Multi-Decadal Skill Variability in Predicting the Spatial Patterns of ENSO Events

M. J. Wright[1] [ORCID]**, A. Weisheimer[1,2,3]** [ORCID]**, and T. Woollings[1]** [ORCID]

[1]Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, UK, [2]European Centre for Medium Range Weather Forecasts, Reading, UK, [3]National Centre for Atmospheric Sciences (NCAS), Leeds, UK

Abstract Seasonal hindcasts have previously been demonstrated to show multi-decadal variability in skill across the twentieth century in indices describing El-Niño Southern Oscillation (ENSO), which drives global seasonal predictability. Here, we analyze the skill of predicting ENSO events' magnitude and spatial pattern, in the CSF-20C coupled seasonal hindcasts in 1901–2010. We find minima in the skill of predicting the first (in 1930–1950) and second (in 1940–1960) principal components of sea-surface temperature (SST) in the tropical Pacific. This minimum is also present in the spatial correlation of SSTs, in 1930–1960. The skill reduction is explained by lower ENSO magnitude and variance in 1930–1960, as well as decreased SST persistence. The SST skill minima project onto surface winds, leading to worse predictions in coupled hindcasts compared to hindcasts using prescribed SSTs. Questions remain about the offset between the first and second principal components' skill minima, and how the skill minima impact the extra-tropics.

Plain Language Summary Seasonal forecasts use slowly changing elements of the climate system to predict average weather and climate several months ahead, and have applications across the energy, agriculture, and insurance industries. The most important signal is El-Niño Southern Oscillation (ENSO), an ocean temperature pattern in the tropical Pacific which varies between warm and cold phases every 2–7 years. ENSO's specific pattern, including how warm/cold it is compared to normal, and where heating/cooling is concentrated, influences global weather. Seasonal forecasting models are tested by running 'reforecasts': comparing a 'forecast' of the past with observations. In this study, we investigate how skilfully reforecasts from a state-of-the-art seasonal forecasting model reproduce observations, for sea-surface temperatures in the tropical Pacific (where ENSO occurs). We analyze reforecast skill in every Northern Hemisphere winter from 1901 to 2010 and find periods of high skill at the start and end of the century, with reduced skill in 1930–1960. This was likely caused by a reduction in the size of ENSO events, and decreased sea-surface temperature memory. This reduction in ocean skill leads to lower skill for surface-level winds. These results are important for understanding whether season-ahead predictability changes over decades/centuries, and how climate change might impact this predictability.

1. Introduction

Skilful seasonal forecasts are of great societal and scientific interest (Shukla et al., 2000; Weisheimer & Palmer, 2014), notably for the energy, utilities, insurance, and agriculture sectors (Lorenz et al., 2021; Soares & Dessai, 2015). The ability to skilfully predict weather and climate on seasonal timescales originates from aspects of the climate system which evolve on monthly and longer timescales, such as the ocean, land surface and stratosphere (Palmer & Anderson, 1994). El-Niño-Southern Oscillation (ENSO) is the most important of these (Jin et al., 2008; Luo et al., 2008), because of its global teleconnections (e.g., Bjerknes (1969); Diaz et al. (2001); Trenberth et al. (1998); Horel and Wallace (1981)).

There is substantial diversity in ENSO events' spatial and temporal characteristics (Ashok et al., 2007; Kao & Yu, 2009; Kug et al., 2009; Okumura, 2019). El Niños, for example, can be categorized as Eastern Pacific (EP) or Central Pacific (CP), named after the location of their maximum positive SST anomalies (Capotondi et al., 2015). EP events tend to have greater SST anomalies than CP events, and they evolve differently in time and space (Capotondi et al., 2020). Causes and classification of ENSO flavors are the subject of active research (see e.g., Capotondi et al. (2020); Dieppois et al. (2021)).

It is important for seasonal forecasts to skilfully predict both an ENSO event's magnitude and spatial pattern, as both are important in determining its remote impacts (e.g., Ashok et al. (2007); Garfinkel and Hartmann (2008)).

For example, Aleutian Low deepening caused by El Niño events occurs closer to the equator in CP events than EP events (Alizadeh et al., 2022), and precipitation impacts in Japan and New Zealand are opposite for EP and CP events (Ashok et al., 2007). Some simple models, and many GCMs, simulate this diversity, implying it is part of ENSO's natural variability (Geng & Jin, 2022). The frequency of CP El Niños has increased in recent years, which is attributed to global warming (Yeh et al., 2009) or natural variability (Cai et al., 2021; Wittenberg, 2009).

Hindcasts, or reforecasts, produced by running a model with past initial conditions, are used to assess a seasonal forecast model's skill (e.g., Weisheimer et al. (2017)). Hindcasts can be compared to observations and/or reanalyses, to assess model skill. Most studies of seasonal forecast skill analyze approximately the last 30 years (e.g., C3S (2018)), failing to sample multi-decadal variability, including in ENSO.

In order to examine multi-decadal variability in seasonal forecast skill, Weisheimer et al. (2017, 2020) produced and analyzed 110 year hindcast datasets (atmosphere-only runs, ASF-20C, and coupled runs, CSF-20C). They found robust decadal variability in seasonal forecast skill in both coupled and uncoupled hindcasts, with mid-century periods of reduced skill in Niño 3.4, North Atlantic Oscillation (NAO) and Pacific-North American (PNA) indices. Since skill does not change monotonically with time, changes in skill are attributed to changes in the climate system's physical state, not just inferior initial conditions. The skill minimum is more pronounced in CSF-20C, but also present for the NAO and PNA in ASF-20C (O'Reilly et al., 2017).

Additionally, studies have looked at multi-decadal variability in ENSO, its diversity, and its predictability. Dieppois et al. (2021) showed multi-decadal variability in type of ENSO event, which may impact predictability. Lou et al. (2023) used model analogs to test ENSO hindcast skill, showing multi-decadal variations in forecast skill from 1800 to the present.

Here, we build on previous work to show that coupled seasonal forecasts exhibit multi-decadal variability in their skill at predicting the spatial pattern of SST anomalies in the ENSO region. We propose reasons for this drop in skill, and examine implications for surface wind skill, by comparing to atmosphere-only hindcasts.

2. Model Experiments, Data, and Skill Metrics

The Coupled Seasonal Forecasts of the Twentieth Century (CSF-20C) (Weisheimer & O'Reilly, 2020) reforecasts were performed with ECMWF's coupled seasonal forecasting model (Johnson et al., 2019), which includes state-of-the-art atmospheric, land surface, oceanic, and sea-ice components. The atmospheric model was fully coupled to the Nucleus for European Modeling of the Ocean (NEMO). Horizontal atmospheric resolution was T_{L255} ($\simeq$80 km), with 91 vertical levels; ocean resolution was 1°, with 42 vertical levels. Initial conditions were from CERA-20C: ECMWF's coupled twentieth century reanalysis (Laloyaux et al., 2018). CERA-20 C assimilates surface pressure and marine wind observations (no satellite data) in the atmosphere, and subsurface temperature and salinity profile observations in the ocean. SSTs are relaxed toward monthly HadISST2 reconstructed data (Rayner et al., 2003).

