



# Evaluating skill in predicting the Interdecadal Pacific Oscillation in initialized decadal climate prediction hindcasts in E3SMv1 and CESM1 using two different initialization methods and a small set of start years

Gerald A. Meehl<sup>1</sup> · Ben Kirtman<sup>2</sup> · Anne A. Glanville<sup>1</sup> · Jadwiga Richter<sup>1</sup> · Nan Rosenbloom<sup>1</sup> · Stephen Yeager<sup>1</sup>

Received: 2 April 2023 / Accepted: 21 September 2023 / Published online: 29 September 2023  
© The Author(s) 2023

## Abstract

It is a daunting challenge to conduct initialized hindcasts with enough ensemble members and associated start years to form a drifted climatology from which to compute the anomalies necessary to quantify the skill of the hindcasts when compared to observations. This limits the ability to experiment with case studies and other applications where only a few initial years are needed. Here we run a set of hindcasts with CESM1 and E3SMv1 using two different initialization methods for a limited set of start years and use the respective uninitialized free-running historical simulations to form the model climatologies. Since the drifts from the observed initial states in the hindcasts toward the uninitialized model state are large and rapid, after a few years the drifted initialized models approach the uninitialized model climatological errors. Therefore, hindcasts from the limited start years can use the uninitialized climatology to represent the drifted model states after about lead year 3, providing a means to compute forecast anomalies in the absence of a large hindcast sample. There is comparable skill for predicting spatial patterns of multi-year Pacific sea surface temperature anomalies in the domain of the Interdecadal Pacific Oscillation using this method compared to the conventional methodology with a large hindcast data set, though there is a model dependence to the drifts in the two initialization methods.

**Keywords** Decadal prediction · Anomaly calculation · Prediction skill · Model error

## 1 Introduction

From the earliest days of initialized multi-year predictions with Earth system models, two of the biggest challenges have involved not only what initialization method to use, but also how to deal with the large drifts that result as a consequence of systematic errors in the models (e.g. Meehl et al. 2009, 2014, 2021). Several initialization methods have been applied (Meehl et al. 2021), and bias and drift can be removed by use of one of several different methods (Meehl et al. 2022). For the latter, the most common is to run a large hindcast set, form a model climatology of drifted states to compare to the lead year predictions to be

evaluated, compute anomalies, do a similar calculation for observations, and evaluate the hindcast skill of the model anomalies compared to the observed anomalies for the same period (e.g. Doblas-Reyes, 2013). Another is to compute mean drift adjustments from observations for the lead years of interest, remove those drifts from the hindcasts, and then compute anomalies from the previous 15 years of observations prior to the initial year of the prediction (e.g. Meehl et al. 2016). A third is a compromise between the first two whereby a mean drift is calculated in the hindcasts for the 15 model years prior to the initial year, and used as a reference to compute anomalies, with a similar calculation for the observations (Meehl et al. 2022). It has been noted that this third method is likely the most “fair” of the three methods in that computing a model climatology over the entire reference period is not what occurs in a real-time prediction exercise where you only have information prior to the initial year (Risbey et al. 2021). In an evaluation of those three methods, it was determined that the “fair” method (comparing to the previous 15 year climatologies) has somewhat

✉ Gerald A. Meehl  
meehl@ucar.edu

<sup>1</sup> National Center for Atmospheric Research, Boulder, CO, USA

<sup>2</sup> University of Miami, Miami, FL, USA

less skill as could be expected (Risbey et al. 2021), though all were roughly comparable with skill varying somewhat differently as a function of time, but with no method clearly superior to the others (Meehl et al. 2022).

To run a large set of hindcasts with which to form the climatologies to compute anomalies of the initialized predictions is a massive computational burden. Consequently, only occasionally is there enough computer time available to perform such a set of simulations. This limits the research that can be done if it is desired to perform case studies to study processes, or experiment with alternate model configurations and resolutions or different initialization methods. Here we first apply two different initialization methods in two different Earth system models (Energy Exascale Earth System Model version 1, E3SMv1; and the Community Earth System Model version 1, CESM1) with a limited set of start years to compare the model-dependence and initialization-dependence of the drifts. As the verification metric, we quantify the skill of predictions of spatial patterns of multi-year Pacific sea surface temperature (SST) anomalies in the domain of the Interdecadal Pacific Oscillation (IPO). The dominant pattern of decadal timescale SST variability in the Pacific is the IPO (Power et al. 1999). The IPO can be represented as the second EOF of low pass filtered non-detrended observed SSTs, with the PC time series from that EOF as an IPO index (e.g. Meehl and Arblaster 2011). The IPO SST anomaly pattern is characterized in its positive phase by positive SST anomalies in the tropical Pacific, and negative SST anomalies in the northwest and southwest Pacific, with opposite sign SST anomalies for the negative phase. The objective of using Pacific region SSTs as the verification metric is to be able to predict elements of this IPO SST pattern. We then use the much larger hindcast data set from the Decadal Prediction Large Ensemble (DPLE) with CESM1, and compute the drifted hindcasts to form a model climatology, and then calculate anomalies to compare to anomalies formed from the fully-drifted free-running historical simulations. The objective is to determine whether a smaller set of start years and computing a model climatology from the historical simulations is comparable to using a much larger set of start years with a model climatology computed from the drifted hindcasts to compute anomalies, and whether either makes a difference when using different initialization methodologies.

## 2 Models, simulations, and initialization methods

The two Earth system models analyzed here are E3SMv1 and CESM1. The components and coupled simulations of E3SMv1 are documented by Golaz et al. (2019). The

horizontal resolution of the E3SMv1 configuration analyzed here consists of a 110-km atmosphere, 165-km land, and a 0.5° river model. The MPAS ocean (Petersen et al. 2019) and sea ice models in E3SMv1 have a mesh spacing that varies from 60 km in the mid-latitudes and 30 km at the equator and poles. The 60 vertical layers use a z-star vertical coordinate.

E3SM was originally branched from CESM but has since evolved with significant developments in all model components. The E3SM atmospheric model (EAMv1) was built upon the CESM atmospheric model (CAM Version 5.3) but the two models diverge in details of tuning and cloud and aerosol formations, and EAMv1 uses a spectral element dynamical core. The ocean represents probably the biggest difference between E3SMv1 and CESM1 in that E3SMv1 uses an unstructured mesh ocean model (MPAS-Ocean) while CESM1 includes a conventional dipolar mesh grid ocean model (POP2) as discussed below. An overall description of CESM1 is provided by Hurrell et al. (2013). The atmospheric model in CESM1, as in E3SMv1, has a nominal 1 degree latitude-longitude resolution. Note the CESM atmospheric model (CAM5.3) uses a finite-volume dynamical core. The ocean model in CESM1 is a version of POP2 (Parallel Ocean Program, version 2). The ocean has a nominal 1 degree horizontal resolution and enhanced meridional resolution in the equatorial tropics, and 60 levels in the vertical with ocean biogeochemistry. While the meridional resolution in the MPAS ocean in E3SMv1 is very similar to POP2 in CESM1, the grid cells are much more isotropic in E3SMv1. This gives MPAS ocean in E3SMv1 a much higher effective resolution than that in CESM1. Other features of CESM1 involving land and sea ice are described in detail by Hurrell et al. (2013).

The predictions in E3SMv1 and CESM1 are initialized on November 1 of the start years listed in Table 1, and each uses two different initialization schemes. The first is the “forced ocean sea ice” (FOSI) method (Yeager et al. 2018). In this method the ocean model is run repeatedly over the historical period (1958–2018) with observed atmospheric forcings. With each forcing cycle, the upper ocean asymptotes closer to the climatological base state such that year-to-year variability increasingly reflects the ocean response to the imposed atmospheric conditions (Tsujino et al. 2020). By the fifth cycle (used for the initial states of the hindcasts), the ocean model evolution is dominated by the observed atmospheric forcing and resembles the observed time evolution of the ocean (Yeager et al. 2018). This method is relatively economical and leaves the ocean in a state that contains some of the model systematic errors but also retains elements of the observed ocean states.

