



Estimating Probabilities of Extreme ENSO Events from Copernicus Seasonal Hindcasts

William J. Merryfield¹ · Woo-Sung Lee¹

Published online: 4 May 2023
© Crown 2023

Abstract

Multi-system seasonal hindcasts supporting operational seasonal forecasts of the Copernicus Climate Change Service (C3S) are examined to estimate probabilities that El Niño and La Niña episodes more extreme than any in the reliable observational record could occur in the current climate. With 184 total ensemble members initialized each month from 1993 to 2016, this dataset greatly multiplies the realizations of ENSO variability during this period beyond the single observed realization, potentially enabling a detailed assessment of the chances of extreme ENSO events. The validity of such an assessment is predicated on model fidelity, which is examined through two-sample Cramér–von Mises tests. These do not detect differences between observed and modeled distributions of the Niño 3.4 index once multiplicative adjustments are applied to the latter to match the observed variance, although differences too small to be detected cannot be excluded. Statistics of variance-adjusted hindcast Niño 3.4 values imply that El Niño and La Niña extremes exceeding any that have been instrumentally observed would be expected to occur with a > 3% chance per year on average across multiple realizations of the hindcast period. This estimation could also apply over the next several decades, provided ENSO variability remains statistically similar to the hindcast period.

Keywords El Niño-southern oscillation · Climate extremes · Seasonal forecasts · Teleconnections

1 Introduction

El Niño and La Niña events in the tropical Pacific, reflecting opposite phases of El Niño–Southern Oscillation (ENSO), can cause large and often adverse climate impacts over much of the globe (Taschetto et al. 2020). These impacts differ among events, both because the ENSO sea surface temperature (SST) anomalies that drive them have a diverse range of magnitudes and patterns (Capotondi et al. 2020), and because atmospheric internal variability imparts randomness to the response (Deser et al. 2017, 2018; Singh et al. 2018). ENSO-driven climate anomalies therefore should be viewed in terms of probabilities

(Mason and Goddard 2001; Davey et al. 2014), rather than as deterministic certainties. Nonetheless, ENSO impacts tend on average to scale with the intensity of El Niño and La Niña events, both in the tropics (Santoso et al. 2017) and extratropics (Hoell et al. 2016; Kumar and Chen 2020).

Direct information about the strongest recorded ENSO events is provided by instrumental observations. Gridded SST products from which ENSO metrics such as the Niño 3.4 index (defined as mean SST anomaly in the region 5°N–5°S, 120°–170°W) can be computed extend back through the second half of the nineteenth century (Rayner et al. 2003; Huang et al. 2017). Although the quality of these products diminishes in the earlier periods for which fewer and less accurate measurements are available, they indicate that the strongest instrumentally recorded El Niño events have occurred in recent decades, most notably in 1982–83, 1997–98 and 2015–16 (Fig. 1), and that no comparable events have been recorded except possibly in the late nineteenth century (Rayner et al. 2003; Chen et al. 2004). Strong La Niña events have also been seen in recent decades (Trenberth 2020).

While skillful ENSO forecasts can provide advance warnings about when large El Niño and La Niña events will occur

Responsible Editor: Yoo-Geun Ham.

Comments from two anonymous reviewers helped to improve the manuscript.

✉ William J. Merryfield
bill.merryfield@ec.gc.ca

¹ Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, 2472 Arbutus Rd., Victoria, British Columbia V8N 1V8, Canada

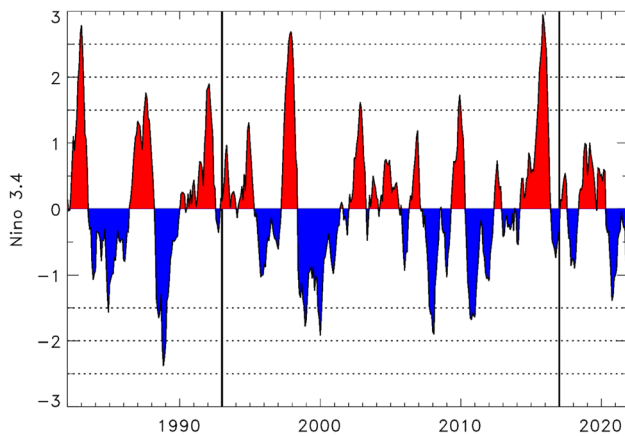


Fig. 1 Time series of monthly Niño 3.4 index in °C from the OIS-STv2 observational analysis, spanning 1982–2021. The horizontal dashed lines indicate thresholds of ± 1.5 , 2.0 and 2.5 for extrema of El Niño (red) and La Niña (blue) events, and the vertical solid lines the 1993–2016 C3S hindcast period that is analysed here and serves as a climatological base period

(L’Heureux et al. 2020), a similarly important matter from the perspective of preparedness and climate change adaptation is what are the largest El Niño and La Niña events that *can* occur. Given the limited instrumental record and inherent uncertainties in pre-instrumental (paleo) determinations (Emile-Geay et al. 2020), applying global climate models to this question would appear to be a logical approach because of the sizeable sample of events represented in the many available historical simulations and climate change projections, including large ensembles from a range of models (Deser et al. 2020). However, climate model simulations continue to exhibit a range of biases and imperfections in their representations of ENSO (Planton et al. 2021), which brings into question their ability to accurately represent the nature and magnitudes of ENSO extremes.

An alternative approach to assessing the potential for unprecedented climate extremes to occur is to examine ensembles of climate model runs initialized from observation-based states, as are applied for subseasonal, seasonal and decadal predictions (Meehl et al. 2021). Although starting from realistic states, such runs gradually develop biases (Saurral et al. 2021) and so after some period tend, like uninitialized simulations, to become inconsistent with observations. One therefore needs ideally to consider a range of lead times that are sufficiently long that the model state is not too strongly constrained by the initial conditions (in which case it is unlikely to simulate unprecedented extremes), and sufficiently short that model states are consistent with observations according to statistical tests. One such methodology, called Unprecedented Simulated Extremes using Ensembles, or UNSEEN, has been applied to assessing probabilities of record regional rainfall and heat extremes (Thompson et al. 2017, 2019). By considering N

(typically 20–40) start dates, M (typically 10–50) ensemble members, and L lead times, a sample of $N \times M \times L$ simulations is obtained that potentially amounts to thousands of realizations of weather and climate for a particular calendar month, from which chances for low-probability events can straightforwardly be inferred. This sample can potentially be increased further by considering simulations from multiple prediction systems (Jain and Scaife 2022).

In this study, a similar approach is applied to estimate the likelihood that extreme El Niño and La Niña events, including ones stronger than any yet observed, could occur in the current climate. Section 2 outlines the seasonal prediction systems, observational data and analysis techniques. Section 3 assesses the extent to which ENSO variability in the hindcasts is consistent with observations, estimates probabilities of unprecedented ENSO extremes, and examines simulated global impacts of such events. Concluding remarks are provided in Section 4.

2 Data and Methods

2.1 Seasonal Prediction Systems

The seasonal predictions analysed here consist of 6-month hindcasts initialized each month in 1993–2016 from eight systems contributing to the Copernicus Climate Change Service (C3S) seasonal forecast suite, obtained via the C3S Climate Data Store (Buontempo et al. 2022). These systems are identified in Table 1, along with corresponding hindcast ensemble sizes and references where comprehensive information about the systems is provided. Altogether, the eight systems provide 184 ensemble members. Data from each system are provided on a common 1° by 1° grid, and for the analysis reported here monthly mean predicted values for sea surface temperature (SST) are considered.

Table 1 Seasonal prediction systems contributing to C3S multi-system seasonal forecasts that are considered in this study, providing a total of 184 ensemble members

Center	System	Ensemble size	Reference
CMCC	SPS3	40	Sanna et al. (2017)
DWD	GCFS2.0	30	Fröhlich et al. (2021)
ECCC	CanCM4i	10	Lin et al. (2020, 2021)
ECCC	GEM5-NEMO	10	Lin et al. (2020, 2021)
ECMWF	SEAS5	25	Johnson et al. (2019)
Météo-France	System 7	25	Batté et al. (2019)
Met Office	GloSea5	28	MacLachlan et al. (2015)
NCEP	CFSv2	16	Saha et al. (2014)
Total		184	

2.2 Observational Data

As an observational reference we consider monthly mean SST values from Version 2 of the Optimum Interpolation Sea Surface Temperature (OISSTv2) dataset (Reynolds et al. 2002). OISSTv2 combines measurements from satellite radiometers and in situ data from ships, buoys and Argo floats, with interpolation applied to fill in spatial gaps. It spans from December 1981 until present and therefore covers the 1993–2016 period considered here. The representation in OISSTv2 of SST in the equatorial Pacific region that is the focus of this study is similar to that of other gridded SST analyses based on remote sensing and in situ measurements (Yang et al. 2021).¹

2.3 Analysis Approach and Methods

The ENSO metric considered here is the Niño 3.4 index, which provides a robust measure of ENSO SST variability and its wider influence (Barnston et al. 1997). Anomalies for each calendar month are computed as differences from the average value for that month during 1993–2016, separately for observations and for each model as a function of lead time, with forecast months 1–6 corresponding to lead times of 0 to 5 months.

Despite being initialized near observed states, seasonal prediction models can exhibit biases in representing the amplitude of predicted ENSO variability (e.g., Johnson et al. 2019). These biases generally depend on calendar month and lead time, and if they are appreciable then inferences about ENSO extremes cannot meaningfully be drawn from the hindcasts unless corrected for. This can be done by multiplying the predicted Niño 3.4 anomalies from a given model by the ratio of their observed standard deviations for the 1993–2016 hindcast period to those from the model, for each calendar month and lead time. Because this ratio can differ somewhat between ensemble members due to sampling variability, it is best to average the standard deviations from each of the ensemble members of the model considered. Such a correction is sometimes applied operationally,² and is effectively equivalent to standardizing the observed and predicted anomalies. The hindcast amplitude biases and resulting corrections are examined in Section 3.1.1.

¹ Yang et al. (2021) consider the OISSTv2.1 dataset (Huang et al. 2020, 2021), which is identical to a high-resolution version of OISSTv2 (Reynolds et al. 2007) that is considered here before 2016.

² For example, amplitude corrections are routinely applied to Niño 3.4 values predicted by models contributing to the North American Multi-Model Ensemble: <https://www.cpc.ncep.noaa.gov/products/NMME/current/plume.html>

To validate the representation of ENSO variability in each model and the multi-model ensemble, we examine whether the observed and forecasted distributions of Niño 3.4 values are statistically distinguishable. For this purpose, a two-sample Cramér–von Mises (CvM) test (Anderson 1962) is applied separately for each calendar month and lead time. Like the frequently applied Kolmogorov–Smirnov test, the CvM test is non-parametric, but since it considers distributional differences in the full joint sample it is more sensitive to differences in higher moments than the mean. The outcomes of these tests are discussed in Section 3.1.2.

