragraph in the Introduction (also see reply to R1.3). Weinberger et al. is also cited in Sect 3.1 (also reply to R1.2).

**2.** I agree with reviewer 1, that it would be good to show the individual monthly mean evolution of the teleconnection from Nov to March.

(Same reply to Reviewer 1) We have now added further explanation in the second paragraph of Sect. 3.1 as well as a new Fig. A1. Details of the teleconnection in the four months individually have also been presented in previous studies (e.g. in the cited Moron and Gouirand 2003, among others). **Tag: R1.1, R2.2**

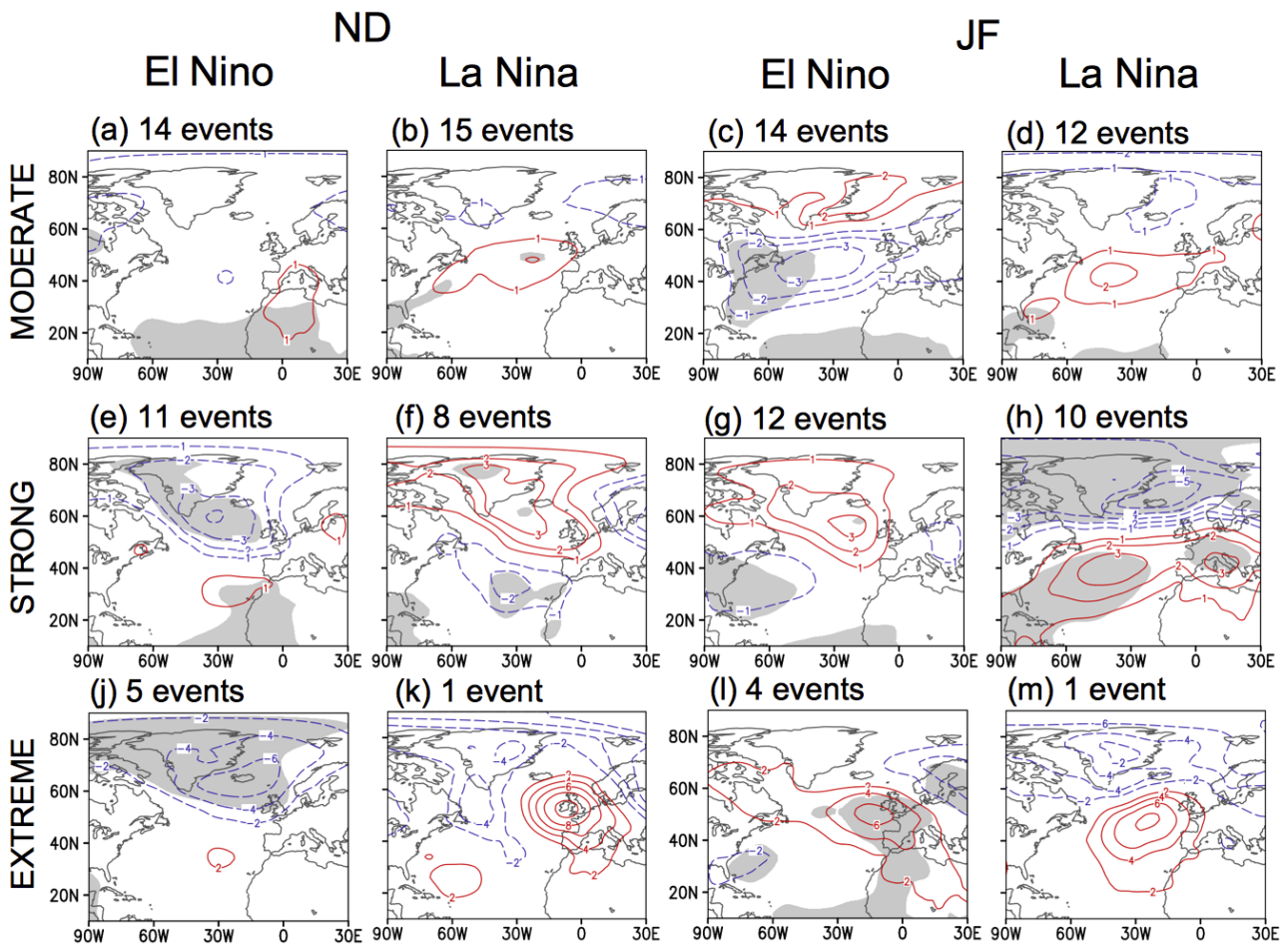
**3.** The authors analyze the effect of separating the ENSO events into CP and EP events. I think, given the nature of the paper, it would be important to analyze what is the effect of the ENSO magnitude, i.e. consider strong ( $Nino3.4 > 1.5/2SD$ ) and moderate ( $1SD > Nino3.4 > 1.5/2SD$ ) ENSO events separately. I think the current length of the paper would allow for the addition of this analysis.

We have now calculated these additional composites based on event strengths as suggested. Additionally, we added an 'extreme' category (see Fig. R1 and caption below). Very roughly, the stronger ENSO events appear to be associated with stronger anomalies over the North Atlantic. However, it is difficult to be certain about this for a number of reasons:

**a.** Sample sizes are smaller after these separations and so statistical significance is severely affected in many cases. (please see also somewhat related point R1.2 above)

**b.** There is only one extreme event for extreme ND or JF La Nina.

**c.** The composite for extreme JF El Nino (Fig. R11, 4 events) appears to be fundamentally different. This is a well-known result shown in Toniazzo and Scaife (2006, <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2006GL027881>)



*Fig. R1 SLP composites for El Niño and La Niña events during 1870-2014 in ND (first and second columns) and JF (third and fourth columns) means. Each row of these panels corresponds to the first row of Fig. 5 except here there is further separation according to the ENSO amplitude in terms of the standard deviation of the Niño2.4 index: moderate ( $1 \text{ SD} \leq |\text{Niño}3.4| < 1.5 \text{ SD}$ ), strong ( $|\text{Niño}3.4| \geq 1.5 \text{ SD}$ ), and extreme ( $|\text{Niño}3.4| \geq 2.0 \text{ SD}$ ) (Note that the 'strong' category also includes the 'extreme' events). There is only one extreme La Niña event each for ND and JF. Red (blue) contours indicate positive (negative) anomalies, with gray shading indicating the 5% significance level for a two-tailed t test. In first and second rows: contour interval = 1hPa, in third row: contour interval = 2hPa.*

We decided not to include this result in the manuscript because we believe these additional calculations do not do justice to the nonlinearity topic nor add to previous findings (such as those cited in R2.1 above, especially Trascasa-Castro et al. 2019). As these authors and you have indicated, which we agree, the results concerning nonlinearity in ENSO teleconnections based on only the reanalysis data are very uncertain (perhaps this shouldn't be a surprise as even the certainty in the linear composites from reanalysis is being questioned).

On the other hand, in addressing your points R2.25 and R2.26 below about the asymmetrical parts in terms of the El Niño+La Niña composites, we decided that the result is a warranted addition to the main text and figure. We recognise that there are a number of other factors that may affect nonlinearity which we cannot deal with in the manuscript (amplitudes of ENSO being one of them) and these are mentioned as a caveat in Sect 3.2. Please also see R2.25 and R2.26 below.

### R2.3



4. Section 3.3 discusses different methods to compute confidence intervals. However, the conclusion of this section is rather “boring”. Figures 7 and 8 look almost the same for all the different methods employed. I understand that the reason why the authors have decided to put these figures in the main text is to show that they are actually very similar. Nevertheless, I would recommend putting these two figures as an appendix as they do not provide a lot of quantitative information, but they are a justification for the methods employed in the first part of the study.

The original complete set of Figs. 7, 8 are now moved to the appendix figures. Now, only the confidence intervals for the JF El Nino composite are shown in the main figure. **Tag: R2.4**

**Specific comments (line by line):**

**Line 10:** What does “some confidence in the signs” means here? Is there a more objective way to state this?

The description is modified to “... we can have confidence (at the 5% significance level)...”.

This statement is based on the results in Sect. 3.3 as well as applying the standard t-test for the composites in Sect. 3.1.

Please see Abstract. **Tag: R2.5**

**Lines 11-12:** Is the lower confidence in the amplitude due to internal North Atlantic variability or due to ENSO diversity (asymmetry and nonlinearity)?

It should be due to the internal North Atlantic SLP variability. Although the ENSO events exhibit variability/diversity, the average of the selected ENSO events (i.e., the composite) from the reanalysis appears to be very stable (do not vary among the bootstrap composites). This is the main thesis of Deser et al. (2017) and our study. Please also refer to the Abstract, Fig. 5 and related text in the cited Deser et al. (2017).

Added explanation in the second paragraph of Introduction. **Tag: R2.6**

**Line 20:** All teleconnections are uncertain in amplitude due to internal variability no? Maybe one should say if this amplitude uncertainty is similar or smaller than in the North Atlantic?

The description is modified according to your suggestion. Please see paragraph 2 in Introduction. **Tag: R2.7**

**Line 22-23:** What do you mean by “detectable response”? Do you mean that the sign of the teleconnection is uncertain?

Sorry for the confusion. We didn’t have a very precise meaning for “detectable” other than to mean the composites (both in sign and amplitude) calculated from the reanalysis data. We changed “detectable” to “observed”. Second paragraph in Introduction. **Tag: R2.8**

**Line 25:** In my opinion, the ND pattern is not similar to the NAO pattern, as one center is much stronger than the other, so it is not just a shift, but more similar to a monopole that resembles the EA pattern (e.g. Wollings et al., 2010)

Woollings, T., Hannachi, A., & Hoskins, B. (2010). Variability of the North Atlantic eddy-driven jet stream. *Quarterly Journal of the Royal Meteorological Society*, 136(649), 856–868. <https://doi.org/10.1002/qj.625>

That part of the sentence is modified to “The Nov-Dec teleconnection resembles the East Atlantic pattern;...”. Please see the third paragraph in Introduction. **Tag: R2.9**

**Line 26:** I think we cannot say that the teleconnection is “the NAO pattern” or NAO-like pattern, but that it projects onto the NAO pattern. Related to this point, the authors might refer to the paper by Mezzina et al. (2020a,b).

Mezzina, B., García-Serrano, J., Bladé, I., & Kucharski, F. (2020). Dynamics of the ENSO Teleconnection and NAO Variability in the North Atlantic–European Late Winter. *Journal of Climate*, 33(3), 907–923. <https://doi.org/10.1175/JCLI-D-19-0192.1>

Mezzina, B., García-Serrano, J., Bladé, I., Palmeiro, F. M., Batté, L., Ardilouze, C., ... Gualdi, S. (2020). Multi-model assessment of the late-winter extra-tropical response to El Niño and La Niña. *Climate Dynamics*, 1, 3. <https://doi.org/10.1007/s00382-020-05415-y>

This part of the sentence is modified to “... while the Jan-Feb teleconnection projects onto the NAO pattern (Mezzina et al. 2020).” Please see the third paragraph in Introduction. **Tag: R2.10**

**Line 30:** Maybe you could add the paper by Jimenez-Esteve and Domeisen (2018) where the tropospheric mechanisms in late winter are analyzed together under the influence of the stratospheric pathway.

Jiménez-Esteve, B., & Domeisen, D. I. V. (2018). The tropospheric pathway of the ENSO-North Atlantic teleconnection. *Journal of Climate*, 31(11), 4563–4584. <https://doi.org/10.1175/JCLI-D-17-0716.1>

Added. See third paragraph in Introduction. **Tag: R2.11**

**Line 34:** Actually, Hardimann et al. (2019) also find that this Rossby wave train is important throughout the season, mainly for strong EN events, but it is more muted for moderate EN or LN events when the stratospheric pathway is active in late winter.

Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Ineson, S., Lim, J., & Fereday, D. (2019). The Impact of Strong El Niño and La Niña events on the North Atlantic. *Geophysical Research Letters*, 46(5), 2874–2883. <https://doi.org/10.1029/2018GL081776>

This point is recognised in the previous paragraph (please see the penultimate sentence in the third paragraph of Introduction). We now added reference to Hardiman et al., 2019. **Tag: R2.12**

Paragraph 4 in the Introduction is describing the teleconnection in Nov-Dec specifically, with a number of recent studies reporting interbasin effects from western tropical Atlantic and the Indian Ocean, etc.

**Line 56:** “larger” than what?

We have changed the description. Our domain includes the midlatitudes and encloses both centres of anomalies in the North Atlantic, in contrast to Deser et al. (2017) who used a polar area. See last paragraph in Introduction. **Tag: R2.13**

**Line 58:** A “more optimistic” perspective compared to? DESER17? I am not sure if “optimistic” is the right term to use here as optimism is subjective and not objective science.

The description/wording is modified. Please see the last sentence in Introduction and the second paragraph in Concluding remarks. Please also see our reply to your related comment below in R2.28. **Tag: R2.14**

**Line 66:** Why are the two citations after a full stop?

The typo is corrected. **Tag: R2.15**

**Line 68-69:** For what months or running mean is the Nino3.4 averaged?

This information is described in the first paragraph of Sect 3.1. Since our analyses involve DJF, ND and JF, (also for periods 1920-2014 or 1870-2014) we feel that it is more convenient to describe the selection of the events closer to where the results are reported. We added a sentence in Sect. 2.2 now (“Table 1 lists the selected ENSO events for calculating the DJF, Nov-Dec and Jan-Feb SLP composites, further explanation is given in the beginning of Sect. 3.1”). **Tag: R2.16**

**Line 72:** How many years do you draw every time from the  $E_0$  and  $L_0$  sets every time (2000 times)? In your expression of  $C^*$ ,  $E^*$  and  $L^*$  are averaged before computing the difference, is this correct? Please clarify.

The bootstrap sample size is the same as the original/observed sample size (same approach in Deser et al. 2017). This is an appropriate choice for investigating the uncertainty of the observed samples with their specific finite sizes. In an added part in Sect 3.1 we also investigate the effect of sample size (from R1.2). A sentence clarifying this point is added in Sect 2.2. **Tag: R2.17**

Yes,  $C^*$  is calculated as the average of  $E^*$  minus the average of  $L^*$  (not different from the common way a composite is calculated). This is already given in the equations with overbar terms, with the meaning of the overbar described. Note that normally we would probably not need to use the notations/equations for the explanation in Sect. 2.2. But in this study we need the notations for explaining the calculations of the different types of confidence intervals in Sect 3.3.

**Lines 89-98:** After reading this subsection, I understand the meaning of what is represented in the Taylor diagram, i.e the amplitude and spatial correlations, but I struggle to understand what exactly each point in the Taylor diagram represents. For example, what do the  $\| \cdot \|_2$  operators symbolize? A better explanation of how the spatial information is averaged in this diagram would be desired.

The symbol means Euclidean norm or L2-norm (ie. the square root of sum of squares). We write out the equations in expanded form now, so the compact form is not needed anymore. The meaning of the polar coordinates  $(r, \alpha)$  of each dot is described in words and equations. Please see Sect. 2.4. **Tag: R2.18**

**Line 103:** Why using different definitions of ENSO depending on the period analyzed? I think the common approach is defining ENSO events based on the NDJ or ONDJF Nino3.4 mean and then use the same years for the 2-months composites (ND or JF). I am aware this will not lead to different results, but I think it complicates the methodology more than necessary.