The second initialization scheme is termed the “brute force” (BF) method (Kirtman and Min 2009; Paolino et

**Table 1** Initial years and number of ensemble members for each model and initialization method. Note that CESM1 FOSI is a subset of the CESM1 DPLE. The initial year refers to the start date of Nov. 1 of that year. Lead month 1 is the December of the initial year, and lead year 1 is the first full calendar year after the Nov. 1 start date. The historical large ensembles use 13 members from E3SMv1, and 33 members from CESM1 averaged from 1985–2017

Initial year	E3SMv1 BF number of ensemble members	E3SMv1 FOSI number of ensemble members	CESM1 BF number of ensemble members	CESM1 FOSI number of ensemble members
1985	5	5	5	5
1990	5	5	5	5
1995	5	5	5	5
2000	5	5	5	5
2005	3	3	5	5
2010	3	3	5	5
2015	3	3	5	5
2016	3	3	5	5
2017	3	3	5	5

al. 2012; Infanti and Kirtman 2017). The data source for the initialization is the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010), and in the above references the ocean, land and sea ice are initialized. In the application with E3SMv1 and CESM1 described here, only the ocean component is initialized. Atmosphere, land and sea-ice states are randomly sampled from long climate simulations.

The following is the procedure to generate the ocean restart file converted from the CFSR ocean data assimilation restart file. The same procedure is used for both E3SM and CESM. Both restart files have different resolutions in the horizontal and vertical. The CFSR fields have been interpolated horizontally and vertically using a bilinear interpolation scheme. Climatological data from long model simulations are used in regions where CFSR data is undefined with respect to the model grids. The surface pressure for the ocean component of the models is estimated using the sea surface height, and the pressure gradient terms are estimated using centered differencing. The salinity initial condition for model restarts is constructed by adding anomalies from CFSR, rescaled to have the same anomaly standard deviation as CESM1, to the model climatology. This approach has proven successful for NMME and SubX forecasts with CCSM4 at ocean eddy parameterized scales (e.g., Infanti and Kirtman 2017) and at ocean eddy resolving scales (e.g., Infanti and Kirtman 2019; Siqueira et al. 2021).

Of these two initialization methods, BF, where reanalyses are interpolated straight to the model grid, is the most economical and the ocean initial state is, by definition, very nearly the reanalysis observed state. But the initial drifts could be expected to be larger and more rapid as the model wants to quickly leave the observed state and get to its systematic error state. FOSI, on the other hand, is more expensive (but still economical compared to a data assimilation method) and the ocean initial state is a compromise between a representation of the observations imprinted on the ocean from the observed atmospheric forcing, and the ocean systematic error state. Thus, coupling shock and initial drifts

could be expected to be less compared to the BF method. But it is unclear if either shows clear advantages over the other.

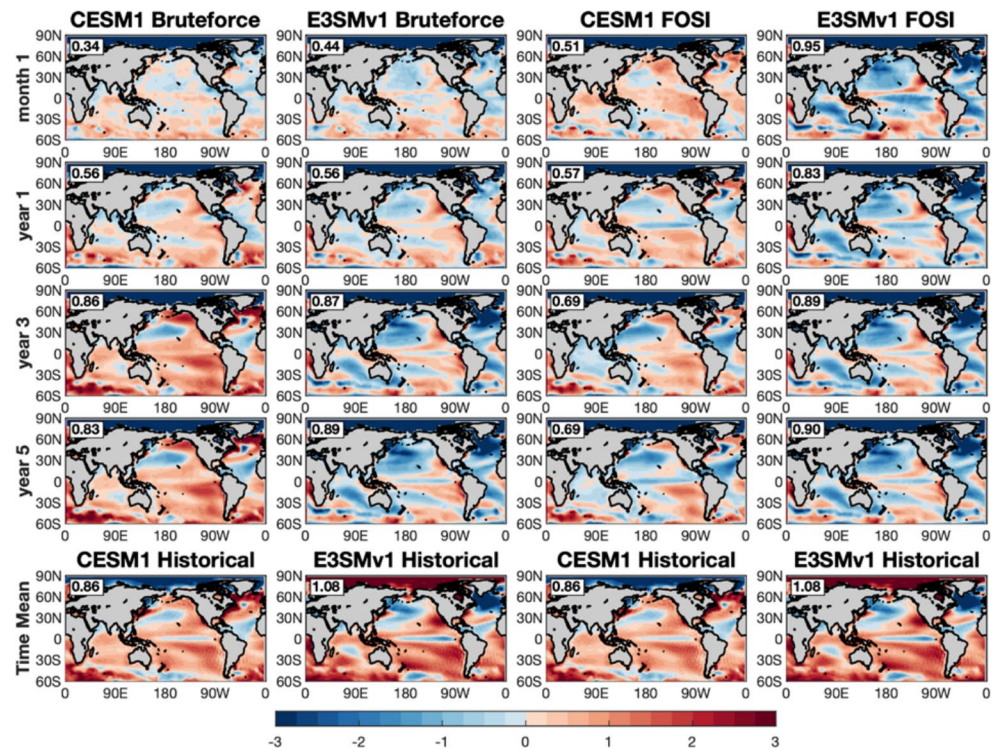
With regards to generating ensembles, the BF ensembles for E3SMv1 and CESM1 are generated using the CFSR atmosphere and land states chosen randomly from the extended free running coupled simulation, using the full fields from that atmosphere and land. For example, randomly chosen end of October states are used for the forecasts initialized at the beginning of November. No perturbations are made to the ocean and ice states. For the FOSI ocean initial states, E3SMv1 and CESM1 both generate different ensemble members through random small perturbations in the atmospheric model (Yeager et al. 2018).

Table 1 shows the initial years and number of ensemble members for the two models and two initialization methods. All cases are run for 5 years. Note that some start years have 5 ensemble members while others have 3. The smaller number of ensemble members in the latter will introduce some additional noise to the results. These will be compared to the much larger number of start years (every year from 1958 to 2017) and ensemble members (40 for each start year) from the CESM1 DPLE (Yeager et al. 2018). The CESM1 FOSI experiments are 5 member and 5 year subsets taken from DPLE. The CESM1 historical simulations are from the large ensemble (LE) described by Kay et al. (2015). Averages are taken from a subset of the LE with ensemble size of 33. The E3SMv1 historical runs are the standard CMIP6 simulations initialized from different years of the 1850 pre-industrial control with ensemble size of 13 described by Golaz et al. (2019).

### 3 Drifts from initialized states

As has been documented previously, models initialized close to some observed state tend to drift away from the initial state, with the drifts large and rapid (e.g. see Fig. 1

**Fig. 1** Ensemble mean SST bias as represented by drifts ( $^{\circ}\text{C}$ ) from the observations for start month of November for CESM1 BF (first column), E3SMv1 BF (second column), CESM1 FOSI (third column), and E3SMv1 FOSI (fourth column) for lead month 1 (first row, averaged for December monthly mean), lead year 1 (row 2), lead year 3 (row 3), and lead year 5 (row 4). Historical large ensemble systematic errors are shown in the fifth row, using 13 ensemble members from E3SMv1 and 33 from CESM1, averaged from 1985–2017. The cosine weighted  $60^{\circ}\text{S}$ – $60^{\circ}\text{N}$  root mean square error is shown in the upper left corner of each panel



in Meehl et al. 2022). To illustrate the drifts in the two models with the two initialization methods, Fig. 1 shows SST differences for lead month 1 (models are initialized on November 1, so lead month 1 is the first December after initialization), and lead years 1 (the first full calendar year after the November 1 initialization time), 3 and 5, for CESM1 BF, E3SMv1 BF, CESM1 FOSI, and E3SMv1 FOSI for all available ensemble members in Table 1. As noted above, since the FOSI method does not bring the model initial state into perfect agreement with observations at the time of initialization, and BF does, the SST errors are larger at lead month 1 (top row of Fig. 1) in both models with the FOSI method compared to the BF method. But even in the space of one month, the BF models have drifts on the order of about  $\pm 0.5$  C. Consequently, root-mean-square (RMS) errors for lead month 1 for CESM1 BF (0.34) and E3SMv1 BF (0.44) are smaller (i.e. closer to the observed state) than CESM1 FOSI (0.51) and E3SMv1 FOSI (0.95), the latter having larger lead month 1 errors than CESM1 FOSI.

