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

Martin P. King[1,2,3], Camille Li[2,3], and Stefan Sobolowski[1,3]

[1]NORCE Norwegian Research Centre, Bergen, Norway
[2]Geophysical Institute, University of Bergen, Bergen, Norway
[3]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 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

5   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

10   the interpretation by DES2017, our results indicate that we can have some confidence in the signs of the teleconnection 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.

1 Introduction

15   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 pattern and amplitude due to sampling fluctuations arising from internal (and non-ENSO forced) variability. The North Atlantic circulation response, in particular, exhibits a large range of amplitudes, different levels of statistical significance, and even anomalies of opposite signs in some locations. In

20   contrast, the North Pacific teleconnection is uncertain in amplitude, but not in sign. 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 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
25   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 the East Atlantic pattern; the Jan-Feb teleconnection is more of a classical NAO-like dipole. Many studies have not considered how this difference may lead to diverse results and conclusions. Recently, King et al. (2018a) highlighted these intra-seasonal differences, and also suggested that the Nov-Dec ENSO teleconnection to the North Atlantic involves a tropospheric pathway, while the Jan-Feb teleconnection
30   involves both tropospheric and stratospheric pathways (also see Ayarzagüena et al. 2018; Domeisen et al. 2018). 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
35   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., 2020; Joshi et al., 2020). A number of studies (Ayarzagüena et al., 2018; Joshi et al., 2020; King et al., 2018a; Molteni et al., 2020) report that models are generally able to simulate the varying teleconnections to the North Atlantic from November
40   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 method-ological differences compared to DES2017 are that we consider a varying ENSO teleconnection through the cold season, as
45   well as a larger North Atlantic domain that includes the mid-latitudes, where 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
50   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 ob-servations, 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
55   of SLP composites associated with ENSO for Jan-Feb and Nov-Dec means separately, in addition to DJF means, and using a larger North Atlantic domain; **(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 perspective on the uncertainty in the ENSO teleconnection to the North Atlantic.

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. 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 $\mathbf{C}_o$ for ENSO is then calculated as $\mathbf{C}_o = \overline{\mathbb{E}_o} - \overline{\mathbb{L}_o}$, where the overbar denotes the mean of the set. In constructing a bootstrap composite $\mathbf{C}^*$, 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 $\mathbf{C}^* = \overline{\mathbb{E}^*} - \overline{\mathbb{L}^*}$ is then calculated. 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 $\mu = \overline{x} \pm t \cdot SE$ (Bulmer, 1979), where μ is the population mean, $\overline{x}$ 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 $\mu = |\overline{x}| - t \cdot SE > 0$ 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 $|\mathbf{C}_o| - t \cdot \mathbf{SE}_o > 0$, 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, α), where the radial distance $r = \|\mathbf{C}^*\|_2 / \|\mathbf{C}_o\|_2$ of each point from the origin indicates the ratio of the spatial amplitude of a bootstrap composite to that of the observed composite; and the cosine of the angle α from the positive x-axis direction $cos(\alpha) = \mathbf{C}^* \cdot \mathbf{C}_o / \|\mathbf{C}^*\|_2 \|\mathbf{C}_o\|_2$ indicates the spatial correlation between the bootstrap and observed composites. Both the spatial amplitude and spatial correlation are area-weighted (by the cosine of latitude). In practice, we compute r and $cos(\alpha)$ using the GrADS functions `atot` and `scorr`. The general interpretation of the Taylor 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 period used in DES2017. In addition to the standard DJF composite, we compute composites for ND and JF. 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. 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.

The value of separating late fall from the rest of winter has been addressed in previous studies concerned with ENSO teleconnections in the North Atlantic (Ayarzagüena et al., 2018; King et al., 2018a, b), and is demonstrated in Fig. 1 for the data set used in this study. 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 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.

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 ($r = 1$,

120   $\alpha = \arccos 1$) indicates the observed composite $\mathbf{C}_o$, and the small dots represent the 2000 bootstrap composites in $\mathbb{C}^*$. 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) is bounded by them. 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

125   blue arc and straight 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 tele-connection 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

130   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

135   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, and the weak bootstrap composites contain very limited areas of statistical significance (Fig. 3).

The spatial patterns of the bootstrap percentiles in Fig. 2 are consistent with the t test results shown in gray shading (Fig.

140   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 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,

145   reduce 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.

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

150   in Fig. A1, 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 reanalysis period for the resampling does not produce markedly different results that affect the conclusions.

155   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, unlike in 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

160   amplitude can be important for climate impacts assessment (Deser et al. 2018; Michel et al. 2020).

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

165   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 $\mathbf{C}^*$ 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

170   La Niña events are different.

The resulting Taylor diagrams are shown in Fig. 5. 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

175   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.

180   ## 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 analyses also allow us to inspect whether there is any asymmetry in the anomaly patterns (our approach does not test for statistical significance in the nonlinearity of the amplitudes).

The observed composites are shown in the top row of Fig. 6, and the corresponding Taylor diagrams for the North Atlantic
185 are shown in the bottom row. The Taylor diagrams do not indicate any serious uncertainty in the observed composite spatial
patterns. 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 reveal any sign asymmetry; 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 the 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
190 b, c), 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 a further study and more observations, it is not possible to provide a physical explanation for the differences in the
uncertainties between these composites in Fig. 6.