3 Results

3.1 Hindcast Analysis and Validation

3.1.1 Niño3.4 Amplitude Biases

The dependence of Niño 3.4 amplitude biases on calendar month and lead time are illustrated in Fig. 2 for the eight C3S systems considered here. In the first forecast month the biases are relatively small although not insignificant (Fig. 2a, b), whereas by the sixth forecast month they have become more substantial, depending on the system and time of year (Fig. 2c, d). The lead-time dependence of these amplitude biases is examined in further detail for December, the month of peak observed Niño 3.4 standard deviation, in Fig. 3. As indicated by Fig. 2, the biases tend to be relatively small (hindcast/observed Niño 3.4 standard deviation close to 1) at a lead time of 0 months, whereas behavior at longer lead times is mixed, with some systems showing generally increasing or decreasing amplitudes and others less systematic behavior. To offset these biases and enable a more plausible estimation of ENSO extremes, we subsequently implement the amplitude correction described in Section 2.3 by multiplying hindcast Niño 3.4 anomalies by the reciprocal of the ratio indicated on the horizontal axis of Fig. 3.

Notably, Fig. 3 shows an evident inverse association between biases in Niño 3.4 amplitude and climatological mean SST in the Niño 3.4 region. The latter are indicative of systematic temperature errors in the equatorial cold tongue region (most often cooler than observed) that are prevalent in many climate simulation and prediction models (e.g., Bayr et al. 2019; Ying et al. 2019; Ma et al. 2020). Quantitatively, the correlation between December Niño 3.4 amplitude and cold tongue biases is -0.84 if all of the systems are considered, or -0.56 if the DWD system that has a growing warm bias in the cold tongue region is excluded. Similar behavior has been noted in an earlier generation of seasonal prediction systems (Vannière et al. 2013).

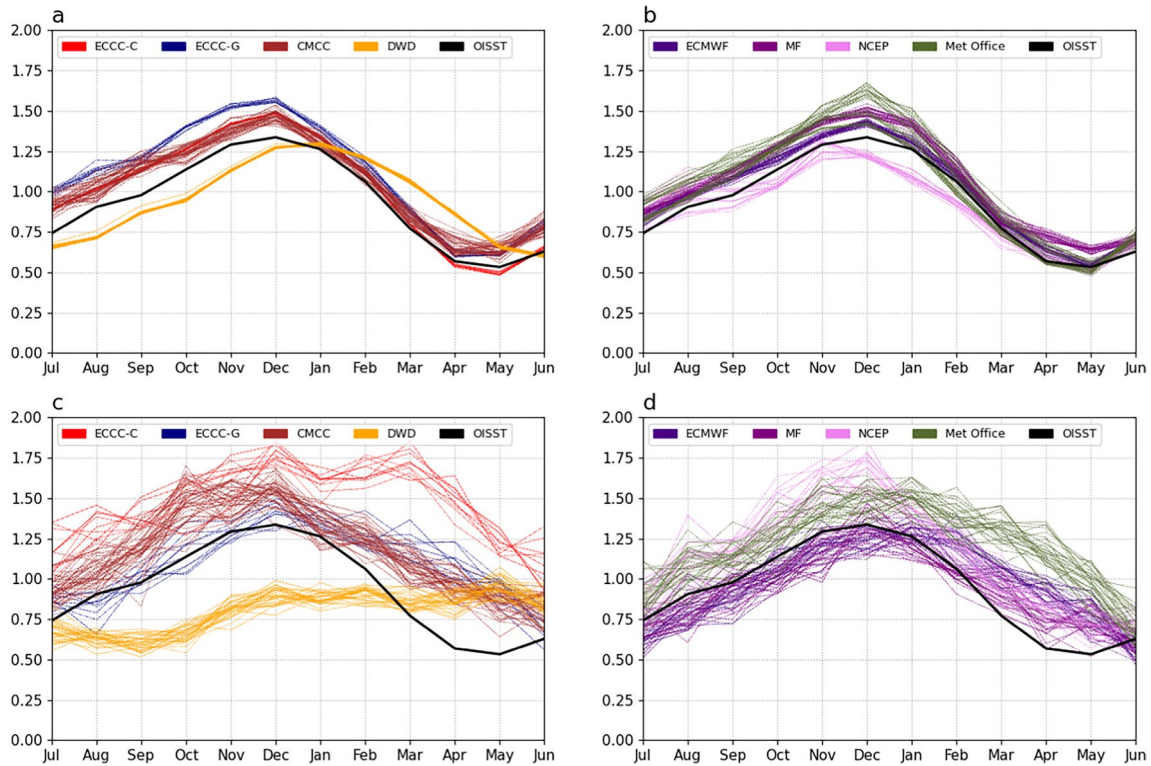


Fig. 2 Standard deviations of the Niño 3.4 index for the indicated months during the 1993–2016 hindcast interval, for lead times of (a)–(b) 0 months, and (c)–(d) 5 months. Colored lines denote values for

individual ensemble members of eight seasonal prediction systems as indicated, and the black line values for the OISSTv2 observational reference

Whether these biases are smaller for this range of lead times than in long climate simulations for which the initial conditions have little influence cannot be determined

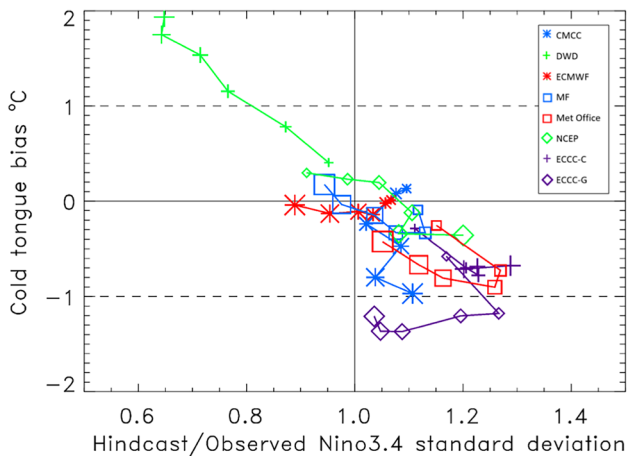


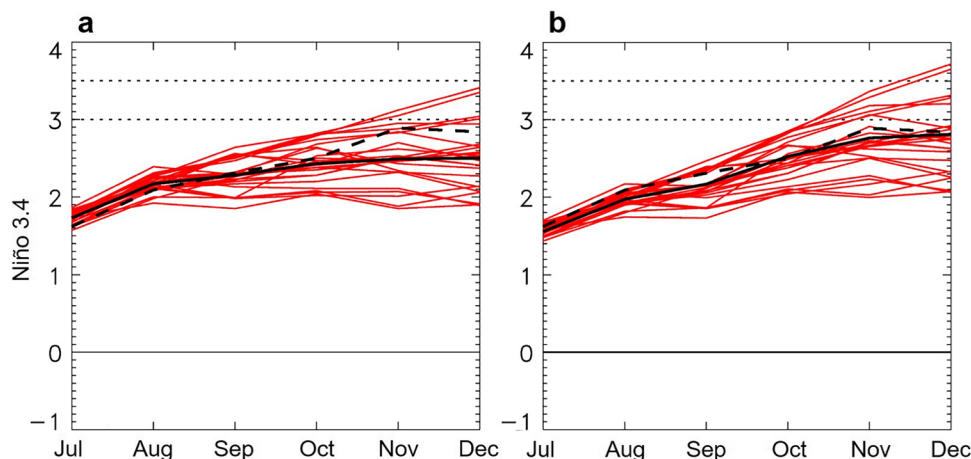
Fig. 3 Relationship between December ENSO amplitude bias (ratio of the uncorrected hindcast Niño 3.4 standard deviation to observed Niño3.4 standard deviation, horizontal axis) and December equatorial Pacific cold tongue bias (hindcast minus observed climatological mean SST in the Niño 3.4 region, vertical axis) during the 1993–2016 hindcast period. Increasing lead time from 0 to 5 months is indicated by increasing symbol sizes for each of the C3S prediction systems (symbols as indicated)

because such simulations are not available for the set of prediction systems considered here. However, for other sets of models it has been found that equatorial Pacific SST biases in seasonal predictions can sometimes overshoot or develop signs opposite to those in long climate simulations from the same model (Hermanson et al. 2018; Ma et al. 2020). These studies also find that SST biases tend to develop more rapidly in the tropics than the extratropics. Equatorial Pacific SST and upper ocean biases in seasonal hindcasts have complex origins (Vannière et al. 2013; Siongco et al. 2020), and complex effects on ENSO forecasts (Ma et al. 2020; Wu et al. 2022). Since it is evident that ENSO-related SST and amplitude biases cannot be avoided for the range of lead times considered here, we proceed pragmatically by implementing the amplitude correction described above and examining extremes in the resulting Niño 3.4 distributions.

3.1.2 Bias-Corrected Niño3.4 Distributions

An example of the ensemble Niño3.4 predictions considered here is depicted in Fig. 4. In this instance, the first forecast month is July 2015, during the lead-up to the very strong El Niño that peaked in late 2015 and early 2016 (McPhaden et al. 2020), and the ensemble hindcast from ECMWF

Fig. 4 Niño 3.4 index from ECMWF SEAS5 hindcast starting from July 2015, showing predicted anomalies (a) before and (b) after variance rescaling. The 25 ensemble members are represented in red, the ensemble mean by the heavy solid black line, and the OISSTv2 verification by the heavy dashed black line. Light dotted black lines indicate Niño 3.4 anomaly thresholds of 3.0 °C and 3.5 °C that are unobserved in the modern era



SEAS5 is shown. According to Fig. 2, SEAS5 hindcasts tend to systematically overestimate interannual Niño3.4 variance in July at lead 0, and slightly underestimate Niño3.4 variance in December at lead 5. After the amplitude correction is applied (Fig. 3b), the mean of the ensemble verifies close to observed values throughout the forecast range. December Niño3.4 values for five ensemble members exceed 3.0 °C, greater than any values in the modern instrumental record, with two ensemble members exceeding the even more extreme threshold of 3.5 °C.