This study analyses hindcasts initialized on 1 November every year from 1901 to 2010. 51 ensemble members per start date were created through a combination of stochastic perturbations to atmospheric model physics, and sampling initial conditions from CERA-20C's 10 realizations. Time-varying forcings from greenhouse gases, the solar cycle, and volcanic and sulfate tropospheric aerosols were prescribed; other tropospheric aerosols were constant. See Johnson et al. (2019) for full details. The linear trend and mean-state bias (against HadISST2) was removed before this analysis. This helps offset the known cold tongue bias in the ECMWF model (Johnson et al., 2019). Some bias may remain, but this should not impact the EOFs or PCs as they are calculated using anomalies.

Data from the equivalent, uncoupled, atmosphere-only hindcasts (ASF-20C) is used for comparison. ASF-20 C used the same atmospheric model as CSF-20C, with prescribed SSTs from HadISST2. The initial conditions were from ECMWF's ERA-20C reanalysis (Poli et al., 2016).

Two SST datasets (both cropped to 1901–2010) are used for validation: HadISST2, and ERSSTv5 (Huang et al., 2017). Both assimilate global ocean observations to produce temporally and spatially consistent reconstructed SST datasets, and incorporate bias-corrected observations from floats and ship-based instruments. HadISST2 uses reduced space optimal interpolation (RSOI) to reconstruct the full SST field at 1° × 1° resolution.

ERSSTv5 uses an EOF-based approach, projecting modern EOFs onto past observations to achieve a full SST field at $2° \times 2°$ resolution. The ERA-20C reanalysis is used to validate atmospheric variables.

The Niño 3.4 Index is defined as the average SST anomaly over 5°N–5°S, 170°–120°W, and the Trans-Niño Index (TNI) as the difference between normalized SST anomalies in Niño 1and2 (0°–10°S, 90°–80°W) and Niño 4 (5°N–5°S, 160°E–150°W) regions (Trenberth & Stepaniak, 2001).

Empirical orthogonal functions (EOFs) are found by computing eigenvectors of the monthly SST anomaly covariance matrix in the Pacific basin (30°N–30°S, 150°E–80°W), and the principle components (PCs) by a projection of the data onto these eigenvectors. The eigenvalues associated with each EOF represent the proportion of the total variance attributable to that mode of variability.

Hindcast skill is evaluated using the ensemble mean anomaly correlation coefficient (ACC) and the root mean square skill score (RMSSS).

The ACC between ensemble mean forecast index $\overline{g_k}$ and observation b_k index pairs over the years $k = 1, …, P$ is calculated as

$$ACC_{temporal} = \frac{\sum_{k=1}^{P} \overline{g_k'} * b_k'}{\sqrt{\sum_{k=1}^{P} \overline{g_k'^2} * \sum_{k=1}^{P} b_k'^2}} \tag{1}$$

using the ensemble mean forecast anomaly $\overline{g_k'} = \overline{g_k} - \frac{1}{P}\sum_{k=1}^{P}\overline{g_k}$, and the observed anomaly $b_k' = b_k - \frac{1}{P}\sum_{k=1}^{P} b_k$.

The spatial ACC between ensemble mean forecast $\overline{f_{ij}}$ and observation a_{ij} pairs over latitudes $i = 1, …, N$ and longitudes $j = 1, …, M$ is defined as

$$ACC_{spatial} = \frac{\sum_{i,j=1}^{N,M} \overline{f_{ij}'} * a_{ij}'}{\sqrt{\sum_{i,j=1}^{N,M} \overline{f_{ij}'^2} * \sum_{i,j=1}^{N,M} a_{ij}'^2}} \tag{2}$$

using the ensemble mean forecast anomaly (relative to climatology calculated over years $k = 1, …, P$) $\overline{f_{ij}'} = \overline{f_{ij}} - \frac{1}{P}\sum_{k=1}^{P}\overline{f_{ijk}}$, and the observed anomaly $a_{ij}' = a_{ij} - \frac{1}{P}\sum_{k=1}^{P} a_{ijk}$ at each gridpoint.

The spatial RMSSS is defined as $RMSSS = 1 - \sqrt{1 - MSSS}$ over latitudes $i = 1, …, N$ and longitudes $j = 1, …, M$, where

$$MSSS = 1 - \frac{\sum_{i,j=1}^{N,M} (\overline{f_{ij}} - a_{ij})^2}{\sum_{i,j=1}^{N,M} (c_{ij} - a_{ij})^2} \tag{3}$$

using, at each gridpoint, the ensemble mean forecast value $\overline{f}$, the observed value a_{ij} and the climatological value c_{ij} (the time-mean across the whole period at ij). Using the climatology across the whole period enables analysis of all years from 1901, but this may slightly increase the observed skill (Risbey et al., 2021). The year of interest has not been excluded; this would only negligibly affect the 110 year mean. The RMSSS expresses the differences between forecasts and observations compared to climatology. Higher values are better: 0 means a forecast adds no value over climatology; one is perfect.

