Evaluating skill in predicting the IPO in initialized decadal climate prediction hindcasts in CESM1 and E3SMv1 using a small set of start years

Gerald Meehl[1]

Ben Kirtman[2], **Sasha Glanville**[1], **Yaga Richter**[1], **Nan Rosenbloom**[1], **and Steve Yeager**[1]

1. National Center for Atmospheric Research, Boulder, CO
2. University of Miami, Miami, FL

NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

U.S. DEPARTMENT OF ENERGY
Office of Science
Biological and Environmental Research
Regional and Global Model Analysis

Challenges in decadal climate prediction have involved choosing an appropriate method for initializing the model, then devising methods for removing the inevitable drift away from the initial state due to model error

Challenges in decadal climate prediction have involved choosing an appropriate method for initializing the model, then devising methods for removing the inevitable drift away from the initial state due to model error

This typically requires running a large set of initialized hindcasts with many ensemble members to evaluate skill of the hindcasts (e.g. the DPLE has 64 start years and 40 member ensembles for each start year)

Challenges in decadal climate prediction have involved choosing an appropriate method for initializing the model, then devising methods for removing the inevitable drift away from the initial state due to model error

This typically requires running a large set of initialized hindcasts with many ensemble members to evaluate skill of the hindcasts (e.g. the DPLE has 64 start years and 40 member ensembles for each start year)

The computational burden of running such large hindcast data sets is daunting and, consequently, only occasionally is there enough computer time available to perform such a set of simulations

Challenges in decadal climate prediction have involved choosing an appropriate method for initializing the model, then devising methods for removing the inevitable drift away from the initial state due to model error

This typically requires running a large set of initialized hindcasts with many ensemble members to evaluate skill of the hindcasts (e.g. the DPLE has 64 start years and 40 member ensembles for each start year)

The computational burden of running such large hindcast data sets is daunting and, consequently, only occasionally is there enough computer time available to perform such a set of simulations

One reason to run such large hindcast data sets is to have enough samples of start years to form a drifted climatology from which to compute anomalies used to compare to observations to quantify skill of the hindcasts

Challenges in decadal climate prediction have involved choosing an appropriate method for initializing the model, then devising methods for removing the inevitable drift away from the initial state due to model error

This typically requires running a large set of initialized hindcasts with many ensemble members to evaluate skill of the hindcasts (e.g. the DPLE has 64 start years and 40 member ensembles for each start year)

The computational burden of running such large hindcast data sets is daunting and, consequently, only occasionally is there enough computer time available to perform such a set of simulations

One reason to run such large hindcast data sets is to have enough samples of start years to form a drifted climatology from which to compute anomalies used to compare to observations to quantify skill of the hindcasts

Is it possible to run a smaller number of start years for hindcasts to perform sensitivity experiments or case studies and use the uninitialized free-running historical simulations to form the model climatologies to compute anomalies and remove drift?

There are a number of ways to deal with bias and drift when computing anomalies to evaluate skill of initialized multi-year hindcasts:

1. **Forecast year differences from a model climatology (e.g. Doblas-Reyes, et al 2013: Boer et al., 2016 for DCPP)**

(trends in bias and drift introduce enhanced skill estimates for earlier and later in the hindcast period)