To assess the fidelity of the hindcasts in representing the distribution of Niño 3.4 values, the CvM test is applied to pairings of hindcast and observed Niño 3.4 samples. One such pairing is illustrated in Fig. 5, which

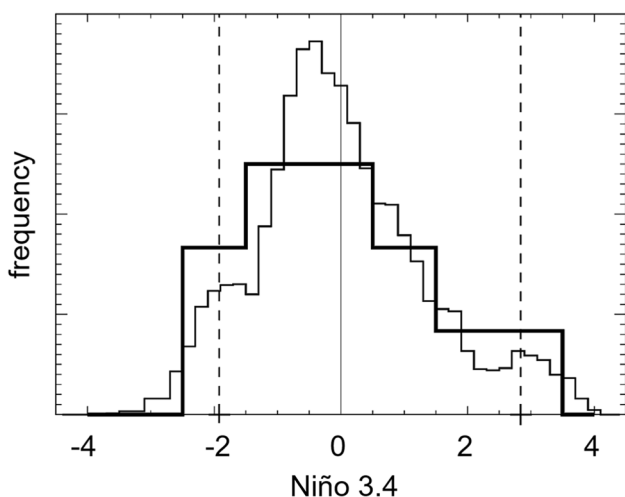


Fig. 5 Distributions of observed December mean Niño 3.4 values during the 1993–2016 hindcast period from OISSTv2 (thick histogram, 24 values) and rescaled values from the 184 C3S ensemble members at 5 month lead, i.e. starting from July (thin histogram, 4416 values). The vertical dashed lines indicate positive and negative extrema of the observed sample

compares the distribution of observed December Niño 3.4, with 24 values during the hindcast period, with that from the combined hindcast ensemble at 5 month lead time. The hindcast sample size is larger by a factor of 184, the aggregate number of ensemble members. Two evident aspects of this comparison are that the two distributions have roughly the same shape, and a non-negligible fraction of hindcast values lie outside the observed positive and negative extrema.

The outcome of the CvM tests is illustrated in Fig. 6, which shows the CvM test statistic³ t_c for each of the individual models and the multi-model ensemble, for all calendar months and lead times. Here the p values represent probabilities that the associated t_c values would be equaled or exceeded by chance if the hindcast and observed samples came from the same distribution. Thus, for small p values this null hypothesis is unlikely to be true, and if the two samples do come from the same distribution then 9 samplings out of 10 will be characterized by $p \geq 0.10$ ($t_c < 0.347$) for example. This is seen to be the case for all but 14, or 2.4% of the raw single-model hindcast distributions (576 cases stemming from 8 models, 12 calendar months and 6 lead times), and none of the rescaled single-model hindcast distributions. In addition, all of the multi-model hindcast distributions (raw and rescaled) meet this criterion, with only the raw distributions showing any t_c values exceeding 0.1. Thus, for the rescaled samples in particular, the most that can be inferred given the implied large p values is that the null hypothesis that the hindcast and observed samples come from a common distribution cannot be rejected, and the test was unable to detect any inconsistency between the two distributions. However, differences too small to be detected might exist, in which case the test fails to reject the

³ The subscript distinguishes the CvM test statistic, defined in Anderson (1962), from other test statistics including that of the well-known parametric Student's t test.

null hypothesis when it is false (Type II error or false negative). Even though the CvM test is relatively powerful (Liu and Chan 2016; DeSole and Tippett 2022), its power may be limited in this instance because the means and standard deviations of the two distributions have been made identical by construction. A further aspect to note is the absence of any rescaled cases for which $p \leq 0.10$ ($t_c > 0.347$) as would be expected by chance with $\sim 10\%$ frequency if the two samples did come from a common distribution. This likely can be explained by some degree of non-independence of the hindcast samples, since they tend to be correlated with the observed samples to the extent that Niño3.4 is predicted with some skill for a given target month and lead time.

Finally, to check if consistency between the hindcast and observed distributions depends on the length of the observational record, December Niño 3.4 values from the ERSSTv5 dataset (Huang et al. 2017) for 1950–2022 were considered, with hindcast values again rescaled to match the standard deviations of the observed time series. Taking December at 5 month lead time as an example, values of t_c are slightly larger when ERSSTv5 is employed as the observational reference, 0.078 using a 1993–2016 base period and 0.094 for centered 30-year base periods updated every 5 years (Lindsey 2013), compared to 0.053 for OISSTv2. Although suggesting that hindcast values are slightly less consistent with the longer record from ERSSTv5 than with OISSTv2, the

ERSSTv5-based t_c continue to imply p values substantially exceeding 0.10 according to Fig. 6, and the conclusions drawn above based on 1993–2016 values of OISSTv2 are unchanged.

3.2 Occurrences of ENSO Extremes

The objective here is to assess the potential for occurrence of unprecedented ENSO extremes based on C3S multi-model seasonal hindcasts. With 184 total ensemble members initialized monthly over the 24-year hindcast period, these comprise approximately 5.3×10^4 6-month model runs, and nearly 3.2×10^5 simulated months. Early in the forecasts, however, Niño 3.4 values are highly constrained by the initial conditions as exemplified in Fig. 4. As a result, unprecedented extremes are less likely to be simulated than later in the forecasts when more substantial ensemble spreads have developed.

Bearing this consideration in mind, along with the property that observed Niño 3.4 tends to peak in December (Fig. 2), Fig. 7 illustrates the frequency with which December Niño 3.4 values exceeding various positive and negative thresholds occur at the longest lead time of 5 months among the 4416 hindcast runs, represented by the 184 C3S ensemble members across the 24 year hindcast period (red circles in Fig. 7). Based on this sample, El Niños having December Niño 3.4 > 1.5 and La Niñas having December Niño 3.4 < -1.5 (sometimes classified as “strong” events; Trenberth 2020) occur more frequently than once in 10 years, both in the hindcasts and observations. More specifically, both types of events were observed to occur on average once in 6 years during the hindcast period (corresponding to four El Niño and four La Niña years meeting this criterion in 1993–2016), whereas the

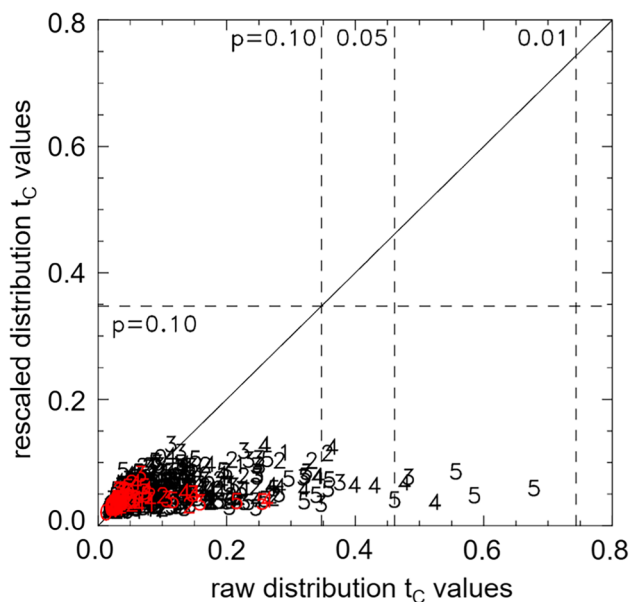


Fig. 6 Values of the two-sample Cramér–von Mises test statistic t_c , based on comparing observed and hindcast distributions of Niño 3.4 for the 1993–2016 hindcast period. Black numerals denote t_c for the indicated lead times for all of the eight C3S prediction systems and initial months, and red numerals corresponding values for the multi-model hindcast distribution. The horizontal axis indicates t_c values for raw hindcast Niño 3.4 anomalies, and the vertical axis values for hindcast distributions to which variance rescaling has been applied

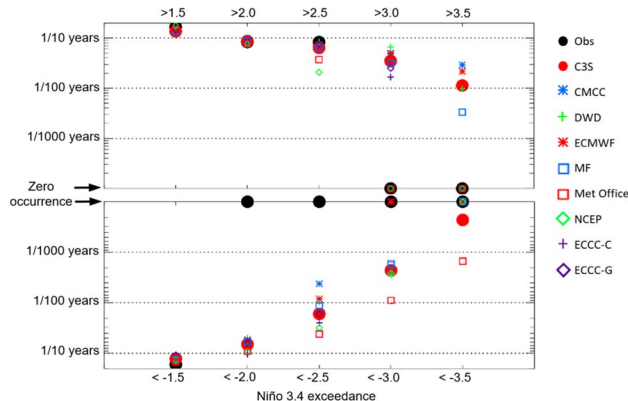
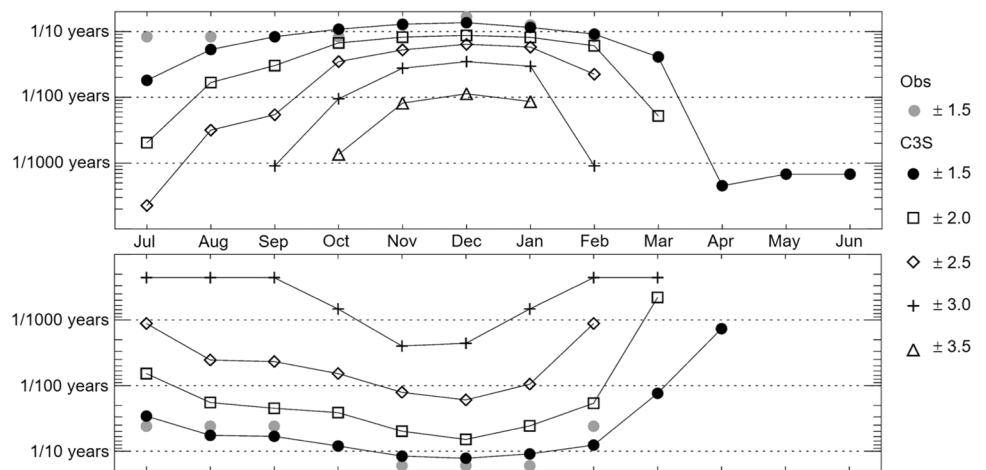


Fig. 7 Frequencies of occurrence of positive (El Niño, upper panel) and negative (La Niña, lower panel) December Niño 3.4 extremes in exceedance of the indicated thresholds, according to 1993–2016 observations (black circles), aggregated C3S ensemble hindcasts (red circles), and hindcasts from individual C3S prediction systems (symbols as indicated). Frequencies are computed as fractions of the observed and lead 5 month hindcast Niño 3.4 values for December that exceed the corresponding thresholds. Absence of any exceedances is indicated by symbols at the edges of each panel

Fig. 8 Frequencies of occurrence of positive (El Niño, upper panel) and negative (La Niña, lower panel) Niño 3.4 extremes in each calendar month exceeding thresholds indicated by the symbols. Frequencies connected by solid lines are for the aggregated C3S ensemble at lead 5 months (sample size 4416), and gray circles are for 1993–2016 observations (sample size 24, ± 1.5 threshold only). Absence of a symbol for a given month and threshold indicates no occurrences in the data considered



hindcast frequency is once in 7.3 years for El Niño and once in 7.8 years for La Niña. By contrast, two “very strong” El Niño events having December Niño 3.4 > 2.0 (and indeed > 2.5) were observed during this period, implying an observed Niño 3.4 > 2.5 exceedance frequency of once in 12 years, whereas the hindcasts imply a frequency of once in 15.7 years.