We added an explanation in the beginning of Sect. 3.1. **Tag: R2.19**

One reason we would use NDJ to define Nino3.4 for selecting DJF extratropics anomalies is to account for a delay (through both the tropospheric and stratospheric pathways) to the extratropical surface anomalies (Deser et al. 2017 use same approach). Following this reasoning for the one-month lag, we use ON Nino3.4 for selecting ND SLP anomalies, and DJ Nino3.4 for selecting JF anomalies. This may not be a widely-used definition of ENSO events per se (as compared to ONDJF, for example). However, as you mentioned this does not affect the results, mainly because ENSO related anomalies often persist from about Oct to Feb. Looking at the years selected in Table 1, we see that in fact NDJ and DJ events are identical. And the ON events are mostly the same ones as well. Only one or two years when the Nino3.4 values are close to the threshold then there are differences.

**Lines 116-117:** It is inaccurate to say that the meridional gradient is opposite for ND than JF, and just say that the centers are shifted, while as you mention in the intro, this pattern, although maybe projecting weakly on the positive NAO, is more closely related to the EA pattern, which usually leads to a strengthening of the North Atlantic jet stream (Woollings et al., 2010).

We deleted "... shows a meridional gradient in the North Atlantic opposite of that for JF, and the centers located south of those for JF." And now simply describe it as resembling the East Atlantic pattern, consistent with Introduction. **Tag: R2.20**

**Figure 3:** Maybe for consistency with Figure 2 you could also show the 5<sup>th</sup> and 95<sup>th</sup> percentile maps when using only the Icelandic low box.

These maps are now added to Fig. 3, and the description at the end of the 4th paragraph in Sect. 3.1 is modified to include the new information. **Tag: R2.21**

**Lines 143-147:** Including a larger area does not decrease the uncertainty in the polar regions (>60degN), but because the uncertainty is much lower over the Azores anticyclone, this compensates the uncertainty of the Icelandic low center when both are averaged in your spatial correlation number.

Yes, you are right. The metric is based on the chosen domain, i.e. it includes the more southern area which has better statistical significance. We deleted this part "... as well as including the mid-latitude North Atlantic in the analysis, ..." **Tag: R2.22**

**Lines 161-170:** I am not very convinced why the authors need to prove that EN is different than LN? Isn't this obvious? What one could test here is if the teleconnection pattern is different (asymmetry). This could be done by multiplying the SLP pattern for LN years by -1, and then perform the analysis that the authors propose here.

We agree that the permutation tests are not strictly necessary after the tests in Figs. 2, A2. However, we think that the test and result are interesting, and would like to use the opportunity to show another method for resampling. We have moved this part to Appendix A and refer to it near the end of Sect. 3.1. **Tag: R2.23**

(Please see reply to R2.25 and R2.26 below regarding the test for asymmetry).

**Lines 171-173:** I do not understand where the 30/2000, 16/2000 and 11/2000 values come from? I also assume in Figure 5 you are using the 1870-2014 period as you refer to figure A1, but this is not mentioned in the text. Could you please clarify?

The numerators of these fractions are the number of red dots in the respective panels (i.e., the red dots are those bootstrap composites having amplitudes at least as large as the amplitudes of the observed composites). And 2000 is just the total number of bootstrap samples. Additional explanation is given now. Please see paragraph 2 in Appendix A. The period "1870-2014" is added now to the figure caption. **Tag: R2.24**

Note that this part is moved to Appendix A per reply R2.23 above.

**Line 187:** I do not agree with the statement that there is no sign of asymmetry between EN and LN response. I can see a more zonally extended NAO-like dipole response for JF LN or a much stronger deepening of the Aleutian low in the North Pacific for EN. The asymmetry seems to be more evident in late winter. This is asymmetry from my understanding. Alternatively, one could do bootstrapping to analyze the asymmetry similar as in Figures 1 and 2 but using the EN+LN composite instead.

We wrote "...do not reveal any sign asymmetry" - meaning there is no asymmetry in signs (not considering asymmetry/nonlinearity in amplitudes) of the SLP anomalies. Also, the analyses using Taylor diagrams are based on the domain as a whole, not specific parts of the domain. This sentence has been edited for clarity. Please see the second paragraph in Sect. 3.2. Additionally, we previously mentioned in the same paragraph the 'apparent' asymmetry in some locations. **Tag: R2.25.**

Regarding testing the EN+LN composites... We had previously performed this analysis but were reluctant to include the result partly because nonlinearity is not our main focus and partly because we think this is a challenging problem that we cannot address comprehensively here. However, we are now persuaded to include this result. It is described in a new paragraph in Sect. 3.2 together with a new main figure. Some caveats on the limitations of the result are included. **Tag: R2.25**

Similar to the papers you also referred to below in R2.26, we could not show statistically significant nonlinearity using the reanalysis data. We do not claim that there is no nonlinearity per se. Only that our specific analyses (domains, variables, ENSO definition, limited reanalysis events) cannot show it. These factors are mentioned as the caveats to the added result. **Tag: R2.25**

**Line 209:** If the purpose of this paper is uncertainty, I think the authors should try to see if the uncertainties in the nonlinearity/asymmetry are significant or not (see my previous comment).

Previous studies identifying nonlinearities (e.g., Trascasa-Castro 2019; and Jimenez-Estevé and Domeisen, 2020) have shown that those are only statistically significant in model simulations where a large sample of strong ENSO events can be simulated.

Please see our reply to R2.25 above. What we have found is consistent with your summary here.  
**Tag: R2.26**

**Lines 259-260:** As I mentioned before I wouldn't call the early-winter response a dipole response.

The word "dipole" is deleted. Please see second paragraph in Sect. 4. **Tag: R2.27**

**Lines 264-265:** What is the thing that makes your view more optimistic than previous studies? This optimistic view that you claim needs a more clear justification. Most previous studies acknowledge that internal variability in the North Atlantic is large and those signals are not always as robust as one would like, but that does not mean that other papers are less optimistic. Nevertheless, optimism is a subjective vision and I think science should try to be objective

We have modified our wording and description, please see also reply to R2.14 above. **Tag: R2.28**

By "more optimistic" we had meant to argue that we can have confidence in the patterns of the teleconnection in the North Atlantic, which is not the message one might get from Deser et al. (2017). See in particular their Figs. 5b, 6, and conclusions:

"Even with nearly 100 years of data comprising 18 El Niño and 14 La Niña events, we have shown that the observed extratropical NH SLP response to ENSO in boreal winter (DJF) is subject to considerable uncertainty in pattern and amplitude..."

"The observed SLP composite shows a robust ENSO response over the North Pacific and North America..."

"Other regions, such as the Arctic, North Atlantic, and Europe, show a larger range of patterns, amplitudes, and statistical significance across the bootstrapped samples."



# Resampling of ENSO teleconnections: accounting for cold season evolution reduces uncertainty in the North Atlantic

Martin P. King<sup>1,2,3</sup>, Camille Li<sup>2,3</sup>, and Stefan Sobolowski<sup>1,3</sup>

<sup>1</sup>NORCE Norwegian Research Centre, Bergen, Norway

<sup>2</sup>Geophysical Institute, University of Bergen, Bergen, Norway

<sup>3</sup>Bjerknes Centre for Climate Research, Bergen, Norway

**Correspondence:** Martin P. King (martin.king@uib.no)

**Abstract.** We re-examine the uncertainty of the El Niño–Southern Oscillation (ENSO) teleconnection to the North Atlantic following the investigation of Deser et al. (2017) (DES2017). Our analyses are performed on the November–December (ND) and January–February (JF) means separately, and for a geographical area that covers a larger extent in the midlatitude North Atlantic than DES2017. The motivation for splitting the cold season in this way arises from the fact that the teleconnection patterns and underlying physical mechanisms are different in late fall compared to mid-winter. As in DES2017, our main technique in quantifying the uncertainty is bootstrap resampling. Amplitudes and spatial correlations of the bootstrap samples are presented together effectively using Taylor diagrams. In addition to the confidence intervals calculated from Student’s  $t$  tests and the percentiles of anomalous sea-level pressure (SLP) values in the bootstrap samples, we also investigate additional confidence intervals using techniques that are not widely used in climate research, but have different advantages. In contrast to the interpretation by DES2017, our results indicate that we can have **some confidence (at the 5% significance level)** R2.5 in the patterns of the teleconnected SLP anomalies. The uncertainties in the amplitudes remain large, with the upper-percentile anomalies at up to 2 times those of the lower percentiles in the North Pacific, and 2.8 times in the North Atlantic.

*Copyright statement.* TEXT

## 1 Introduction

During the 1980’s–90’s, dynamical understanding of tropical–extratropical atmospheric teleconnections saw substantial progress, with observational analyses beginning in earnest as data from in-situ and satellite measurements became available. Trenberth et al. (1998) provides an excellent review of many foundational studies performed during this period establishing the idea that tropical SST perturbations such as El Niño–Southern Oscillation (ENSO) can drive upper-tropospheric divergence, which in turn acts as a source for planetary scale Rossby waves with far-reaching climate impacts. Since then, a major development was the discovery of ENSO’s influences on the stratosphere, and the stratosphere as a pathway to influence the surface climate of the extratropics (reviews by Domeisen et al. 2019; Baldwin et al. 2019). More recent efforts in the field include studies on the effects of ENSO sea-surface temperature (SST) diversity and amplitude, and nonlinearity of the teleconnections (e.g.,

Feng et al. 2017; Frauen et al. 2014; Garfinkel et al. 2013; Toniazzo and Scaife 2006; Trascasa-Castro et al. 2019; Weinberger et al. 2019; Zhang et al. 2018). The role of tropical-extratropical teleconnections in the predictability of Northern Hemisphere climate, and how prediction systems can be improved by including these processes correctly, was recognized early on. Attention to this aspect has intensified in recent years (e.g., Hardiman et al. 2020; Scaife et al. 2014, 2016). However, establishing the teleconnection patterns and amplitudes as well as the physical mechanisms involved remains a challenge owing to large atmospheric internal variability in the extratropics. R1.3

The issue of uncertainty in the ENSO teleconnection to the Northern Hemisphere during wintertime (DJF) was raised by Deser et al. (2017) (DES2017 hereafter). They used bootstrapping (random sampling with replacement) to show that the sea level pressure (SLP) composites associated with ENSO vary considerably in both pattern and amplitude, due primarily to sampling fluctuations arising from internal (non-ENSO forced) atmospheric variability rather than ENSO variability, as the ENSO SST composites (average of selected events) are relatively robust and hence have very narrow confidence intervals (see Fig. 5 in DES2017). R2.6 The North Atlantic circulation response exhibits a large range of amplitudes, different levels of statistical significance, and even anomalies of opposite signs in some locations. In contrast, the North Pacific teleconnection ~~is~~ has lower uncertainty in amplitude than the North Atlantic, ~~but not~~ with a high confidence in sign. R2.7 Through the use of large multi-model ensembles DES2017 also examined model biases in both the forced response to ENSO and the internal atmospheric variability. A key message from their paper is that even with nearly 100 years of observations, the ~~detectable~~ observed R2.8 response to ENSO over the North Atlantic sector in boreal winter is very uncertain.

The late-fall to early-winter (Nov-Dec) ENSO teleconnection in the North Atlantic is notably different spatially from its late-winter (Jan-Feb) counterpart (e.g., Moron and Gouirand 2003). The Nov-Dec teleconnection ~~has a dipolar pattern similar to the NAO but with its northern and southern centers shifted southward, resembling~~ resembles R2.9 the East Atlantic pattern, while the Jan-Feb teleconnection ~~is more of a classical NAO-like dipole~~ projects onto the NAO pattern (King et al. 2018a; Mezzina et al. 2020) R2.10. Consistently, Molteni et al. (2020) examined the PDFs of NAO index obtained through resampling a reanalysis record and found a change from positive to negative (negative to positive) NAO index associated with the warm (cold) ENSO phase during the cold season (although the use of a single index loses some accuracy in the spatial patterns). The fall-to-winter evolution has implications for the underlying teleconnection mechanism and associated impacts. Recently, King et al. (2018a) highlighted these intra-seasonal differences, and suggested that the Nov-Dec ENSO teleconnection to the North Atlantic involves a tropospheric pathway, unlike the Jan-Feb teleconnection which involves both tropospheric and stratospheric pathways (also see Ayarzagüena et al. 2018; Domeisen et al. 2019; Hardiman et al. 2019; Jiménez-Estève and Domeisen 2018) R2.11, R2.12. A statistically significant relationship between ENSO and western European temperatures in Nov-Dec was also presented by King et al. (2018a).

The mechanisms behind and model performance in representing the varying ENSO teleconnection during the extended cold season have been explored. Ayarzagüena et al. (2018) suggest that Rossby wave propagation in the troposphere excited by atmospheric diabatic heating (a result of latent heating related to precipitation) in the western tropical Atlantic is responsible for the Nov-Dec teleconnection. The late fall response may also be modulated by SST forcing from the tropical west Pacific (Bladé et al., 2008) or atmospheric diabatic heating over the Indian Ocean, which is itself amplified by ENSO (Abid et al., 2021; Joshi

et al., 2021). A number of studies (Ayarzagüena et al., 2018; Joshi et al., 2021; King et al., 2018a; Molteni et al., 2020) report that models are generally able to simulate the varying teleconnection to the North Atlantic from November through February in initialized hindcasts, while “free-running” coupled-model experiments have less success, producing an NAO-like pattern similar to the observed late-winter teleconnection through all these months with at best a weak signature of the transition.

In this study, we revisit the question of uncertainty in the ENSO teleconnection to the North Atlantic. The major methodological differences compared to DES2017 are that we consider a varying ENSO teleconnection through the cold season, as well as a larger North Atlantic domain that includes the mid-latitudes, where an important part of the teleconnection signal resides. As in DES2017, our main technique for estimating uncertainty is bootstrap resampling (see Hesterberg (2015) and chapters 10, 11 in Efron and Hastie (2016) for good general introductions). This technique is widely employed in climate research for estimating uncertainties in teleconnections (e.g., Michel et al. 2020), estimating prediction skill in large ensembles (e.g., Stockdale et al. 2015), and investigating the effects of sampling variability (e.g., Cash and coauthors 2017). Bootstrapping artificially increases the number of samples or sample size. It is simple to implement and it does not make any assumption about the underlying distribution. However, it should not be regarded as a substitute for producing larger model ensembles or making longer observations, because certain properties of the bootstrap distributions (such as the location parameter) are still dependent on the original samples (Hesterberg, 2015).

The three subsections in Results correspond to the steps of our investigation: (a) Analyze the sampling variability problem of SLP composites associated with ENSO for Jan-Feb and Nov-Dec means separately in addition to DJF means, using a larger North Atlantic domain that includes the midlatitude region in contrast to the polar area used by DES2017; (b) investigate asymmetry in El Niño and La Niña-related SLP composites; and (c) investigate several types of confidence intervals that are available via bootstrapping in addition to the ordinary  $t$  intervals. In designing the bootstrap tests with these factors in mind, we provide a more optimistic aim to present a more nuanced perspective on the uncertainty in the of the ENSO teleconnection to the North Atlantic with reduced uncertainty compared to earlier studies which considered the standard DJF winter season view.