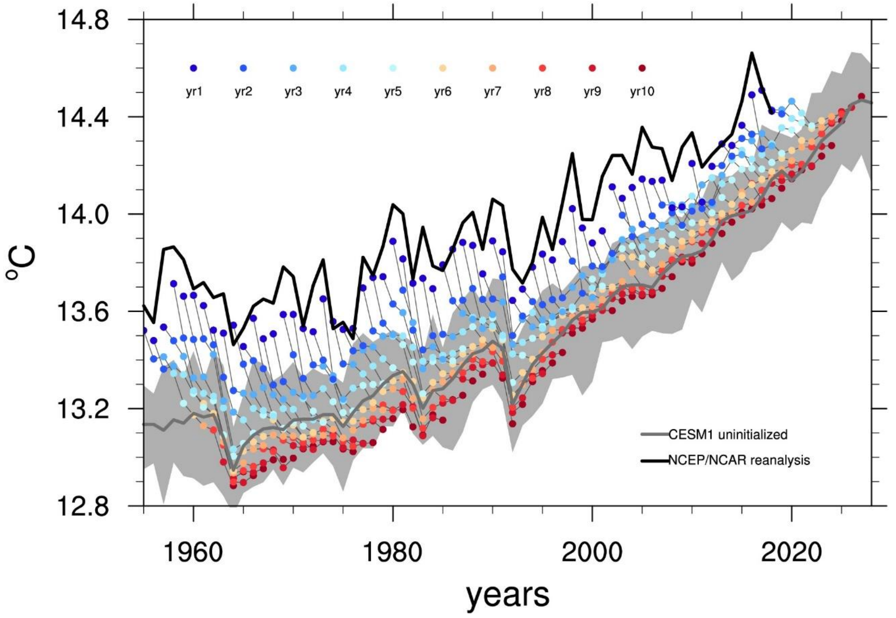
2. **Bias-adjusted lead year differences from the previous 15 year average from observations (e.g. Meehl et al., 2016)** (unrealistic skill can be introduced when low frequency variability in the observations is large compared to the hindcasts on timescales greater than 15 years)

3. **Forecast year differences from the previous 15 year average of model initial states (Meehl et al., 2021)**