The observed sample contains no events exceeding the higher thresholds in Fig. 7, and thus becomes inadequate for even crudely estimating the frequency of such events by direct means. Notably, no La Niña events having Niño 3.4 < -2.0 were observed in 1993–2016, reflecting the inherent asymmetry of ENSO (An et al. 2020). According to the hindcasts, however, such events should be expected approximately once in 15 years. This can be somewhat reconciled with longer SST records such as ERSSTv5, indicating that the Niño 3.4 < -2.0 threshold may have been crossed, or nearly so, by observed La Niñas in 1973–74 and 1988–89 (e.g., Santoso et al. 2017), and was nearly crossed in 1998 when December Niño 3.4 reached -1.92.⁴ Even stronger La Niñas are implied to be progressively rarer, with exceedances of -2.5 and -3.0 occurring approximately once every 60 and 440 years respectively, whereas the -3.5 threshold is crossed in only one of the 4416 lead 5 hindcast realizations.

By contrast, hindcast El Niño events having lead 5 December Niño 3.4 values that exceed 3.0, unprecedented in the reliable instrumental record, are not exceptionally rare, with an implied occurrence rate of once in 29 years. The 3.5 threshold is exceeded once in 88 years (50 realizations), and 4.0 is exceeded in one hindcast realization.

⁴ These values are somewhat sensitive to the base period considered due to long-term variability and warming in the Niño 3.4 region (Lindsey 2013). For example, the December 1973 Niño 3.4 index is -2.23 according to ERSSTv5 using the 1993–2016 base period of the C3S hindcasts, and -2.06 using a centered 30-year base period. The relative Niño3.4 index of van Oldenborgh et al. (2021) offers another means to account for such changes.

As seen in Fig. 7, these frequencies are somewhat system dependent, with four systems (CMCC, DWD, ECMWF, Météo-France) accounting for the 50 realizations having Niño 3.4 > 3.5, and a different set of 4 (DWD, MF, Met Office, NCEP) accounting for the 10 realizations having Niño 3.4 < -3.0. The Met Office system in particular stands out as having particularly frequent large negative Niño 3.4 values, despite having a rescaled Niño 3.4 distribution not demonstrably different from observations (Section 3.1).

Exceedance frequencies of the aggregated C3S ensemble at 5 month lead considering all calendar months are shown in Fig. 8. As expected, the frequencies are generally highest in December (an exception is slightly more frequent Niño 3.4 < -3.0 exceedances in November based on a small number of cases). Exceedance frequencies are markedly lower in spring and early summer months, with virtually no exceedances even of ± 1.5 occurring in May and June.

3.3 Interpretation

As suggested by Fig. 4, the occurrences of extremes shown in Figs. 7 and 8 are mainly associated with hindcasts of the largest El Niño and La Niña events during the 1993–2016 hindcast period, in particular the El Niños of 1997–98 and 2015–2016, and La Niñas of 1998–2000 and 2010–2011. This is a consequence of the hindcasts being skillful, and hence significantly correlated with the observed Niño 3.4 values, even at the longest available lead time of 5 months. Figure 9 shows the dependence of predicted December Niño 3.4 values on lead time for the two strongest El Niño events (Niño 3.4 > 2.0) and four strongest La Niña events (Niño 3.4 < -1.5) during the hindcast period. It is seen that the strongest unprecedented extremes having Niño 3.4 > 3.5 and < -3.0 occur at lead times of 3 and 4 months (and to some extent 2 months) in addition to the longest lead time of 5 months, but are largely absent at the shortest 0 and 1 month lead times where the prediction is more strongly

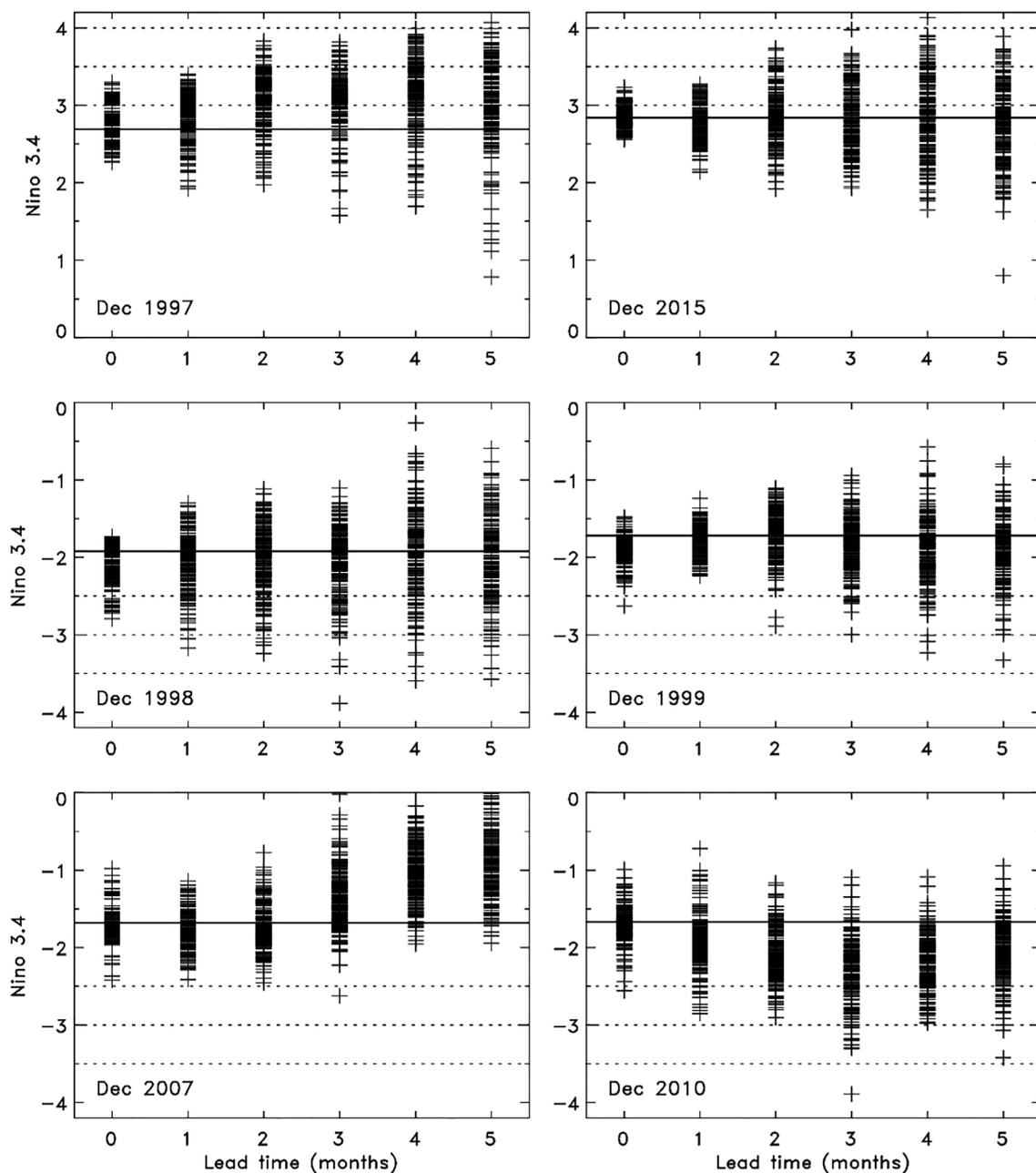


Fig. 9 Predicted December Niño 3.4 values from the 184 C3S ensemble members at lead times from 0 to 5 months in the years indicated, for the two strongest El Niño events ($\text{Niño } 3.4 > 2.0$, top row) and

four strongest La Niña events ($\text{Niño } 3.4 < -1.5$, middle and bottom rows) during the hindcast period. Thick lines in each panel indicate the observed December Niño 3.4 values

constrained by initial conditions. It is noteworthy that no such December Niño 3.4 extremes occurred in the hindcasts for the La Niña of 2007–2008, which was not predicted well beyond 2-month lead, or for any other hindcast years not represented in Fig. 9.

The estimates presented here for frequencies of occurrence of ENSO extremes are thus specific to the period considered, and are conditioned by the major El Niños and La Niñas that occurred within it (Fig. 1). By contrast,

based on available measurements ENSO activity tended to be weaker during a multidecadal period before 1970, whereas several large ENSO events appear to have occurred in the late 19th and early 20th centuries (Chen et al. 2004; Giese et al. 2010). Whether this is due to interdecadal to intercentennial variability in ENSO intensity as has been noted in long climate model control simulations in which radiative forcings are fixed (e.g., Wittenberg 2009), or is driven by changes in radiative forcings

during the historical period, is not currently understood (Fedorov et al. 2020). In either case, the estimates for extreme ENSO frequency presented here can be viewed as characterizing recent decades, and framing them as “present day” probabilities, e.g. applicable to the next decade or so, implicitly assumes that this recent level of ENSO variability will continue in the near term. This assumption is not guaranteed of course, and may not apply later in the twenty-first century, particularly if the characteristics of ENSO are appreciably modified by climatic warming (Collins et al. 2010; Cai et al. 2021). The brevity of the 24-year hindcast record and the small number of ENSO events from which the exceedance statistics are drawn implies that sampling is also a source of uncertainty.

3.4 Global Influences

The above results suggest that the occurrence of an El Niño or La Niña far more extreme than any in the instrumental record is not beyond the realm of possibility in the current epoch. This raises the question of what the associated global impacts could be, considering that major regional societal and ecosystem disruptions have been caused by lesser ENSO events in the past (e.g., McPhaden et al. 2006).

Figure 10 compares global composites of December near-surface 2 m air temperature (T2m) and precipitation anomalies for the historically strong 1997–1998 and 2015–2016 El Niños having peak Niño 3.4 values between 2.5 and 3.0 that were observed during the hindcast period, and composites for the 66 occurrences of Niño 3.4 index >3.5 prior to rescaling in lead 5 predictions of these events. For this hindcast composite and a corresponding one for La Niña events described below, ensemble members are selected based on non-rescaled Niño 3.4 values, so as to maintain physical consistency between Niño 3.4 and the two-dimensional anomalies that are shown which also are not rescaled. In addition, the years from which these ensemble members are drawn are weighted equally in order to reflect the diversity of observed ENSO events in the same manner as the observed composites.