For spatial measures of skill, 5–95% confidence intervals are calculated using the standard error of the ensemble mean. For temporal measures of skill, 5–95% confidence intervals are calculated using bootstrapping with replacement: sampling 21 years within the period 10,000 times and computing statistics on the resulting distribution.
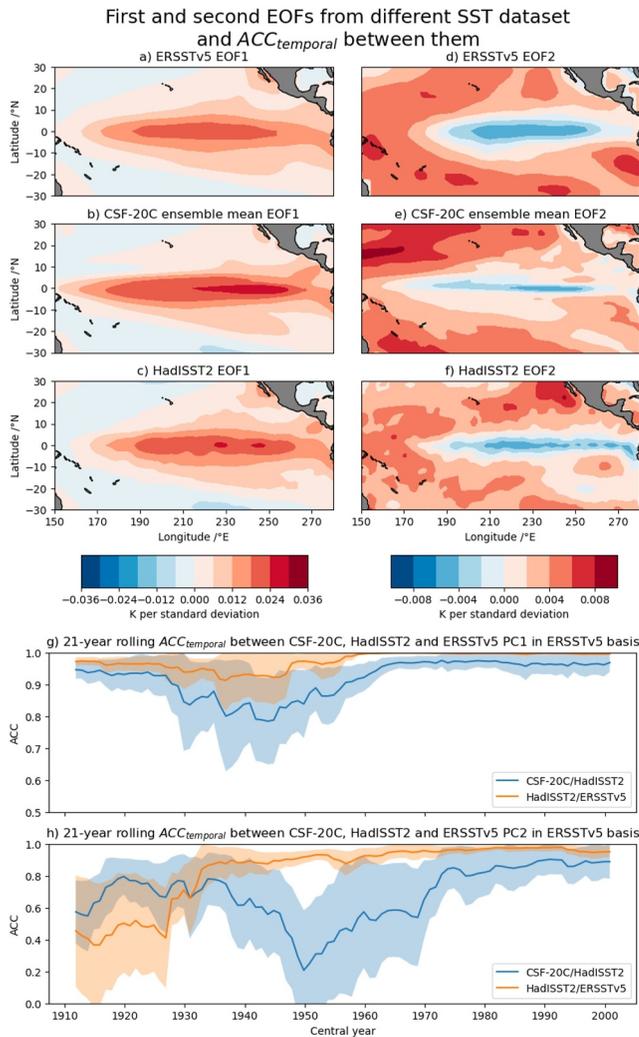
Figure 1. (a)-(e) The first and second EOFs from: (a) and (d) ERSSTv5; (b) and (e) CSF-20C ensemble mean; (c) and (f) HadISST2. (g)&(h) 21 year rolling $ACC_{temporal}$, plotted against the central year of the 21 year period, of the (g) first, and (h) second principal components. The EOFs and PCs were calculated in the domain 30°N–30°S, 150°–280°E from the DJF-mean SSTs from each data set. The PCs are calculated by projecting each data set's EOFs onto the EOFs from ERSSTv5. 5–95% confidence intervals in g&h were calculated using bootstrapping with replacement.

3. Results

3.1. First and Second Principal Components

To examine how well CSF-20C represents ENSO events' spatial pattern, DJF-mean tropical Pacific SST EOFs were calculated from 1901 to 2010. The first two EOFs from CSF-20C, HadISST2 and ERSSTv5 are shown in Figures 1a–1e.

Whilst the EOFs appear smoother in ERSSTv5 (due to coarser resolution) and CSF-20C (due to being an ensemble mean) than in HadISST2, the first EOF is similar in all three. All show warming through the Pacific cold tongue, with cooling at westerly longitudes and in the subtropics. The warming associated with the first EOF is further east in CSF-20C than HadISST2 or ERSST, and warm anomalies transition to cold anomalies further west. This is a known issue in models (e.g., Luo et al. (2005)), including SEAS5 (Beverley et al. (2023), Figure S3 in Supporting Information S1), and suggests the hindcast model may have a slight bias for EP over CP El Niño events.

There are more obvious differences in the second EOF, which can be interpreted as representing ENSO's spatial variability. The overall pattern is similar between datasets: cooling in the cold tongue, accompanied by warming in the rest of the basin. Compared to the two observational datasets, especially ERSSTv5, CSF-20C's second EOF cools over a much smaller range of latitudes, and warms more strongly in the west Pacific. Also note that the second EOFs of HadISST2 and ERSSTv5 do not entirely agree. This is likely due to worse observational coverage toward the start of the century, meaning different reconstruction techniques (see Section 2) play a larger role in determining the SST pattern.

The skill (as 21 year rolling $ACC_{temporal}$) of the coupled hindcasts at predicting the DJF-mean value of the PCs associated with the EOFs is shown in Figure 1g and 1h, alongside the rolling correlation between the observational datasets. To ease comparison, the ERSSTv5 EOF patterns were used for all, with time series of other datasets formed by projection onto these. The analysis was also performed using the CSF-20C and HadISST2 patterns (Figure S1 in Supporting Information S1), with qualitatively similar results. Note that PC1 is approximately equal to the Niño 3.4 index, and PC2 to the Trans-Niño index (TNI).

There is clear multi-decadal variability in both PC1 and PC2 skill, but PC1 is better predicted than PC2 throughout the whole century. The changes in skill are non-monotonic, suggesting they are not simply due to changes in initial condition (and reference data set) quality, but changes in the coupled atmosphere-ocean system's physical state. PC1's minimum in skill in 1930–1950 is consistent with the minimum in Niño 3.4 Index skill shown by Weisheimer et al. (2022). The correlation between the two reconstructed datasets also decreases in 1930–1950, and the uncertainty in that correlation increases.

There are also multi-decadal variations in PC2 skill, which is important for determining ENSO events' spatial pattern. There is still a drop in skill in the middle of the century, but this is not in-phase with PC1: the minimum in PC2 skill occurs in 1940–1960, lagging the PC1 minimum by 10 years. This slightly later period is coincident with minima in NAO and PNA skill shown by O'Reilly et al. (2017).

