

# A Century in Hindcast

## Building a Suitable Test for Seasonal Forecasts

Adapted from “Seasonal Forecasts of the Twentieth Century,” by **Antje Weisheimer** (University of Oxford and European Centre for Medium-Range Weather Forecasts), **Daniel J. Bafort**, **Dave MacLeod**, **Tim Palmer**, **Chris O’Reilly**, and **Kristian Strømmen**. Published online in *BAMS*, August 2020. For the full, citable article, see [DOI:10.1175/BAMS-D-19-0019.1](https://doi.org/10.1175/BAMS-D-19-0019.1).

A crucial component of any seasonal forecast system is the set of retrospective forecasts, or hindcasts, used for calibration and estimating skill. These hindcasts are usually produced over a period of around 20–30 years.

However, this time span of hindcasts can be too short for the evaluation of sporadic drivers such as El Niño, the single most important source of predictability on seasonal time scales. Furthermore, seasonal forecast skill can undergo pronounced multidecadal variations.

Because the relatively short hindcasts are not adequate for reliably testing seasonal forecasts and producing robust skill estimates, we have instead created hindcasts based on climate reanalyses for a much longer period, 110 years in this case. These Coupled Seasonal Forecasts of the 20th Century (CSF-20C), for the years 1901–2010, are ideal datasets to initialize, verify, and test seasonal forecasts.

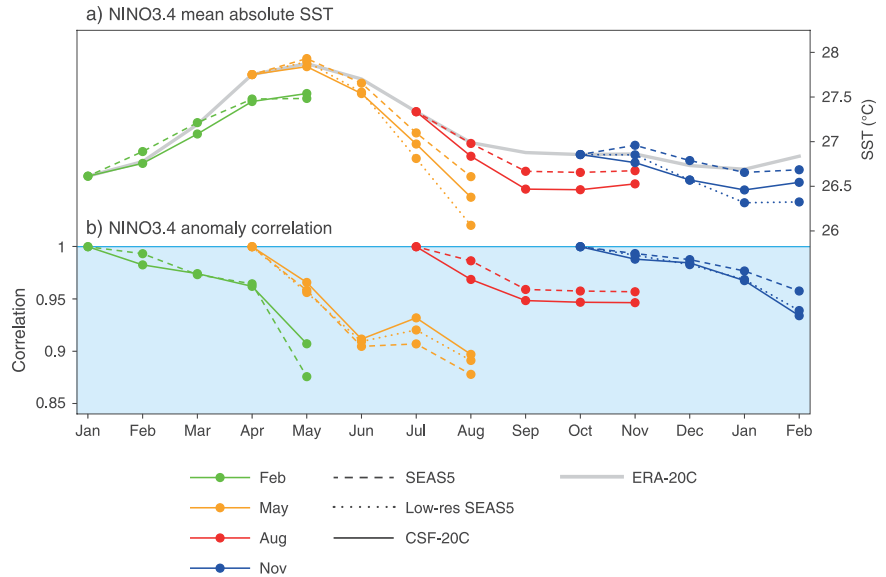
### The CSF-20C hindcasts

We performed CSF-20C experiments with ECMWF’s coupled model, including state-of-the-art atmospheric, land surface, oceanic, and sea ice components. The atmospheric resolution for these runs is approximately 80 km