The observed T2m composite (Fig. 10a), although based on only two events, is broadly similar to the pattern of linear dependence of T2m on the Niño 3.4 index over a 70 year period shown in Taschetto et al. (2020). Exceptions include north-central North America which was exceptionally warm even for an El Niño winter in 1997–98 (Smith et al. 1999), and mid-to-high latitude regions, particularly in the Eastern Hemisphere, where ENSO influences are less pronounced compared to internal atmospheric variability (e.g., McPhaden et al. 2020). Other differences may be attributable to model biases. For example, the excessive warm anomalies over Brazil in the C3S composite compared to the observed composite may reflect a tendency for models to overestimate

temperature variability in this region (e.g., Merryfield et al. 2013). In addition, the unrealistic equatorial Pacific warm anomalies to the west of the Dateline in Fig. 10b are likely due to a tendency for excessive westward extension of ENSO SST anomalies in global climate models, which in turn is related to the tendency for cool biases in the equatorial Pacific cold tongue discussed in Section 3.1.1 (Jiang et al. 2021). Overall, while the T2m anomalies in the Niño 3.4 region are higher in the C3S composite than the observed composite due to the stronger El Niño events represented in the former, the associated global temperature influences are not conspicuously stronger in Fig. 10b than in Fig. 10a, except where noted above. This is likely at least partially due to the extremely limited sample of observed very strong El Niño events, together with differences in atmospheric responses to ENSO, particularly in the extratropics, caused by atmospheric internal variability (Kumar and Hoerling 1997). By contrast, averaging over the much larger sample of 66 simulated exceptional El Niño events more effectively distills the average response, while filtering out the atmospheric internal variability that accompanies individual events. However, model inadequacies in representing El Niño global temperature influences could also be a factor.

Similar considerations apply in comparing the observed and C3S composites of El Niño precipitation influences in Fig. 10c–d. While both patterns again broadly resemble that of a regression on Niño 3.4 based on long climate records (Taschetto et al. 2020), some regional differences are evident. For example, the C3S composite shows a precipitation enhancement of up to 2 mm/day along the California coast in qualitative agreement with long climate records, whereas the observed composite for the two very strong El Niño shows a mixed influence due to an atypical response to the 2015–2016 El Niño in this region, likely as a result of atmospheric internal variability (Kumar and Chen 2017; Chen and Kumar 2018). Northern Australia is another region where the C3S ensemble agrees qualitatively with long climate records, but the observed composite of the 1997–1998 and 2015–2016 El Niños is opposite in sign. By contrast, Fig. 10c qualitatively agrees with long climate records in eastern and southern Africa whereas Fig. 10d does not, which may be indicative of model errors in representing El Niño precipitation influences in these areas.

Figure 11 presents similar comparisons between observed December T2m and precipitation composites for the four strongest La Niña winters of 1998–1999, 1999–2000, 2007–2008 and 2010–2011 for which Niño 3.4 <-1.5 (Fig. 1), and 11 ensemble members for which Niño 3.4 <-3.5 prior to rescaling in December 1998 and 2010. In this instance the pronounced difference between the strength of the observed and simulated La Niña events is especially evident, although the global temperature influences are again mostly similar except in mid-to-high

latitude regions of the Eastern Hemisphere where atmospheric internal variability is likely dominant in the small observed sample. Where there is qualitative agreement, the signal in the C3S composite is generally not larger except in northern Brazil, where a tendency for at least some models to overestimate temperature variability was previously noted, and in Australia.

The picture is somewhat different for the La Niña precipitation composites, in which the simulated wet influence over the warm pool and adjacent regions including the Philippines and northern Australia in the C3S composite far exceeds that in the observed composite (Fig. 11d). This suggests that if such a La Niña event having Niño 3.4 < -3.5 were to occur (which is exceedingly unlikely in a given year according to Fig. 6 based on rescaled Niño 3.4 values) then the excess December rainfall in these regions would be comparably extraordinary. Elsewhere, the dry anomalies in southwestern North America are in qualitative agreement with long climate records as in Taschetto et al. (2020), but not the observed composite for the hindcast period (Fig. 11c), due to atmospheric internal variability evidently

playing a confounding role in this smaller sample. In some regions such as southeastern South America and East Africa, precipitation anomalies in the C3S composite are in qualitative disagreement both with the hindcast-period composite and with longer climate records, pointing to model biases as a likely cause for the disagreement.

A further important aspect of these comparisons is that the most severe ENSO impacts, exceeding the mean ENSO influences represented by the composites in Figs. 10 and 11, accompany extreme weather and climate events caused by a combination of ENSO and other influences (Goddard and Gershunov 2020). For example, particularly severe instances of low northern Brazil rainfall induced by major El Niños in boreal winter have been linked in ensemble hindcasts to warm SST anomalies in the northeastern subtropical Pacific and Atlantic, accompanied by regions of anomalously low sea level pressure further north (Kay et al. 2022). Further examinations of the C3S seasonal hindcast dataset considered here could support additional studies of the diversity of ENSO influences, and of factors associated with particularly severe impacts.

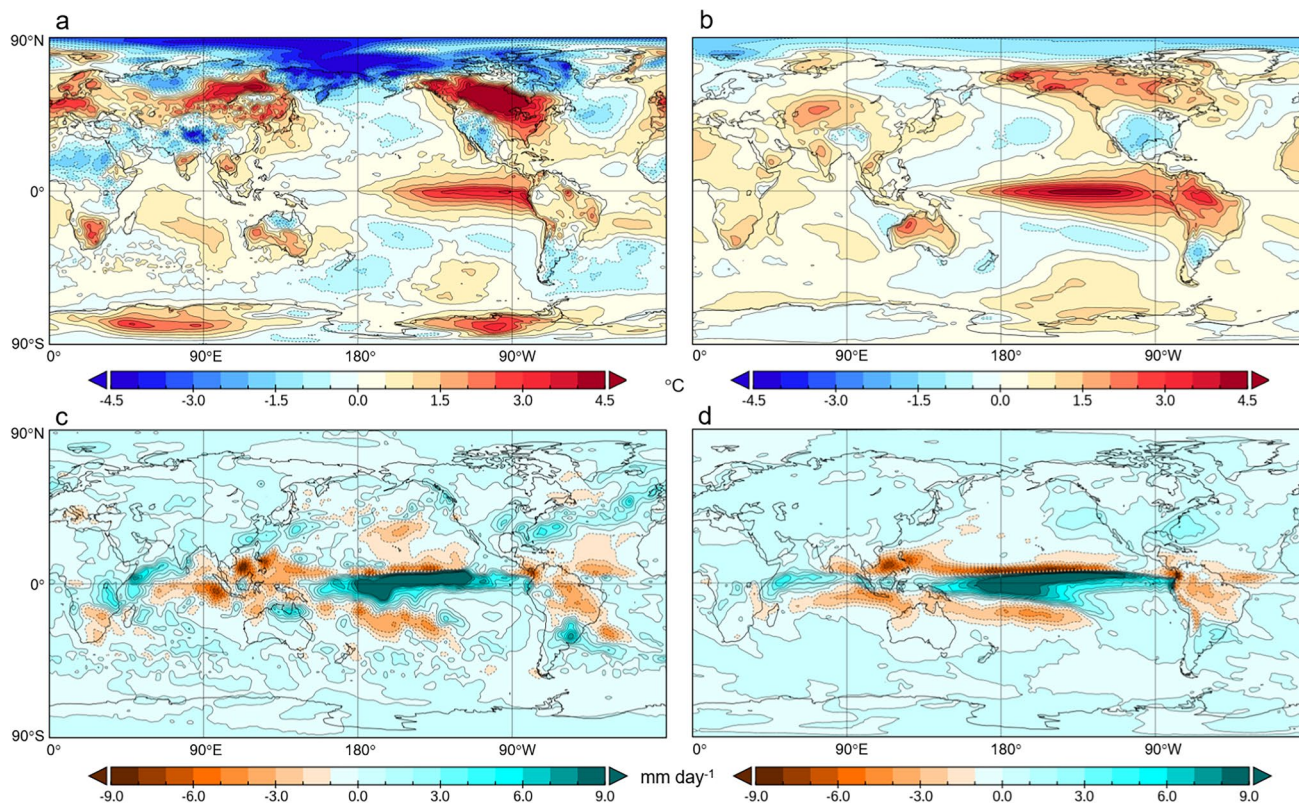


Fig. 10 Composites of December (a) observed T2m anomalies in 1997 and 2015, during the two strongest El Niño events of the hindcast period, (b) T2m anomalies of 66 C3S ensemble members for which December Niño 3.4 index > 3.5 at lead 5 prior to rescaling, (c)

observed precipitation anomalies for the same years as in (a), (d) C3S precipitation anomalies for the same C3S ensemble members and lead time as in (b)