## 2 Data and methods

### 2.1 Data

To be consistent with DES2017, we use the same datasets. The SLP data is from the NOAA-CIRES Twentieth Century Reanalysis V2c (Compo and Coauthors, 2011). The sea surface temperature (SST) used to construct the Nino3.4 index is from HadISST1.1 (Rayner and Coauthors, 2003). DES2017 verified that their results are robust across different datasets; our experience has been similar when dealing with SST, SLP, and geopotential heights in monthly to seasonal means for tropical-extratropical teleconnections. (e.g., King et al. 2018a, b).

## 2.2 SLP composites and bootstrapping

El Niño and La Niña events are identified according to when the magnitude of the Nino3.4 index is greater than or equal to one standard deviation. Table 1 lists the selected ENSO events for calculating the DJF, Nov-Dec and Jan-Feb SLP composites, with further explanation given at the beginning of Sect. 3.1. R2.16 For each category, we have a set of SLP anomaly fields denoted  $\mathbb{E}_o$  or  $\mathbb{L}_o$ , which are approximations of the populations associated with El Niño and La Niña, respectively. The observed (i.e., ‘original’ to differentiate it from ‘bootstrap’) SLP composite  $\mathbb{C}_o$  for ENSO is then calculated as  $\mathbb{C}_o = \overline{\mathbb{E}_o} - \overline{\mathbb{L}_o}$ , where the overbar denotes the mean of the set. In constructing a bootstrap composite  $\mathbb{C}^*$ , we draw randomly with replacement from the sets of  $\mathbb{E}_o$  or  $\mathbb{L}_o$  to form  $\mathbb{E}^*$  or  $\mathbb{L}^*$ , respectively; and  $\mathbb{C}^* = \overline{\mathbb{E}^*} - \overline{\mathbb{L}^*}$  is then calculated. The sample size of each bootstrap  $\mathbb{E}^*$  and  $\mathbb{L}^*$  composite is the same as the observed  $\mathbb{E}_o$  and  $\mathbb{L}_o$  respectively, meaning that the bootstrap composites contain the same number of El Niño or La Niña events as the observations (except when investigating the effect of sample size in Sect. 3.1). R2.17 Multiple bootstrap composites are generated in this way to form a set  $\mathbb{C}^*$ . As in DES2017, we typically create 2000 bootstrap composites (i.e., there are 2000 members in  $\mathbb{C}^*$ ). Larger numbers of bootstrap samples have been tested for some cases and the results are not so different that the findings are affected.

## 2.3 Confidence intervals and significance testing

The ordinary  $t$  confidence intervals are written conventionally as  $\mu = \bar{x} \pm t \cdot SE$  (Bulmer, 1979), where  $\mu$  is the population mean,  $\bar{x}$  a sample mean calculated from  $n$  elements, and  $SE = s/\sqrt{n}$  the standard error estimated from  $s$  the sample standard deviation. In the two-tailed  $t$ -test with the null hypothesis  $\mu = 0$ , we check if  $\mu = |\bar{x}| - t \cdot SE > 0$  is true, in which case we can reject the null hypothesis. The  $t$  value is usually obtained from a table (e.g., p. 265 in Spiegel and Liu 1999). For example, for a 5% significance level, we look up  $t_{0.025}$  for a two-tailed test – that is, a 5% probability split evenly between two tails, which is one reason that a normal (unskewed) distribution is required for the  $t$  test. Rewriting the  $t$  test using the notation introduced in the previous paragraph for the observed composite gives  $|\mathbb{C}_o| - t \cdot \mathbf{SE}_o > 0$ , which is checked gridpoint-wise to see where the null hypothesis can be rejected, and hence the signal can be considered statistically significant. The method described here is widely used in climate research (and also here in Sects. 3.1 and 3.2 to check statistical significance). It is useful to describe this general approach because in Sect. 3.3, we discuss three additional types of confidence interval that bootstrapping allows us to calculate.

## 2.4 Taylor diagram

The number of bootstrap composites in  $\mathbb{C}^*$  is rather large, and the composites themselves exhibit varying spatial patterns and amplitudes. An effective way to summarize this information is through the Taylor diagram (Taylor, 2001). A Taylor diagram is plotted in polar coordinates  $(r, \alpha)$ , where we define the radial distance  $r = \|\mathbb{C}^*\|_2 / \|\mathbb{C}_o\|_2$  of each plotted point from the origin as the ratio of the spatial amplitude (Euclidean norm) of a bootstrap composite to that of the observed composite: R2.18

$$r = \sqrt{\sum_{i,j} \cos(\phi_j) \mathbb{C}^*(i,j)^2} / \sqrt{\sum_{i,j} \cos(\phi_j) \mathbb{C}_o(i,j)^2},$$

and the cosine of the angle  $\alpha$  from the positive  $x$ -axis direction  $\cos(\alpha) = \mathbf{C}^* \cdot \mathbf{C}_o / \|\mathbf{C}^*\|_2 \|\mathbf{C}_o\|_2$  as the spatial correlation  
 120 between the bootstrap and observed composites: R2.18

$$\cos(\alpha) = \frac{\sum_{i,j} \cos(\phi_j) \mathbf{C}^*(i,j) \mathbf{C}_o(i,j)}{\sqrt{\sum_{i,j} \cos(\phi_j) \mathbf{C}^*(i,j)^2} \sqrt{\sum_{i,j} \cos(\phi_j) \mathbf{C}_o(i,j)^2}}.$$

Both the spatial amplitude and spatial correlation are area-weighted (by the cosine of latitude,  $\cos(\phi_j)$  above). The symbol  $\sum_{i,j} \equiv \sum_i \sum_j$  represents the summation over grid points in a chosen spatial domain. R2.18 For the North Pacific analysis, we select the domain 150°E–120°W, 30°–60°N; while for the North Atlantic the domain 50°W–10°E, 20°–60°N is used for  
 125 Nov-Dec, and 60°W–0°E, 30°–70°N for DJF and Jan-Feb. The exact choices of the boundaries might be somewhat subjective but are chosen to enclose the important anomalies in the observed composites within the geographical area of interest (i.e., North Pacific, North Atlantic). R1.4

In practice, we compute  $r$  and  $\cos(\alpha)$  using the GrADS functions `atot` and `scorr`. The general interpretation of the Taylor  
 130 diagram is that the radial distance of a point indicates how strong the anomalies in a bootstrap composite are compared to the observed composite, and the proximity from the  $x$ -axis indicates how closely the pattern of a bootstrap composite resembles the observed composite.

### 3 Results

#### 3.1 Uncertainties of ENSO teleconnections in Nov/Dec and Jan/Feb

First, the observed SLP composites ( $\mathbf{C}_o$  described in Sect. 2) for ENSO are calculated for the period 1920–2013, the same  
 135 period used in DES2017. In addition to the standard DJF composite, we compute composites for Nov-Dec (ND) and Jan-Feb (JF). The events included in the various seasonal composite calculations in this study are listed in Table 1. Also similar to the DES2017 approach, the Nino3.4 indices for NDJ, ON and DJ are used to identify SLP anomalies for DJF, ND, and JF, respectively. One reason for using this one-month lag is to account for a delay in the influence of tropical SST anomalies on the extratropics through both the tropospheric and stratospheric pathways. It turns out that exactly the same years are selected  
 140 for NDJ and DJ, and they are also mostly the same events as for ON. R2.19 For Nino3.4 in NDJ and DJ, our method selects 18 El Niños with 1 event being different to DES2017 (we do not have the 1968/69 event, but instead 1977/78), and 16 La Niñas instead of 14 (we do not have the 1938/39 event, but instead 1950/51, 1970/71, and 2011/12). For the ON Nino3.4, there are 18 and 17 events selected for El Niño and La Niña, respectively. ~~The events included in the various seasonal composite calculations in this study are listed in Table 1.~~

145 The value of separating the cold season into late fall (ND) ~~from~~ and winter (JF) has been addressed by previous studies ~~investigating statistical and dynamical properties of ENSO teleconnections to the North Atlantic (Abid et al., 2021; Ayarzagüena et al., 2018; King et al., 2018a, b; Moron and Gouirand, 2003; Toniazzo and Scaife, 2006), and is also demonstrated in Fig. 1 for the data set used in this study.~~ Shown in Fig. A1 are composites of the four months individually (also see Moron and Gouirand 2003 in particular) which indicate the differences qualitatively and confirm that it is reasonable to perform analyses

150 on ND and JF means. R1.1, R2.2 The DJF SLP composite as well as the spatial pattern of significant anomalies (at the 5% significance level) are very similar to those presented by DES2017 (Fig. 1a). There is a strengthening of the Aleutian Low with amplitude up to about 8 hPa; in the North Atlantic the anomalies project onto the negative NAO, with about 4 hPa weakening of the southern center (Azores High) and up to 3 hPa weakening of the northern center (Icelandic Low, but with a very small area of statistical significance). Panel c shows the composite for JF. The main difference compared to DJF is that the anomaly  
155 in the northern center in the Atlantic sector strengthens to 4 hPa with a larger area of statistical significance. The composite for ND (panel b) instead ~~shows a meridional gradient in the North Atlantic opposite of that for JF, and the centers located south of those for JF~~ resembles the East Atlantic pattern. R2.20

Next, bootstrapping is carried out to produce resampled ENSO composites as described in Sect. 2. The corresponding Taylor diagrams for the North Atlantic sector are shown in the top row of Fig. 2. The large black dot on the  $x$ -axis at coordinates ( $r = 1$ ,  
160  $\alpha = \arccos 1$ ) indicates the observed composite  $C_o$ , and the small dots represent the 2000 bootstrap composites in  $C^*$ . The blue and red lines indicate the 5th and 95th percentiles of the amplitude ratios or spatial correlations: we refer to these as the *bootstrap percentiles* bracketing the *bootstrap confidence interval*, which contains 90% of the bootstrap composites (see also caption). The second and third rows in Fig. 2 show the spatial pattern of bootstrap composites close to the confidence interval bounds in both the amplitude ratio *and* spatial correlation (i.e., the point closest to the intersection of the blue arc and straight  
165 blue line, or of the red arc and red straight red line).

According to the Taylor diagrams, separating the cold season into ND and JF reduces the uncertainty in the ENSO teleconnection to the North Atlantic. Specifically, the cloud of bootstrap composites becomes denser, contracting towards the  $r = 1$  semi-circle and the  $x$ -axis. In terms of the amplitude ratios, the bootstrap confidence intervals for DJF, ND and JF are (0.57,1.59), (0.69,1.56) and (0.58,1.57) respectively, meaning that the upper bounds are about 2.8, 2.3, and 2.7 times the lower  
170 ones. In terms of spatial correlations, the improvements are clear for both ND and JF over the traditional DJF winter season definition. Specifically, the bootstrap confidence intervals are narrower and the 5th percentiles are larger, sitting close to 0.8 (straight blue lines in panels b, c) compared to 0.69 for the DJF bootstrap distribution. As a visual guide for the comparison, the thick gray arc indicating the confidence interval for DJF in panel a is repeated in panels b and c. Even for DJF, our analysis, which includes a larger mid-latitude North Atlantic domain than in DES2017, indicates some certainty in the spatial pattern  
175 because the confidence interval spans only positive spatial correlations (0.69 to 0.99 given in Fig. 2a). If we instead construct a DJF Taylor diagram for a “polar” North Atlantic area (poleward of 60°N) similar to DES2017, the uncertainty is larger in both the spatial correlations and amplitude ratios (Fig. 3a), ~~and the weak bootstrap composites contain~~ The 5th to 95th percentile interval contains bootstrap composites at the lower limit that have limited areas of statistical significance (Fig. 3b); the midlatitude region that exhibits larger patches of statistically significant SLP anomalies (first column in Fig. 2) are not included, and  
180 hence, cannot contribute to narrowing the uncertainty. R2.21

The spatial patterns of the bootstrap percentiles in Fig. 2 are consistent with the  $t$  test results shown in gray shading (Fig. 1), indicating robust signals where the confidence interval is bounded by significant anomalies of the same sign. For DJF, the southern center has statistically significant negative anomalies at both the 5th and 95th bootstrap percentiles, while the northern center has a very small area of statistical significance at the lower 5th percentile (comparing panels d, g). For ND and JF, both



185 the southern and northern centers are bounded by bootstrap composites showing the same (expected) sign. To summarize, these results show that separating the DJF winter into JF and ND, ~~as well as including the mid-latitude North Atlantic in the analysis~~, R2.22 reduces the uncertainty in the pattern of the ENSO teleconnection (as indicated by the narrower intervals for spatial correlations in panels b, c), with a slight improvement in the certainty of the amplitude for ND only. This is examined further in sections 3.2 and 3.3.

190 DES2017 used data in 1920–2013 to coincide with the availability of model data they also analyzed. We have checked the longer period of 1870–2014 available in the reanalysis data (ENSO events selected are listed in Table 1). The result is shown in Fig. A2, with the top row being the observed composites, and the bottom row the corresponding Taylor diagrams. The main difference compared to Figs. 1, 2 is that for the JF analysis, the bootstrap confidence intervals are narrower for both amplitude ratio and spatial correlation, with the 5th percentile for the latter increasing from 0.77 to 0.84. Otherwise, using the longer  
195 reanalysis period *with more available events* for the resampling does not produce markedly different results that affect the conclusions.

The effect of sample size (number of ENSO events) on the robustness and uncertainty range of ENSO signals is often estimated using sub-sampling techniques, especially in large ensembles of modelling experiments (e.g., Michel et al. 2020; Weinberger et al. 2019). Although the total number of available observed events is greatly limited compared to a large model  
200 ensemble, we can still examine the effect of sample size on the uncertainty by sub-sampling with replacement. We perform this test using sample sizes of 5, 10, 15, ..., etc. for the cases in Fig. A2 (see Fig. A3). As expected, the teleconnection becomes more robust with sample size: the uncertainty range for amplitude decreases and the spatial correlations increases as the number of events increases from 5 to 25. At spatial correlations of about 0.6, the bootstrap composites begin to consistently have similar structures to the observed ones. Using this information, our estimate of the minimum sample sizes required for  
205 robust signals are 20, 15, and 15 for DJF, JF and ND, respectively (panels j, h, i in Fig. A3, marked with \*). This result is consistent with the estimates of Weinberger et al. (2019) for a 41-member model ensemble, even given differences in terms of ENSO definitions, variables, months and regions analysed. As the sample size continues to increase beyond the original observed number of events, the confidence intervals narrow further and the bootstrap composites, unsurprisingly, converge to the observed composites (not shown). Note that Taylor diagrams provide only a general diagnostic for this sample size test; the  
210 spatial patterns of the bootstrap composites at the lower limits have to be examined visually (as in the second row of Fig. 2) to provide more details. R1.2

While the focus of this study is the North Atlantic, we briefly touch on ENSO's effects in the North Pacific. The analyses in Fig. 2 are repeated for the North Pacific, and shown in Fig. 4. In general, the North Pacific is less affected by sampling variability in both the amplitude ratios and spatial correlations. The teleconnection pattern is consistent across all three seasonal means,  
215 unlike the North Atlantic, where ND is different. The clouds of bootstrap composites are tighter and closer to the  $x$ -axis, with less uncertainty for DJF than JF or ND. For JF, the bootstrap confidence interval for the amplitude ratios is (0.75,1.31), where the upper bound is 1.75 times the lower one (close to the factor of 2 reported by DES2017). The uncertainty in the amplitude can be important for climate impacts assessment (Deser et al. 2018; Michel et al. 2020).