The relationship between PC2 in the two reconstructed datasets is also different to PC1. At the start of the century, PC2 in HadISST2 and ERSSTv5 are much less well correlated than PC1, but have good agreement from 1940 onwards. Unlike for PC1, there is no observational disagreement coincident with the skill minimum, suggesting the reasons for the minimum in PC2 skill are independent of observational quality.
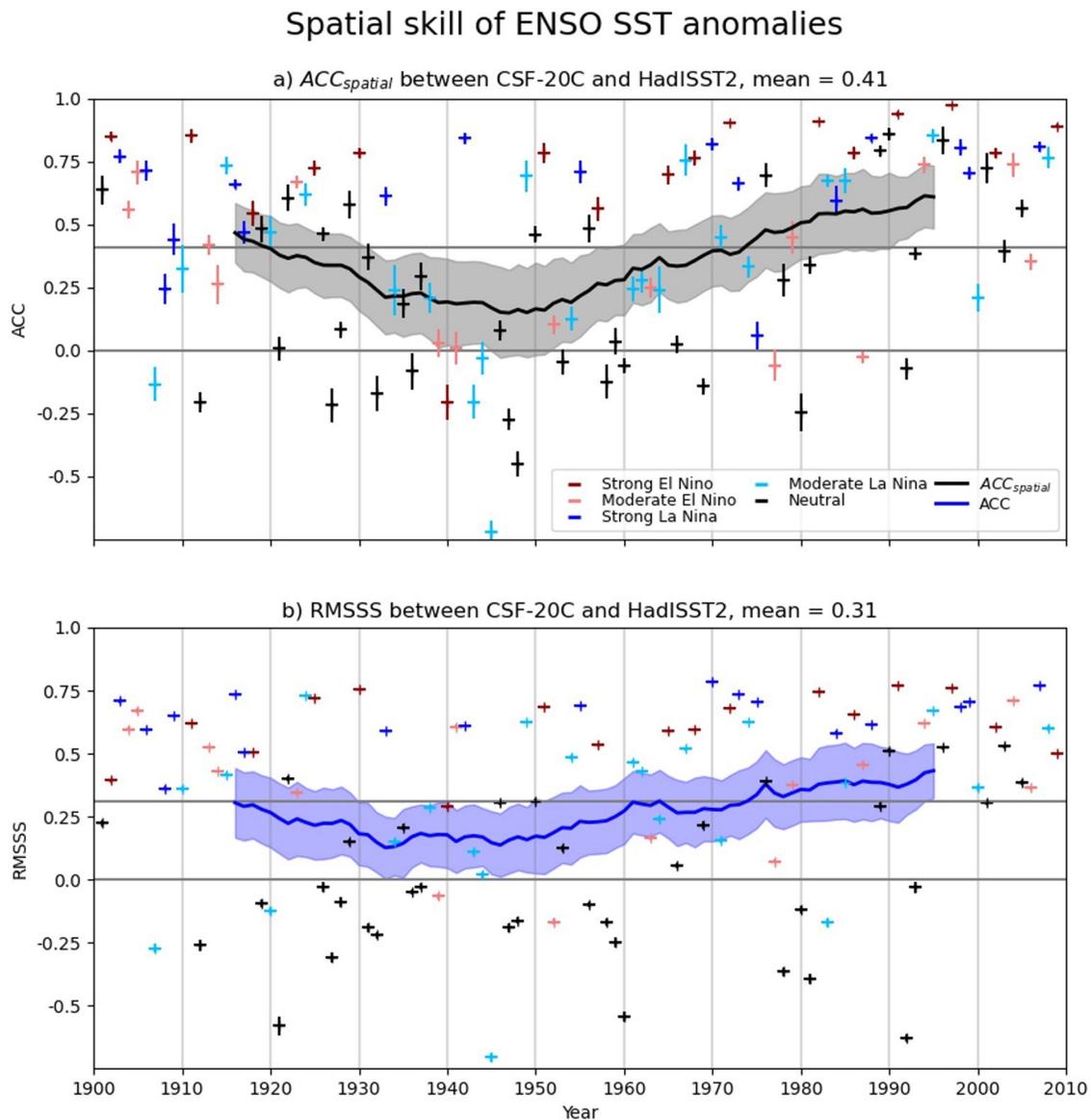
Figure 2. The (a) $ACC_{spatial}$, and (b) RMSSS between the spatial fields of DJF-mean SST anomalies in CSF-20C ensemble mean and HadISST2. The domain was 5°N–5°S, 160°–270°E: the Niño 3 and Niño 4 regions. The thick lines (black in (a); blue in (b)) show rolling 21 year mean spatial skill. 5–95% confidence intervals are calculated using bootstrapping with replacement. Individual crosses show each year's spatial skill, with error bars showing the 5–95% confidence interval within CSF-20C ensemble members. Neutral events are defined as having a DJF-mean Nino 3.4 index of between −0.5 and +0.5 K; moderate events as between ±0.5 and ±1 K; strong events as exceeding ±1 (k).

3.2. ENSO's Spatial Pattern

We also calculated the hindcast $ACC_{spatial}$ (Figure 2a, black line) and RMSSS (Figure 2b, blue line) at predicting the spatial pattern of SST anomalies in the tropical Pacific. Both measures of skill ($ACC_{spatial}$ based on correlations, RMSSS on the absolute values of hindcast anomalies) show similar behavior. The smoothed lines show an extended period of reduced skill (1930–1960), which encompasses the minima in skill of both PC1 and PC2 (Figure 1).

There is substantial variation in individual years' skill scores. Whilst some events for example, 1997./98, have high $ACC_{spatial}$ and RMSSS (approaching 'perfect' prediction of ENSO's spatial pattern), others have spatial skill scores close to zero, or negative. These low/negative scores are clustered in the middle period. They tend to be moderate or neutral ENSO years (DJF-mean Nino 3.4 between −0.5 and +0.5 K). Whilst we may expect

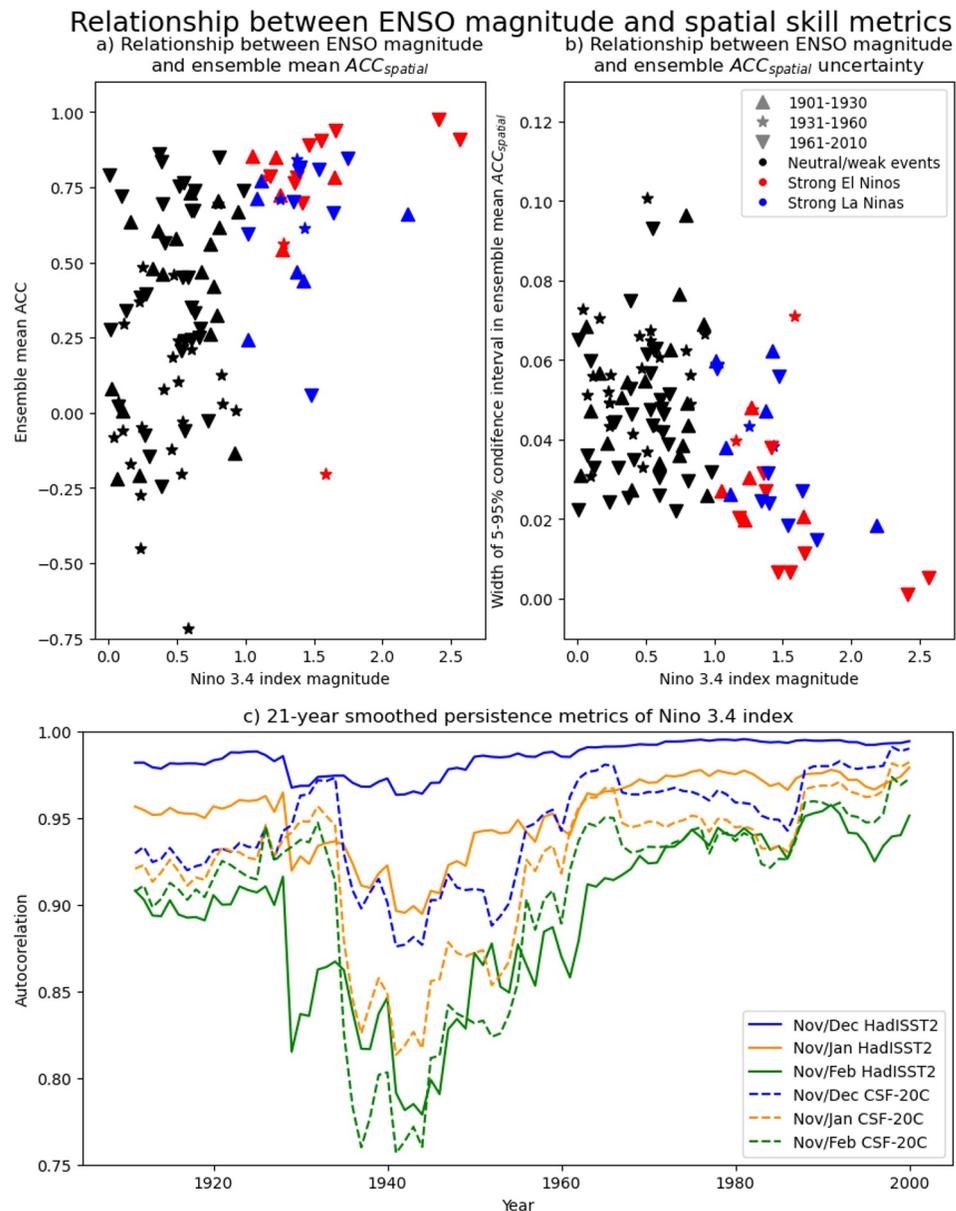## Relationship between ENSO magnitude and spatial skill metrics