Many previous studies (e.g., Feng et al. 2017; Frauen et al. 2014; Garfinkel et al. 2013; Toniazzo and Scaife 2006; Zhang
195 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.
A2e, g), while CP-La Niña and EP-La Niña (Fig. A2c, h) teleconnections in the Euro-Atlantic area are not the same either
spatially or in sign. 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
200 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. A2h), which is not symmetrical to El Niño's.

Here, we find that the neither CP- nor EP-ENSO teleconnections in ND result in any large areas with statistical significance
(Fig. A2a–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 for all events pooled together. Studies
205 such as those cited in the previous paragraph do imply 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.

We do not further investigate questions related to asymmetrical teleconnections due to different flavours of ENSO, nor to
nonlinearity of the SLP amplitudes associated with them. Our analysis does not reveal asymmetry in El Niño and La Niña
210 (based on Nino3.4) teleconnections in terms of the pattern of the anomalies for JF, nor 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
215 significance (see Sect. 2.3 and Table 2). The lower and upper limits (at 95% confidence) of this interval are shown in the first
column of Figs. 7, 8 for the SLP composites.

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 Figs. 7, 8. 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 2 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 $\mathbf{t}^* = (\mathbf{C}^* - \overline{\mathbb{C}^*})/\mathbf{SE}^*$, where $\mathbf{C}^*$ is a bootstrap composite, $\overline{\mathbb{C}^*}$ 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 2), and these are shown in the third column of Figs. 7, 8.

Comparing the $ordt$, $perc$ and $bootT$ confidence intervals, we note a few interesting aspects in Figs. 7, 8. Firstly, $ordt$ intervals are consistent (and must be by definition) with the absence or presence of statistically significant areas shown in Fig. 6. For example, Fig. 8a, 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. 6). 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. 6c 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 nothern 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

250   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

255   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. 7, 8) are generally reliable for assessing the statistical significance of ENSO teleconnection patterns.

The key results of this study are encapsulated in Figures 2 and 6. Based on our analyses of the observational record, we find that there is reasonable confidence in the the dipole pattern of SLP anomalies associated with ENSO in the North Atlantic as

260   it changes from November through February. This is more certainty 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 fairly 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, our analyses and conclusions permit a more complete and optimistic view of

265   the uncertainty of the ENSO teleconnection in the North Atlantic.

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. 2018). 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.

270   2018a) or diabatic heating over the Indian Ocean (Abid et al. 2020; Joshi et al. 2020). 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, reproduce the ENSO teleconnection to the North Atlantic through the cold season, while intialized 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 perfor-

275   mance 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 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 https://github.com/martin-king/kingetal_2020_resamplingenso.

280 *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.

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
285 Centre for Climate Research.

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, EOR, https://doi.org/10.1175/JCLI-D-20-0075.1, 2020.

290 Ayarzagüena, B., Ineson, S., Dunstone, N. J., Baldwin, M. P., and Scaife, A. A.: Intraseasonal effects of El Nino-Southern Oscillation on North Atlantic climate, J. Climate, 31, 8861–8873, https://doi.org/10.1175/JCLI-D-18-0097.1, 2018.

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.

Brönnimann, S., Xoplaki, E., Casty, C., Pauling, A., and Luterbacher, J.: ENSO influence on Europe during the last centuries, Climate Dyn.,
295 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.

Compo, G. P. and Coauthors: The Twentieth Century Reanalysis Project, Quart. J. Roy. Meteor. Soc., 137, 1–28,
300 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.

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
305 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, 2018.

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
310 of ENSO, Climate Dyn., 48, 2135–2151, https://doi.org/10.1007/s00382-016-3196-2, 2017.

Frauen, C., Dommenget, D., Tyrrell, N., Rezny, M., and Wales, S.: Analysis of nonlinearity of El Nino-Southern Oscillation teleconnections, J. Climate, 27, 6225–6244, https://doi.org/10.1175/JCLI-D-13-00757.1, 2014.

Garfinkel, C. I., Hurwitz, M. M., Waugh, D. W., and Butler, A. H.: Are the teleconnections of central Pacific and eastern Pacific El Nino distinct in boreal winter, Climate Dyn., 41, 1835–1852, https://doi.org/10.1007/s00382-012-1570-2, 2013.

315 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/10.1080/0031305.2015.1089789, 2015.

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, EOR, https://doi.org/10.1175/JCLI-D-20-0269.1, 2020.

320 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.

325 King, M. P., Yu, E., and Sillmann, J.: Impact of strong and extreme El Ni nos on European hydroclimate, Tellus A, 72, 1704 342, https://doi.org/10.1080/16000870.2019.1704342, 2020.

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,
330 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.

Moron, V. and Gouirand, I.: Seasonal modulation of the El Niño–Souther 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
335 century, J. Geophys. Res., 108(D14), 4407, https://doi.org/10.1029/2002JD002670, 2003.

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.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106(D7), 7183–7192,
340 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.

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.

345 Volpi, D., Batté, L., Guérémy, J.-F., and Déqué, M.: Teleconnection-based evaluation of seasonal forecast quality, Climate Dyn., https://doi.org/10.1007/s00382-020-05327-x, 2020.