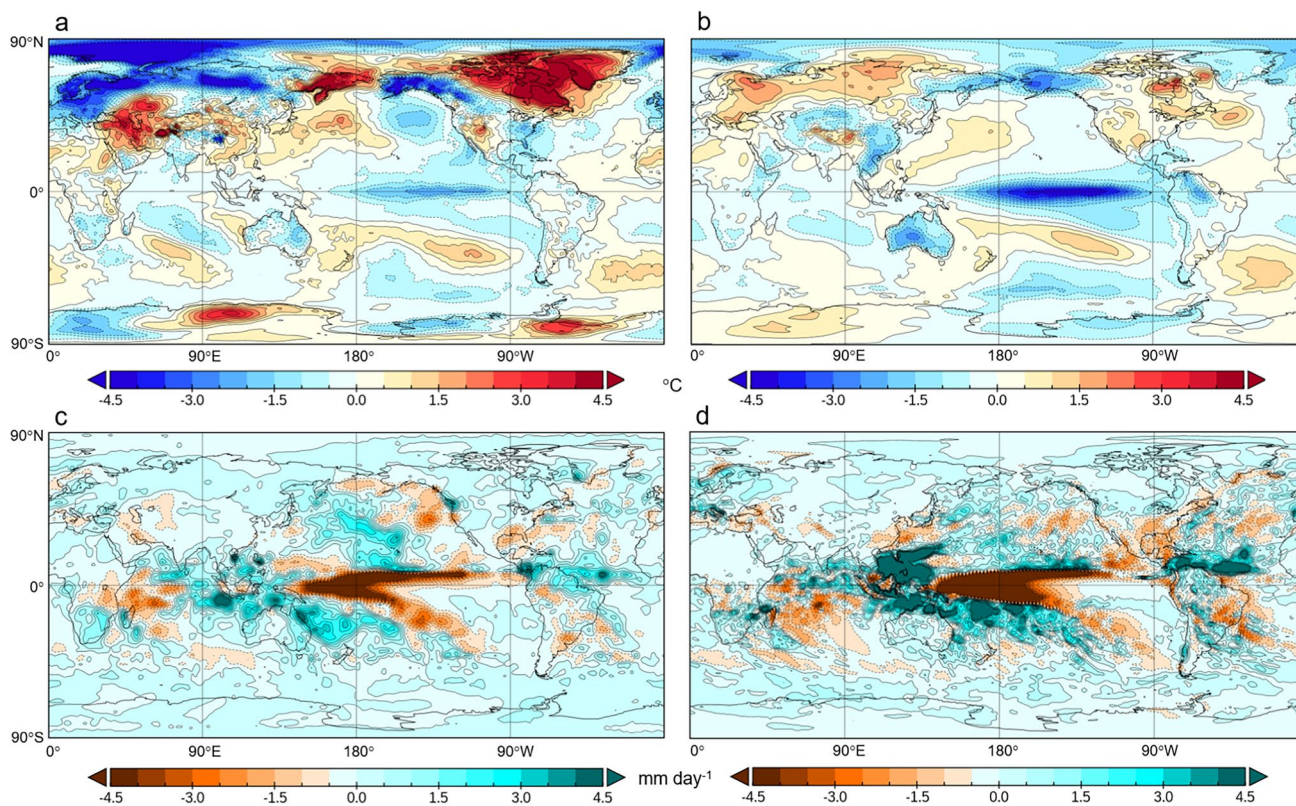


Fig. 11 Composites of December (a) observed T2m anomalies in 1998, 1999, 2007 and 2015, during the four strongest La Niña events of the hindcast period, (b) T2m anomalies of 11 C3S ensemble members for which December Niño 3.4 < -3.5 at lead 5 prior to rescal-

ing, (c) observed precipitation anomalies for the same years as in (a), (d) C3S precipitation anomalies for the same C3S ensemble members and lead time as in (b)

4 Summary and Conclusions

This study has examined the occurrence of ENSO extremes, especially El Niño and La Niña episodes more extreme than any in the reliable instrumental record, in seasonal hindcasts from eight systems contributing to the C3S multi-system seasonal forecast ensemble. With a total of 184 ensemble members and monthly initializations across the 24-year hindcast period 1993–2016, these hindcasts represent over 5×10^4 initialized model runs, each having a 6-month range.

The Niño 3.4 index, which represents monthly-mean SST anomalies in the central-eastern equatorial Pacific where SST variance is particularly high and is especially well correlated with wider ENSO phenomena (Barnston et al. 1997), is used as an ENSO metric. To reduce the impact of model biases affecting simulated ENSO amplitude, hindcast Niño 3.4 values were multiplied by a correction factor so that for each model, Niño 3.4 variance for each calendar month and lead time matches observed Niño 3.4 variance in the corresponding calendar month. Within these adjusted Niño 3.4 hindcasts numerous instances of unprecedented extremes occur, mainly in years when the strongest El Niño and La Niña events are observed and

predicted, and for longer lead times of up to 5 months when the simulations are less constrained by their observation-based initial conditions than at shorter lead times. Implied occurrence rates for unprecedented extremes in December, when ENSO SST anomalies typically peak, are approximately once in 30 years for Niño 3.4 > 3.0, once in 90 years for Niño 3.4 > 3.5, once in 60 years for Niño 3.4 < -2.5, and once in 440 years for Niño 3.4 < -3.0. These results, including asymmetry between the likelihood of El Niño and La Niña extremes, are qualitatively similar to inferences drawn from statistics of El Niño and La Niña magnitudes inferred from existing instrumental records dating to the mid-1800s (Douglass 2010), although the implied frequencies for unprecedented El Niño events found here are somewhat larger.

As an exploration of typical global impacts of unprecedented ENSO extremes, T2m and precipitation composites for the strongest observed El Niño and La Niña events in the hindcast period were compared with composites for extreme simulated El Niños having unadjusted Niño 3.4 > 3.5, and extreme simulated La Niñas having unadjusted Niño 3.4 < -3.5. Although impacts in the tropical Pacific region that is the seat of ENSO are generally larger in the

composites of simulated ENSO extremes than the observed composites, the magnitudes of remote influences are less distinguishable. This reflects the significant role of internal variability in determining remote climate anomalies for individual ENSO events (or averages over a small number of events as for the observational composites considered here), particularly in mid-to-high latitudes.

Although C3S seasonal hindcasts provide a large sample for examining extreme ENSO events, certain assumptions, caveats and limitations of this study should be kept in mind. First, the hindcasts are assumed to represent 184 simulated realizations of what could have happened if conditions at each hindcast start date had been slightly different from “truth”. In reality these realizations differ from the observed one for a variety of reasons, including observational, assimilation and model errors, and in some cases stochastic model physics. They also are not entirely independent of the state of the climate system at each start date, even at the longest lead time, because the hindcasts have some skill. Therefore, characterizing the derived probabilities of ENSO extremes as present-day values implicitly assumes a near-term continuance of the level of ENSO variability that occurred during the hindcast period. As a result, the analysis and conclusions drawn, while pertaining essentially to present-day climate, are somewhat conditioned on the single observed realization of ENSO activity during 1993–2016, which included two historically large El Niños and several significant but more moderate La Niñas (Fig. 1). Subsequent years through 2022 have yet to see any comparably strong events, and a continuation of that circumstance would suggest that the hindcast period was an interval of unusually high ENSO activity, possibly due to natural multi-decadal variation (e.g., Rodrigues et al. 2019). Longer hindcast periods, spanning even up to a century or more enabled by corresponding reanalyses (Weisheimer et al. 2020), could potentially provide more robust estimates, although climate nonstationarity and the general degradation of observational coverage and accuracy more than a few decades in the past would remain considerations. In addition, any significant future changes to ENSO properties caused by anthropogenic forcings could alter the statistics of extreme ENSO events.

A further caveat is that the conclusions drawn are reliant on model fidelity in representing ENSO. Corrections were applied to the mean (through the calculation of anomalies) and variance of the Niño 3.4 index, and a sensitive statistical test, the two-sample Cramér–von Mises test, did not detect gross differences between the observed and variance-adjusted Niño 3.4 distributions. However, some distributional differences could have escaped detection due to the limited observational sample, particularly for individual models as suggested by the differences between models in representing the frequencies of ENSO events for high exceedance thresholds (Fig. 6), although sampling fluctuations may be another

contributing factor. As well, systematic model errors such as an excessive westward extent of equatorial Pacific temperature and unrealistic rainfall anomalies associated with El Niño episodes are evident to some degree in the model composites (Fig. 9), despite the initialization of these simulations from observation-based states.

Overall, this study demonstrates the potential for applying large ensembles of initialized climate predictions to explore probabilities of unprecedented extremes of ENSO variability and remote influences. In particular, much further potential exists for assessing chances of extreme ENSO impacts, which typically arise from a combination of direct ENSO influence and other factors including internal atmospheric variability (Goddard and Gershunov 2020). Such studies, like that of Kay et al. (2022), require additional tests of model fidelity for the variables and regions considered to assure that the many simulated realizations of climate variability represent a plausible proxy for the natural system. The C3S seasonal hindcast ensemble considered here could provide a valuable resource for such studies, as well as for broader studies of the likelihood of unprecedented climate extremes across the globe.

Acknowledgements This paper is an outcome of the “Risks of Extremes” initiative of the World Climate Research Programme’s Working Group on Subseasonal to Seasonal Prediction (WGSIP). NOAA Optimum Interpolation (OI) SST V2 data were downloaded from the NOAA Physical Sciences Laboratory, Boulder, Colorado, USA THREDDS Data Server <https://psl.noaa.gov/thredds/catalog/Datasets/noaa.oisst.v2/catalog.html?dataset=Datasets/noaa.oisst.v2/sst.mnmean.nc>, accessed from <https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html>. NOAA Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5) monthly Niño 3.4 values were obtained for a fixed 1991–2020 base period from <https://www.cpc.ncep.noaa.gov/data/indices/ersst5.nino.mth.91-20.ascii>, and for a centered 30-year base period updated every 5 years from https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensos_tuff/detrend.nino34.ascii.txt. Data from seasonal hindcasts of the eight systems contributing to this study were obtained from the C3S Climate Data Store at <https://cds.climate.copernicus.eu/>.