(somewhat lower skill compared to each of the previous methods, but less difficulties with long term trends in the model climatology, and no unrealistic situational skill from using observations as a reference)

4. **Form anomalies from a sensitivity hindcast experiment for the same time period as a reference hindcast**

(unambiguously removes bias and drift, but can only be used in a sensitivity experiment context)



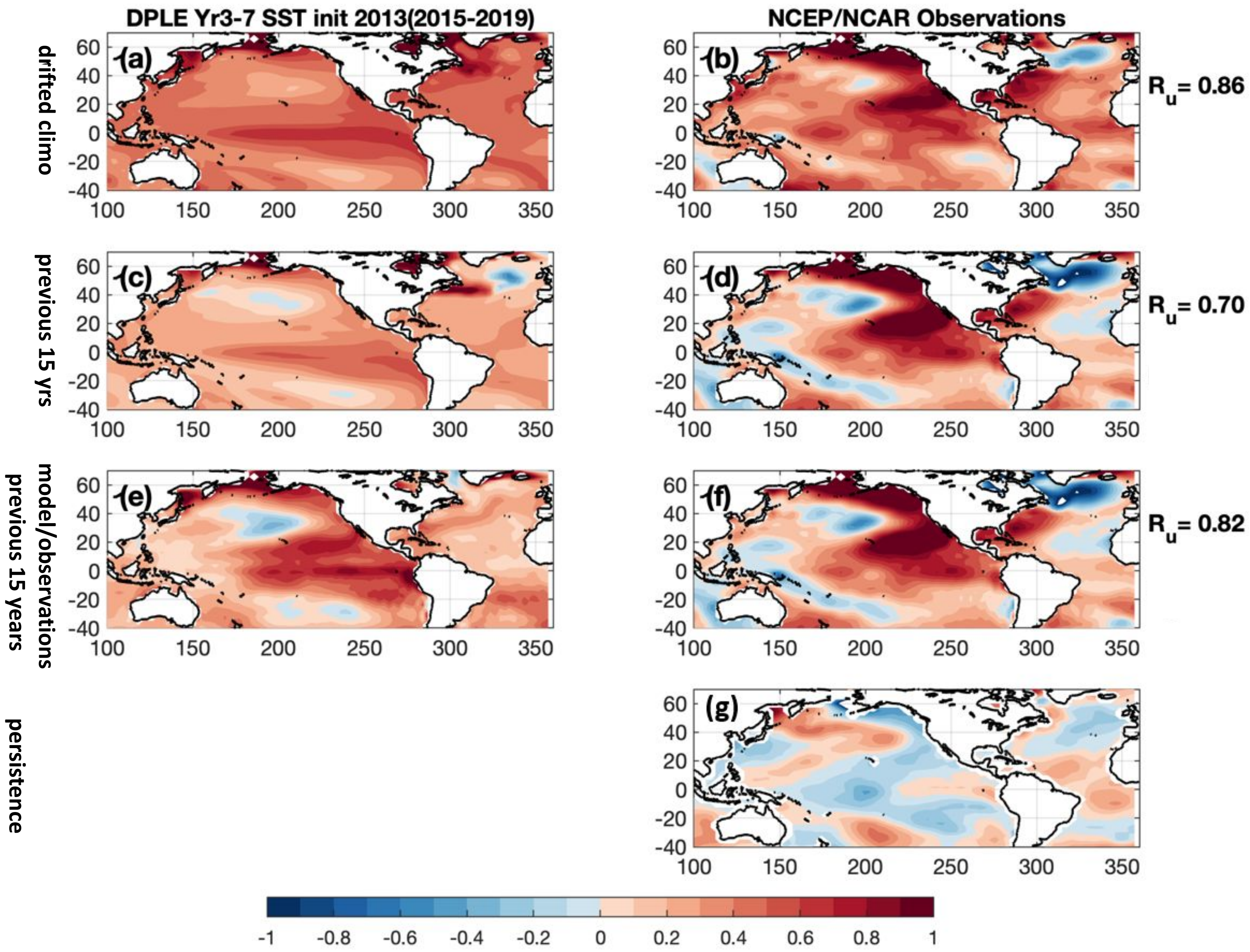
(Meehl, Teng, Smith, Yeager, Merryfield, Doblas-Reyes, and Glanville, 2021, *Cli. Dyn.*)