Zhang, W., Wang, Z., Stuecker, M., Turner, A. G., Jin, F.-F., and Geng, X.: Impact of ENSO longitudinal position on teleconnections to the NAO, Climate Dyn., https://doi.org/10.1007/s00382-018-4135-1, 2018.
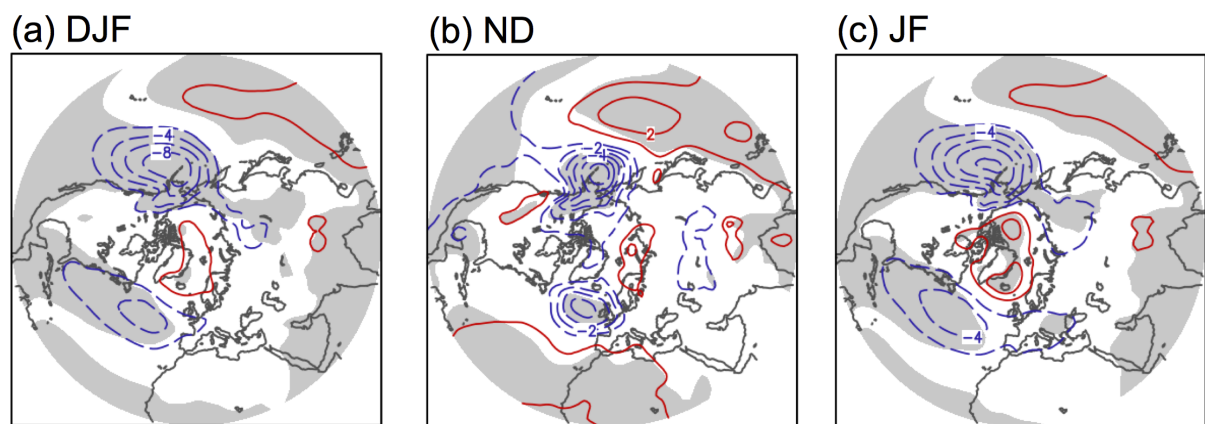
Figure 1. Sea level pressure (SLP) composites for El Niño minus La Niña events during 1920–2014 for the months shown on the panels. The ENSO events are defined by Nino3.4 amplitudes greater than 1 standard deviation. Contour interval is 2 hPa in (a) and (c) and 1 hPa in (b), with red (blue) contours 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.
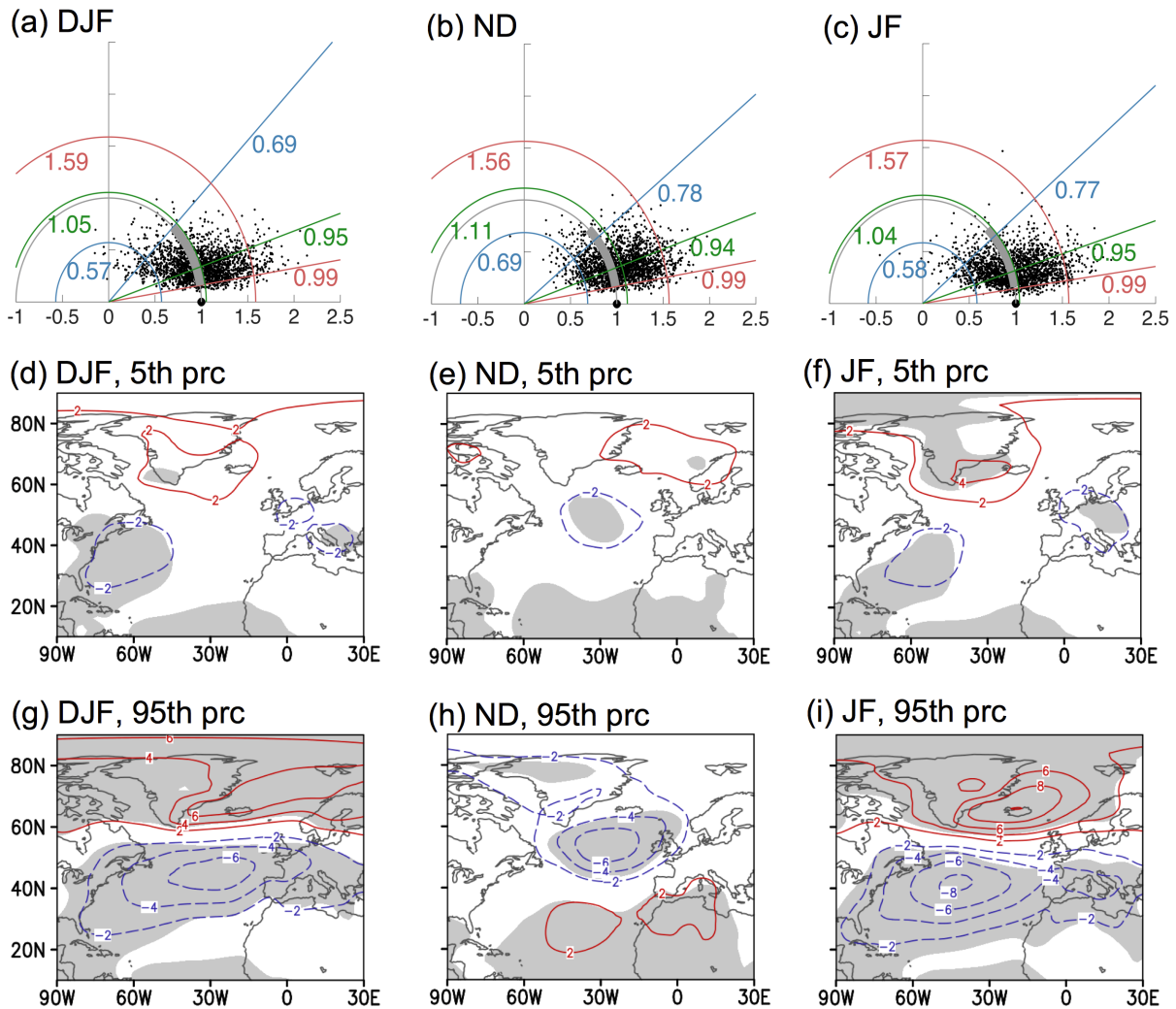