Figure 3. (a)&(b) Scatter plots demonstrating the relationship between the magnitude of an ENSO event in HadISST2, as measured by the Niño 3.4 Index, and (a) ensemble mean $ACC_{spatial}$ for CSF-20C; (b) 5–95% confidence intervals for $ACC_{spatial}$ within ensemble members. Colors denote different types of ENSO events: red shows strong El Niño events (DJF-mean Niño 3.4 index $>+1$ (k); blue shows strong La Niña events <-1 (k); black shows all others). Symbols represent different periods: upwards triangles show years from 1901 to 1930; stars show 1931–1960; downwards triangles show 1961–2010. (c) Correlation between the Niño 3.4 index in November and (blue line) December (orange) January and (green) February, for (solid lines) HadISST2 and (dashed lines) CSF-20C. Correlations are taken in rolling 21 year periods, and the value plotted against the central year.

$ACC_{spatial}$ to be very sensitive in moderate/neutral years, the fact that the RMSSS has the same results shows that these years are, indeed, being more poorly predicted than years in the early and late periods. Uncertainties also increase significantly in the mid-century, shown by a two-sided student's t-test, particularly for the strongest El Niño and La Niña events.

Figures 3a and 3b show a strong, significant positive relationship between the magnitude of an ENSO event and its $ACC_{spatial}$, and a strong negative relationship between the magnitude and the uncertainty in $ACC_{spatial}$ (i.e., strong ENSO events have high skill and low ensemble spread). The relationship is consistent for strong events, but

weaker ENSO events show a larger range of $ACC_{spatial}$ values and uncertainties. This analysis was repeated for the Niño 1and2 index and the TNI, but no strong relationship was found (Figures S2 and S3 in Supporting Information S1).

This relationship between ENSO magnitude and $ACC_{spatial}$ suggests that ENSO variability plays a role in seasonal hindcast skill of ENSO's spatial pattern. ENSO variability is reduced in the middle of the century, and this contributes to the minimum in spatial skill. The standard deviation in the DJF-mean Niño 3.4 Index in HadISST2 decreased from 1.0 K in 1910–1930 to 0.8 K in 1940–1960, before rising again to 1.1 K in 1990–2010. This was tested further by taking 10,000 random 30-year samples from the whole data set: this showed a strong, statistically significant relationship between ENSO variance and $ACC_{spatial}$ (not shown).

As shown by the location of the stars in Figures 3a and 3b, the mid-century reduction in ENSO variability led to lower spatial skill, and larger uncertainties. However, lower variability does not explain the whole picture: even the neutral and weak ENSO events, shown in black in Figure 3, have a significantly lower $ACC_{spatial}$ in mid-century compared to the early and late periods, as tested by a two-sided student's t-test (not shown).

This suggests there are other factors contributing to the mid-century minimum in skill. One factor is likely the reduction in SST persistence in the middle of the century, shown in Figure 3c. The autocorrelation of November SSTs with December, January and February SSTs has a minimum in 1930–1960. This reduction in persistence is also present in CSF-20C, and is observed for the TNI too (Figure S3 in Supporting Information S1). For the Niño 1and2 index (Figure S2 in Supporting Information S1), whilst there is a reduction in persistence in the observations in 1940–1970, this is not captured by the hindcast model; this likely reduced skill even further. We propose that this change in behavior, with less persistence of SSTs throughout the season (which may be linked to lower inter-annual variability), contributes to the mid-century skill drop.

3.3. Wind Anomaly Skill

ENSO interacts with the tropical atmosphere through surface wind anomalies. When the east-west SST gradient across the tropical Pacific is strengthened (La Niña events), this leads to easterly wind anomalies. During El Niño events, when the SST gradient is weakened, westerly wind anomalies are observed. Figure 4 shows a Hövmoller plot of the 21 year smoothed meridionally averaged (5°N–5°S) zonal surface wind anomalies from CSF-20C, ASF-20C, and ERA-20C, along with hindcast skill.

Whilst the magnitude of the anomalies are much smaller (as expected from an ensemble mean), ASF-20C (Figure 4b) predicts the smoothed, decadally varying zonal wind anomalies very well compared to ERA-20C (Figure 4c) throughout the whole century. This high skill is clearly visible in Figure 4e: 21 year smoothed $ACC_{temporal}$ is very high throughout the whole century, especially in the central Pacific, and always statistically significant. This is perhaps expected, as ASF-20C uses prescribed SSTs to force the atmosphere. However, other atmospheric signals, including the PNA and NAO, show a minimum in skill in ASF-20C in the middle of the century, despite prescribed SSTs (Weisheimer et al., 2017).

CSF-20C's surface wind anomalies (Figure 4a) show much worse agreement with ERA-20C. At the start of the century agreement is high, before there is a mid-century period of persistent, strong easterly anomalies which is not present in the reanalysis. This leads to multi-decadal variations in the surface wind $ACC_{temporal}$ (Figure 4d). In the central Pacific – the region best predicted in ASF-20C – there is a period in 1940–1960 of reduced skill, which coincides with the skill minimum in SST PC2. There is also a band around 230°E where the surface wind anomalies are almost always badly predicted. This could be due to differences in the location of the maximum/minimum SST anomalies, compared to ERA-20C. As with the SST metrics, there is a strong relationship between the magnitude of an ENSO event and the skill at predicting its spatial pattern in CSF-20C (Figure S4 in Supporting Information S1). This is also true in ASF-20C, even though skill is much higher overall (Figure S5 in Supporting Information S1). We conclude that lower SST spatial skill in the middle of the century leads to worse predictions of surface wind anomalies in the tropical Pacific, compared to hindcasts with prescribed SSTs.