DPLE prediction initialized 2013 for lead years 3-7 (2015-2019)

Observations (2015-2019)

An example of the use of three different methods to compute anomalies

Comparable hindcast skill with some regional differences



(Meehl, Teng, Smith, Yeager, Merryfield, Doblas-Reyes, and Glanville, 2022, *Climate Dynamics*)

Does initialization scheme make a difference if a small number of start years for initialized hindcasts use the uninitialized historical large ensemble to form model climatologies from which to compute anomalies to evaluate skill of IPO predictions?

Perform two sets of initialized hindcasts with CESM1 and E3SMv1 (1 degree versions) using two different initialization schemes:

1. forced ocean sea ice (FOSI)—run the ocean model through five cycles using observed forcing from the atmosphere such that the ocean acquires the imprint of those observations, and use the final cycle as initial states

2. "brute force"—use observed reanalysis products for ocean, atmosphere and land and simply interpolate to model grid for initial states

These initial years have 5 ensemble members with CESM1 brute force and FOSI, (and more with DPLE), and 5 ensemble members with E3SMv1 brute force and FOSI

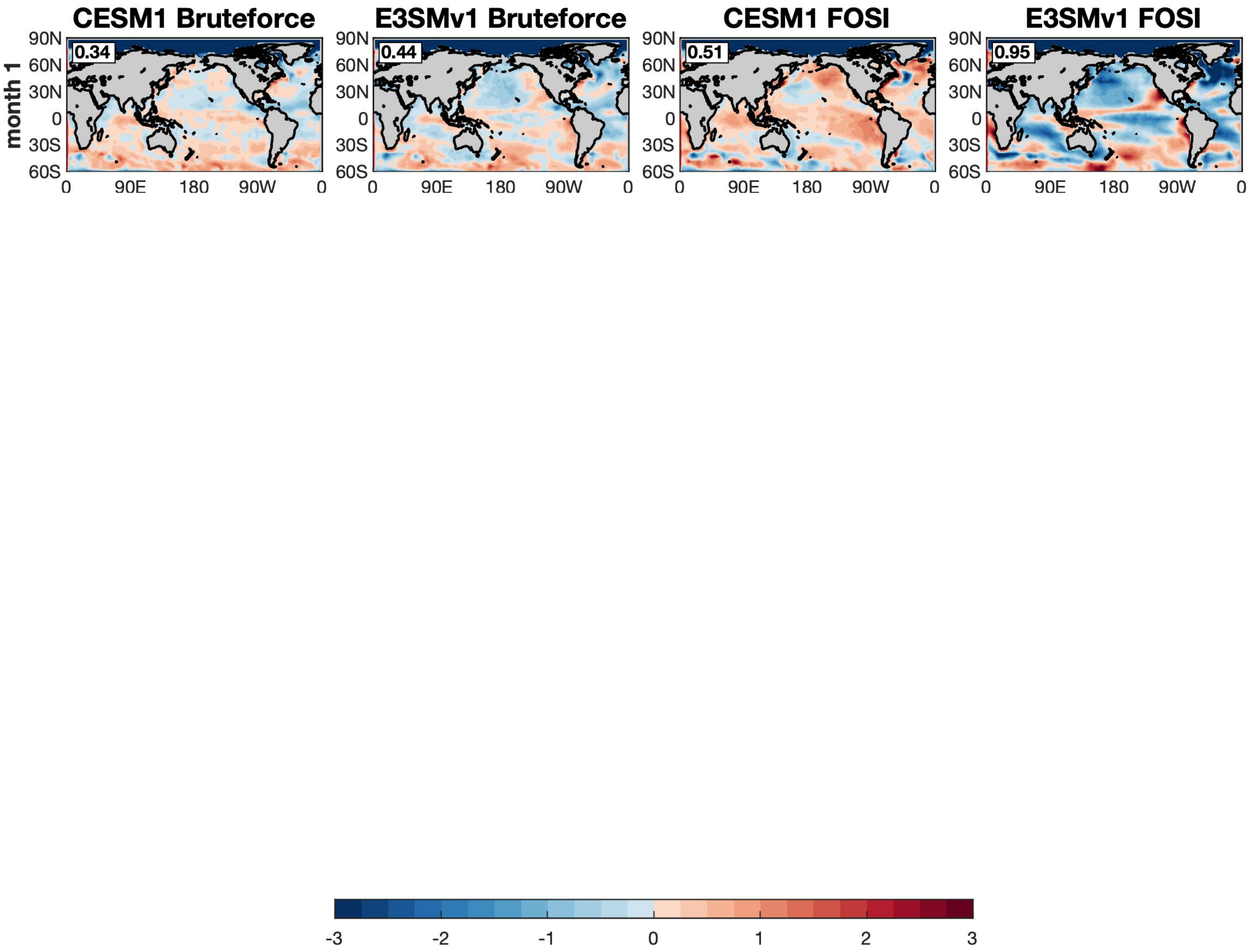
1985, 1990, 1995, 2000

These initial years have 5 ensemble members with CESM1 brute force and FOSI (and more with DPLE), and 3 ensemble members with E3SMv1 brute force and FOSI