What is apparent from Fig. 1 is that each model and initialization method settles into a preferred error pattern that is already recognizable at lead year 1 (second row of Fig. 1) as being similar to the systematic error pattern in the historical simulations (bottom row of Fig. 1). In fact, three (with the exception of CESM1 BF) have very similar drift patterns (the three rightmost columns in Fig. 1), with negative SST anomalies in the northwest Pacific, subtropical southwest Pacific, tropical North Atlantic, subtropical southern Indian Ocean, and equatorial Pacific. There are

positive anomalies in those three in the subtropical south-east and northeast Pacific, and subtropical eastern Atlantic. The CESM1 BF has some similarities to the other three in the Northern Hemisphere, but with mostly positive SST anomalies across the Southern Hemisphere oceans. Additionally, CESM1 BF SST biases asymptote more rapidly (by year 3) to the historical systematic errors in the bottom row than the other hindcast sets, particularly the E3SMv1 sets. This could be partially explained by the fact that the POP ocean is more diffusive, as noted earlier. Thus, the CESM1 drift from the observed BF state would be more rapid than the E3SMv1 BF or E3SMv1 FOSI. Meanwhile, the biggest difference in drifts between CESM1 BF and CESM1 FOSI lie in the Southern Hemisphere where CESM1 FOSI drifts less rapidly toward the eventual systematic error state in that region. This is likely due in part to the FOSI initialization fixing an initial error state that is farther from observations represented by BF in the Southern Hemisphere. Therefore, the BF state can drift more rapidly to the eventual error state than FOSI in CESM1.

Though the drift patterns set up early, the magnitudes take until about lead year 3 to stabilize. This can be seen by the RMS errors increasing from lower values in lead month 1, to being larger, as the errors grow, but very similar in lead years 3 and 5 in all four configurations, with differences in RMS errors all less than 4%. Thus, in Fig. 1, RMS errors in lead year 3 and 5 for CESM1 BF are 0.86 and 0.83; in E3SMv1 BF they are 0.87 and 0.89; in CESM1 FOSI they are both 0.69; and in E3SMv1 FOSI they are 0.89 and 0.90.

Another way of quantifying how the drifts stabilize in pattern and magnitude is by calculating the drift in lead month 1, lead year 1, and lead year 3 as a percent of the drift in lead year 5 for the four configurations (Fig. 2). The darker and more extensive the dark red colors in Fig. 2, the closer the drifts are to the eventual drift values in lead year 5. The globally averaged percentages of drift for ocean grid points are given at upper left of each panel. For lead year 3, these values range from 74.7% for CESM1 FOSI to 86.2% for CESM1 BF (where a value of 100% would indicate drift numbers equaling drift in lead year 5). For the latter, in lead year 3 almost all the global oceans are close to the eventual drifted values in lead year 5, but for the other three configurations there are notable areas where the drift is not yet very close to the eventual drift values in lead year 5. These areas include the equatorial, eastern subtropical, and the South Pacific Convergence Zone (SPCZ) areas of the Pacific Ocean. There are some interesting fluctuations of the drift in the two FOSI methods (the rightmost two columns) where drift percentages are near 80% in the equatorial eastern Pacific in lead year 1, and then are reduced to less than 30% in lead year 3. This suggests that the time-dependence of the drifts in those regions using the FOSI method in both models includes substantial non-monotonicity in the equatorial Pacific, possibly through ocean dynamics related to initialization shock (Teng et al. 2017), with this variation more pronounced in CESM1 FOSI.

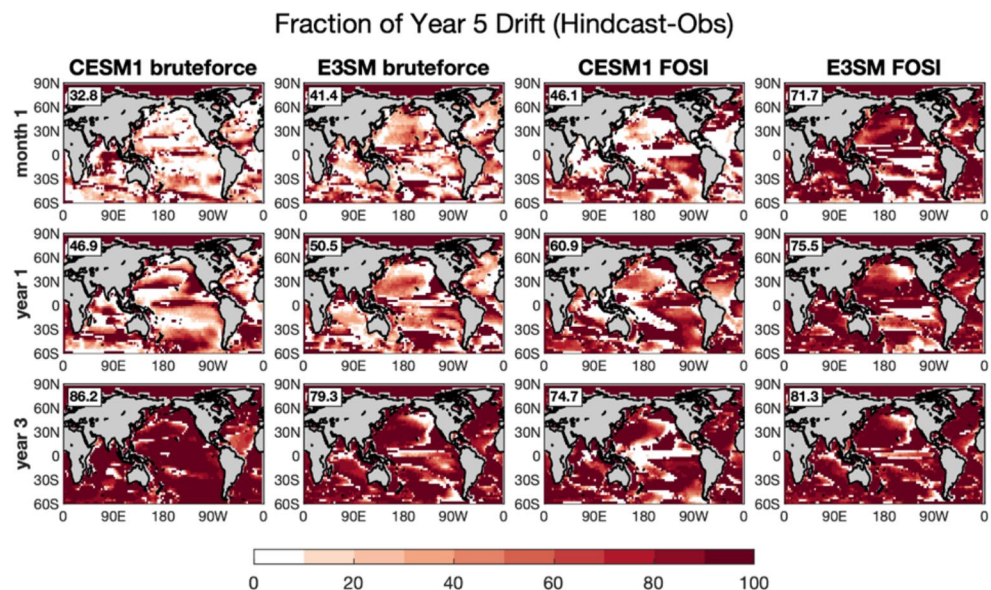
To further highlight differences in drift between the two initialization methods and two models, Fig. 3 shows when BF and FOSI drifts are the same sign (blue colors indicate both are negative or both are positive), while red colors denote drifts of opposite sign. For E3SMv1 in the right column, both for lead year 3 and lead year 5, nearly all ocean areas are the same sign (blue) indicating very similar

drift characteristics between the BF and FOSI initialization schemes. However, CESM1 (left column) shows differences of drift sign between the two initialization schemes mainly in areas of tropical precipitation maxima in the Indian and Pacific Oceans. This was noted in Fig. 1 where the lead year 5 drifts from CESM1 BF (bottom of left column in Fig. 1) show nearly all positive values in the Intertropical Convergence Zone (ITCZ), SPCZ and tropical Indian Ocean compared to negative values in CESM1 FOSI (bottom of second from right column). One possibility for this result is that the hindcast sample is too small to isolate drift, and the differences highlighted in Fig. 3 are due to noise. However, using all 64 start years for the hindcasts from DPLE for CESM1 shows nearly identical patterns to CESM1 FOSI with the smaller number of start years in the third column of Fig. 3 (not shown). Thus, the drifts are robust even when using a smaller number of start years, so the differences in Fig. 3 are not due to noise. This indicates that there likely is a difference in interannual drift variability (ENSO evolution) that is systematic and likely related to how initialization shock manifests in FOSI vs. BF which also has been seen in other models (e.g. the CanCM4, Fig. 15 in Saurral et al. 2021).