4. Discussion and Conclusions

Section 3 shows multi-decadal variations in the skill of seasonal hindcasts at predicting ENSO events and their spatial pattern throughout the twentieth century. There are two distinct low skill periods: in 1930–1950 for PC1 (Niño 3.4 Index), and 1940–1960 for PC2 (Trans-Niño Index). The PC1 low-skill period is consistent with the
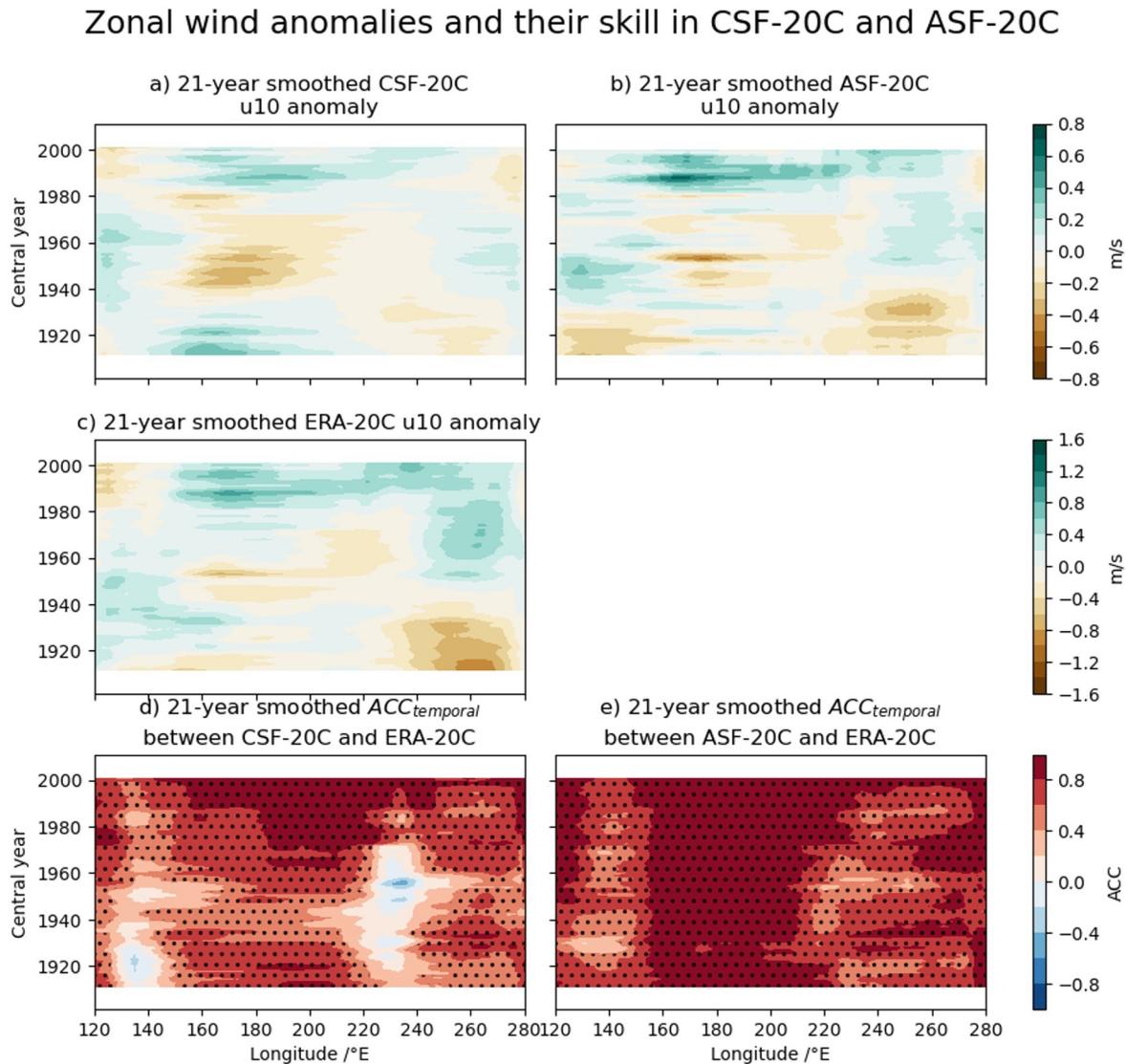
Zonal wind anomalies and their skill in CSF-20C and ASF-20C

Figure 4. 21 year smoothed DJF-mean anomaly of the meridionally averaged (5°N–5°S) zonal wind at 10 m height: (a) CSF-20C ensemble mean; (b) ASF-20C ensemble mean; (c) ERA-20C reanalysis. The rolling 21 year $ACC_{temporal}$ between the hindcasts and reanalysis is shown in (d) for CSF-20C, and (e) for ASF-20C. Values are plotted at the central year of the 21 year period. Stippling in (d) and (e) shows significance at the 5% level. Subplots (a) and (b) share a color bar; note this is different to the color bar for (c).

Niño 3.4 Index skill minimum shown by Weisheimer et al. (2020, 2022). The PC2 low-skill period is consistent with the PNA and NAO skill minima shown by O'Reilly et al. (2017). There is a prolonged minimum (1930–1960) in the spatial skill of predicting tropical Pacific SST anomalies.

The skill minima for SST indices are related to the agreement between different reconstructed SST datasets. The correlation between PC1 in HadISST2 and ERSSTv5 also has a minimum in the middle of the century, which coincides with the skill minimum. For PC2, there is no period of observational disagreement coincident with the minimum, suggesting it is caused by other factors.

The variation in correlation between the two reconstructed datasets' PC1s adds evidence that mid-century seasonal prediction is hampered by lower quality and coverage of initialization and observational data. The different reconstruction methods used by HadISST2 and ERSSTv5 may contribute to their differences in this period when they are less constrained by sparser observations. The EOF approach of ERSSTv5 assumes stationarity of the SST EOF patterns, which may not be true throughout the whole century, and the interpolation used by HadISST2 is

less accurate with fewer observations. Sparser observations may also contribute to the increased ensemble spread in the mid-century, with seasonal hindcasts less constrained by initial conditions.

Inferior quality initialization data does not explain the full story: at the start of the century - when there are just as few observations - the skill in PC1, PC2 and spatial pattern is very high, comparable to recent years. Further, whilst the PC1 minimum in skill coincides with observational disagreement, the PC2 minimum does not. Other factors contribute to the mid-century skill minimum. The hindcast skill at predicting the spatial pattern of SST anomalies is strongly related to the magnitude of ENSO events that year. During strong ENSO events, whether El Niño or La Niña, the hindcasts accurately predict the spatial pattern of SST anomalies in the tropical Pacific, with a lower ensemble spread. During weak or neutral events, the skill is lower, with higher spread. This is perhaps expected, as weak/neutral initial condition anomalies can lead to El Niños, La Niñas, or neither; so small perturbations can dramatically alter the DJF-mean outcome. If initial conditions have strong anomalies, the signal-to-noise ratio is larger, leading to a more predictable system. In the mid-century period, ENSO events have a lower magnitude, which contributes to the skill minima.