2005, 2010, 2015, 2016, 2017, 2018

The pattern of the drifts sets up early in both models

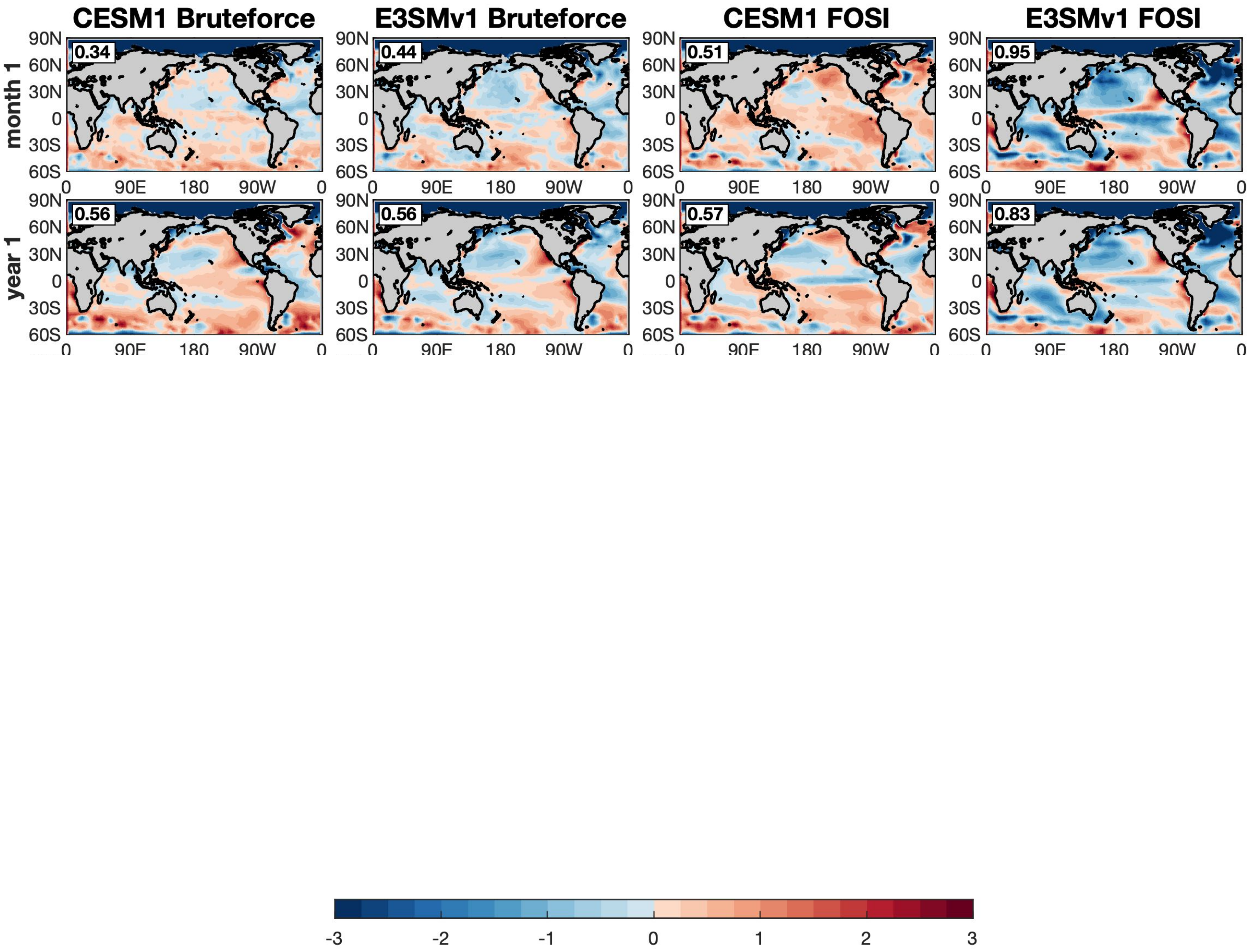
Initially there are lower amplitude errors in the two brute force methods