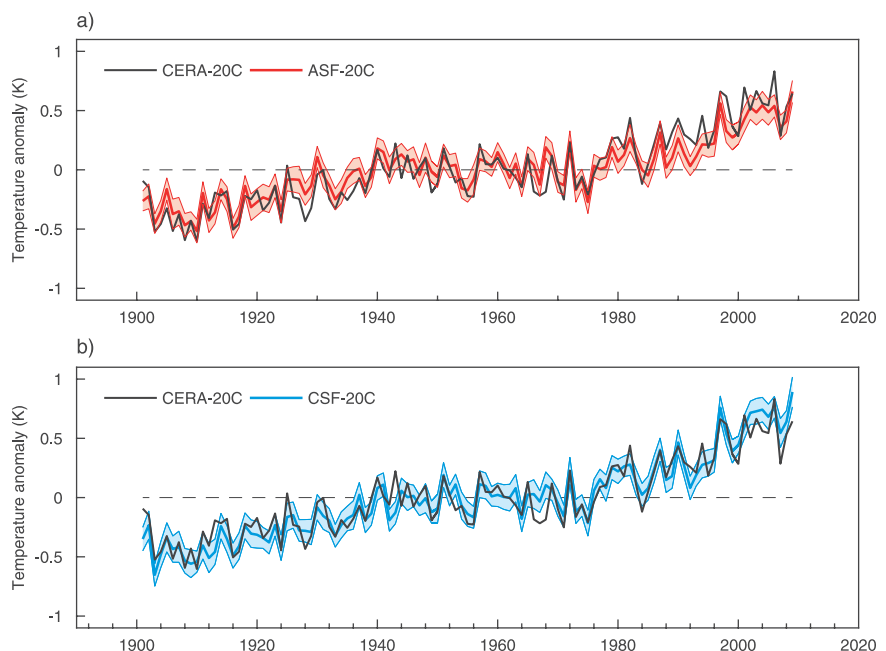
horizontally with 91 vertical levels. The ocean resolution is 1° horizontally with 42 vertical levels. All model components of CSF-20C were initialized with ECMWF's first coupled reanalysis of the twentieth century, CERA-20C, which provides data from 1901 to 2010. Only conventional surface observations in the atmosphere were assimilated in CERA-20C. The hindcasts were initialized on the first of February, May, August, and November from 1901 to 2010 and run for 4 forecast months so that they cover all seasons. They consist of ensembles of 51 members for the May and November start dates and 25 members for the February and August start dates. The experiment we set up mimics ECMWF's operational forecasts as much as possible to enable a clear comparison with a real-time forecasting system.

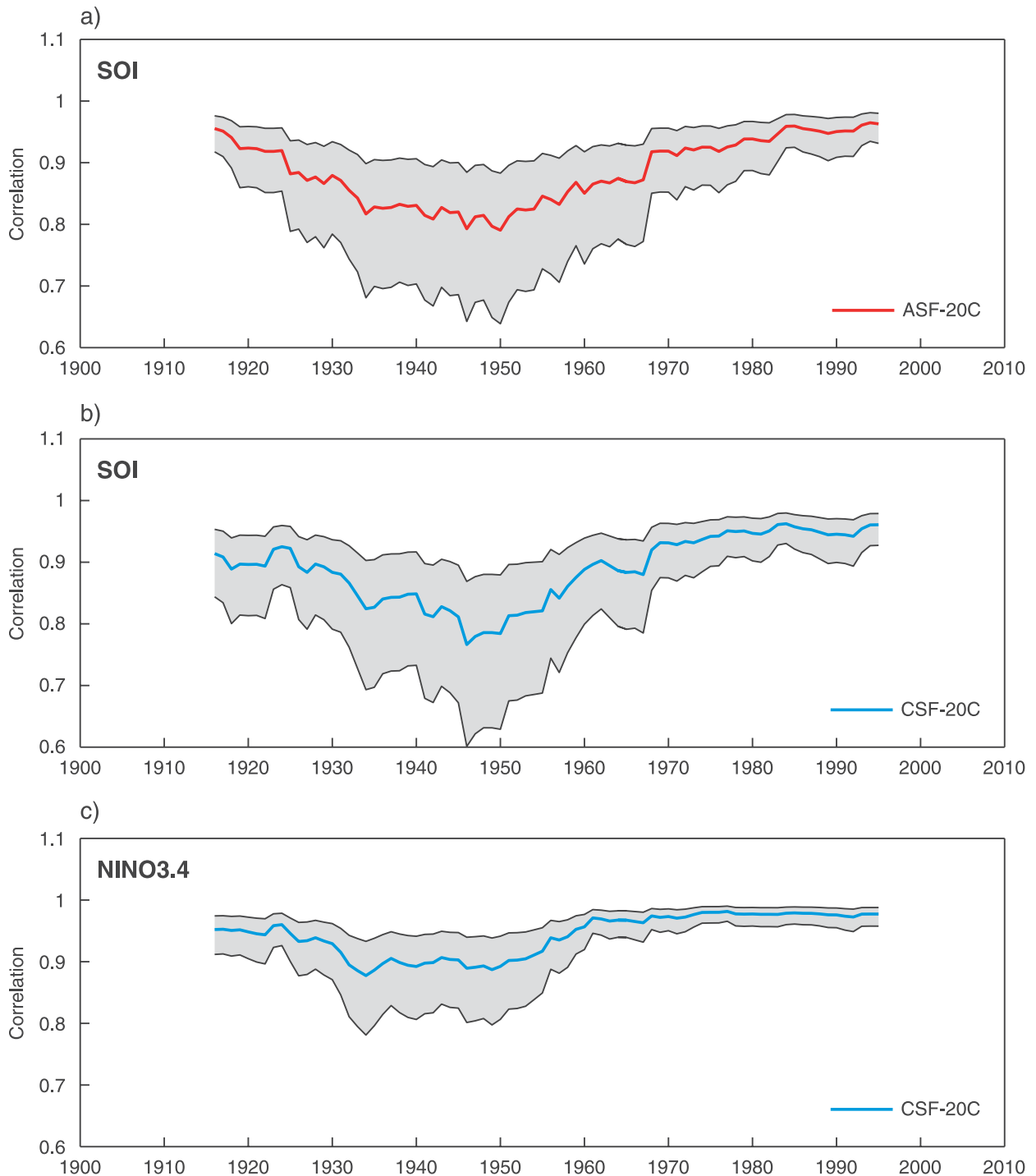
For comparison, a companion hindcast experiment uses an identical model setup but with prescribed observed sea surface temperatures (SSTs) instead of a fully coupled ocean. These Atmospheric Seasonal Forecasts of the 20th Century (ASF-20C) have been initialized using ERA-20C, ECMWF's