Another resampling technique is known as the permutation test. We have performed this test with an aim to demonstrate its workings in quantifying uncertainty of the systematic difference between North Atlantic SLP anomalies associated with El Niño and La Niña. This is essentially equivalent to the above assessments which examine the hypothesis of El Niño minus La Niño composites are significantly different from zero. The result is reported in Appendix A. R2.23

In the above, the uncertainties of the composites are investigated, in essence, by questioning the sampling variability of SLP anomalies associated with El Niño and La Niña. In addition, we can also ask: Is there any systematic difference between El Niño and La Niña SLP anomalies such that calculating a linear composite makes sense? Here, we make a quick detour to demonstrate another resampling technique called the permutation test. The null hypothesis is that the SLP anomalies under El Niño and La Niña events originate from the same population. First, the El Niño and La Niña years are put into the same pool. For example, for NDJ, the 26 El Niño and 22 La Niña years (see Table 1) are pooled together. Second, we randomly draw 26 years from the pool and reassign (also called relabel) them as El Niño, and the remaining 22 years as La Niña. The composite  $G^*$  is then calculated, and the steps are repeated to obtain 2000 composites as before. If the reassigned bootstrap composites are distinguishable from the original bootstrap composites, then the null hypothesis can be rejected, meaning that El Niño and La Niña events are different.

The resulting Taylor diagrams are shown in Fig. A4. For the amplitude, we estimate a  $p$  value for the SLP composite based on the proportion of bootstrap composites at least as large as the original composite: 30/2000, 16/2000, and 11/2000 (indicated by the red dots) for DJF, ND, and JF, respectively. Another way to describe this  $p$  value is that it is the probability of obtaining an SLP composite at least as large as the observed one if El Niño and La Niña related SLP anomalies were from the same population. These  $p$  values are small enough (following common practice) to reject the null hypothesis. Note also that none of the bootstrap confidence intervals for the amplitude ratios, indicated with the blue and red semi-circles, crosses over the gray semi-circles (therefore radii for red semi-circles  $< 1$ ) representing the observed amplitudes. The permutation tests carried out here suggest that the SLP anomalies associated with El Niño and La Niña in the North Atlantic can be distinguished from each other with a high confidence. R2.23 //Moved to Appendix A//

### 3.2 Uncertainties in El Niño and La Niña teleconnections

The aim of this subsection is to examine the uncertainties for SLP anomalies associated with El Niño and La Niña separately (based on Niño3.4 as above). These Additional analyses also allow us to inspect whether there is any asymmetry in the anomaly patterns associated with these teleconnections. (our approach does not test for statistical significance in the nonlinearity of the amplitudes) R2.25, R2.26.

The observed composites are shown in the top row of Fig. 5, and the corresponding Taylor diagrams for the North Atlantic are shown in the bottom row. These Taylor diagrams do not indicate any serious problem with the uncertainty in the observed composite spatial patterns as even the 5th percentiles of the bootstrap spatial correlations are quite high at 0.59 and higher. Furthermore, as also reported by DES2017 for DJF, we find that the SLP anomalies during El Niño and La Niña events in the ND and JF seasons do not indicate any sign asymmetry in terms of the sign of the anomalies over the domains used in the Taylor diagrams R2.25; this is true for both the North Atlantic and North Pacific sectors. However, the spatial extents of

significant teleconnections in the North Atlantic are different. During El Niño for ND and La Niña for JF, the SLP anomalies in the North Atlantic have larger areas of statistical significance (panels a, d compared with b, c) [R2.25](#), as well as narrower bootstrap confidence intervals in the spatial correlations (panels e, h compared with f, g). Note also the smaller scatter of the bootstrap composites and their proximity to the observed composite on the  $x$  axis in panels e, h. Without further study and more observations, it is not possible to provide a physical explanation for the differences in the uncertainties between these composites (but see Hardiman et al. 2019).

We performed additional analyses to investigate if there is any statistical significance for the asymmetry seen in the composite anomalies in the first row of Fig. 5. This is done by assessing the composites for the asymmetrical portion of the ENSO teleconnection (El Niño+La Niña) against the null hypothesis that they are statistically indistinguishable from zero. The results are shown in Fig. 6. Firstly, in the top row, it is noted that there are no statistically significant values in the North Pacific and North Atlantic areas of interest in this study. In other words, we cannot reject the null hypothesis at the 5% significance level based on  $t$  tests, meaning that we find no significant asymmetry in these regions. There are some locations at lower latitudes over the extratropical Pacific and Atlantic oceans which exhibit statistically significant nonlinearity. Secondly, the second row of Fig. 6 shows the Taylor diagrams for the asymmetrical portion of the teleconnection over the North Atlantic. The confidence intervals for the spatial correlations of the bootstrap composites and the observed composites have lower limits at -0.09 and -0.22 (blue lines), which are weak and even negative, indicating again that nonlinearity is not detectable, at least according to these domain-wise metrics. Note that these analyses are performed using observed SLP anomalies identified with Niño3.4 events as in other parts of this study. We do not investigate nonlinearity related to other factors such as ENSO SST types (except briefly below) or amplitudes, other atmospheric variables and regions, as this is a challenging question in its own right, and is addressed by other studies (e.g., Garfinkel et al. 2013; Jiménez-Esteve and Domeisen 2020; Trascasa-Castro et al. 2019; Weinberger et al. 2019). Consistent with our results, these studies also generally agree that elucidation of nonlinearities using the limited sample of events in the reanalysis record is difficult, and therefore this research requires larger samples of model data. [R2.1](#), [R2.25](#), [R2.26](#)

Many previous studies (e.g., Feng et al. 2017; Frauen et al. 2014; Garfinkel et al. 2013; Toniazzo and Scaife 2006; Zhang et al. 2018) investigated asymmetry in ENSO teleconnections arising from central Pacific (CP) or eastern Pacific (EP) events, or due to varying ENSO strength. Zhang et al. (2018) show that CP-El Niño and CP-La Niña events are symmetrical (see Fig. A5e, g), while CP-La Niña and EP-La Niña (Fig. A5g, h) teleconnections in the Euro-Atlantic area are not the same either spatially or in sign (the CP- and EP-ENSO events selected are listed in Table 2) [R1.5](#). They postulate that the asymmetry can be due to the fact that as the EP-La Niña develops, the eastern tropical Pacific becomes progressively colder, thus reducing the chances of overcoming the convection threshold to influence the overlying atmosphere and trigger teleconnections. They note, however, that they are unable to provide an explanation for the strong anomaly in the North Atlantic-Europe during EP-La Niña (Fig. A5h), which is not symmetrical with El Niño's. A brief check of the ND composites associated with these previously identified CP and EP events [R1.5](#) show very little in terms of robust teleconnection signals (Fig. A5a-d), especially over the North Atlantic. This is likely due to a combination of weaker ENSO SST anomalies in ND and smaller areas over which SST anomalies occur during CP or EP events compared to all events pooled together. Studies such as those cited in the previous

paragraph do suggest, however, that for JF, mixing CP- and EP-ENSO (as the Nino3.4 index does) may result in cancellation of signals in the North Atlantic, thus affecting the statistical significance and increasing the uncertainty of the signals. Therefore, asymmetry is perhaps a more important factor for JF than for ND.

### 3.3 Confidence intervals and *t*-test

Confidence intervals are important for quantifying uncertainty. Different types of confidence intervals have different accuracies depending on the properties of the data, therefore it is informative to consider more than one type (Hesterberg, 2015). The most common (without bootstrapping) is the “ordinary *t* interval” (*ordt*) which is directly related to the *t* test for statistical significance (see Sect. 2.3 and Table 3). The lower and upper limits (at 95% confidence) of this interval are shown in the first column of Fig. 7 for the SLP composite for JF El Niño (the full set for different events is presented in Figs. A6, A7).

Bootstrapping allows for confidence intervals that do not rely on assumptions about the sample distribution. The most intuitive type of confidence interval from bootstrapping is the “bootstrap percentile interval” (*perc*), where the percentiles are obtained directly from the values of the bootstrap composites in  $\mathbb{C}^*$ , and shown in the second column of Fig. 7 (also Figs. A6, A7). The confidence intervals used throughout Sect. 3.1 and by DES2017 are essentially of this type.

Further types of confidence intervals which are not commonly considered in the climate research literature can be calculated from bootstrapping. We describe two additional ones – *tBoot* and *bootT* (Table 3 and Hesterberg 2015). The *tBoot* interval is also quite straightforward: the standard error ( $\mathbf{SE}$  in  $\mathbf{C}_o \pm t_{\alpha/2} \cdot \mathbf{SE}$ ) is calculated as the standard deviation of the bootstrap composites. This comes from the definition of standard error being the standard deviation of the sampling distribution, and using the bootstrap distribution as a substitute for the sampling distribution. For our SLP composites, these *tBoot* intervals (not shown) are virtually identical to the *perc* intervals. The *bootT* interval is less immediately obvious but important because it has higher accuracy (Hesterberg 2015) and, like *perc*, it allows for asymmetrical intervals about the mean. Instead of using a table, the *t* values themselves are calculated from the bootstrap samples (of which we have 2000) and then the 2.5th and 97.5th percentiles of the *t* values are obtained. To do this, from each sample we calculate a value of  $\mathbf{t}^* = (\mathbf{C}^* - \overline{\mathbf{C}^*})/\mathbf{SE}^*$ , where  $\mathbf{C}^*$  is a bootstrap composite,  $\overline{\mathbf{C}^*}$  indicates the mean of all the bootstrap composites, and  $\mathbf{SE}^* = \mathbf{s}^*/\sqrt{n}$  is the standard error for this bootstrap sample. This is performed for all samples to obtain 2000 *t* values. The percentiles of the  $\mathbf{t}^*$  values are then used to determine the *bootT* confidence intervals (see equations in Table 3), and these are shown in the third column of Figs. 7, A6, A7).

Comparing the *ordt*, *perc* and *bootT* confidence intervals, we note a few interesting aspects in Figs. A6, A7. Firstly, *ordt* intervals are consistent (and must be by definition) with the absence or presence of statistically significant areas shown in Fig. 5. For example, Fig. 7a, d show that *ordt* in the North Pacific center and the southern center of the North Atlantic are bounded by negative values at the 95% confidence level, and the signals in these regions are indeed largely significant (Fig. 5). In contrast, the confidence interval for the northern center of the North Atlantic crosses “0” (includes negative and positive values), thus implying that the sign and amplitude of the signal are uncertain, consistent with the absence of statistical significance in Fig. 5c over this area. Secondly, all estimates of the confidence interval agree very well with each other, implying

that the *ordt* intervals are good enough for our purposes (although we would not know this *a priori*). Inspecting the different types of confidence intervals for these cases shows, however, that the *perc* and *bootT* intervals yield slightly larger areas of statistical significance in the northern center North Atlantic teleconnections than *ordt*, suggesting that the bootstrap composites exhibit a small amount of skewness here.

Examining different types of confidence intervals, three of which are obtained from the bootstrap test, provides further support for the results shown in previous subsections. In particular, despite uncertainty in the amplitudes, we have reasonably high certainty in the signs of the teleconnection anomalies in the main centers of action shown by the ordinary *t* test in Fig. 1. The bootstrap confidence intervals may also provide new information for regions where uncertainty in the ENSO response is not normally distributed, such as the northern center of the North Atlantic during late winter.

#### 4 Concluding remarks

This study clarifies the uncertainty in the ENSO teleconnection in North Atlantic by considering the early (Nov-Dec) and late (Jan-Feb) parts of the cold season separately, as well as using a geographical area that better covers both anomaly centers in the North Atlantic. The motivation for separating the seasons in this way follows logically from previous studies that find different teleconnection signals and mechanisms at work during ND compared to JF. Various confidence intervals were used to assess uncertainty, including the widely used Student's *t* test as well as several types of intervals calculated from the bootstrapping analysis (Sect. 3.3). These produce nearly the same results in most cases, thus indicating that the conventional Student's *t* test (Fig. 1) and the equivalent *ordt* confidence intervals (first columns of Figs. A6, A7) are generally reliable for assessing the statistical significance of ENSO teleconnection patterns.

The key results of this study are encapsulated in Figures 2 and 5. Based on our analyses of the observational record, we find that there is confidence in the ~~the dipole spatial~~ R2.27 pattern of SLP anomalies associated with ENSO in the North Atlantic as it changes from November through February. There is an improved uncertainty than suggested by traditional DJF winter analyses, which average over opposite-signed anomalies in early and late winter. In agreement with DES2017, the analyses performed here indicate that uncertainty in the amplitudes of the teleconnection signals is indeed high, even when the cold-season transition is accounted for, with a 95th-to-5th percentiles ratio reaching a maximum factor of 2.8 in the North Atlantic (compared to a factor of 2.0 in the North Pacific). Overall, ~~we argue that~~ our analyses and conclusions permit ~~a more complete and optimistic view of the~~ understanding of the ENSO teleconnection signals in the North Atlantic ~~with reduced uncertainty compared to earlier studies which considered only the standard DJF winter season view.~~ R2.14, R2.28

After many decades of research, interest in ENSO teleconnections in the North Atlantic remains high and new results continue to appear (e.g., recent review by Domeisen et al. 2019). A number of recent studies have contributed to the continuous improvement in understanding the mechanisms behind the ENSO teleconnection to the North Atlantic, including details about how the late fall signal may arise from an atmospheric bridge via the tropical Atlantic (Ayarzagüena et al. 2018; King et al. 2018a) or diabatic heating over the Indian Ocean (Abid et al. 2021; Joshi et al. 2021). Such studies (and Molteni et al. 2020) have also found that “free-running” coupled model or SST-forced atmospheric model experiments do not, or only weakly,

355 reproduce the ENSO teleconnection to the North Atlantic through the cold season, while initialized hindcasts perform better. Other studies have focused on the effects of the ENSO teleconnection on surface climate in Europe (precipitation, temperature, drought indices, e.g., Brönnimann et al. 2007; King et al. 2020; van Oldenborgh and Burgers 2005), for which model performance is less well documented than for atmospheric circulation anomalies (but see e.g., King et al. 2020; Volpi et al. 2020). Further research on surface climate impacts in models and prediction skills should consider the varying nature of the ENSO  
360 teleconnection through the cold season (discussion in King et al. 2018a).

*Code and data availability.* Supplementary code and data that may be used to reproduce the analyses and figures are archived at <http://doi.org/10.5281/zenodo.4587405>.