The pattern of the drifts sets up early in both models

Initially there are lower amplitude errors in the two brute force method,

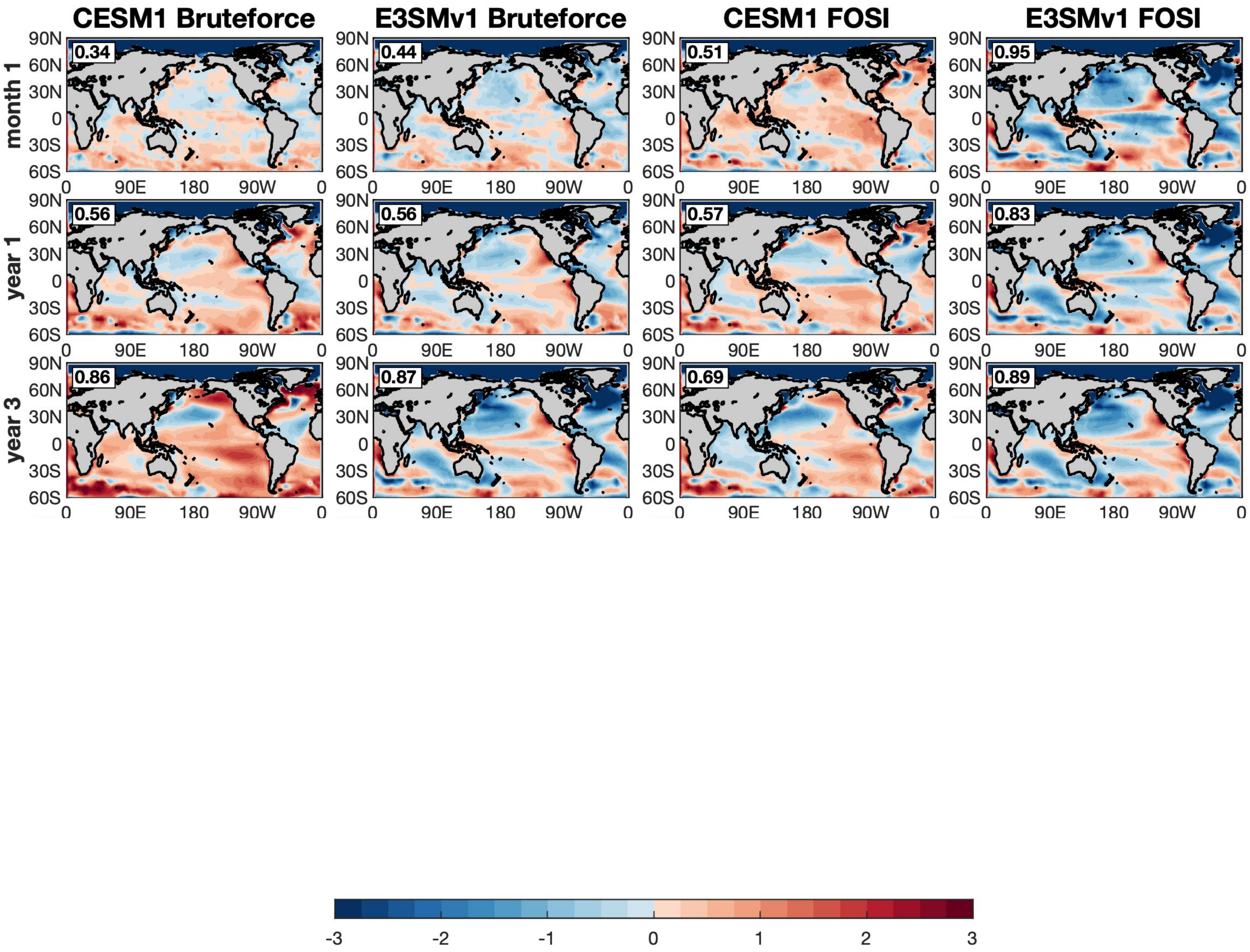
By year 1 the patterns are evident that then grow in amplitude through lead year 5;