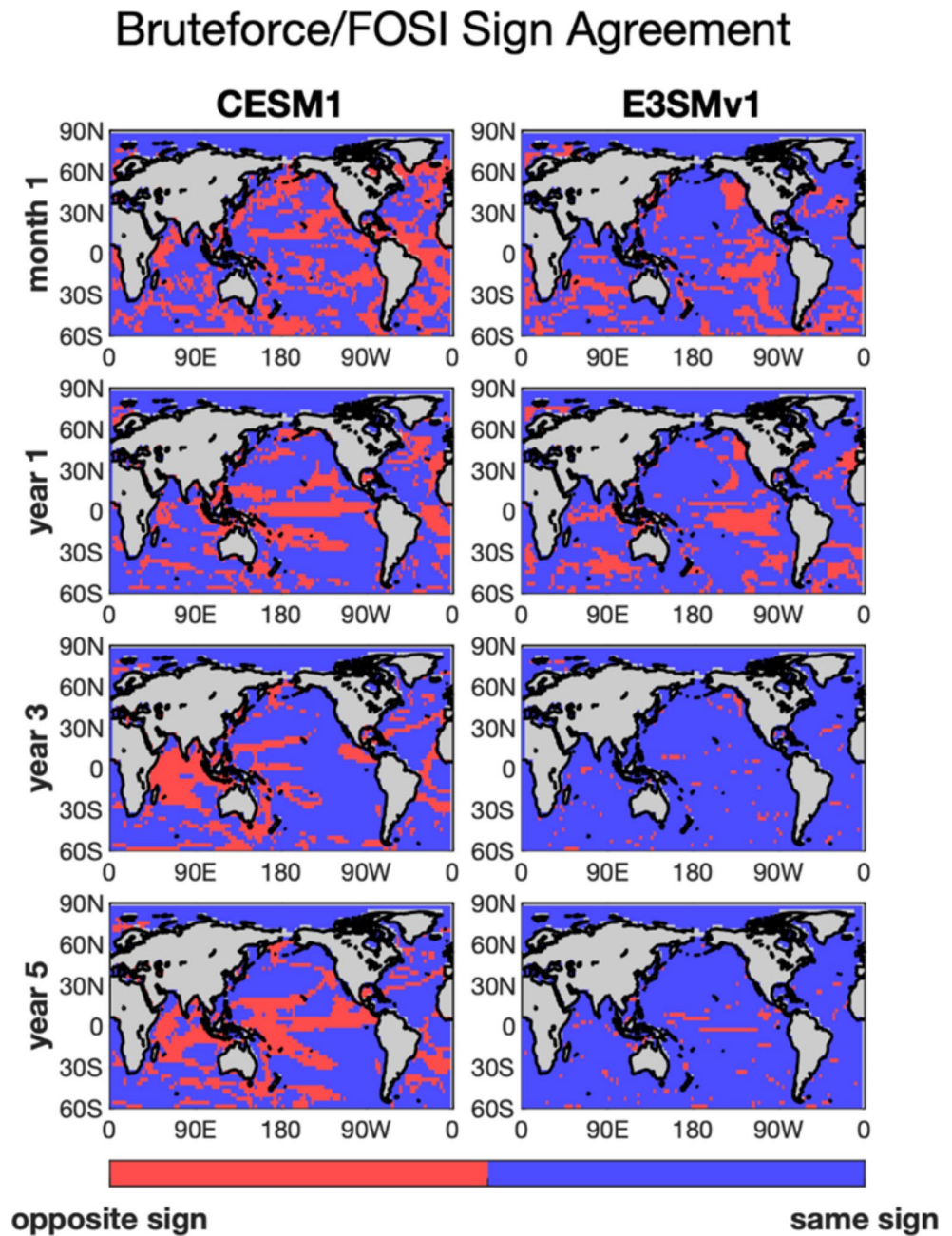
#### 4 Computing anomalies to evaluate hindcast skill

Removing drift from initialized hindcasts usually involves computing some sort of drifted climatology from the model with which to compute anomalies that can be compared to comparable anomalies over those time periods from observations (Doblas-Reyes et al. 2013; Meehl et al. 2022). However, since the patterns of model drifts shown in Fig. 1 are comparable for each model and initialization method,

**Fig. 2** Fraction of lead year 5 SST drift ( $^{\circ}\text{C}$ ) computed as a % for lead year drift divided by year 5 drift for CESM1 BF (first column), E3SMv1 BF (second column), CESM1 FOSI (third column), and E3SMv1 FOSI (fourth column) for lead month 1 (first row), lead year 1 (row 2), lead year 3 (row 3). Average percent of drift for total ocean area is at upper left of each panel. White areas denote values with opposite sign

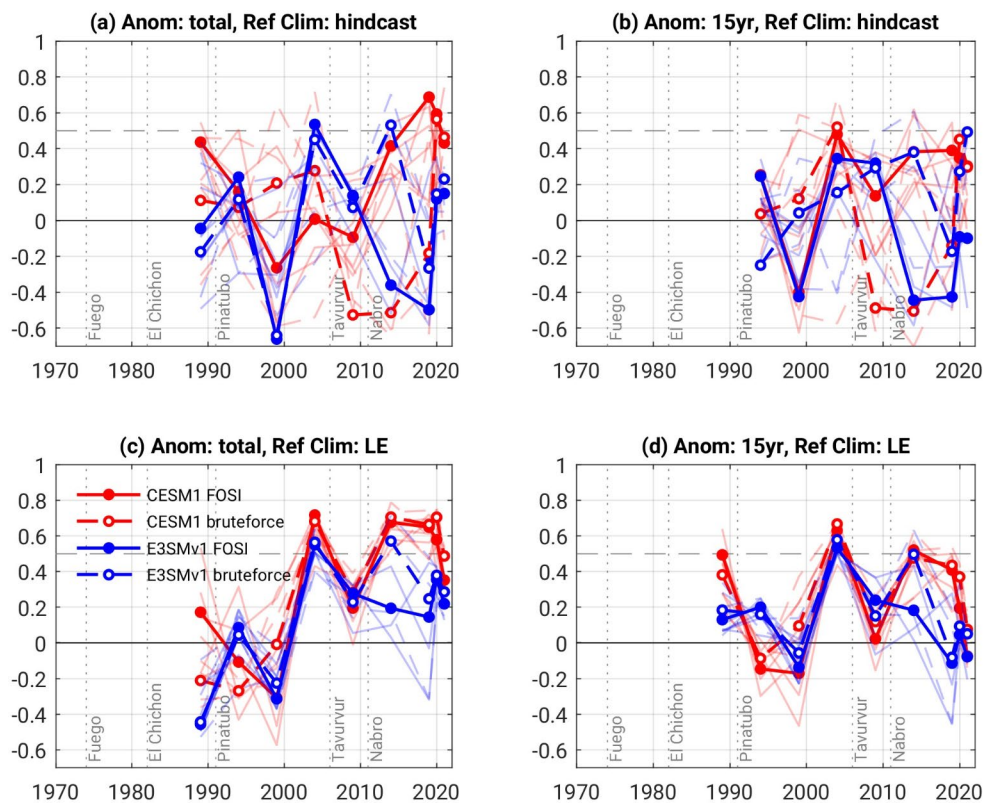


**Fig. 3** Comparison of sign of the drift between BF and FOSI methods as the fraction of when BF and FOSI are the same sign (blue = both negative or both positive) or opposite sign (red) for a given model (CESM1, left column; and E3SMv1, right column) for lead month 1 (first row), lead year 1 (row 2), lead year 3 (row 3), and lead year 5 (row 4)



and approach the mean systematic errors of the historical ensemble averages from the two models (e.g. pattern correlations between lead year 5 drifts and historical large ensemble climatological errors are all greater than +0.7), it may be possible to use ensemble mean systematic errors from the uninitialized historical simulations to represent the mean drifts toward that systematic error state in the initialized hindcasts from the models. This would allow the creation of many more samples of model climatology to calculate anomalies from the limited set of initial years of the initialized hindcasts.