*Author contributions.* MPK conceived the initial concepts and performed the analyses. CL and SS provided suggestions for the analyses and figures. All authors contributed to the writing.

365 *Competing interests.* Camille Li is an Executive Editor of this journal.

*Acknowledgements.* This study is supported by Research Council of Norway (project nos. 255027, 275268). Stefan Sobolowski was partly supported by the EMULATE project, funded through basic institutional support from the Norwegian Department of Education to the Bjerknes Centre for Climate Research.

## Appendix A: Permutation test

370 In Sect. 3.1, the assessment of the composites is essentially performed by quantifying the uncertainty of the El Niño minus La Niña composites. ~~In addition, we can also ask: Is there any systematic difference between El Niño and La Niña SLP anomalies such that calculating a linear composite makes sense?~~ Here, we demonstrate an ~~other~~ alternative resampling technique called the permutation test. The null hypothesis is that the SLP anomalies under El Niño and La Niña events originate from the same population. First, the El Niño and La Niña years are put into the same pool. For example, for NDJ, the 26 El Niño and  
375 22 La Niña years (see Table 1) are pooled together. Second, we randomly draw 26 years from the pool and reassign (also called relabel) them as El Niño, and the remaining 22 years as La Niña. The composite  $C^*$  is then calculated, and the steps are repeated to obtain 2000 composites as before. If the reassigned bootstrap composites are distinguishable from the original bootstrap composites, then the null hypothesis can be rejected, meaning that El Niño and La Niña events are different.

The resulting Taylor diagrams are shown in Fig. A4. For the spatial amplitude, we estimate a  $p$  value for the SLP composite  
380 based on the proportion of bootstrap composites at least as large as the original composite:  $30/2000 = 0.015$ ,  $16/2000 =$



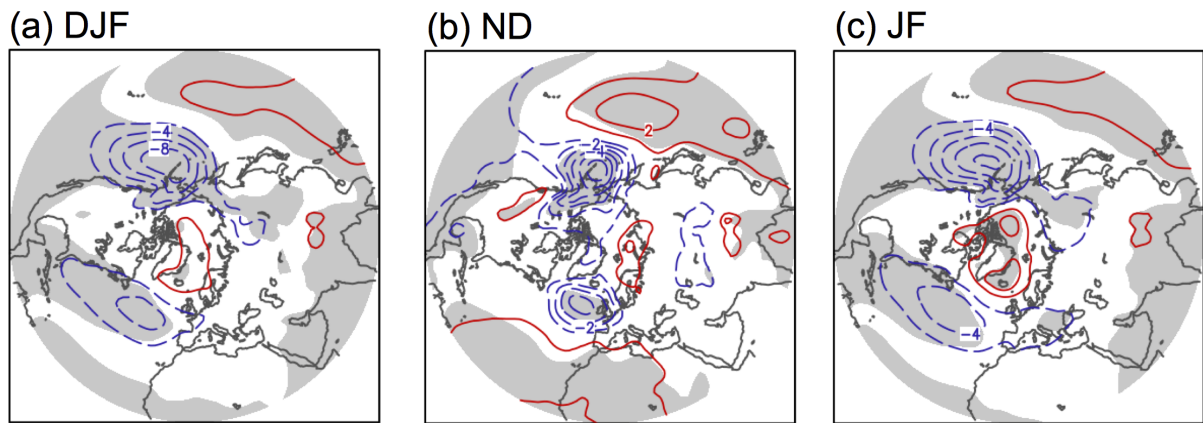
0.008, and  $11/2000 \approx 0.005$  (indicated by the numerators of the fractions are the numbers of red dots in the panels, and the denominators are just the total number of bootstrap composites) R2.24 for DJF, ND, and JF, respectively. Another way to describe these  $p$  values is that each one is the probability of obtaining an SLP composite at least as large as the observed one if El Niño and La Niña related SLP anomalies were from the same population. These  $p$  values are small enough (following common practice of requiring for e.g.  $p = 0.05$ ) to reject the null hypothesis. Note also that none of the bootstrap confidence intervals for the amplitude ratios, indicated with the blue and red semi-circles, crosses over the gray semi-circles representing the observed amplitudes (therefore radii for red semi-circles  $< 1$ ). The permutation tests carried out here suggest that the SLP anomalies associated with El Niño and La Niña in the North Atlantic can be distinguished from each other with a high confidence. This is equivalent and consistent to the result in Sect. 3.1. R2.23 //Moved from end of Sect. 3.1//

## 390 **References**

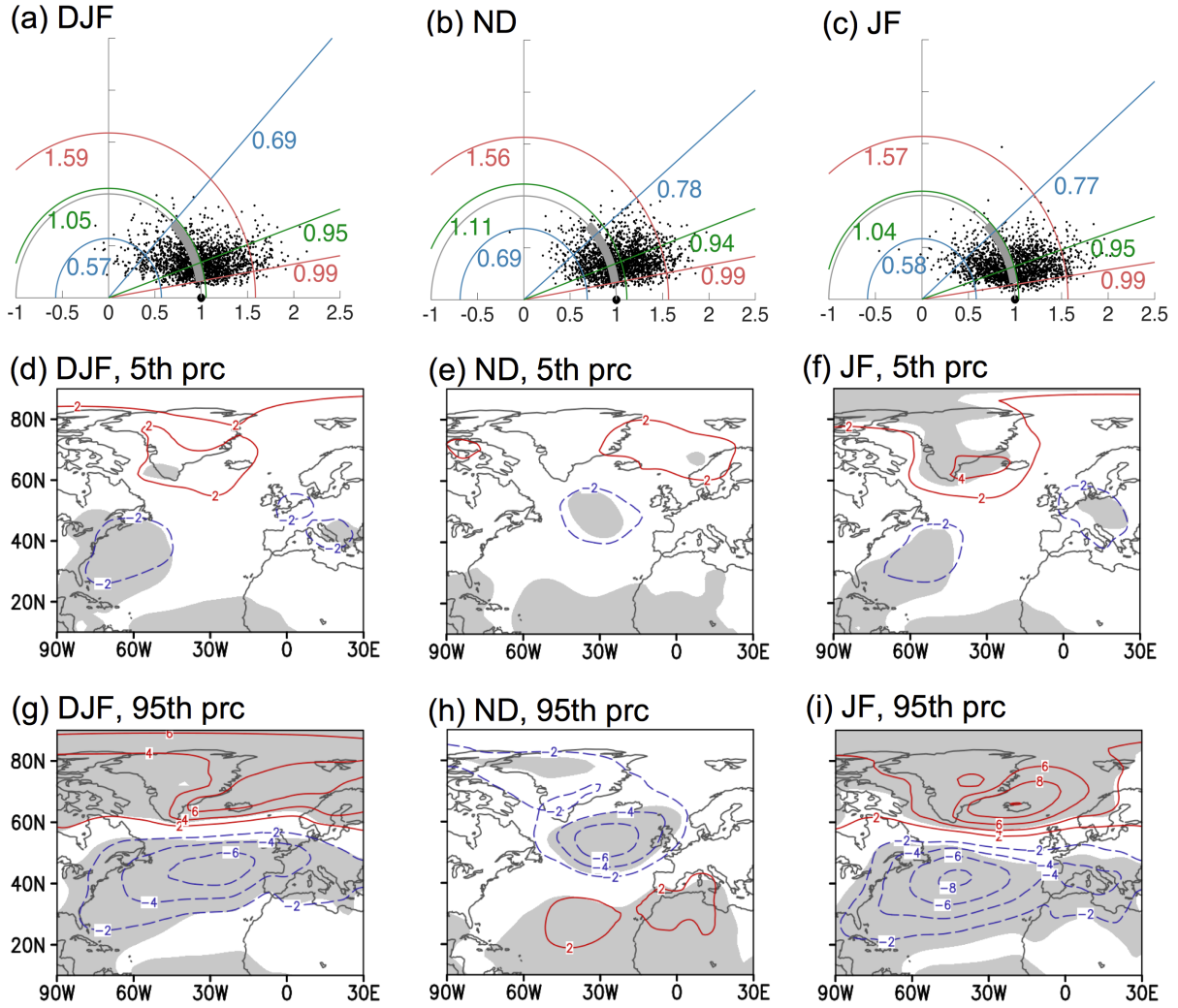
- Abid, M. A., Kucharski, F., Molteni, F., Kang, I.-S., Tompkins, A. M., and Almazroui, M.: Separating the Indian and Pacific Ocean impacts on the Euro-Atlantic response to ENSO and its transition from early to late winter, *J. Climate*, 34, 1531–1548, <https://doi.org/10.1175/JCLI-D-20-0075.1>, 2021.
- Ayarzagüena, B., Ineson, S., Dunstone, N. J., Baldwin, M. P., and Scaife, A. A.: Intraseasonal effects of El Niño-Southern Oscillation on North Atlantic climate, *J. Climate*, 31, 8861–8873, <https://doi.org/10.1175/JCLI-D-18-0097.1>, 2018.
- 395 Baldwin, M. P., Birner, T., Brasseur, G., Burrows, J., Butchart, N., Garcia, R., Geller, M., Gray, L., Hamilton, K., Harnik, N., Hegglin, M. I., Langematz, U., Robock, A., Sato, K., and Scaife, A. A.: 100 years of progress in understanding the stratosphere and mesosphere, *Meteor. Monogr.*, 59, 27.1–27.62, <https://doi.org/10.1175/AMSMONOGRAPHS-D-19-0003.1>, 2019.
- Bladé, I., Newman, M., Alexander, M. A., and Scott, J. D.: The late fall extratropical response to ENSO: Sensitivity to coupling and convection in the tropical West Pacific, *J. Climate*, 21, 6101–6118, <https://doi.org/10.1175/2008JCLI1612.1>, 2008.
- 400 Brönnimann, S., Xoplaki, E., Casty, C., Pauling, A., and Luterbacher, J.: ENSO influence on Europe during the last centuries, *Climate Dyn.*, 28, 181–197, <https://doi.org/10.1007/s00382-006-0175-z>, 2007.
- Bulmer, M. G.: *Principles of statistics*, Dover Publications, 2 edn., 1979.
- Cash, B. A. and coauthors: Sampling variability and the changing ENSO-monsoon relationship, *Climate Dyn.*, 48, 4071–4079, <https://doi.org/10.1007/s00382-016-3320-3>, 2017.
- 405 Compo, G. P. and Coauthors: The Twentieth Century Reanalysis Project, *Quart. J. Roy. Meteor. Soc.*, 137, 1–28, <https://doi.org/10.1002/qj.776>, 2011.
- Deser, C., Simpson, I. R., McKinnon, K. A., and Phillips, A. S.: The Northern Hemisphere extratropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly?, *J. Climate*, 30, 5059–5082, <https://doi.org/10.1175/JCLI-D-16-0844.1>, 2017.
- 410 Deser, C., Simpson, I. R., Phillips, A. S., and McKinnon, K. A.: How well do we know ENSO's climate impacts over North America, and how do we evaluate models accordingly?, *J. Climate*, 31, 4991–5014, <https://doi.org/10.1175/JCLI-D-17-0783.1>, 2018.
- Domeisen, D. I. V., Garfinkel, C. I., and Butler, A. H.: The teleconnection of El Niño Southern Oscillation to the stratosphere, *Rev. Geophys.*, 57, 5–47, <https://doi.org/10.1029/2018RG000596>, 2019.
- 415 Efron, B. and Hastie, T.: *Computer age statistical inference*, Cambridge University Press, 2016.
- Feng, J., Chen, W., and Li, Y.: Asymmetry of the winter extra-tropical teleconnections in the Northern Hemisphere associated with two types of ENSO, *Climate Dyn.*, 48, 2135–2151, <https://doi.org/10.1007/s00382-016-3196-2>, 2017.
- Frauen, C., Dommenges, D., Tyrrell, N., Rezny, M., and Wales, S.: Analysis of nonlinearity of El Niño-Southern Oscillation teleconnections, *J. Climate*, 27, 6225–6244, <https://doi.org/10.1175/JCLI-D-13-00757.1>, 2014.
- 420 Garfinkel, C. I., Hurwitz, M. M., Waugh, D. W., and Butler, A. H.: Are the teleconnections of central Pacific and eastern Pacific El Niño distinct in boreal winter, *Climate Dyn.*, 41, 1835–1852, <https://doi.org/10.1007/s00382-012-1570-2>, 2013.
- Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Ineson, S., Lim, J., and Fereday, D.: The impact of strong El Niño and La Niña events on the North Atlantic, *GRL*, 46, 2874–2883, <https://doi.org/10.1029/2018GL081776>, 2019.
- Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Knight, J. R., Davies, P., Claus, M., and Greatbatch, R. J.: Predictability of European winter 2019/20: Indian Ocean dipole impacts on the NAO, *Atmospheric Science Letters*, 21, 1–10, <https://doi.org/10.1002/asl.1005>, 2020.
- 425