**There is, in general, very good \* agreement of the interannual variability of global-mean (60°N–60°S) 2-m temperature in the global coupled reanalysis (CERA-20C; black) and both seasonal forecast hindcast systems, (a) the atmosphere-only seasonal hindcast (ASF-20C) and (b) the coupled seasonal hindcast (CSF-20C). Hindcasts were initialized in November. Ensemble means are plotted with thick lines and the ensemble spread around the mean is given by the shaded band.**



**▲ \* (a) Biases, or drift, and (b) anomaly correlation skill of predictions of the Niño-3.4 SST index for different start dates throughout the calendar year during the hindcast period 1981–2009. CSF-20C data are shown in solid colored lines, ECMWF's currently operational seasonal forecasting system, SEAS5, in dashed lines, and a lower-resolution version of SEAS5 in dotted lines (only available for May and November start dates). Different colors indicate different start dates of the hindcasts. The gray curve in (a) shows the climatological mean evolution of the SSTs in ERA-20C over this period. The ECMWF (the model that's the basis for all of these systems) has a cold bias that varies throughout the calendar year and which is smallest for the high-resolution SEAS5. CSF-20C shows a slightly reduced cold bias compared with the low-resolution SEAS5 hindcasts, which could be the benefit of initializing with a coupled atmosphere–ocean reanalysis in CSF-20C, rather than using separate reanalyses for the atmosphere and the ocean as in SEAS5.**





**▲ \* Decadal variability of forecast skill in December–February from hindcasts initialized in November, for atmospheric (SOI) and oceanic (NINO3.4SST) ENSO indices. Values express the 30-year moving-window correlation coefficients of the hindcast ensemble mean with the global coupled reanalysis CERA-20C, plotted in the central year for each 30-year window. (The gray-shaded bands indicate the 5%–95% confidence intervals.) For the period 1980–2009, the hindcast skill for all three indices is very high, including the atmosphere-only seasonal hindcasts (ASF-20C). The coupled CSF-20C system (in blue, panel b) maintained these high levels of skill for several decades in the past with a remarkable constant level of very high NINO3.4 SST skill (panel c) until the early 1960s, implying (as shown elsewhere) that ENSO skill was largely unaffected by the fewer ocean observations during these decades. We hypothesize instead that these changes in model predictability are linked to intrinsic changes of the coupled climate system.**

first atmospheric reanalysis of the twentieth century.