The pattern of the drifts sets up early in both models

Initially there are lower amplitude errors in the two brute force method,

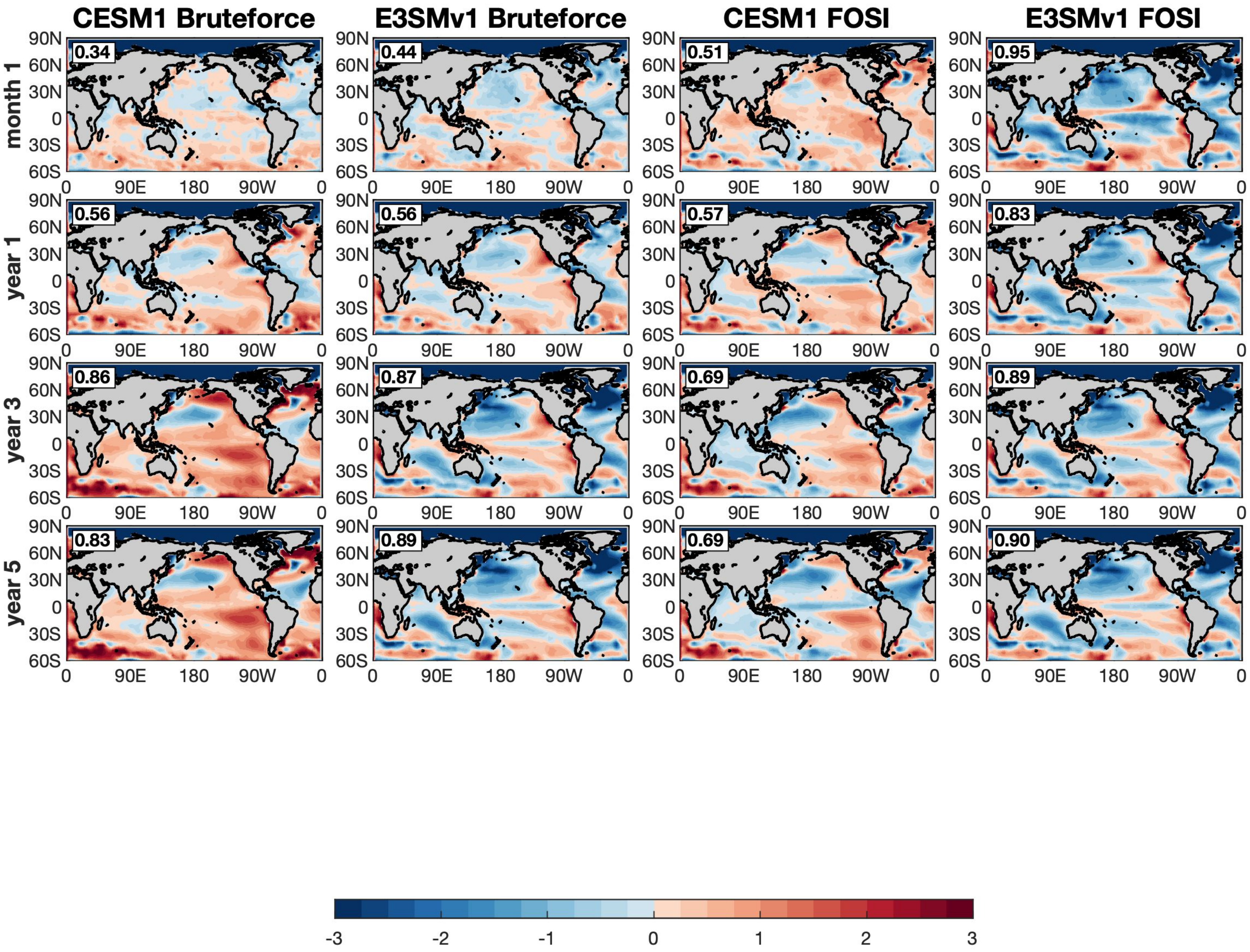
By year 1 the patterns are evident that then grow in amplitude through lead year 5;



The pattern of the drifts sets up early in both models

Initially there are lower amplitude errors in the two brute force method,

By year 1 the patterns are evident that then grow in amplitude through lead year 5;