- Hesterberg, T. C.: What teachers should know about the bootstrap: resampling in the undergraduate statistics curriculum, *The American Statistician*, 69, 371–386, <https://doi.org/10.1080/00031305.2015.1089789>, 2015.
- Jiménez-Esteve, B. and Domeisen, D. I. V.: The tropospheric pathway of the ENSO-North Atlantic teleconnection, *J. Climate*, 31, 4563–4584, <https://doi.org/10.1175/JCLI-D-17-0716.1>, 2018.
- 430 Jiménez-Esteve, B. and Domeisen, D. I. V.: Nonlinearity in the tropospheric pathway of ENSO to the North Atlantic, *Weather and Climate Dynamics*, 1, 225–245, <https://doi.org/10.5194/wcd-1-225-2020>, 2020.
- Joshi, M. K., Abid, M. A., and Kucharski, F.: The role of an Indian Ocean heating dipole in the ENSO teleconnection to the North Atlantic European region in early winter during 20th century in Reanalysis and CMIP5 simulations, *J. Climate*, 34, 1047–1060, <https://doi.org/10.1175/JCLI-D-20-0269.1>, 2021.
- 435 King, M. P., Herceg-Bulic, I., Blade, I., Garcia-Serrano, J., Keenlyside, N., Kucharski, F., Li, C., and Sobolowski, S.: Importance of late fall ENSO teleconnection in the Euro-Atlantic Sector, *Bull. Amer. Meteor. Soc.*, pp. 1337–1343, <https://doi.org/10.1175/BAMS-D-17-0020.1>, 2018a.
- King, M. P., Herceg-Bulic, I., Kucharski, F., and Keenlyside, N.: Interannual tropical Pacific sea surface temperature anomalies teleconnection to Northern Hemisphere atmosphere in November, *Climate Dyn.*, 50, 1881–1899, <https://doi.org/10.1007/s00382-017-3727-5>, 2018b.
- 440 King, M. P., Yu, E., and Sillmann, J.: Impact of strong and extreme El Ni nos on European hydroclimate, *Tellus A*, 72, 1704–1709, <https://doi.org/10.1080/16000870.2019.1704342>, 2020.
- Mezzina, B., García-Serrano, J., Bladé, I., and Kucharski, F.: Dynamics of the ENSO teleconnection and NAO variability in the North Atlantic-European late winter, *J. Climate*, 33, 907–923, <https://doi.org/10.1175/JCLI-D-19-0192.1>, 2020.
- 445 Michel, C., Simpson, I. R., Bethke, I., King, M. P., and Sobolowski, S.: The change in the ENSO teleconnection under a low global warming scenario and the uncertainty due to internal variability, *J. Climate*, 33, 4871–4889, <https://doi.org/10.1175/JCLI-D-19-0730.1>, 2020.
- Molteni, F., Roberts, C. D., Senan, R., Keeley, S. P. E., Belluci, A., Corti, S., Franco, R. F., Haarsma, R., Levine, X., Putrasahan, D., Roberts, M. J., and Terray, L.: Boreal-winter teleconnections with tropical Indo-Pacific rainfall in HighResMIP historical simulations from the PRIMAVERA project, *Climate Dyn.*, 55, 1843–1873, <https://doi.org/10.1007/s00382-020-05358-4>, 2020.
- 450 Moron, V. and Gouirand, I.: Seasonal modulation of the El Niño–Southern Oscillation relationship with sea level pressure anomalies over the North Atlantic in October–March 1873–1996, *Int. J. Climatol.*, 23, 143–155, <https://doi.org/10.1002/joc.868>, 2003.
- Rayner, N. A. and Coauthors: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, 108(D14), 4407, <https://doi.org/10.1029/2002JD002670>, 2003.
- Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, N., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A.: Skillful long-range prediction of European and North American winters, *Geophys. Res. Lett.*, 41, 2514–2519, <https://doi.org/10.1002/2014GL059637>, 2014.
- 455 Scaife, A. A., Comer, R. E., Dunstone, N. J., Knight, J. R., Smith, D. M., MacLachlan, C., Martin, N., Peterson, K. A., Rowlands, D., Carroll, E. B., Belcher, S., and Slingo, J.: Tropical rainfall, Rossby waves and regional winter climate predictions, *Quart. J. Roy. Meteor. Soc.*, 143, 1–11, <https://doi.org/10.1002/qj.2910>, 2016.
- 460 Spiegel, M. R. and Liu, J.: *Mathematical handbook of formulas and tables*, McGraw-Hill, 2 edn., 1999.
- Stockdale, T. N., Molteni, F., and Ferranti, L.: Atmospheric initial conditions and the predictability of the Arctic Oscillation, *Geophys. Res. Lett.*, 42, 1173–1179, <https://doi.org/10.1002/2014GL062681>, 2015.

- 465 Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, 106(D7), 7183–7192,  
<https://doi.org/10.1029/2000JD900719>, 2001.
- Toniazzo, T. and Scaife, A. A.: The influence of ENSO on winter North Atlantic climate, *Geophys. Res. Lett.*, 33, L24 704,  
<https://doi.org/10.1029/2006GL027881>, 2006.
- Trascasa-Castro, P., Maycock, A. C., Yiu, Y. Y. S., and Fletcher, J. K.: On the linearity of the stratospheric and Euro-Atlantic response to  
ENSO, *J. Climate*, 32, 6607–6626, <https://doi.org/10.1175/JCLI-D-18-0746.1>, 2019.
- 470 Trenberth, K. E., Branstator, G. W., Karoly, D., Kumar, A., Lau, N.-C., and Ropelewski, C.: Progress during TOGA in understand-  
ing and modeling global teleconnections associated with tropical sea surface temperatures, *J. Geophys. Res.*, 103(C7), 14 291–14 324,  
<https://doi.org/10.1029/97JC01444>, 1998.
- van Oldenborgh, G. J. and Burgers, G.: Searching for decadal variations in ENSO precipitation teleconnections, *Geophys. Res. Lett.*, 32,  
L15 701, <https://doi.org/10.1029/2005GL023110>, 2005.
- 475 Volpi, D., Batté, L., Guérémy, J.-F., and Déqué, M.: Teleconnection-based evaluation of seasonal forecast quality, *Climate Dyn.*, 55, 1353–  
1365, <https://doi.org/10.1007/s00382-020-05327-x>, 2020.
- Weinberger, I., Garfinkel, C. I., White, I. P., and Oman, L. D.: The salience of nonlinearities in the boreal winter response to ENSO, *Climate  
Dyn.*, 53, 4591–4610, <https://doi.org/10.1007/s00382-019-04805-1>, 2019.
- Zhang, W., Wang, Z., Stuecker, M., Turner, A. G., Jin, F.-F., and Geng, X.: Impact of ENSO longitudinal position on teleconnections to the  
480 NAO, *Climate Dyn.*, <https://doi.org/10.1007/s00382-018-4135-1>, 2018.

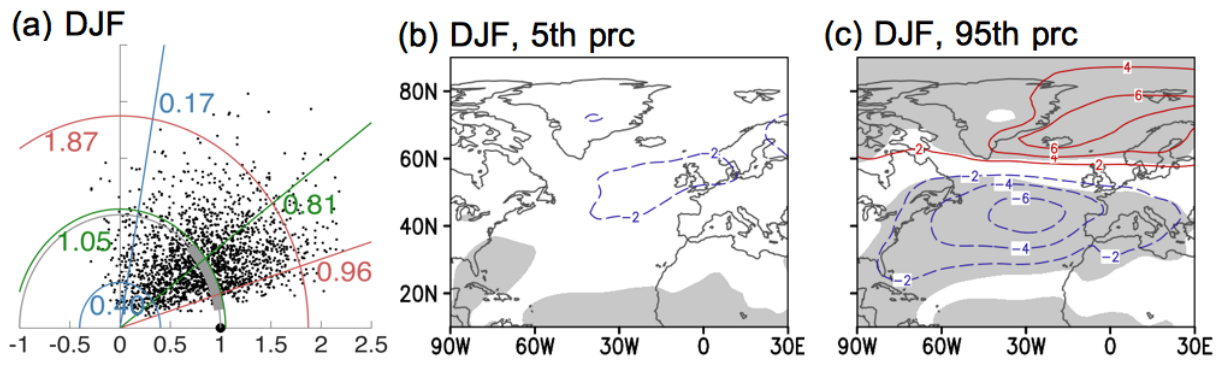


**Figure 1.** Sea level pressure (SLP) composites for El Niño minus La Niña events during 1920–2014 for the months shown on the panels. The ENSO events are defined by Nino3.4 amplitudes greater than 1 standard deviation. Contour interval is 2 hPa in (a) and (c) and 1 hPa in (b), with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates the 5% significance level for a two-tailed  $t$  test. All analyses in this paper use data from HadSST and NOAA-CIRES 20CR V2c.



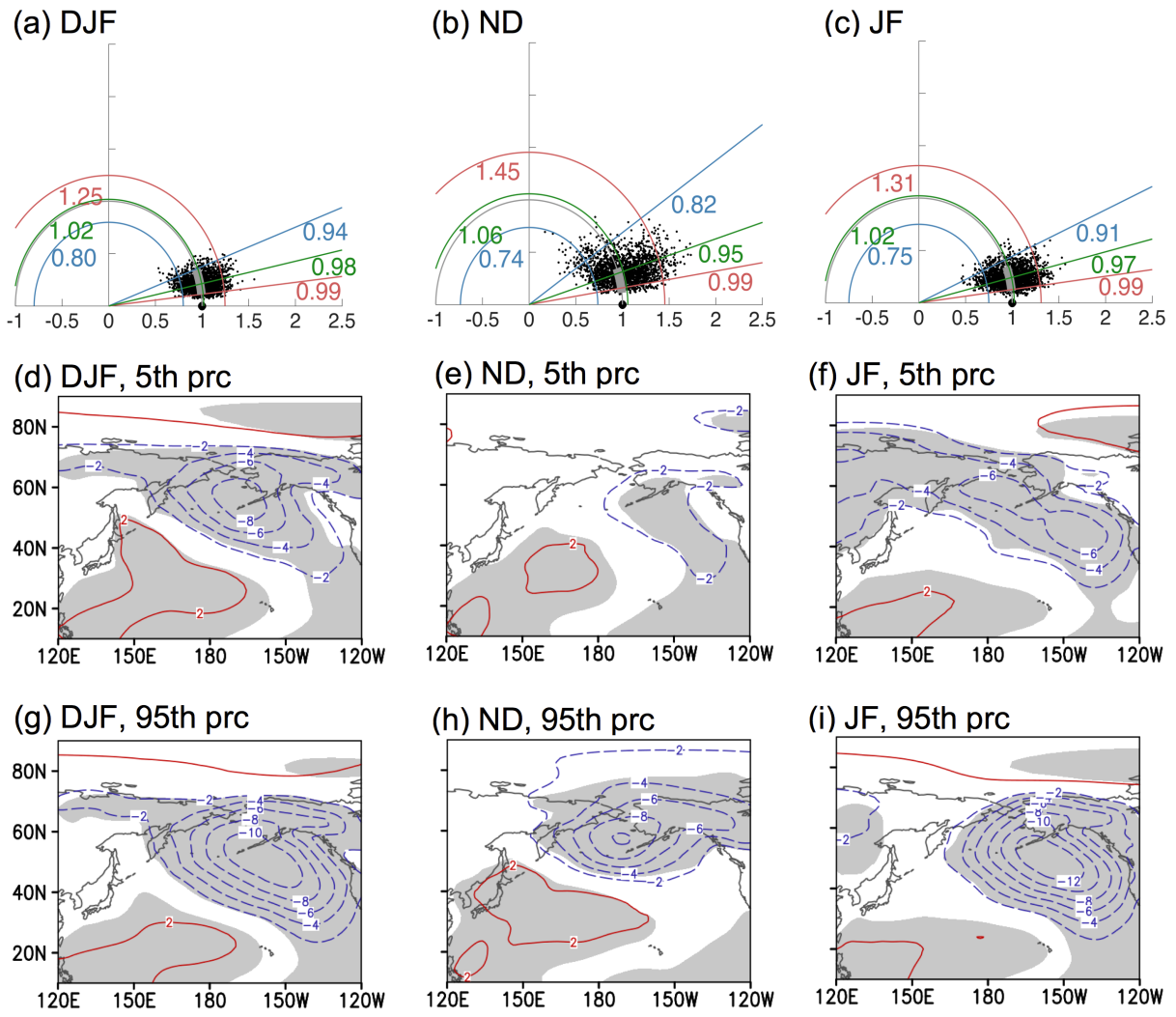
**Figure 2.** Uncertainty of ENSO (El Niño minus La Niña) teleconnection in the North Atlantic during 1920-2014 for the months shown. Top row: Taylor diagrams for North Atlantic with 2000 bootstrap SLP composites (represented by the small dots). The blue, green, and red semi-circles indicate the 5th, median, and 95th percentiles of the amplitude ratios. Blue, green, and red lines indicate the 5th, median, and 95th percentiles of the spatial correlation with the observed composite (represented by the large dot). Second and third rows: Bootstrap SLP composites corresponding to the 5th or 95th percentile based on both amplitude ratio and spatial correlation. Contour interval is 2 hPa, with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates the 5% significance level for a two-tailed  $t$  test. See Sects. 2.4, 3.1 for more details.



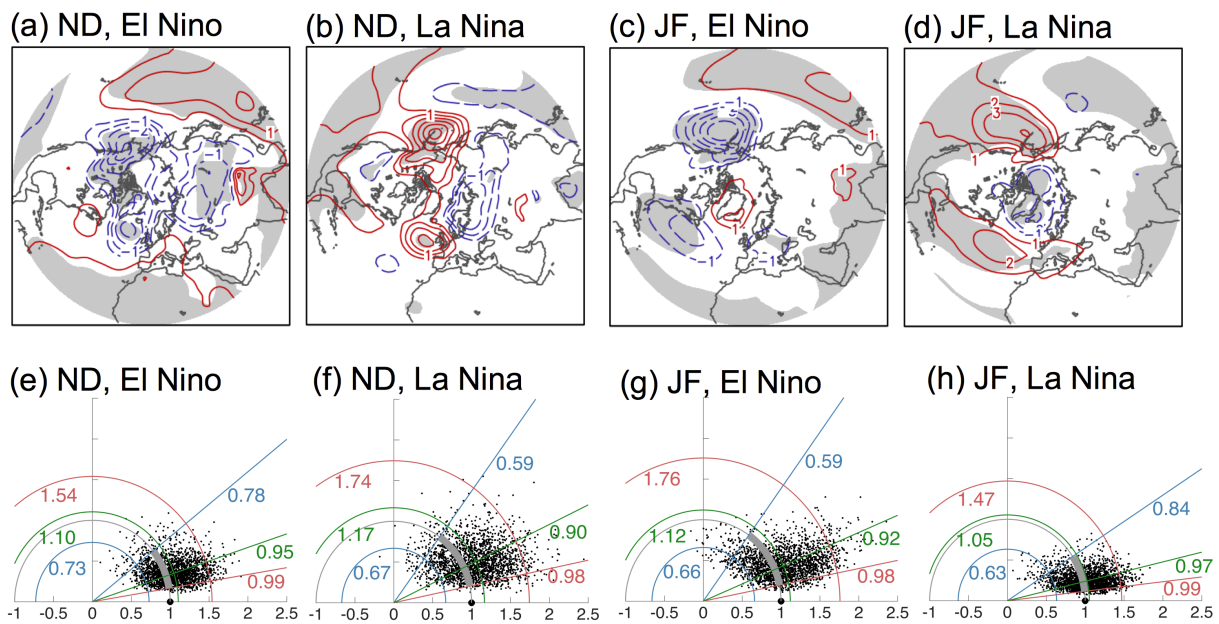


**Figure 3.** Taylor diagram for polar (poleward of 60°N) North Atlantic in DJF with 2000 bootstrap SLP composites, and the 5th and 95th percentile maps. These panels correspond to the first column in Fig. 2. Features in the Taylor diagram are as described in Fig. 2. See Sect. 3.1 for more details.