Figure 2. Uncertainty of ENSO (El Niño minus La Niña) teleconnection in the North Atlantic during 1920-2014 for the months shown. Top row: Taylor diagrams for North Atlantic with 2000 bootstrap SLP composites (represented by the small dots). The blue, green, and red semi-circles indicate the 5th, median, 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 Sect. 3.1 for more details.
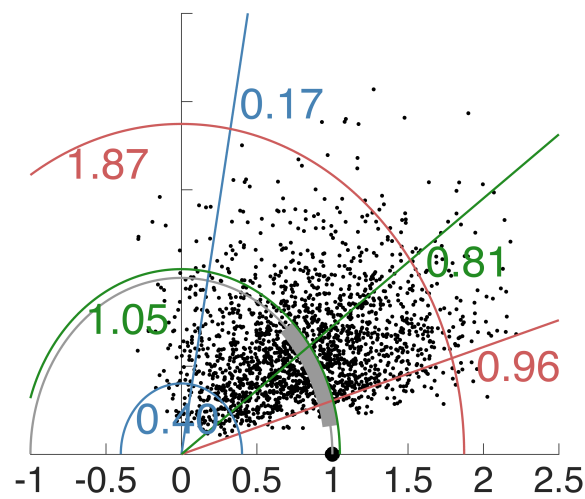
Figure 3. Taylor diagram for polar (poleward of 60°N) North Atlantic in DJF with 2000 bootstrap SLP composites, to be compared with Fig. 2a. Features in the Taylor diagram are as described in Fig. 2. See Sect. 3.1 for more details.
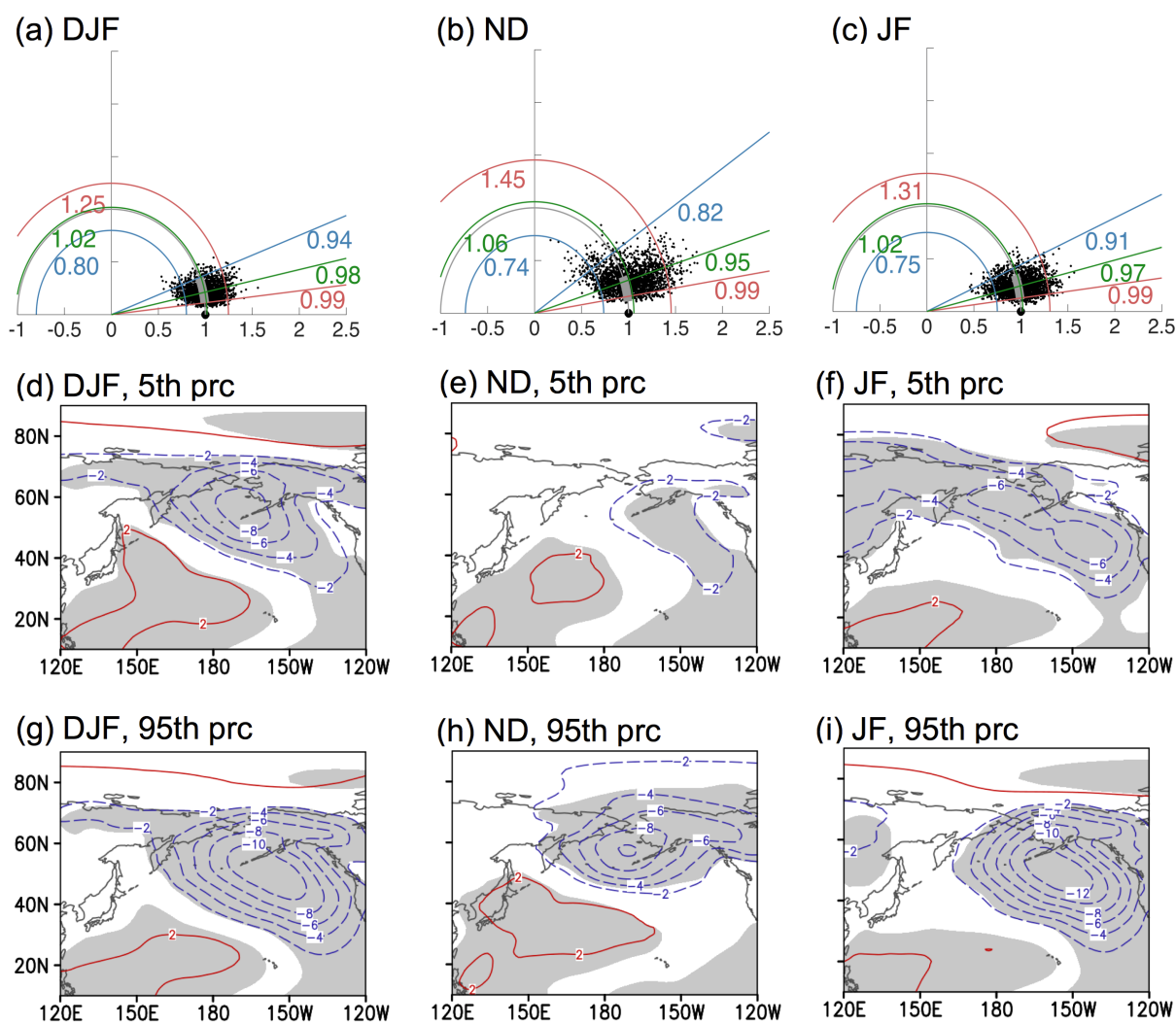
Figure 4. Same as Fig. 2 except for the North Pacific.
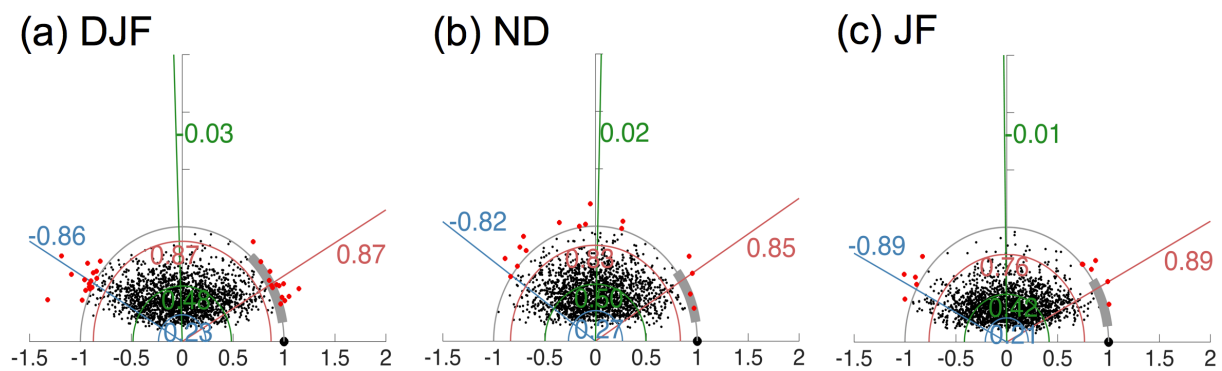