Funding Open Access provided by Environment & Climate Change Canada.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- An, S.-I., Tziperman, E., Okumura, Y.M., Li, T.: ENSO irregularity and asymmetry. In: McPhaden, M.J., Santoso, A., Cai, W. (eds.) *El Niño Southern Oscillation in a Changing Climate*. AGU Monograph. American Geophysical Union, Washington, D.C. (2020). <https://doi.org/10.1002/9781119548164.ch7>
- Anderson, T.W.: On the distribution of the two-sample Cramér-von Mises criterion. *Ann. Math. Stat.* **33**, 1148–1159 (1962). <https://doi.org/10.1214/aoms/1177704477>
- Barnston, A.G., Chelliah, M., Goldenberg, S.B.: Documentation of a highly ENSO-related SST region in the equatorial Pacific. *Atmosphere-Ocean*. **3**(35), 367–383 (1997). <https://doi.org/10.1080/07055900.1997.9649597>
- Batté, L., Dorel, L., Ardilouze, C., Guérémy, J.F.: Documentation of the METEO-FRANCE Seasonal Forecasting System 7. Technical Report, Copernicus Climate Change Service (2019). <http://www.umr-cnrm.fr/IMG/pdf/system7-technical.pdf>. Accessed 1 May 2023
- Bayr, T., Domeisen, D., Wengel, C.: The effect of the equatorial Pacific cold SST bias on simulated ENSO teleconnections to the North Pacific and California. *Clim. Dyn.* **53**, 3771–3789 (2019). <https://doi.org/10.1007/s00382-019-04746-9>
- Buontempo, C., Burgess, S.N., Dee, D., Pinty, B., Thépaut, J.N., Rixen, M., Almond, S., Armstrong, D., Brookshaw, A., Alos, A.L., Bell, B., Bergeron, C., Cagnazzo, C., Comyn-Platt, E., Damasio-Da-Costa, E., Guillory, A., Hersbach, H., Horányi, A., Nicolas, J., Obregon, A., Penabad Ramos, E., Raoult, B., Muñoz-Sabater, J., Simmons, A., Soci, C., Suttie, M., Vamborg, F., Varndell, J., Vermoote, S., Yang, X., Garcés de Marcilla, J.: The Copernicus Climate Change Service: Climate Science in Action. *Bull. Amer. Meteor. Soc.* (2022). <https://doi.org/10.1175/BAMS-D-21-0315.1>
- Cai, W., Santoso, A., Collins, M., Dewitte, B., Karamperidou, C., Kug, J.S., Lengaigne, M., McPhaden, M.J., Stuecker, M.F., Taschetto, A.S., Timmermann, A., Wu, L., Yeh, S.-W., Wang, G., Ng, B., Jia, F., Yang, Y., Ying, J., Zheng, X.-T., et al.: Changing El Niño–Southern Oscillation in a warming climate. *Nat. Rev. Earth Environ.* **2**, 628–644 (2021). <https://doi.org/10.1038/s43017-021-00199-z>
- Capotondi, A., Wittenberg, A.T., Kug, J.S., Takahashi, K., McPhaden, M.J.: ENSO diversity. In: McPhaden, M.J., Santoso, A., Cai, W. (eds.) *El Niño Southern Oscillation in a Changing Climate*. AGU Monograph. American Geophysical Union, Washington, D.C. (2020). <https://doi.org/10.1002/9781119548164.ch4>
- Chen, M., Kumar, A.: Winter 2015/16 atmospheric and precipitation anomalies over North America: El Niño response and the role of noise. *Mon. Weather Rev.* **146**, 909–927 (2018). <https://doi.org/10.1175/MWR-D-17-0116.1>
- Chen, D., Cane, M.A., Kaplan, A., Zebiak, S.E., Huang, D.: Predictability of El Niño over the past 148 years. *Nature*. **428**, 733–736 (2004). <https://doi.org/10.1038/nature02439>
- Collins, M., An, S.I., Cai, W., Ganachaud, A., Guilyardi, E., Jin, F.F., Jochum, M., Lengaigne, M., Power, S., Timmermann, A., Vecchi, G.: The impact of global warming on the tropical Pacific and El Niño. *Nat. Geosci.* **3**, 391–397 (2010). <https://doi.org/10.1038/ngeo868>
- Davey, M.K., Brookshaw, A., Ineson, S.: The probability of the impact of ENSO on precipitation and near-surface temperature. *Clim. Risk Manag.* **1**, 5–24 (2014). <https://doi.org/10.1016/j.crm.2013.12.002>
- DelSole, T., Tippett, M.: *Statistical Methods for Climate Scientists*. Cambridge University Press, Cambridge (2022)
- Deser, C., Simpson, I.R., McKinnon, K.A., 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. Clim.* **30**, 5059–5082 (2017). <https://doi.org/10.1175/JCLI-D-16-0844.1>
- Deser, C., Simpson, I.R., Phillips, A.S., McKinnon, K.A.: How well do we know ENSO's climate impacts over North America, and how do we evaluate models accordingly? *J. Clim.* **31**, 4991–5014 (2018). <https://doi.org/10.1175/JCLI-D-17-0783.1>
- Deser, C., Lehner, F., Rodgers, K.B., Ault, T., Delworth, T.L., DiNezio, P.N., Fiore, A., Frankignoul, C., Fyfe, J.C., Horton, D.E., Kay, J.E., Knutti, R., Lovenduski, N.S., Marotzke, J., McKinnon, K.A., Minobe, S., Randerson, J., Screen, J.A., Simpson, I.R., Ting, M.: Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Chang.* **10**, 277–286 (2020). <https://doi.org/10.1038/s41558-020-0731-2>
- Douglass, D.H.: El Niño–southern oscillation: magnitudes and asymmetry. *J. Geophys. Res.* **115**, D15111 (2010). <https://doi.org/10.1029/2009JD013508>
- Emile-Geay, J., Cobb, K.M., Cole, J.E., Elliot, M., Zhu, F.: Past ENSO variability: reconstructions, models, and implications. In: McPhaden, M.J., Santoso, A., Cai, W. (eds.) *El Niño Southern Oscillation in a Changing Climate*. AGU Monograph. American Geophysical Union, Washington, D.C. (2020). <https://doi.org/10.1002/9781119548164.ch5>
- Fedorov, A.V., Hu, S., Wittenberg, A.T., Levine, A.F., Deser, C.: ENSO low-frequency modulation and mean state interactions. In: McPhaden, M.J., Santoso, A., Cai, W. (eds.) *El Niño Southern Oscillation in a Changing Climate*. AGU Monograph. American Geophysical Union, Washington, D.C. (2020). <https://doi.org/10.1002/9781119548164.ch8>
- Fröhlich, K., Dobrynin, M., Isensee, K., Gessner, C., Paxian, A., Pohlmann, H., Haak, H., Brune, S., Fruh, B., Baehr, J.: The German climate forecast system: GCFS. *J. Adv. Model. Earth Syst.* **13**, e2020MS002101 (2021). <https://doi.org/10.1029/2020MS002101>
- Giese, B.S., Compo, G.P., Slowey, N.C., Sardeshmukh, P.D., Carton, J.A., Ray, S., Whitaker, J.S.: The 1918/19 El Niño. *Bull. Am. Meteor. Soc.* **91**, 177–183 (2010). <https://doi.org/10.1175/2009BAMS2903.1>
- Goddard, L., Gershunov, A.: Impact of El Niño on Weather and Climate Extremes. El Niño Southern Oscillation in a Changing Climate, 361–375 (2020). In: M.J. McPhaden, A. Santoso and W. Cai (Eds.) *El Niño Southern Oscillation in a Changing Climate*. AGU Monograph. American Geophysical Union, Washington, D.C. (2020). <https://doi.org/10.1002/9781119548164.ch16>
- Hermanson, L., Ren, H.-L., Vellinga, M., Dunstone, N.D., Hyder, P., Ineson, S., Scaife, A.A., Smith, D.M., Thompson, V., Tian, B., Williams, K.D.: Different types of drifts in two seasonal forecast systems and their dependence on ENSO. *Clim. Dyn.* **51**, 1411–1426 (2018). <https://doi.org/10.1007/s00382-017-3962-9>
- Hoell, A., Hoerling, M., Eischeid, J., Wolter, K., Dole, R., Perlwitz, J., Xu, T., Cheng, L.: Does El Niño intensity matter for California precipitation? *Geophys. Res. Lett.* **43**, 819–825 (2016). <https://doi.org/10.1002/2015GL067102>
- Huang, B., Thorne, P.W., Banzon, V.F., Boyer, T., Chepurin, G., Lawrimore, J.H., Menne, M.J., Smith, T.M., Vose, R.S., Zhang, H.M.: Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *J. Clim.* **30**, 8179–8205 (2017). <https://doi.org/10.1175/JCLI-D-16-0836.1>
- Huang, B., Liu, C., Banzon, V.F., Freeman, E., Graham, G., Hankins, B., Smith, T.M., Zhang, H.-M.: NOAA 0.25-Degree Daily Optimum Interpolation Sea Surface Temperature (OISST), Version 2.1. Monthly Values. NOAA National Centers for Environmental Information (2020). <https://doi.org/10.25921/RE9P-PT57>
- Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., Zhang, H.-M.: Improvements of the daily optimum Interpolation Sea surface temperature (DOISST) version