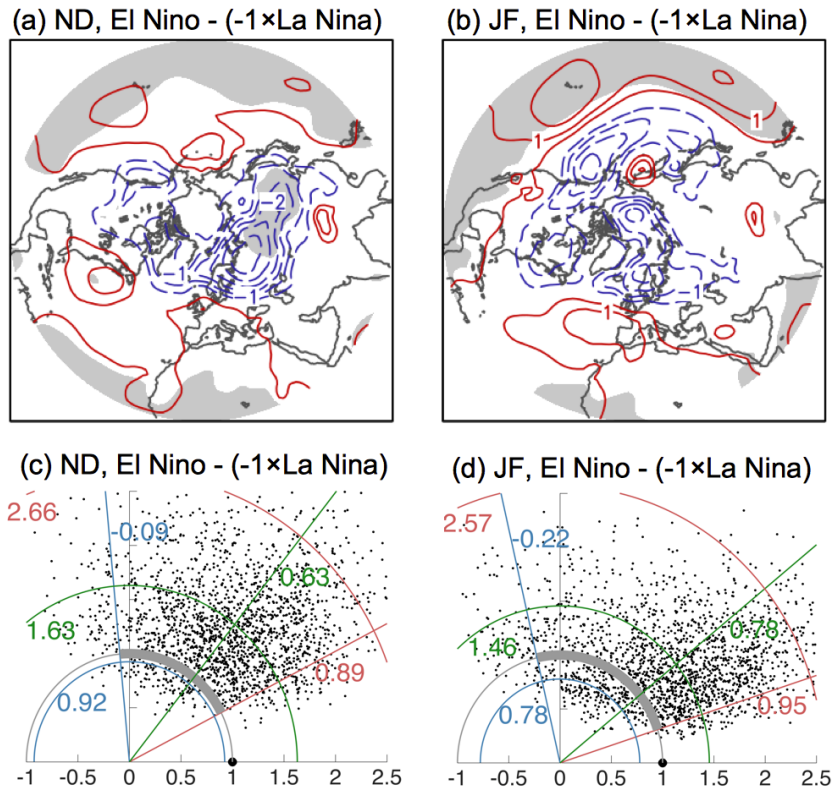
R2.21



**Figure 4.** Same as Fig. 2 except for the North Pacific.

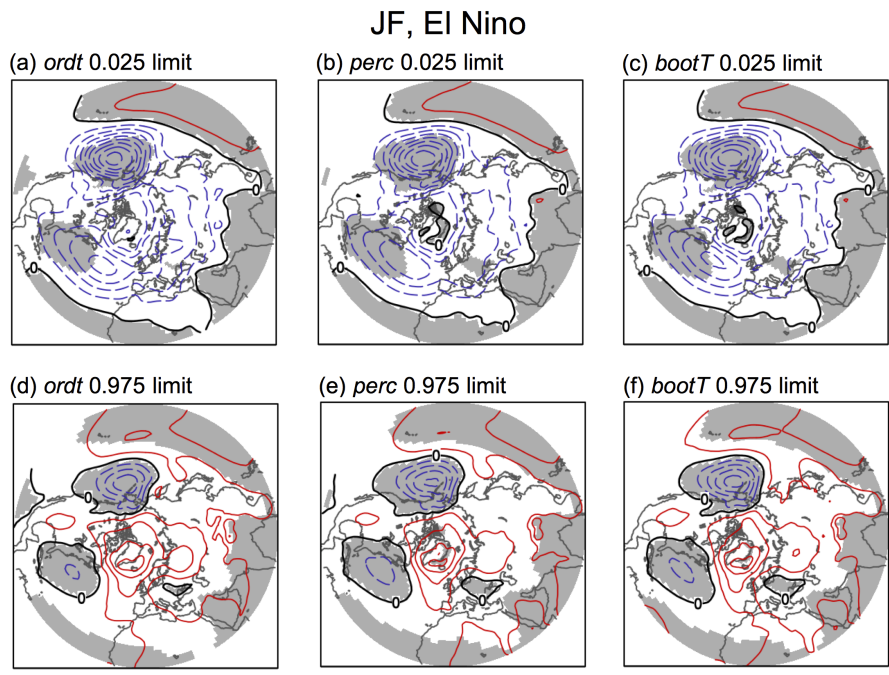


**Figure 5.** Uncertainty of ENSO (for El Niño or La Niña separately) teleconnection in the North Atlantic during 1870-2014 for the months shown. Top row: SLP composites for El Niño or La Niña events. Contour interval is 0.5 hPa for ND and 1 hPa for JF, with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates 5% significance level for a two-tailed  $t$  test. Second row: Corresponding Taylor diagrams for North Atlantic with 2000 bootstrap SLP composites. Features in the Taylor diagrams are as described in Fig. 2. See Sects. 2.4, 3.2 for more details.



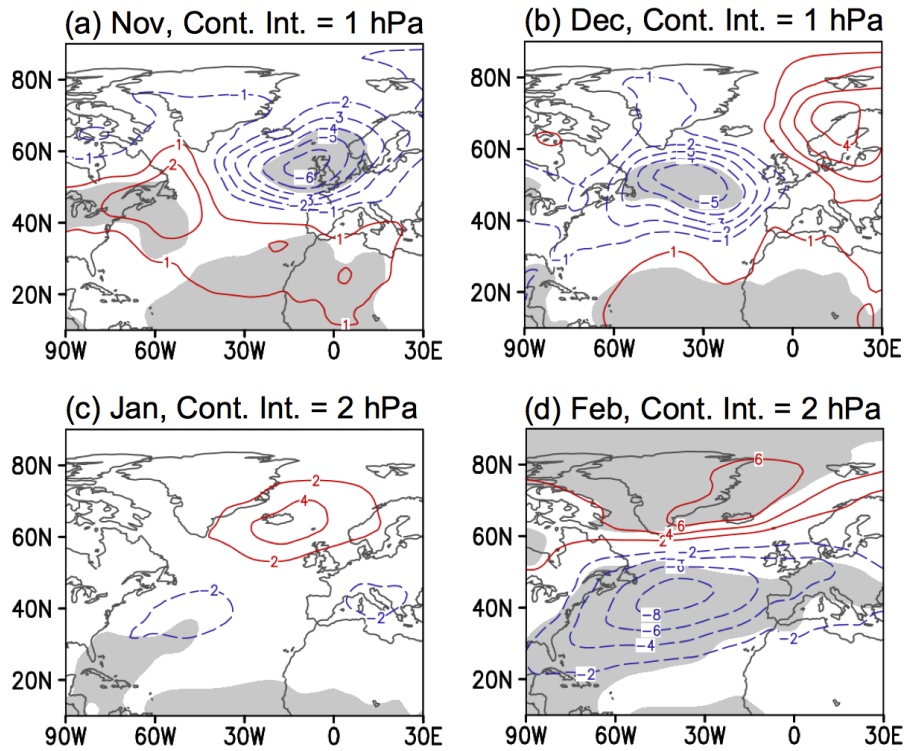
**Figure 6.** Top row: Asymmetrical portion (El Niño- $(-1 \times \text{La Niña})$ ) of the ENSO teleconnection composites in Fig. 5. Red (blue) contours indicate positive (negative) anomalies, with gray shading indicating the 5% significance level for a two-tailed  $t$  test (the null hypothesis is the difference of El Niño and  $-1 \times \text{La Niña}$  is statistically indistinguishable from zero). Contour interval is 0.5 hPa. Second row: Taylor diagrams for North Atlantic with 2000 bootstrap composites of El Niño -  $(-1 \times \text{La Niña})$ , as shown in a, b. Features in the Taylor diagram are as described in Fig. 2. See Sect. 3.2 for more details.

R2.25, R2.26



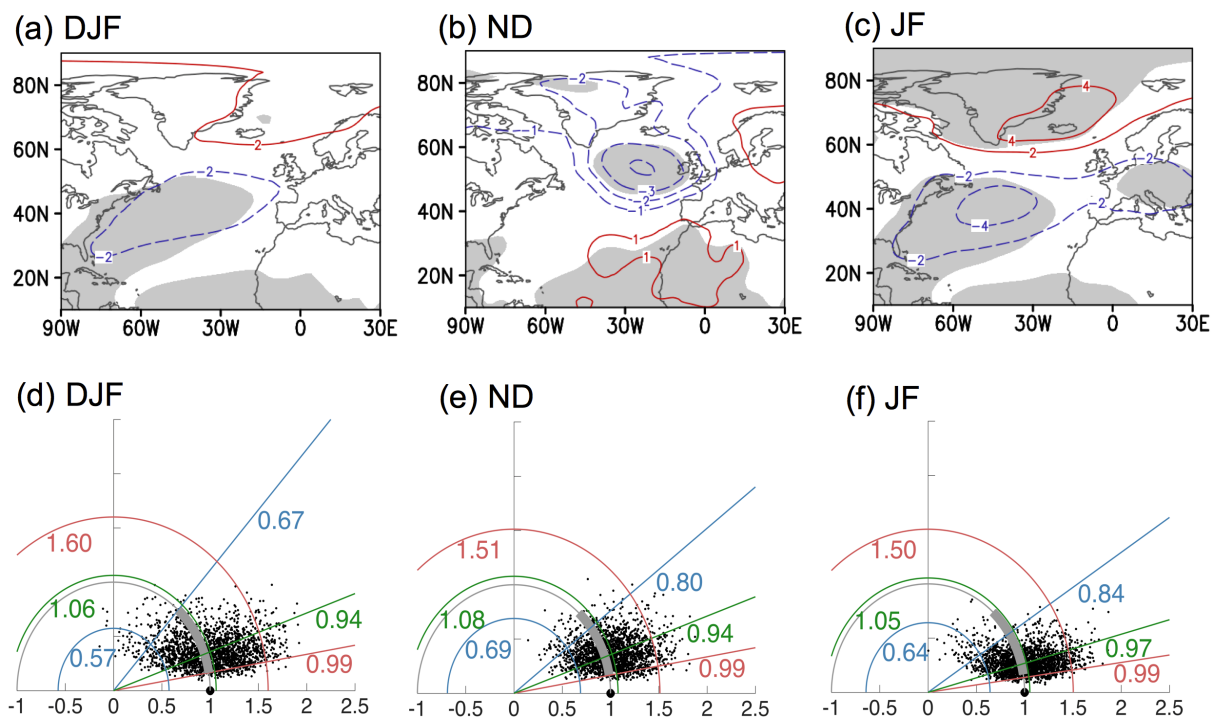
**Figure 7.** Lower and upper limits of confidence intervals (at 95% level) for JF El Niño SLP composite corresponding to Fig. 5c. The *ordt*, *perc*, and *bootT* intervals are described in Sect. 3.3. Contour interval is 1 hPa, with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates locations where the SLP anomalies have the same sign within the entire interval.

R2.4



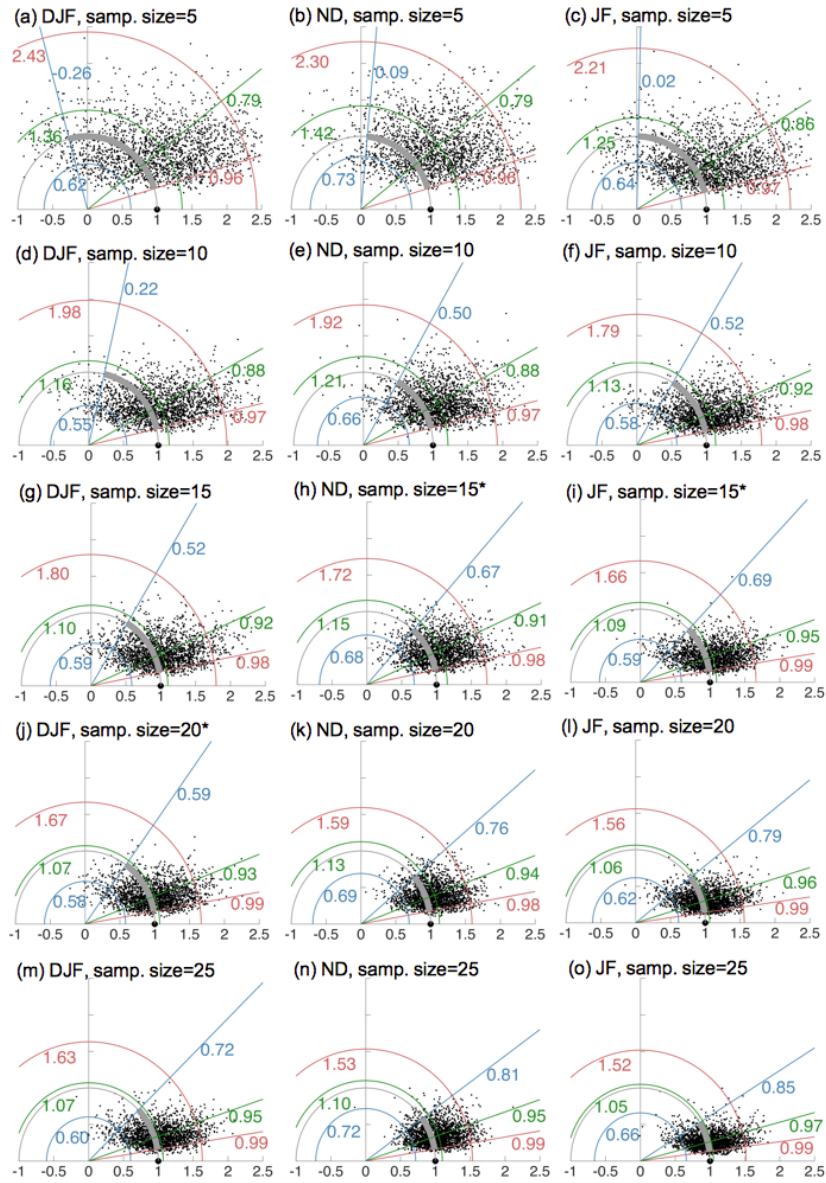
**Figure A1.** SLP composites for El Niño minus La Niña events during 1920–2014 for the months shown on the panels. These are similar to Fig. 1 except here composites for the 4 months are calculated individually. Red (blue) contours indicating positive (negative) anomalies, with gray shading indicates the 5% significance level for a two-tailed t test.

R1.1, R2.2

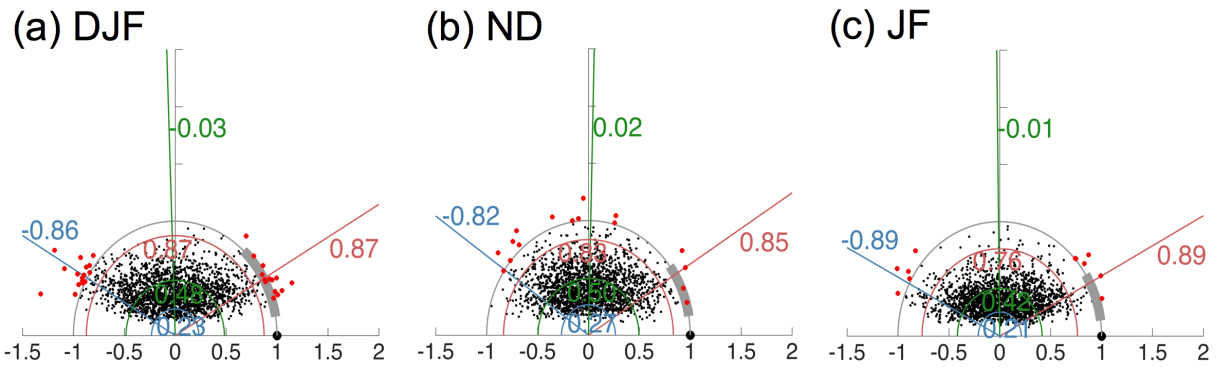


**Figure A2.** Similar to Figs. 1, 2 except for a longer period of 1870–2014. Top row: SLP composites for El Niño minus La Niña events. Contour interval in (a) and (c) is 2 hPa, and is 1 hPa in (b), with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates 5% significance level for a two-tailed  $t$  test. Second row: Corresponding Taylor diagrams for North Atlantic with 2000 bootstrap SLP composites. Features in the Taylor diagrams are as described in Fig. 2. See Sects. 2.4, 3.1 for more details.