Figure 5. Taylor diagrams for permutation tests corresponding to events in Fig. A1. The thick gray arc marks the 5th-to-95th percentile range in the corresponding panel in Fig. A1. Other features in the Taylor diagrams are as described in Fig. 2. See last part of Sect. 3.1 for more details.
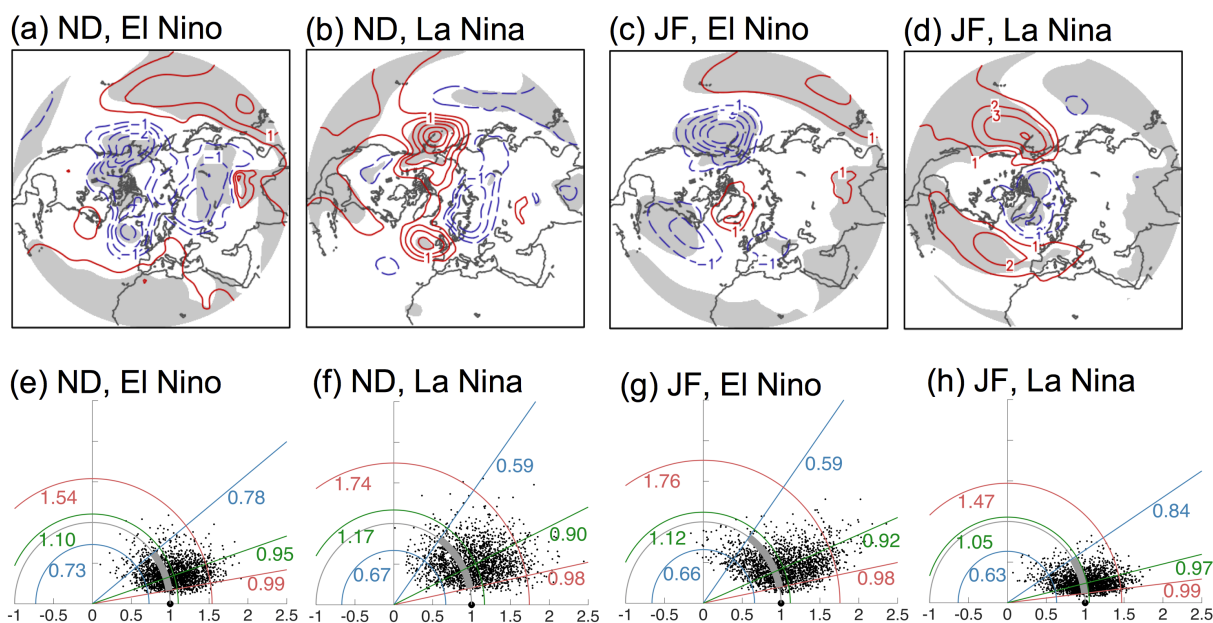
Figure 6. Uncertainty of ENSO (for El Niño or La Niña separately) teleconnection in the North Atlantic during 1870-2014 for the months shown. Top row: SLP composites for El Niño or La Niña events. Contour interval is 0.5 hPa for ND and 1 hPa for JF, with red (blue) contours indicating positive (negative) 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 Sect. 3.2 for more details.