To illustrate the advantage of a larger number of samples from which to compute a model climatology, Fig. 4 shows anomaly pattern correlations for the IPO region in the Pacific (40°S–70°N, 100°E–80°W) which has been a target for previous initialized hindcasts (e.g. Meehl et al. 2016; 2022). If all the limited hindcast drifted states are used to compute the model climatology, Fig. 4a shows much scatter among ensemble members and relatively low positive or even negative anomaly pattern correlations for both models and both initialization methods. Using the available initialized hindcasts from the 15 year periods prior to the initial years to compute the model climatology (Fig. 4b) as in

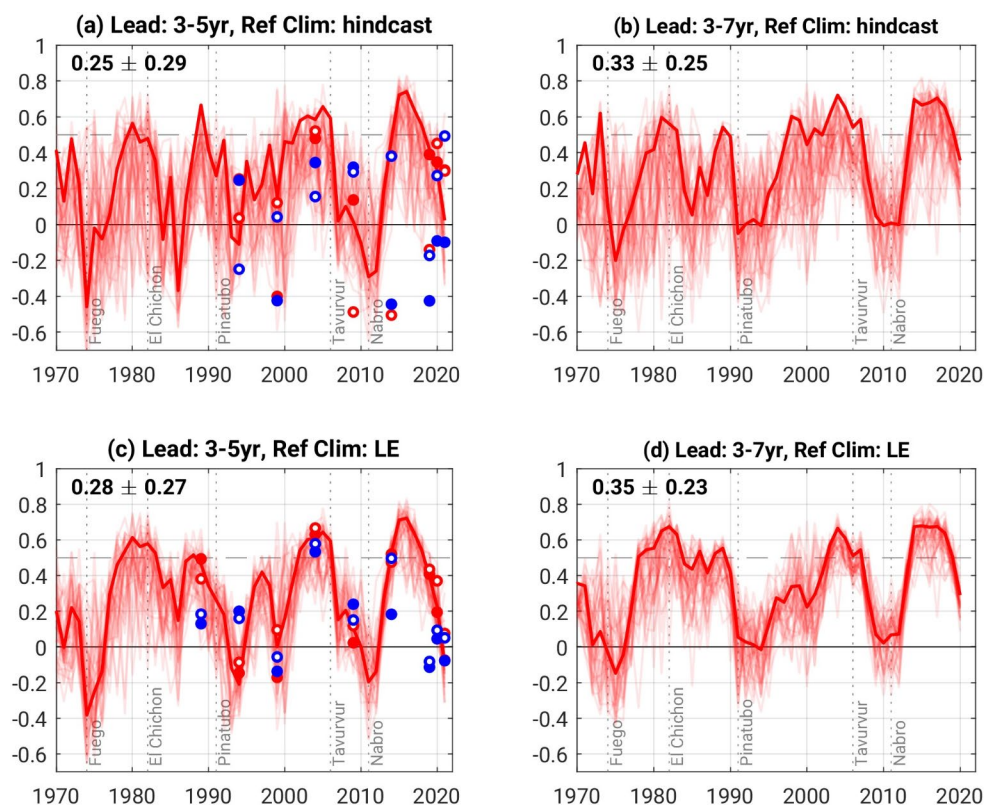


**Fig. 4** Anomaly pattern correlations of SST for the IPO region of the Pacific ( $40^{\circ}\text{S} - 70^{\circ}\text{N}$ ;  $100^{\circ}\text{E} - 80^{\circ}\text{W}$ ) for the limited set of initial years (Table 1) and 3–5 year average leads. Reference climatology used to calculate anomalies to compare to observations is (a) average of drifted initialized hindcasts for those start years and comparable periods from observations, (b) same as (a) except using drifted hindcasts in the 15 years prior to the initial years using the smaller number of CESM1 DPLE hindcasts according to Table 1, (c) same as (a) except using the climatology from the CESM1 and E3SMv1 historical large

ensembles, (d) same as (a) except for climatology from previous 15 years of the CESM1 and E3SMv1 historical large ensembles. Solid lines are ensemble averages (model key at upper left in panel c), and thin lines are individual ensemble members. Horizontal dashed line is a nominal reference value for anomaly pattern correlation of 0.5. The year labels on the x axis refer to lead years 3–5 of each hindcast that are centered on year 4, so, for example, 1989 on the x axis signifies the hindcast initialized in November 1985

Meehl et al. (2022), there is still considerable noise among ensemble members and some large negative anomaly pattern correlations. In Fig. 4b, given the small number of initial years, the 1990 set is dedrifted using only the 1985 set, and the 1995 set uses 1985 and 1990, so that the first year plotted is 1994 representing the start year 1990, thus introducing additional noise for those first few initial years. By using the available drifted hindcasts for the climatology (Table 1), there are so few samples that the average can be far from the average of drifted hindcasts from a much larger ensemble. This introduces bigger anomalies in the former, with some far from the observations, which produces greater noise in the correlations. Even with the same number of samples used from the LEs as noted in Table 1, these ensemble members represent a more equilibrated drifted state and the errors are better sampled so that the ensemble average is closer to the actual drift from the 13 member (E3SMv1) or 33 member (CESM1) LEs. If all 33 members from the CESM1 LE are used as in Fig. 5 discussed

below, there are many more samples, the errors are better sampled, and the anomalies are smaller and closer to the observations with better skill and reduced noise. Therefore, if the ensemble averages from the respective model historical uninitialized simulations are used to compute the model climatology (Fig. 4c), noise among the ensemble members is reduced as could be expected by using a larger number of samples from the historical ensemble averages compared to Fig. 4a. However, there is evidence of a possible trend being introduced into the climatology (as noted by Meehl et al. 2022) and consequent anomaly pattern correlations, with mostly negative values in the early period and positive values in the later period. Additionally, computing anomalies by subtracting the uninitialized climatology (lower row of Fig. 1) will tend to produce negative IPO-like patterns since the hindcast climatology in years 3–5 is generally less positive IPO-like than uninitialized climatology (Fig. 1). Even if drift pattern correlations with uninitialized errors are fairly high, the amplitude in years 3–5 is quite a bit lower than



**Fig. 5** Anomaly pattern correlations for the IPO region of the Pacific ( $40^{\circ}\text{S} - 70^{\circ}\text{N}$ ;  $100^{\circ}\text{E} - 80^{\circ}\text{W}$ ) for the full set of initial years, each of 64 initial years from 1954 to 2017, reference climatology uses the previous 15 years prior to the initial years to calculate anomalies to compare to observations, so first year plotted is 1954+15 or 1969, from the CESM1 DPLE for (a) average 3–5 year leads using climatology from the drifted initialized hindcasts; (b) same as (a) except for 3–7 year leads; (c) 3–5 year average leads using climatology from the

large ensemble; (d) same as (c) except for 3–7 year average leads. Dots on panels a and c are taken from Fig. 4b and d, respectively for the smaller number of start years and ensemble members. Thin red lines are individual ensemble members, and bold red line is the pattern correlation for the ensemble mean, and the mean and standard deviation are in the upper left of each panel. Horizontal dashed line is a nominal reference value for anomaly pattern correlation of 0.5

the uninitialized (except CESM1 BF). Consequently, the negative-IPO anomalies will show better pattern correlation skill after  $\sim 2000$  when the observations switched to negative IPO.