Multi-decadal variation in the persistence of SSTs may also contribute to the observed variations in skill and explain why weak/neutral ENSO events in 1930–1960 are still worse predicted than weak/neutral events in the rest of the century. The reduction in autocorrelation of November SSTs with December-February SSTs in the reconstructed datasets in 1930–1960 signifies a reduction in predictability, which contributes to low hindcast skill. The skill at predicting the atmospheric component of ENSO (here surface winds) was shown to also be impacted in 1930–1960 by this skill minimum, with coupled hindcasts performing much worse than atmosphere-only equivalents.

ENSO influences the extra-tropics through remote teleconnections to the NAO (Hurrell & Deser, 2009) and PNA (Horel & Wallace, 1981). Whilst they have high skill throughout the century in the surface wind anomalies, ASF-20C does show a skill minimum in 1940–1960 for the PNA and the NAO (O'Reilly et al., 2017; Weisheimer et al., 2017). These indices do not have lower variability than average in 1940–1960. Instead, O'Reilly et al. (2017) show that real-world teleconnections weakened during this period, whilst model-world teleconnections stay constant. The impact of ENSO skill variations on extra-tropical skill is another aspect that warrants further investigation, particularly comparing responses in coupled and uncoupled hindcasts.

There is still work to do examining and interpreting multi-decadal variations in seasonal hindcast skill, to fully quantify how much of the skill minimum is caused by inferior initialization data, and how much by reduced predictability in the middle of the century. It is also important to understand why different signals show different periods of reduced skill (e.g., why the PC1 skill minimum leads the PC2 skill minimum by 10 years). Other suggestions for multi-decadal variability in hindcast skill include poor representation of air-sea coupling in seasonal hindcast models (Yao et al., 2022), which was important in the real world in the middle of the century.

In the future, we intend to examine further causes of the skill minima. For example, tropospheric aerosol forcings vary on multi-decadal timescales, so we will quantify the effects of changing this external forcing on seasonal hindcasts and their skill. We would also like to extend this analysis to longer lead times and May initialisations, to further understand variability in spatial skill.

Data Availability Statement

Data from ASF-20C and CSF-20C have become publicly available through a dedicated online dissemination platform hosted by the CEDA archive at https://catalogue.ceda.ac.uk/uuid/6e1c3df49f644a0f812818080-bed5e45. A set of standard monthly mean atmospheric variables including temperature, precipitation, mean sea level pressure, geopotential height, wind, and thermal and radiative fluxes have been provided as global gridded data in netCDF format.

The HadISST2 reconstructed SST data set used in this study was taken from the ERA-20C data set, which uses this SST data set to force it. The ERA-20C data set is available on the ECMWF MARS data store at https://apps.ecmwf.int/mars-catalogue/?class=e2. The catalogue is public but to download data, access must be requested from ECMWF.

The ERSSTv5 reconstructed SST data is available for public download on the NOAA website at https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html.

The CERA-20C reanalysis, used for initial conditions for the hindcasts, is available on the ECMWF MARS data store at https://apps.ecmwf.int/mars-catalogue/?class=ep. The catalogue is public but to download data, access must be requested from ECMWF.

References

Alizadeh, O., Qadimi, M., Zolghadrshojaee, M., & Irannejad, P. (2022). Frequency of different types of El Niño events under global warming. *International Journal of Climatology*, 42(16), 9697–9709. https://doi.org/10.1002/joc.7858

Ashok, K., Behera, S. K., Rao, S. A., Weng, H., & Yamagata, T. (2007). El Niño modoki and its possible teleconnection. *Journal of Geophysical Research*, 112(C11). https://doi.org/10.1029/2006JC003798

Beverley, J. D., Newman, M., & Hoell, A. (2023). Rapid development of systematic enso-related seasonal forecast errors. *Geophysical Research Letters*, 50(10), e2022GL102249. https://doi.org/10.1029/2022GL102249

Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Reyiew*, 97(3), 163–172. https://doi.org/10.1175/1520-0493(1969)097⟨0163:ATFTEP⟩2.3.CO;2

C3S. (2018). Copernicus climate change service hindcasts.

Cai, W., Santoso, A., Collins, M., Dewitte, B., Karamperidou, C., Kug, J. S., et al. (2021). Changing El niño–southern oscillation in a warming climate. *Nature Reviews Earth & Environment*, 2(9), 628–644. https://doi.org/10.1038/s43017-021-00199-z

Capotondi, A., Wittenberg, A. T., Kug, J.-S., Takahashi, K., & McPhaden, M. J. (2020). ENSO diversity. In *El Niño southern oscillation in a changing climate* (pp. 65–86). American Geophysical Union (AGU). https://doi.org/10.1002/9781119548164.ch4

Capotondi, A., Wittenberg, A. T., Newman, M., Lorenzo, E. D., Yu, J. Y., Braconnot, P., et al. (2015). Understanding ENSO diversity. *Bulletin of the American Meteorological Society*, 96(6), 921–938. https://doi.org/10.1175/BAMS-D-13-00117.1

Diaz, H. F., Hoerling, M. P., & Eischeid, J. K. (2001). ENSO variability, teleconnections and climate change. *International Journal of Climatology*, 21(15), 1845–1862. https://doi.org/10.1002/joc.631

Dieppois, B., Capotondi, A., Pohl, B., Chun, K. P., Monerie, P.-A., & Eden, J. (2021). ENSO diversity shows robust decadal variations that must be captured for accurate future projections. *Communications Earth & Environment 2021*, 2(1, 2), 1–13. https://doi.org/10.1038/s43247-021-00285-6

Garfinkel, C. I., & Hartmann, D. L. (2008). Different ENSO teleconnections and their effects on the stratospheric polar vortex. *Journal of Geophysical Research*, 113(D18), D18114. https://doi.org/10.1029/2008JD009920

Geng, L., & Jin, F. F. (2022). ENSO diversity simulated in a revised cane-zebiak model. *Frontiers in Earth Science*, 10. https://doi.org/10.3389/feart.2022.899323

Horel, J. D., & Wallace, J. M. (1981). Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Monthly Weather Review*, 109(4), 813–829. https://doi.org/10.1175/1520-0493(1981)109⟨0813:PSAPAW⟩2.0.CO;2

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended reconstructed sea surface temperature, version 5 (ersstv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, 30(20), 8179–8205. https://doi.org/10.1175/JCLI-D-16-0836.1

Hurrell, J. W., & Deser, C. (2009). North atlantic climate variability: The role of the North atlantic oscillation. *Journal of Marine Systems*, 78(1), 28–41. https://doi.org/10.1016/j.jmarsys.2008.11.026