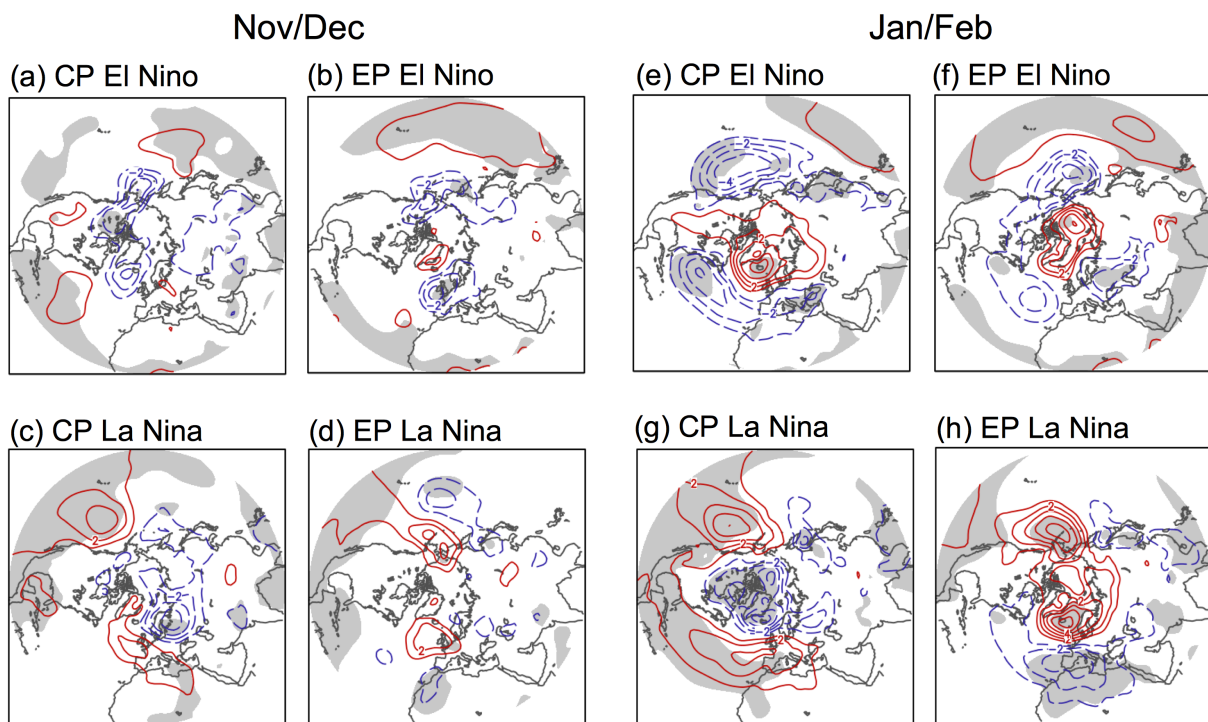


**Figure A3.** Each panel shows a Taylor diagram for North Atlantic with 2000 bootstrap SLP composites, of which each has a sample size shown in the panel title. Each sample is bootstrapped (sub-sampling with replacements) from events in 1870-2014 (Table 1). These panels correspond to the second row of Fig. A2 except for different sample sizes. Panels j, h, i indicate the estimated minimum sample sizes required for robust signals for spatial patterns of the composites. See Sect. 3.1 for more details.

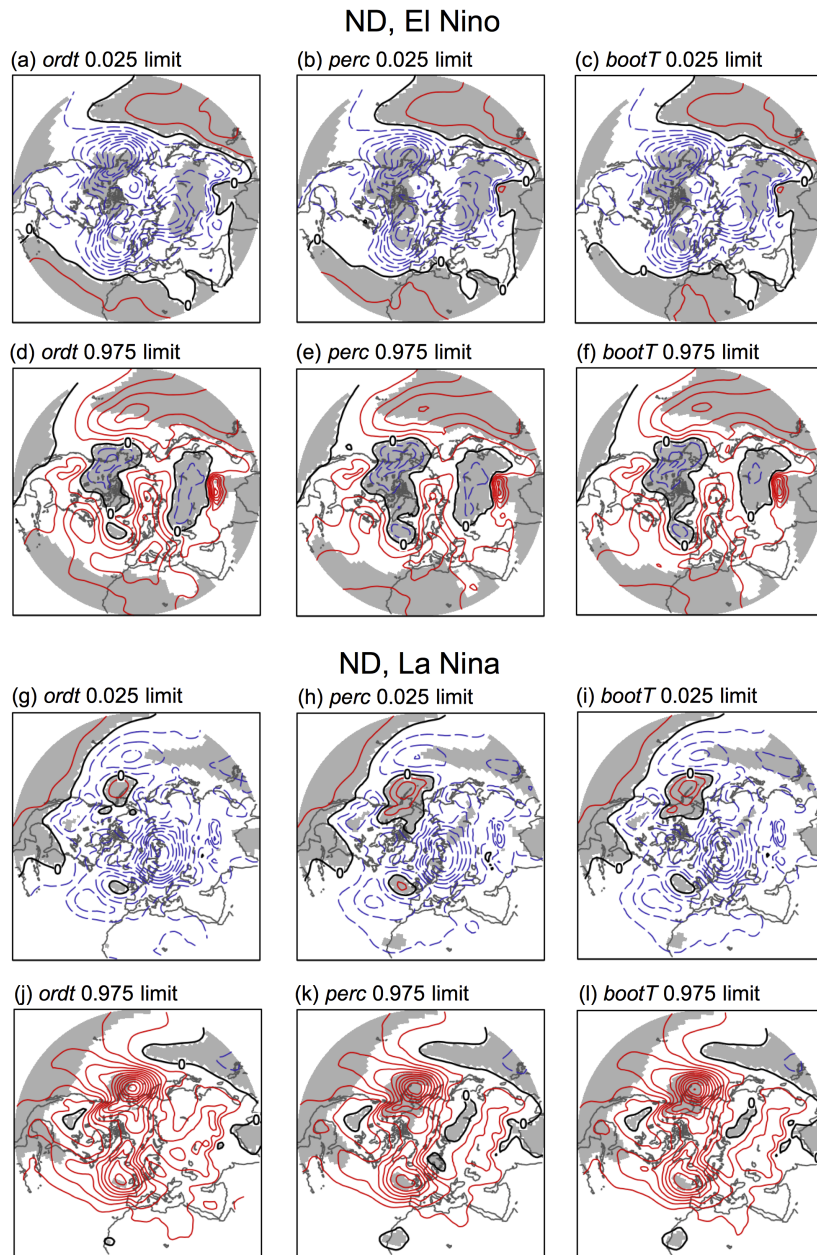


**Figure A4.** Taylor diagrams for permutation tests corresponding to the same events in Fig. A2 (1870-2014). The thick gray arc marks the 5th-to-95th percentile range in the corresponding panel in Fig. A2. Other features in the Taylor diagrams are as described in Fig. 2. See Appendix A for more details.

R2.23 //Moved from main figures//

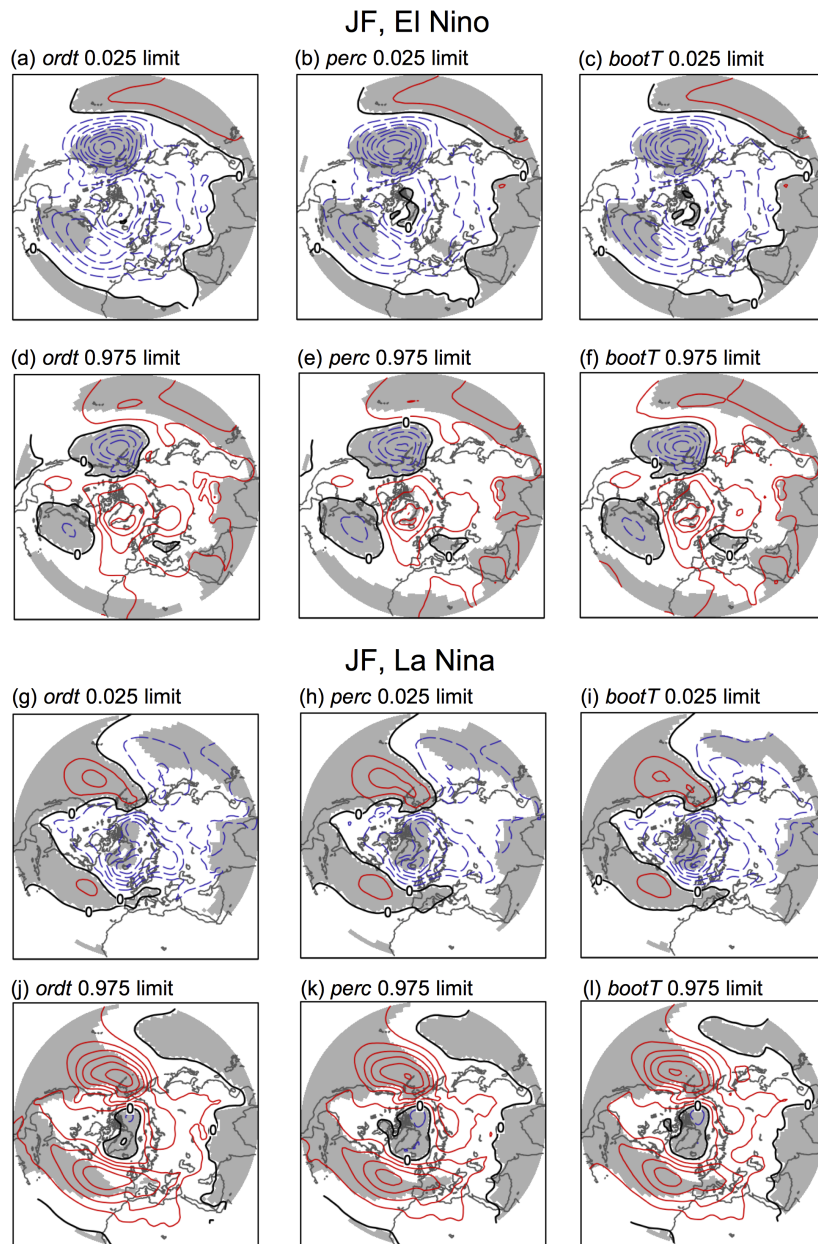


**Figure A5.** SLP composites for central- and eastern-Pacific El Niño, and La Niña, events in the ND or JF means. Contour interval is 1 hPa in all panels, with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates 5% significance level for a two-tailed  $t$  test. See Sect. 3.2 for more details.



**Figure A6.** Lower and upper limits of confidence intervals (at 95% level) for ND El Niño and La Niña SLP composites corresponding to Fig. 5a, b. The *ordt*, *perc*, and *bootT* intervals are described in Sect. 3.3. Contour interval is 0.5 hPa, with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates locations where the SLP anomalies have the same sign within the entire interval.

R2.4 //Moved from main figures//



**Figure A7.** Lower and upper limits of confidence intervals (at 95% level) for JF El Niño and La Niña SLP composites corresponding to Fig. 5c, d. The *ordt*, *perc*, and *bootT* intervals are described in Sect. 3.3. Contour interval is 1 hPa, with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates locations where the SLP anomalies have the same sign within the entire interval.

R2.4 //Moved from main figures//

**Table 1.** ENSO events based on Niño3.4 used in the composite calculations.

Type (number of events): Years
<p><b>El Niño, NDJ or DJ, 1920–2014 (18):</b> 1925/26, 1930/31, 1940/41, 1941/42, 1957/58, 1963/64, 1965/66, 1972/73, 1977/78, 1982/83, 1986/87, 1987/88, 1991/92, 1994/95, 1997/98, 2002/03, 2006/07, 2009/10.</p> <p><b>La Niña, NDJ or DJ, 1920–2014 (16):</b> 1924/25, 1933/34, 1942/43, 1949/50, 1950/51, 1955/56, 1970/71, 1973/74, 1975/76, 1984/85, 1988/89, 1998/99, 1999/00, 2007/08, 2010/11, 2011/12.</p> <p><b>El Niño, ON, 1920–2014 (18):</b> 1923, 1925, 1930, 1941, 1957, 1963, 1965, 1972, 1977, 1982, 1986, 1987, 1991, 1994, 1997, 2002, 2006, 2009.</p> <p><b>La Niña, ON, 1920–2014 (17):</b> 1933, 1942, 1949, 1950, 1954, 1955, 1964, 1973, 1975, 1983, 1984, 1988, 1998, 1999, 2007, 2010, 2011.</p>
<p><b>El Niño, NDJ or DJ, 1870–2014 (26):</b> 1877/78, 1888/89, 1896/97, 1899/00, 1902/03, 1905/06, 1911/12, 1918/19, in addition to those in 1920–2014 given above.</p> <p><b>La Niña, NDJ or DJ, 1870–2014 (22):</b> 1886/87, 1889/90, 1892/93, 1893/94, 1909/10, 1916/17, in addition to those in 1920–2014 given above.</p> <p><b>El Niño, ON, 1870–2014 (25):</b> 1877, 1885, 1888, 1896, 1899, 1902, 1905, 1918, in addition to those in 1920–2014 given above except 2006.</p> <p><b>La Niña, ON, 1870–2014 (23):</b> 1874, 1886, 1889, 1892, 1893, 1909, 1916, in addition to those in 1920–2014 given above except 1954.</p>

**Table 2.** Central–Pacific (CP) and Eastern–Pacific (EP) ENSO events (following Zhang et al. 2018) used for calculating the composites in Fig. A5. See also Sect. 3.2.

---

---

Type (number of events): Years
<b>CP El Niño (9):</b> 1953/54, 1957/58, 1968/69, 1977/78, 1979/80, 1986/87, 2002/03, 2004/05, 2009/10.
<b>EP El Niño (8):</b> 1951/52, 1952/53, 1963/64, 1965/66, 1969/70, 1972/73, 1976/77, 1997/98.
<b>CP La Niña (8):</b> 1973/74, 1974/75, 1975/76, 1988/89, 1998/99, 2000/01, 2010/11, 2011/12.
<b>EP La Niña (8):</b> 1954/55, 1955/56, 1964/65, 1967/68, 1971/72, 1984/85, 1995/96, 2005/06.

---

---

R1.5



**Table 3.** Types of confidence intervals described Sect. 3.3 and Hesterberg (2015). The subscript  $\alpha/2$  indicates a two-tailed  $t$  value at  $\alpha$  significance level. e.g. The current studies uses  $\alpha = 0.05$ .

Long name	Short name	General equation/descriptions
Ordinary $t$ interval	<i>ordt</i>	$\mu = C_o \pm t_{\alpha/2} \cdot \mathbf{SE}_o$ , where $\mathbf{SE}_o$ is estimated from original sample
Bootstrap percentile (or confidence) interval	<i>perc</i>	Intervals are obtained from bootstrap composites directly
$t$ interval with bootstrap SE	<i>tBoot</i>	$\mathbf{SE}$ is estimated as standard deviation of bootstrap composites
Bootstrap $t$ interval	<i>bootT</i>	$\mu = C_o - bootT_{1-\alpha/2} \cdot \mathbf{SE}_o$ , $\mu = C_o - bootT_{\alpha/2} \cdot \mathbf{SE}_o$ , where $\mathbf{SE}_o$ is estimated from original sample, and <i>bootT</i> values are calculated from all bootstrap samples