If the previous 15 year periods are used to compute the model climatology from the uninitialized historical simulations (Fig. 4d), the possible effect of a trend in the historical climatology evident in Fig. 4c is reduced as documented by Meehl et al. (2022). This is because the previous 15 years have a much smaller trend than the entire historical period. Consequently, the spread among ensemble members in the comparable calculation using the small number of drifted hindcasts in Fig. 4b is reduced, and the hindcast periods that are not negatively affected by volcanic eruptions (which reduce hindcast skill after Pinatubo, and the cumulative effects of the Tavurvur and Nabro eruptions in the early 2000s, e.g. Santer et al. 2014; Meehl et al. 2015; Wu et al. 2023) show anomaly pattern correlations near or above 0.5. The drop in skill for years after 2020 is likely related to the multi-year La Niña event that started in 2020

that could have been externally forced by smoke from the catastrophic large scale wildfires in Australia in 2019–2020 (Fasullo et al. 2021, 2023). Since none of the hindcasts considered here, including DPLE, included the Australian wildfire smoke, and if the multi-year La Niña event was at least partly externally forced, the verification observational data would include the effects of the smoke, thus reducing hindcast skill of the model simulations that did not include the smoke.

In any case, comparing the four sets of results in Fig. 4, the method that appears to exhibit the greatest skill and least noise among ensemble members uses the previous 15 year method from the ensemble average of the uninitialized historical simulations from the two models in Fig. 4d.

To provide greater context from which to evaluate the results in Fig. 4, Fig. 5 shows similar anomaly pattern correlations for the Pacific region SSTs for the larger number of initialized hindcasts in the DPLE (using CESM1 and the full 40 members from initial years overlapping those in the limited hindcast set; recall that CESM1 FOSI used only 5



ensemble members from DPLE). Because previous work used the average 3–7 lead years (e.g. Meehl et al. 2016, 2022), and recalling that the limited initial year model simulations only extend to lead year 5, the somewhat longer 3–7 year leads are included with the 3–5 year leads to judge how a longer prediction average (lead years 3–7) relates to the lead year 3–5 average analyzed earlier. Values from each initial year and each ensemble member are plotted in Fig. 5 as the thin red lines. The thick red line is the pattern correlation of the 40-member ensemble mean. Note that the ensemble mean will not necessarily fall in the middle of the spread of individual ensemble members since the correlation is not a linear operation. In other words, if the ensemble mean is closer to observations, there is a higher likelihood that the ensemble mean will have the highest correlation. The dots on Fig. 5a,c are the same as those in Fig. 4b,d for the limited set of start years in both models and both initialization schemes to allow direct comparison with the larger number of start years and ensemble members from the DPLE.

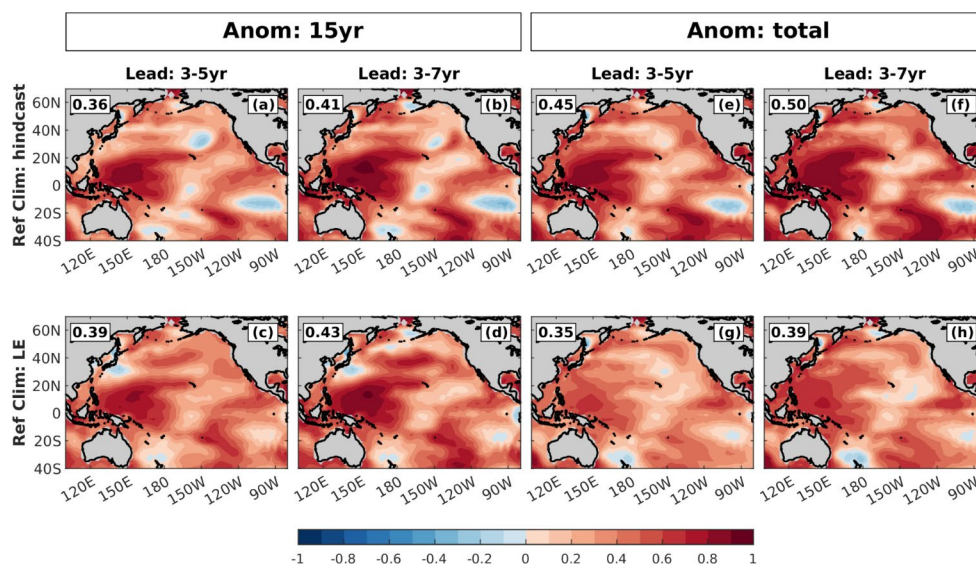
Figure 5a,b shows anomaly pattern correlation values for 3–5 year leads using the ensemble average of the drifted hindcasts from the CESM1 DPLE and the 3–7 year leads, respectively. As expected, the average over the shorter forecast period of 3–5 year leads in Fig. 5a produces lower pattern correlations on average (0.25 in Fig. 5a, 0.33 in Fig. 5b) and introduces greater noise (average standard deviation of 0.29 in Fig. 5a; compared to 0.25 for the 3–7 year leads in Fig. 5b). Both show a reduction of skill due to the Pinatubo and Tavurvur/Nabro volcanic eruptions as noted in earlier studies (e.g. Meehl et al. 2015; Wu et al. 2023). Figure 5c shows the average 3–5 year lead anomaly pattern correlations that can be compared to the average 3–7 year lead time series in Fig. 5d when using the large ensemble to compute the model climatologies. As in the first two panels, use of the shorter forecast averaging period in Fig. 5c has lower average pattern correlation values compared to Fig. 5d for the 3–7 lead year average (0.28 compared to 0.35). Additionally, there is more noise on average in the anomaly pattern correlation values in Fig. 5c (standard deviation of 0.27) compared to 0.23 in Fig. 5d with the average 3–7 year lead values. Comparing the two 3–7 year lead averages in Fig. 5b and d, use of the historical average from the large ensemble compared to the drifted hindcasts produces a somewhat larger average pattern correlation value of +0.35 in Fig. 5d compared to +0.33 in Fig. 5b. This indicates that, for DPLE, using the historical large ensemble compared to the drifted hindcasts produces somewhat more skillful results for the Pacific region.

The 3–5 year leads from the DPLE in Fig. 5a,c can be compared to dots for the reduced set of initial years from CESM1 FOSI taken from Fig. 4b and d. The overall higher anomaly pattern correlation values noted for the previous

historical 15 year climatology in Fig. 4d in CESM1 FOSI compared to the initialized hindcast climatology in Fig. 4b is reflected in similar results for the larger set of hindcasts and ensemble members for DPLE in Fig. 5a,c. That is, the 3–5 year leads using the historical large ensemble average to compute the model climatology for the previous 15 years is less noisy and has overall larger positive anomaly pattern correlations in Fig. 5c, in comparison to Fig. 5a that uses the drifted hindcasts as the reference climatology. The consistency of the larger set of initial years and ensemble members from the DPLE in Fig. 5c, compared to the smaller set of start years with fewer ensemble members (dots in Fig. 5c), supports the notion that using the average of the historical large ensemble as the reference for the previous 15 years for the model climatology presents a viable option for assessing the skill of IPO predictions in the Pacific for a small set of start years.

Another way to check for consistency in the use of the uninitialized historical large ensemble compared to the initialized drifted hindcasts is shown in Fig. 6 where the time series anomaly correlation coefficients at each grid point are calculated from the DPLE. The drifted hindcasts as model climatology (top row) and the uninitialized historical large ensemble as model climatology (bottom row) are shown for lead years 3–5 (first and third columns) and lead years 3–7 (second and fourth columns) for a climatology computed using the 15 years prior to the initial year (left two columns, panels a,b,c,d) and the total available model climatology (right two columns, panels e,f,g,h). Darker red colors indicate greater forecast skill.