## Performance

The hindcasts have comparable skill to state-of-the-art operational seasonal forecasts. The skill of these coupled seasonal forecasts undergoes pronounced decadal variability, confirming the experiments with atmosphere-only models that used prescribed SSTs.

The ability of any seasonal forecasting system to realistically reproduce the behavior of ENSO is of paramount importance, and the new coupled system makes very skillful predictions of ENSO during recent periods and at the beginning of the twentieth Century but has reduced skill during the 1920s and 1930s. Forecast skill also dropped noticeably during the middle of the twentieth century for extratropical circulation indices such as the North Atlantic Oscillation (NAO) and the Pacific–North American (PNA) pattern. This fluctuation in skill cannot be simply explained by the poor quality of past observational datasets.

## Conclusions

The ASF-20C and CSF-20C datasets, with their many hindcast years and large ensemble size, apply to the study of a wide range of predictability problems with a higher level of robustness and confidence than would otherwise be possible. In addition to investigating the nonstationary forecast skill behavior, such prospective studies could, for example, involve the assessment of extremes, exploit a range of applications of seasonal forecasts, or train modern machine learning techniques. We invite the scientific community to explore the datasets for their specific research interests, which could include regional assessments of predictability, physical mechanisms in large-scale variability, or the use of modern machine learning techniques.

Data from ASF-20C and CSF-20C have become publicly available through a dedicated online dissemination platform hosted by the CEDA Archive.<sup>1</sup> We invite the wider international scientific community to explore the plethora of outputs from these novel simulations. ●●

<sup>1</sup> The Natural Environment Research Council Data Repository for Atmospheric Science and Earth Observation, <https://catalogue.ceda.ac.uk/uuid/76e1c3df49f644a0f812818080bed5e45>.

## METADATA

**BAMS:** What would you like readers to learn from this article?

**Antje Weisheimer (University of Oxford and European Centre for Medium-Range Weather Forecasts):** *Confidence in seasonal forecasts can be built from the performance of these forecasts in the past. We normally test our models over hindcast periods of the recent 20–30 years in order to estimate the skill of the forecasts. However, seasonal forecast skill varies over time, and a good performance during the last 30 years is no guarantee for a similar good performance in the years to come. In this article we introduce a new set of unprecedented long seasonal hindcasts that cover the 110-year period, 1901 to 2010. These hindcasts show that although forecast skill at the beginning of the twentieth century is similarly high as for recent decades, there are epochs in midcentury when our abilities to predict ENSO as well as extratropical modes of variability were greatly reduced. These fluctuations in skill cannot be simply explained by the poorer quality of past observational data. Understanding the reasons behind these changes is paramount to improve the forecast quality in the future.*

**BAMS:** How did you become interested in the topic of this article?

**AW:** *I have been doing research on seasonal forecasts for 15 years, and over that period of time I have experienced how challenging it is to improve the forecasts. ENSO in the tropical Pacific is known to be a major source of seasonal predictability, but its interactions with remote areas in the extratropics are the subject of a lot of ongoing research. The advent of century-long coupled reanalyses at ECMWF a few years ago provided an exciting opportunity to perform much longer seasonal hindcasts and to test the representation of global teleconnections within these simulations.*

**BAMS:** What surprises you the most about the work you document in this article?

**AW:** *Seasonal forecast skill for ENSO at the beginning of the twentieth century is nearly as high as it is for recent decades—this was certainly a surprising finding. The consistency of reduced skill in midcentury periods across the modes of variability we have looked at (ENSO, PNA, NAO) has also been a fascinating result.*