ND, El Nino

(a) *ordt* 0.025 limit   (b) *perc* 0.025 limit   (c) *bootT* 0.025 limit

(d) *ordt* 0.975 limit   (e) *perc* 0.975 limit   (f) *bootT* 0.975 limit

ND, La Nina

(g) *ordt* 0.025 limit   (h) *perc* 0.025 limit   (i) *bootT* 0.025 limit

(j) *ordt* 0.975 limit   (k) *perc* 0.975 limit   (l) *bootT* 0.975 limit
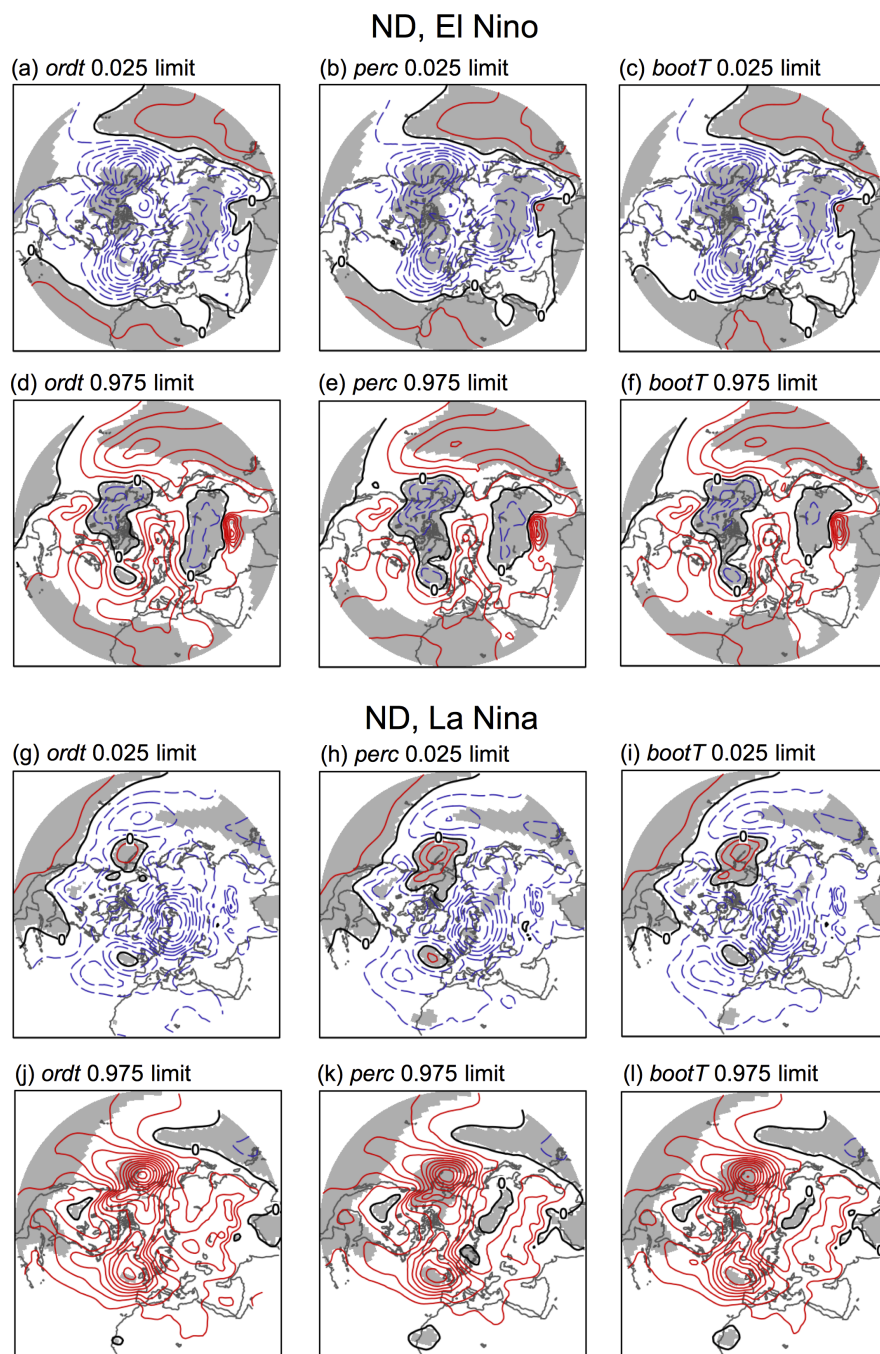
Figure 7. Lower and upper limits of confidence intervals (at 95%) for ND El Niño and La Niña SLP composites corresponding to Fig. 6a, b. The *ordt*, *perc*, and *bootT* intervals are described in Sect. 3.3. Contour interval is 0.5 hPa, with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates locations where the SLP anomalies have the same sign within the entire interval.