First, in agreement with previous results (e.g. Yeager et al. 2018), there is greater skill in the western tropical Pacific (with anomaly correlation coefficient values greater than +0.8) and less skill in the eastern tropical Pacific (with anomaly correlation coefficient values less than +0.3) using all methods and all lead years. For the Pacific basin as a whole, the average correlations over the domain (in the top left corner of each panel) indicate that there is somewhat better skill at lead years 3–5 and 3–7 using the total model climatology (Fig. 6e and f) compared to the previous 15 year climatology (Fig. 6a,b) with average pattern correlations for the former of +0.45 and +0.50, compared to +0.36 and +0.41 for the latter. As noted earlier, this is because the externally-forced trend contributes skill when anomalies are computed relative to the long-term climatology, but this trend-related skill is considerably reduced when using the previous 15 year method (Meehl et al. 2022). The bottom row panels show that there is actually greater overall hindcast skill using the historical ensemble averages for the previous 15 years as climatology compared to using the entire historical period as climatology (average anomaly correlation coefficients of +0.39 and +0.43 for the former,



**Fig. 6** Anomaly correlation time series for full 64 initial years from the DPLE for 1954 to 2017, top row using the drifted hindcasts as the model climatology for previous 15 years (panel a for lead years 3–5, panel b for lead years 3–7), and for the full drifted hindcasts for the entire DPLE period (panel e for lead years 3–5, panel f for lead years 3–7); bottom row using the historical uninitialized large ensemble

from CESM1 as the model climatology for previous 15 years (panel c for lead years 3–5, panel d for lead years 3–7), and for the historical uninitialized large ensemble from CESM1 for the entire DPLE period (panel g for lead years 3–5, panel h for lead years 3–7); numbers at top left of each panel are the average anomaly correlation coefficients over the entire domain plotted

and +0.35 and +0.39 for the latter). This is an indication that the trend in the historical large ensemble is different from the trend in the initialized hindcasts in DPLE (Meehl et al. 2022) and the two different trends reduce overall skill. However, of note is that the overall anomaly correlation coefficients using the historical large ensemble simulations and the previous 15 years as climatology (Fig. 6c,d, with average anomaly correlation coefficients of +0.39 and +0.43) are comparable to the same anomaly calculation using the drifted initialized hindcasts (Fig. 6a,b, with average anomaly correlation coefficients of +0.36 and +0.41). This indicates that if the 15 years prior to the initial years are used as reference climatology, there is comparable skill from using the historical large ensemble simulation to compute the model climatology compared to using the drifted hindcasts as climatology for the Pacific Ocean region.

## 5 Discussion

The fact that there are differences in drift between initialization schemes in CESM1 and not in E3SMv1 points to differences between the two models. For example, as noted by Golaz et al. (2019), the POP2 ocean in CESM1 is more diffusive than MPAS ocean in E3SMv1 because of various model development choices and due to details of the numerics. This could partly explain why the CESM1 drift from the observed BF state is more rapid than the E3SMv1 BF or E3SMv1 FOSI. Second, POP2 in CESM1 has a

lower effective resolution in the tropics than MPAS ocean in E3SMv1, and the latter has a very robust simulation of tropical instability wave (TIW) activity. Additionally, POP2 in CESM1 has a larger Redi mixing coefficient than MPAS ocean in E3SMv1. This has implications for the ENSO amplitude in CESM1 that is about twice that in E3SMv1 (Golaz et al. 2019) since a reduced Redi coefficient can decrease the ENSO amplitude (Gnanadesikan et al. 2017). In addition, the anemic AMOC in E3SMv1 (about 11 Sv at 26.5°N below 500-m depth, Golaz et al. 2019) compared to CESM1 (about 26 Sv at 26.5°N below 500-m depth, Danabasoglu et al. 2020) dominates a cold SST bias in the North Atlantic in that model since weaker AMOC in E3SMv1 transports less heat northwards and contributes to colder SSTs in the North Atlantic (Golaz et al. 2019). This bias pattern sets up quickly and robustly (regardless of initialization) as seen for the E3SMv1 results in Fig. 1. BF seems to enhance AMOC in CESM1 whereas FOSI tends to lead to AMOC weakening.

There is evidence presented here that anomalies computed using the limited start years and the uninitialized climatology can adequately represent the patterns of drifted model states after about lead year 3. Compared to the CESM1 DPLE using the conventional methodology with the large hindcast data sets, there is comparable skill in CESM1 for predicting Pacific region SST patterns indicative of the Interdecadal Pacific Oscillation with this method using the anomaly pattern correlation metric. This works in part because even if anomalies relative to the historical

large ensemble have a uniform offset compared to what they would have relative to full DPLE, the skill metrics examined here (uncentered spatial or temporal correlation; centered produces similar results) remove that mean offset (i.e. correlations don't care about temporal or spatial mean values). This method would be less effective when looking at individual forecast anomaly maps or looking at other skill metrics like mean square skill score (MSSS).

## 6 Conclusions

A set of hindcasts for a limited set of start years are run with CESM1 and E3SMv1 using two different initialization methods to assess whether uninitialized free-running historical simulations can be used to form the model climatologies from which to compute anomalies to evaluate prediction skill. After a few years, the drifted initialized models approach the uninitialized model climatological errors in both magnitude and pattern such that after about lead year 3, the anomalies from the limited start years can use the uninitialized climatology to represent the patterns of drifted model states in the Pacific region. Compared to the conventional methodology with large hindcast data sets, there is comparable skill for predicting the pattern of Pacific region SSTs representative of the Interdecadal Pacific Oscillation using this method. With regards to the two initialization methods, the drifts are somewhat different but are so large that by about lead year 3 the two methods are roughly comparable, though there is some model dependence to the drifts especially in the equatorial Pacific region in CESM1.

We conclude that the use of a limited number of start years for initialized hindcasts can be evaluated after about lead year 3 using the historical large ensemble as a reference climatology if using the 15 years prior to the initial years to compute the reference model climatology. Thus, for these models and for the IPO in the Pacific, indications are that it is likely not necessary to run a complete hindcast data set for every new decadal climate prediction experiment. Additionally, after about lead year 3 the differences between the two initialization methods are minimal in most regions, though there is model dependence for the agreement between the two methods.

**Acknowledgements** Portions of this study were supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) under Award Number DE-SC0022070. This work also was supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation (NSF) under Cooperative Agreement No. 1852977. The Energy Exascale Earth System Model (E3SM) project is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental

Research (BER). E3SM simulations, as well as post-processing and data archiving of production simulations used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The CESM project is supported primarily by the National Science Foundation (NSF). Computing and data storage resources, including the Cheyenne supercomputer, were used for the CESM simulations (doi:<https://doi.org/10.5065/D6RX99HX>).

**Funding** Portions of this study were supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) under Award Number DE-SC0022070. This work also was supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation (NSF) under Cooperative Agreement No. 1852977. The Energy Exascale Earth System Model (E3SM) project is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER). E3SM simulations, as well as post-processing and data archiving of production simulations used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The CESM project is supported primarily by the National Science Foundation (NSF). Computing and data storage resources, including the Cheyenne supercomputer, were used for the CESM simulations (doi:<https://doi.org/10.5065/D6RX99HX>).

**Data Availability** E3SM model code and tools may be accessed on the GitHub repository at <https://github.com/E3SM-Project/E3SM>. A maintenance branch (maint-2.0; <https://github.com/E3SM-Project/E3SM/tree/maint-2.0>) has been specifically created to reproduce E3SMv2 simulations. Complete native model output is accessible directly on NERSC at <https://portal.nersc.gov/archive/home/projects/e3sm/www/WaterCycle/E3SMv2/LR>. A subset of the data reformatted following CMIP conventions is available through the DOE Earth System Grid Federation (ESGF) at <https://esgf-node.llnl.gov/projects/e3sm>. The CESM solutions / datasets used in this study are available from the Earth System Grid Federation (ESGF) at [esgf-node.llnl.gov/search/cmip6](https://esgf-node.llnl.gov/search/cmip6) or from the NCAR Digital Asset Services Hub (DASH) at [data.ucar.edu](http://data.ucar.edu) or from the links provided from the CESM web site at [www.cesm.ucar.edu](http://www.cesm.ucar.edu). HADiSST data are available from: <https://www.metoffice.gov.uk/hadobs/hadisst/>. The ERA-5 data are available from: <https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5>. The output from CESM-DPLE (as well as from the CORE\* simulation used to initialize the ocean and sea ice components) is available as raw, single-variable time series files. A web page ([www.cesm.ucar.edu/projects/community-projects/DPLE](http://www.cesm.ucar.edu/projects/community-projects/DPLE)) provides specifics about the simulations, links to the data, a publication list, and additional overview diagnostics for select fields. The companion CESM-LE simulation set has similar web documentation ([www.cesm.ucar.edu/projects/community-projects/LENS](http://www.cesm.ucar.edu/projects/community-projects/LENS)), including links to output and relevant publications. HADiSST observations are available at <https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>. NCEP/NCAR reanalyses files are by anonymous FTP from <ftp.cdc.noaa.gov/Datasets/ncep.reanalysis.derived/sigma>.