Jin, E. K., Kinter, J. L., Wang, B., Park, C. K., Kang, I. S., Kirtman, B. P., et al. (2008). Current status of ENSO prediction skill in coupled ocean-atmosphere models. *Climate Dynamics*, 31(6), 647–664. https://doi.org/10.1007/s00382-008-0397-3

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2019). Seas5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3), 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019

Kao, H.-Y., & Yu, J.-Y. (2009). Contrasting eastern-pacific and central-pacific types of ENSO. *Journal of Climate*, 22(3), 615–632. https://doi.org/10.1175/2008JCLI2309.1

Kug, J.-S., Sooraj, K. P., Kim, D., Kang, I.-S., Jin, F.-F., Takayabu, Y. N., & Kimoto, M. (2009). Simulation of state-dependent high-frequency atmospheric variability associated with ENSO. *Climate Dynamics*, 32(5), 635–648. https://doi.org/10.1007/s00382-008-0434-2

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J. R., Broennimann, S., Buizza, R., et al. (2018). CERA-20C: A coupled reanalysis of the twentieth century. *Journal of Advances in Modeling Earth Systems*, 10(5), 1172–1195. https://doi.org/10.1029/2018MS001273

Lorenz, C., Portele, T. C., Laux, P., & Kunstmann, H. (2021). Bias-corrected and spatially disaggregated seasonal forecasts: A long-term reference forecast product for the water sector in semi-arid regions. *Earth System Science Data*, 13(6), 2701–2722. https://doi.org/10.5194/essd-13-2701-2021

Lou, J., Newman, M., & Hoell, A. (2023). Multi-decadal variation of ENSO forecast skill since the late 1800s. *npj Climate and Atmospheric Science*, 6(1), 89. https://doi.org/10.1038/s41612-023-00417-z

Luo, J.-J., Masson, S., Behera, S., Shingu, S., & Yamagata, T. (2005). Seasonal climate predictability in a coupled oagcm using a different approach for ensemble forecasts. *Journal of Climate*, 18(21), 4474–4497. https://doi.org/10.1175/JCLI3526.1

Luo, J. J., Masson, S., Behera, S. K., & Yamagata, T. (2008). Extended ENSO predictions using a fully coupled ocean-atmosphere model. *Journal of Climate*, 21(1), 84–93. https://doi.org/10.1175/2007JCLI1412.1

Okumura, Y. M. (2019). ENSO diversity from an atmospheric perspective. *Current Climate Change Reports*, 5(3), 245–257. https://doi.org/10.1007/s40641-019-00138-7

O'Reilly, C. H., Heatley, J., MacLeod, D., Weisheimer, A., Palmer, T. N., Schaller, N., & Woollings, T. (2017). Variability in seasonal forecast skill of Northern Hemisphere winters over the twentieth century. *Geophysical Research Letters*, 44(11), 5729–5738. https://doi.org/10.1002/2017GL073736

Palmer, T. N., & Anderson, D. L. (1994). The prospects for seasonal forecasting—A review paper. *Quarterly Journal of the Royal Meteorological Society*, 120(518), 755–793. https://doi.org/10.1002/qj.49712051802

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., et al. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, 29(11), 4083–4097. (Reanalysis dataset description). https://doi.org/10.1175/JCLI-D-15-0556.1

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002jd002670

Risbey, J. S., Squire, D. T., Black, A. S., DelSole, T., Lepore, C., Matear, R. J., et al. (2021). Standard assessments of climate forecast skill can be misleading. *Nature Communications*, *12*(1), 4346. https://doi.org/10.1038/s41467-021-23771-z

Shukla, J., Anderson, J., Baumhefner, D., Brankovic, C., Chang, Y., Kalnay, E., et al. (2000). Dynamical seasonal prediction. *Bulletin of the American Meteorological Society*, *81*(11), 2593–2606. https://doi.org/10.1175/1520-0477(2000)081⟨2593:DSP⟩2.3.CO;2

Soares, M. B., & Dessai, S. (2015). Exploring the use of seasonal climate forecasts in europe through expert elicitation. *Climate Risk Management*, *10*, 8–16. https://doi.org/10.1016/j.crm.2015.07.001

Trenberth, K. E., Branstator, G. W., Karoly, D., Kumar, A., Lau, N. C., & Ropelewski, C. (1998). Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research*, *103*(C7), 14291–14324. https://doi.org/10.1029/97jc01444

Trenberth, K. E., & Stepaniak, D. P. (2001). Indices of El Niño evolution. *Journal of Climate*, *14*(8), 1697–1701. https://doi.org/10.1175/1520-0442(2001)014⟨1697:LIOENO⟩2.0.CO;2

Weisheimer, A., Balmaseda, M. A., Stockdale, T. N., Mayer, M., Sharmila, S., Hendon, H., & Alves, O. (2022). Variability of ENSO forecast skill in 2-year global reforecasts over the 20th century. *Geophysical Research Letters*, *49*(10), e2022GL097885. https://doi.org/10.1029/2022GL097885

Weisheimer, A., Befort, D. J., MacLeod, D., Palmer, T., O'Reilly, C., & Strømmen, K. (2020). Seasonal forecasts of the twentieth century. *Bulletin of the American Meteorological Society*, *101*(8), E1413–E1426. https://doi.org/10.1175/BAMS-D-19-0019.1

Weisheimer, A., & O'Reilly, C. (2020). Initialised seasonal forecast of the 20th Century. *Centre for Environmental Data Analysis*. Retrieved from https://catalogue.ceda.ac.uk/uuid/6e1c3df49f644a0f812818080bed5e45

Weisheimer, A., & Palmer, T. N. (2014). On the reliability of seasonal climate forecasts. *Journal of The Royal Society Interface*, *11*(96), 20131162. https://doi.org/10.1098/rsif.2013.1162

Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A., & Palmer, T. (2017). Atmospheric seasonal forecasts of the twentieth century: Multidecadal variability in predictive skill of the winter North Atlantic oscillation (NAO) and their potential value for extreme event attribution. *Quarterly Journal of the Royal Meteorological Society*, *143*(703), 917–926. https://doi.org/10.1002/qj.2976

Wittenberg, A. T. (2009). Are historical records sufficient to constrain ENSO simulations? *Geophysical Research Letters*, *36*(12), L12702. https://doi.org/10.1029/2009GL038710

Yao, W., Yan, X., Tang, Y., Yang, D., Tan, X., Song, X., & Liu, T. (2022). Multidecadal variation in the seasonal predictability of winter pna and its sources. *Geophysical Research Letters*, *49*(21), e2022GL099393. (e2022GL099393 2022GL099393. https://doi.org/10.1029/2022GL099393

Yeh, S. W., Kug, J. S., Dewitte, B., Kwon, M. H., Kirtman, B. P., & Jin, F. F. (2009). El Nĩno in a changing climate. *Nature*, *461*(7263), 511–514. https://doi.org/10.1038/nature08316