JF, El Nino

(a) *ordt* 0.025 limit   (b) *perc* 0.025 limit   (c) *bootT* 0.025 limit

(d) *ordt* 0.975 limit   (e) *perc* 0.975 limit   (f) *bootT* 0.975 limit

JF, La Nina

(g) *ordt* 0.025 limit   (h) *perc* 0.025 limit   (i) *bootT* 0.025 limit

(j) *ordt* 0.975 limit   (k) *perc* 0.975 limit   (l) *bootT* 0.975 limit
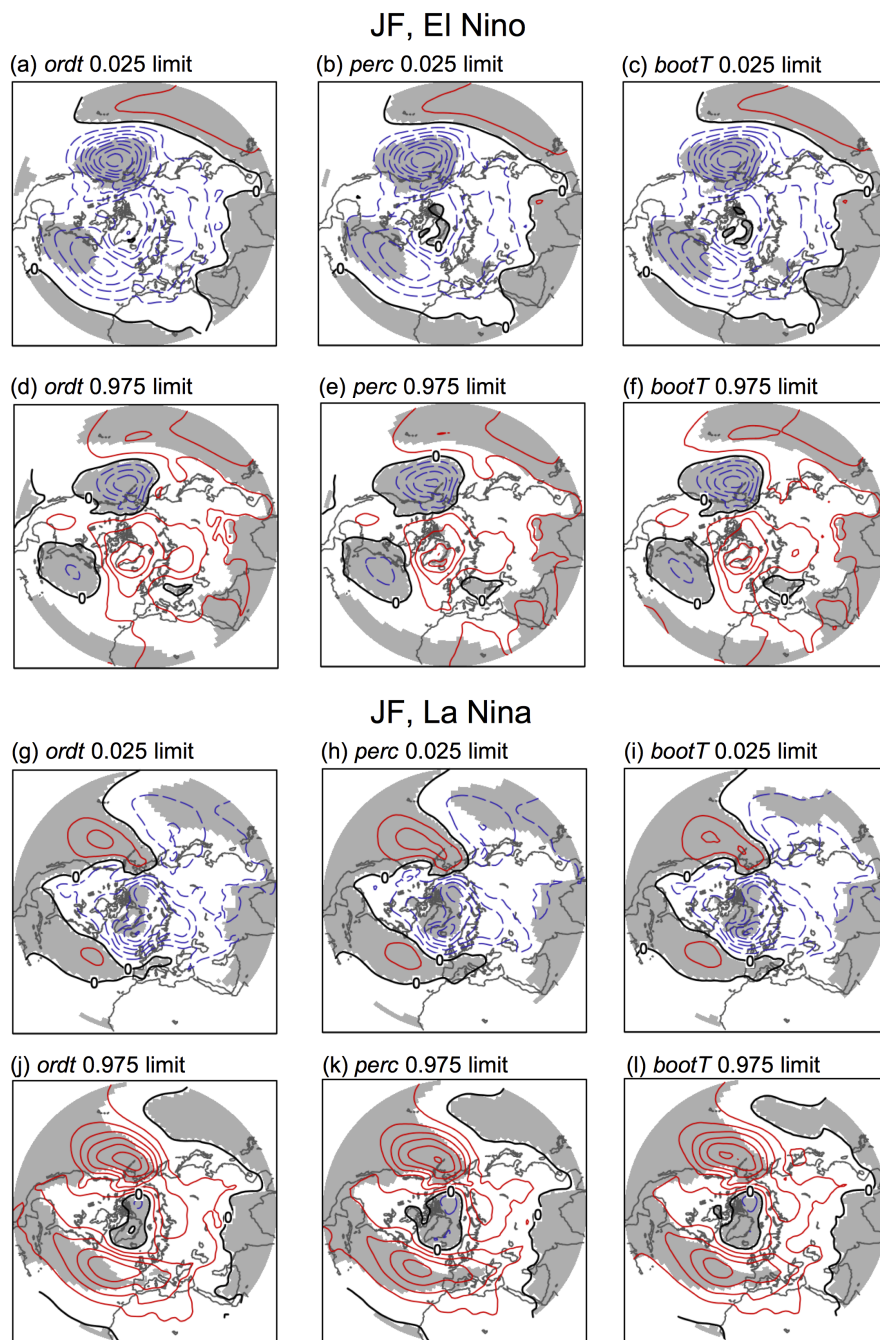
Figure 8. Lower and upper limits of confidence intervals (at 95%) for JF El Niño and La Niña SLP composites corresponding to Fig. 6c, d. The *ordt*, *perc*, and *bootT* intervals are described in Sect. 3.3. Contour interval is 1 hPa, with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates locations where the SLP anomalies have the same sign within the entire interval.
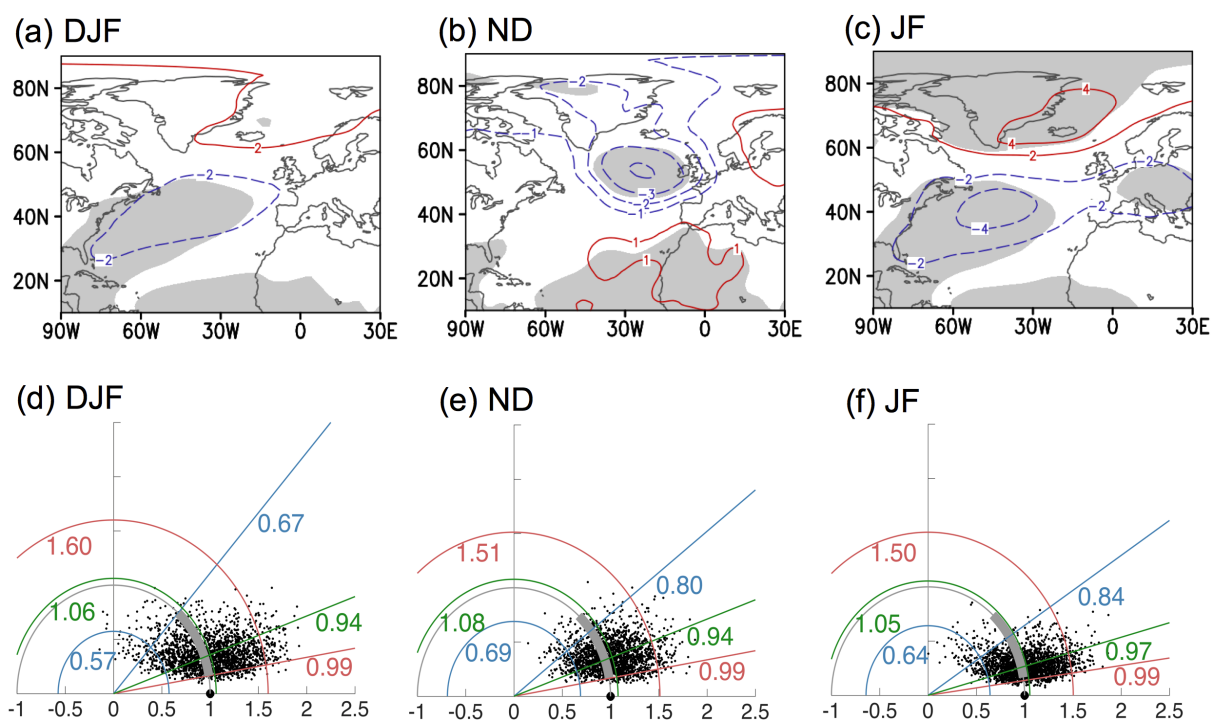
Figure A1. Similar to Figs. 1, 2 except for a longer period of 1870–2014. Top row: SLP composites for El Niño minus La Niña events. Contour interval in (a) and (c) is 2 hPa, and is 1 hPa in (b), with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates 5% significance 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 Sect. 3.1 for more details.
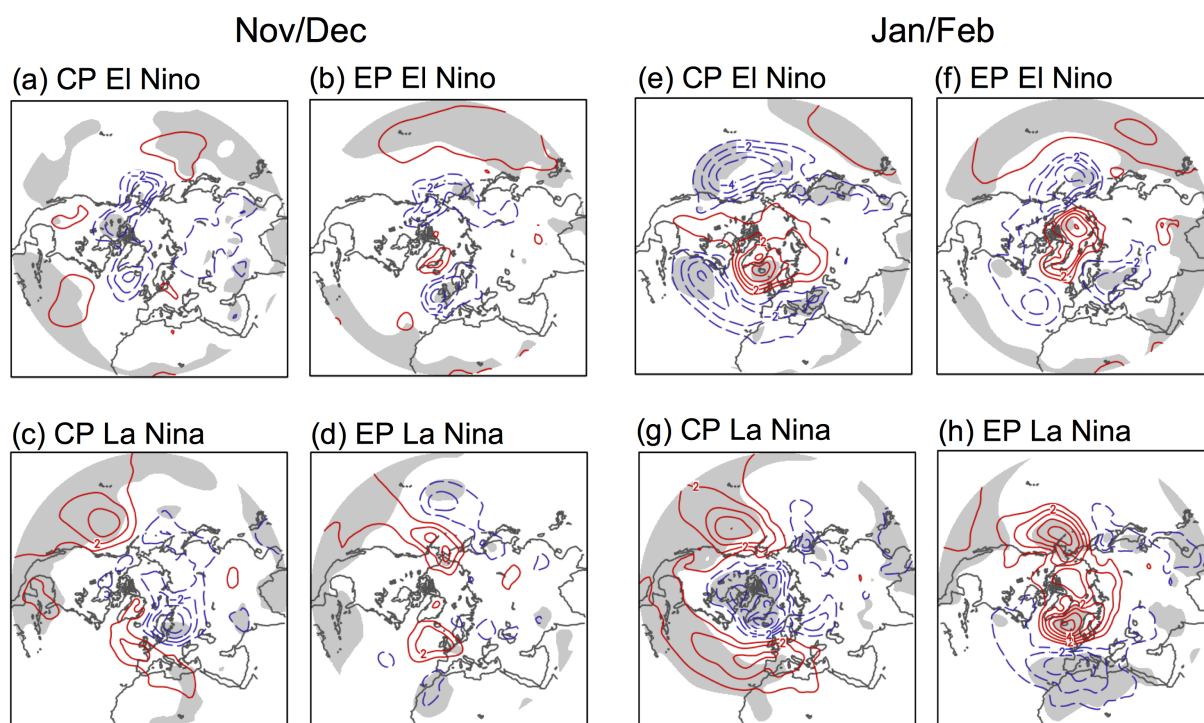