- 2.1. *J. Clim.* **34**, 2923–2939 (2021). <https://doi.org/10.1175/JCLI-D-20-0166.1>
- Jain, S., Scaife, A.A.: How extreme could the near term evolution of the Indian summer monsoon rainfall be? *Environ. Res. Lett.* **17**, 034009 (2022). <https://doi.org/10.1088/1748-9326/ac4655>
- Jiang, W., Huang, P., Huang, G., Ying, J.: Origins of the excessive westward extension of ENSO SST simulated in CMIP5 and CMIP6 models. *J. Clim.* **34**, 2839–2851 (2021). <https://doi.org/10.1175/JCLI-D-20-0551.1>
- Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L., Tietsche, S., Decremier, D., Weisheimer, A., Balsamo, G., Keeley, S.P.E., Mogensen, K., Zuo, H., Monge-Sanz, B.M.: SEAS5: the new ECMWF seasonal forecast system. *Geosci. Model Dev.* **12**, 1087–1117 (2019). <https://doi.org/10.5194/gmd-12-1087-2019>
- Kay, G., Dunstone, N.J., Smith, D.M., Betts, R.A., Cunningham, C., Scaife, A.A.: Assessing the chance of unprecedented dry conditions over North Brazil during El Niño events. *Environ. Res. Lett.* **17**, 064016 (2022). <https://doi.org/10.1088/1748-9326/ac6df9>
- Kumar, A., Chen, M.: What is the variability in U.S. west coast winter precipitation during strong El Niño events? *Clim. Dyn.* **49**, 2789–2802 (2017). <https://doi.org/10.1007/s00382-016-3485-9>
- Kumar, A., Chen, M.: Understanding skill of seasonal mean precipitation prediction over California during boreal winter and role of predictability limits. *J. Clim.* **33**, 6141–6163 (2020). <https://doi.org/10.1175/JCLI-D-19-0275.1>
- Kumar, A., Hoerling, M.: Interpretation and implications of observed inter-El Niño variability. *J. Clim.* **10**, 83–91 (1997). [https://doi.org/10.1175/1520-0442\(1997\)010%3C0083:IAIOTO%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010%3C0083:IAIOTO%3E2.0.CO;2)
- L'Heureux, M.L., Levine, A.F., Newman, M., Ganter, C., Luo, J.J., Tippett, M.K., Stockdale, T.N.: ENSO prediction. In: McPhaden, M.J., Santoso, A., Cai, W. (eds.) *El Niño Southern Oscillation in a Changing Climate*. AGU Monograph. American Geophysical Union, Washington, D.C. (2020). <https://doi.org/10.1002/9781119548164.ch10>
- Lin, H., Merryfield, W.J., Muncaster, R., Smith, G.C., Markovic, M., Dupont, F., Roy, F., Lemieux, J.-F., Dirkson, A., Kharin, V.V., Lee, W.-S., Charron, M., Erfani, A.: The Canadian seasonal to interannual prediction system version 2 (CanSIPSv2). *Weather Forecast.* **35**, 1317–1343 (2020). <https://doi.org/10.1175/WAF-D-19-0259.1>
- Lin, H., Muncaster, R., Diro, G.T., Merryfield, W.J., Smith, G., Markovic, M., Erfani, A., Kharin, S., Lee, W.-S., Parent, R., Pavlovic, R., Charron, M.: The Canadian Seasonal to Interannual Prediction System version 2.1 (CanSIPSv2.1). Technical Report, Canadian Centre for Meteorological and Environmental Prediction (2021). https://collaboration.cmc.ec.gc.ca/cmc/cmoe/product_guide/docs/tech_notes/technote_cansips-210_e.pdf. Accessed 27 Oct 2022
- Lindsey, R.: In watching for El Niño and La Niña, NOAA adapts to global warming. *ClimateWatch Magazine*. <https://www.climate.gov/news-features/understanding-climate/watching-el-ni%C3%B1o-and-la-ni%C3%B1a-noaa-adapts-global-warming> (2013). Accessed 5 March 2023
- Liu, Z., Chan, H.: Power analysis for testing two independent groups of likert-type data. *Proceedings of the 2015 5th International Conference on Computer Sciences and Automation Engineering*, 34–39 (2016). <https://doi.org/10.2991/iccsae-15.2016.8>
- Ma, H.-Y., Cheska Siongco, A., Klein, S.A., Xie, S., Karspeck, A.R., Raeder, K., Anderson, J.L., Lee, J., Kirtman, B.P., Merryfield, W.J., Murakami, H., Tribbia, J.J.: On the correspondence between seasonal forecast biases and long-term climate biases in sea surface temperature. *J. Clim.* **34**, 427–446 (2020). <https://doi.org/10.1175/JCLI-D-20-0338.1>
- MacLachlan, C., Arribas, A., Peterson, K.A., Maidens, A., Fereday, D., Scaife, A.A., Gordon, M., Vellinga, M., Williams, A., Comer, R.E., Camp, J., Xavier, P., Madec, G.: Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Q. J. Roy. Meteor. Soc.* **141**, 1072–1084 (2015). <https://doi.org/10.1002/qj.2396>
- Mason, S.J., Goddard, L.: Probabilistic precipitation anomalies associated with ENSO. *Bull. Amer. Meteor. Soc.* **82**, 619–638 (2001). [https://doi.org/10.1175/1520-0477\(2001\)082%3C0619:PPAAWE%3E2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082%3C0619:PPAAWE%3E2.3.CO;2)
- McPhaden, M.J., Zebiak, S.E., Glantz, M.H.: ENSO as an integrating concept in earth science. *Science*. **314**, 1740–1745 (2006). <https://doi.org/10.1126/science.1132588>
- McPhaden, M.J., Santoso, A., Cai, W.: Introduction to El Niño southern oscillation in a changing climate. In: McPhaden, M.J., Santoso, A., Cai, W. (eds.) *El Niño Southern Oscillation in a Changing Climate*. AGU Monograph. American Geophysical Union, Washington, D.C. (2020). <https://doi.org/10.1002/9781119548164.ch1>
- Meehl, G.A., Richter, J.H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., Donat, M.G., England, M.H., Fyfe, J.C., Han, W., Kim, H., Kirtman, B.P., Kushnir, Y., Lovenduski, N.S., Mann, M.E., Merryfield, W.J., Nieves, V., Kathy, P., Rosenbloom, N., et al.: Initialized earth system prediction from subseasonal to decadal timescales. *Nat. Rev. Earth Environ.* **2**, 340–357 (2021). <https://doi.org/10.1038/s43017-021-00155-x>
- Merryfield, W.J., Lee, W.-S., Boer, G.J., Kharin, V.V., Scinocca, J.F., Flato, G.M., Ajayamohan, R.S., Fyfe, J.C., Tang, Y., Polavarapu, S.: The Canadian seasonal to interannual prediction system. Part I: models and initialization. *Mon. Weather Rev.* **141**, 2910–2945 (2013). <https://doi.org/10.1175/MWR-D-12-00216.1>
- Planton, Y.Y., Guilyardi, E., Wittenberg, A.T., Lee, J., Gleckler, P.J., Bayr, T., McGregor, S., McPhaden, M.J., Power, S., Roehrig, R., Vialard, J., Voltaire, A.: Evaluating climate models with the CLIVAR 2020 ENSO metrics package. *Bull. Am. Meteorol. Soc.* **102**, E193–E217 (2021). <https://doi.org/10.1175/BAMS-D-19-0337.1>
- Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C., Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407 (2003). <https://doi.org/10.1029/2002JD002670>
- Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., Wang, W.: An improved in situ and satellite SST analysis for climate. *J. Clim.* **15**, 1609–1625 (2002). [https://doi.org/10.1175/1520-0442\(2002\)015%3C1609:AIISAS%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015%3C1609:AIISAS%3E2.0.CO;2)
- Reynolds, R.W., Smith, T.M., Liu, C., Chelton, C.B., Casey, K.S., Schlax, M.G.: Daily high-resolution-blended analyses for sea surface temperature. *J. Clim.* **20**, 5473–5496 (2007). <https://doi.org/10.1175/2007JCLI1824.1>
- Rodrigues, R.R., Subramanian, A., Zanna, L., Berner, J.: ENSO bimodality and extremes. *Geophys. Res. Lett.* **46**, 4883–4893 (2019). <https://doi.org/10.1029/2019GL082270>
- Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-Y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M.P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., Becker, E.: The NCEP climate forecast system version 2. *J. Clim.* **27**, 2185–2208 (2014). <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Sanna, A., Borrelli, A., Athanasiadis, P., Matera, S., Storto, A., Navarra, A., Tibaldi, S., Gualdi, S.: CMCC-SPS3: The CMCC Seasonal Prediction System 3. CMCC Technical Note RP0285 (2017)
- Santoso, A., McPhaden, M.J., Cai, W.: The defining characteristics of ENSO extremes and the strong 2015/2016 El Niño. *Rev. Geophys.* **55**, 1079–1129 (2017). <https://doi.org/10.1002/2017RG000560>
- Saurral, R.I., Merryfield, W.J., Tolstykh, M.A., Lee, W.-S., Doblas-Reyes, F.J., García-Serrano, J., Massonnet, F., Meehl, G.A., Teng, H.: A data set for intercomparing the transient behavior of dynamical model-based subseasonal to decadal climate predictions. *J. Adv. Model. Earth Syst.* **13**, e2021MS002570 (2021). <https://doi.org/10.1029/2021MS002570>

- Singh, D., Ting, M., Scaife, A.A., Martin, N.: California winter precipitation predictability: insights from the anomalous 2015–2016 and 2016–2017 seasons. *Geophys. Res. Lett.* **45**, 9972–9980 (2018). <https://doi.org/10.1029/2018GL078844>
- Siongco, A.C., Ma, H.-Y., Klein, S.A., Xie, S., Karspeck, A.R., Raeder, K., Anderson, J.L.: A hindcast approach to diagnosing the equatorial Pacific cold tongue SST bias in CESM1. *J. Clim.* **33**, 1437–1453 (2020). <https://doi.org/10.1175/JCLI-D-19-0513.1>
- Smith, S.R., Legler, D.M., Remigio, M.J., O'Brien, J.J.: Comparison of 1997–98 US temperature and precipitation anomalies to historical ENSO warm phases. *J. Clim.* **12**, 3507–3515 (1999). [https://doi.org/10.1175/1520-0442\(1999\)012%3C3507:COUSTA%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012%3C3507:COUSTA%3E2.0.CO;2)
- Taschetto, S.A., Ummerhofer, C.C., Stuecker, M.F., Dommenget, D., Ashok, K., Rodrigues, R.R., Yeh, S.-W.: ENSO atmospheric teleconnections. In: McPhaden, M.J., Santoso, A., Cai, W. (eds.) *El Niño Southern Oscillation in a Changing Climate*. AGU Monograph. American Geophysical Union, Washington, D.C. (2020). <https://doi.org/10.1002/9781119548164.ch14>
- Thompson, V., Dunstone, N., Scaife, A.A., Smith, D., Slingo, J., Brown, S., Belcher, S.E.: High risk of unprecedented UK rainfall in the current climate. *Nat. Comm.* **8**, 107 (2017). <https://doi.org/10.1038/s41467-017-00275-3>
- Thompson, V., Dunstone, N.J., Scaife, A.A., Smith, D.M., Hardiman, S.C., Ren, H.-L., Lu, B., Belcher, S.E.: Risk and dynamics of unprecedented hot months in south East China. *Clim. Dyn.* **52**, 2585–2596 (2019). <https://doi.org/10.1007/s00382-018-4281-5>
- Trenberth, K.E.: ENSO in the global climate system. El Niño southern oscillation in a changing climate. In: McPhaden, M.J., Santoso, A., Cai, W. (eds.) *El Niño Southern Oscillation in a Changing Climate*. AGU Monograph. American Geophysical Union, Washington, D.C. (2020). <https://doi.org/10.1002/9781119548164.ch2>
- van Oldenborgh, G.J., Hendon, H., Stockdale, T., L'Heureux, M., De Perez, E.C., Singh, R., Van Aalst, M.: Defining El Niño indices in a warming climate. *Environ. Res. Lett.* **16**, 044003 (2021). <https://doi.org/10.1088/1748-9326/abe9ed>
- Vannière, B., Guilyardi, E., Madec, G., Doblas-Reyes, F.J., Woolnough, S.: Using seasonal hindcasts to understand the origin of the equatorial cold tongue bias in CGCMs and its impact on ENSO. *Clim. Dyn.* **40**, 963–981 (2013). <https://doi.org/10.1007/s00382-012-1429-6>
- Weisheimer, A., Befort, D.J., MacLeod, D., Palmer, T., O'Reilly, C., Strømmen, K.: Seasonal forecasts of the twentieth century. *Bull. Am. Meteorol. Soc.* **101**, E1413–E1426 (2020). <https://doi.org/10.1175/BAMS-D-19-0019.1>
- Wittenberg, A.T.: Are historical records sufficient to constrain ENSO simulations? *Geophys. Res. Lett.* **36**, L12702 (2009). <https://doi.org/10.1029/2009GL038710>
- Wu, X., Okumura, Y.M., DiNezio, P.N., Yeager, S.G., Deser, C.: The equatorial Pacific cold tongue bias in CESM1 and its influence on ENSO forecasts. *J. Clim.* **35**, 3261–3277 (2022). <https://doi.org/10.1175/JCLI-D-21-0470.1>
- Yang, C., Leonelli, F.E., Marullo, S., Artale, V., Beggs, H., Nardelli, B.B., Chin, T.M., De Toma, V., Good, S., Huang, B., Merchant, C.J.: Sea surface temperature intercomparison in the framework of the Copernicus Climate Change Service (C3S). *J. Clim.* **34**, 5257–5283 (2021). <https://doi.org/10.1175/JCLI-D-20-0793.1>
- Ying, J., Huang, P., Lian, T., Tan, H.: Understanding the effect of an excessive cold tongue bias on projecting the tropical Pacific SST warming pattern in CMIP5 models. *Clim. Dyn.* **52**, 1805–1818 (2019). <https://doi.org/10.1007/s00382-018-4219-y>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