**Code Availability** NCL plotting routines are available at [https://www.ncl.ucar.edu/Document/Functions/list\\_alpha.shtml](https://www.ncl.ucar.edu/Document/Functions/list_alpha.shtml).

## Declarations

**Conflicts of interest** The authors declare that there are no conflicts of interest or competing interests regarding the publication of this paper.

**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

- Doblas-Reyes FJ et al (2013) Initialized near-term regional climate change prediction. *Nat Commun* <https://doi.org/10.1038/ncomms2704>
- Fasullo JT, Rosenbloom N, Buchholz RR, Danabasoglu G, Lawrence DM, Lamarque J-F (2021) Coupled climate responses to recent australian wildfire and COVID-19 emissions anomalies estimated in CESM2. *Geophys Res Lett*, 48, e2021GL093841.
- Fasullo JT, Rosenbloom N, Buchholz R (2023) A multi-year tropical Pacific cooling response to recent australian wildfires in CESM2. *Sci Adv* 9(19). <https://doi.org/10.1126/sciadv.adg1213>
- Gnanadesikan A, Russell A, Pradal M-A, Abernathy R (2017) Impact of lateral mixing in the ocean on El Niño in a suite of fully coupled climate models. *J Adv Model Earth Syst* 9:2493–2513. <https://doi.org/10.1002/2017MS000917>
- Golaz J-C, Caldwell PM, Van Roekel LP, Petersen MR, Tang Q, Wolfe J et al (2019) The DOE E3SM coupled model version 1: overview and evaluation at standard resolution. *J Adv Model Earth Syst* 11. <https://doi.org/10.1029/2018ms001603>
- Hurrell JW et al (2013) The Community Earth System Model: a framework for collaborative research. *Bull Amer Meteor Soc* 94:1339–1360
- Infanti JM, Kirtman BP (2017) CGCM and AGCM seasonal climate predictions: a study in CCSM4. *J Geophys Res Atmos* 122. <https://doi.org/10.1002/2016JD026391>
- Infanti JM, Kirtman BP (2019) A comparison of CCSM4 high-resolution and low resolutions predictions for south Florida and south-east United States drought. *Clim Dyn*. <https://doi.org/10.1007/s00382-018-4553-0>
- Kay JE et al (2015) The Community Earth System Model (CESM) large Ensemble Project: a community resource for studying climate change in the presence of internal climate variability. *Bull Am Meteorol Soc* 96(1):1333–1349. <https://doi.org/10.1175/BAMS-D-13-00255>
- Kirtman BP, Min D (2009) Multi-model ensemble ENSO prediction with CCSM and CFS. *Mon Wea Rev*. <https://doi.org/10.1175/2009MWR2672.1>
- Meehl GA, Arblaster JM (2011) Decadal variability of asian-australian monsoon-ENSO-TBO relationships. *J Clim* 24:4925–4940. <https://doi.org/10.1175/2011JCLI4015.1>
- Meehl GA et al (2009) Decadal prediction: can it be skillful? *Bull Amer Meteorol Soc* 90:1467–1485
- Meehl GA et al (2014) Decadal climate prediction: an update from the trenches. *Bull Amer Meteorol Soc* 95:243–267. <https://doi.org/10.1175/BAMS-D-12-00241.1>
- Meehl GA, Teng H, Maher N, England MH (2015) Effects of the Mt. Pinatubo eruption on decadal climate prediction skill. *Geophys Res Lett*. 42, 10,840 – 10,846 <https://doi.org/10.1002/2015GL066608>
- Meehl GA, Hu A, Teng H (2016) Initialized decadal prediction for transition to positive phase of the Interdecadal Pacific Oscillation. *Nat Commun* 7. <https://doi.org/10.1038/NCOMMS11718>
- Meehl GA et al (2021) Initialized Earth system prediction from sub-seasonal to decadal timescales. *Nat Rev Earth Environ* <https://doi.org/10.1038/s43017-021-00155-x>
- Meehl GA, Teng H, Smith D, Yeager S, Merryfield W, Doblas-Reyes F, Gnanville AA (2022) The effects of bias, drift, and trends in calculating anomalies for evaluating skill of seasonal-to-decadal initialized climate predictions. *Clim Dynamics*. <https://doi.org/10.1007/s00382-022-06272-7>
- Paolino DA, Kinter JL III, Kirtman BP, Min D, Straus DM (2012) The impact of land surface initialization on seasonal forecasts with CCSM. *J Clim* 25:1007–1021
- Petersen MR et al (2019) An evaluation of the ocean and sea ice climate of E3SM using MPAS and interannual CORE-II forcing. *J Adv Model Earth Syst* 11:1438–1458. <https://doi.org/10.1029/2018MS001373>
- Power S, Casey T, Folland C, Colman A, Mehta V (1999) Interdecadal modulation of the impact of ENSO on Australia. *Clim Dyn* 15:319–324
- Risbey J et al (2021) Standard assessments of climate forecast skill can be misleading. *Nat Commun* 12(4346):1–14
- Saha S et al (2010) The NCEP Climate Forecast System Reanalysis. *Bull Amer Meteor Soc* 91:1015–1057. <https://doi.org/10.1175/2010BAMS3001.1>
- Santer BD et al (2014) Volcanic contribution to decadal changes in tropospheric temperature. *Nat Geo*. <https://doi.org/10.1038/NNGEO2098>
- Saurral RI, Merryfield WJ, Tolstykh MA, Lee W-S, Doblas-Reyes FJ, García-Serrano J et al (2021) A data set for intercomparing the transient behavior of dynamical model-based subseasonal to decadal climate predictions. *J Adv Model Earth Syst* 13:e2021MS002570. <https://doi.org/10.1029/2021MS002570>
- Siqueira L, Kirtman BP, Laurindo LC, Climate J (2021) <https://doi.org/10.1175/JCLI-D-20-0139.1>
- Teng H, Branstator G, Karspeck A, Yeager S, Meehl GA (2017) Initialization shock in CCSM4 decadal prediction experiments, CLIVAR Exchanges. <https://doi.org/10.22498/pages.25.1.41>
- Tsujino H et al (2020) Evaluation of global ocean–sea-ice model simulations based on the experimental protocols of the Ocean Model Intercomparison Project phase 2 (OMIP-2), *Geo*. <https://doi.org/10.5194/gmd-13-3643-2020>. *Model Dev*.
- Wu X, Yeager SG, Deser C, Rosenbloom N, Meehl GA (2023) Model response to volcanic forcing degrades multiyear-to-decadal prediction skill in the tropical Pacific. *Sci Rep* 9:eadd9364. <https://doi.org/10.1126/sciadv.add9364>
- Yeager SG et al (2018) Predicting near-term changes in the Earth System: a large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bull Amer Meteorol Soc* 99:1867–1886. <https://doi.org/10.1175/BAMS-D-17-0098.1>

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