Figure A2. SLP composites for central- and eastern-Pacific El Niño, and La Niña, events in the ND or JF means. Contour interval is 1 hPa in all panels, with red (blue) contours indicating positive (negative) anomalies. Gray shading indicates 5% significance level for a two-tailed t test. See Sect. 3.2 for more details.

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

Type (number of events): Years
El Niño, NDJ or DJ, 1920–2014 (18): 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.
La Niña, NDJ or DJ, 1920–2014 (16): 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.
El Niño, ON, 1920–2014 (18): 1923, 1925, 1930, 1941, 1957, 1963, 1965, 1972, 1977, 1982, 1986, 1987, 1991, 1994, 1997, 2002, 2006, 2009.
La Niña, ON, 1920–2014 (17): 1933, 1942, 1949, 1950, 1954, 1955, 1964, 1973, 1975, 1983, 1984, 1988, 1998, 1999, 2007, 2010, 2011.
El Niño, NDJ or DJ, 1870–2014 (26): 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.
La Niña, NDJ or DJ, 1870–2014 (22): 1886/87, 1889/90, 1892/93, 1893/94, 1909/10, 1916/17, in addition to those in 1920–2014 given above.
El Niño, ON, 1870–2014 (25): 1877, 1885, 1888, 1896, 1899, 1902, 1905, 1918, in addition to those in 1920–2014 given above except 2006.
La Niña, ON, 1870–2014 (23): 1874,1886,1889,1892,1893,1909,1916, in addition to those in 1920–2014 given above except 1954.

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

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