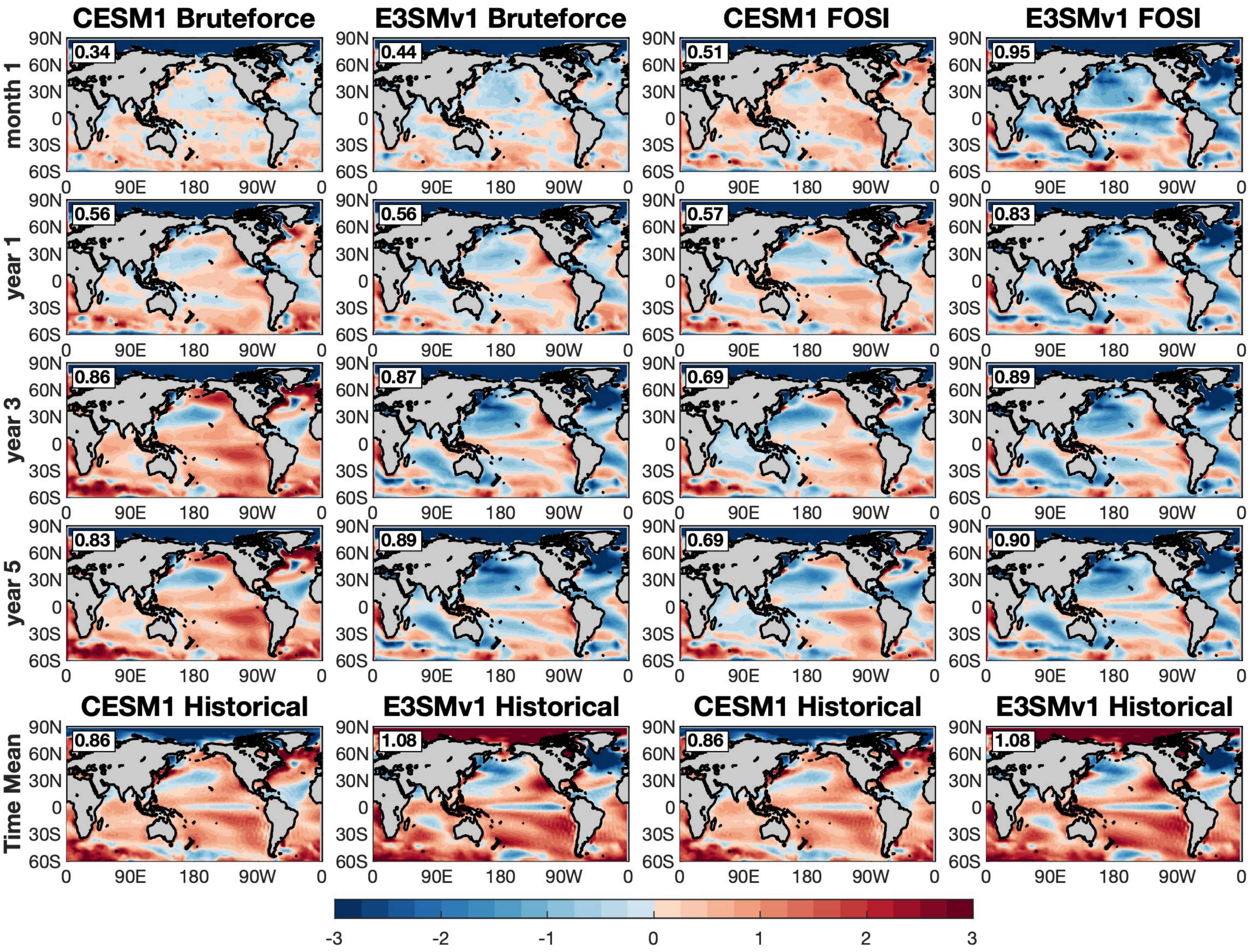


The pattern of the drifts sets up early in both models

Initially there are lower amplitude errors in the two brute force method,

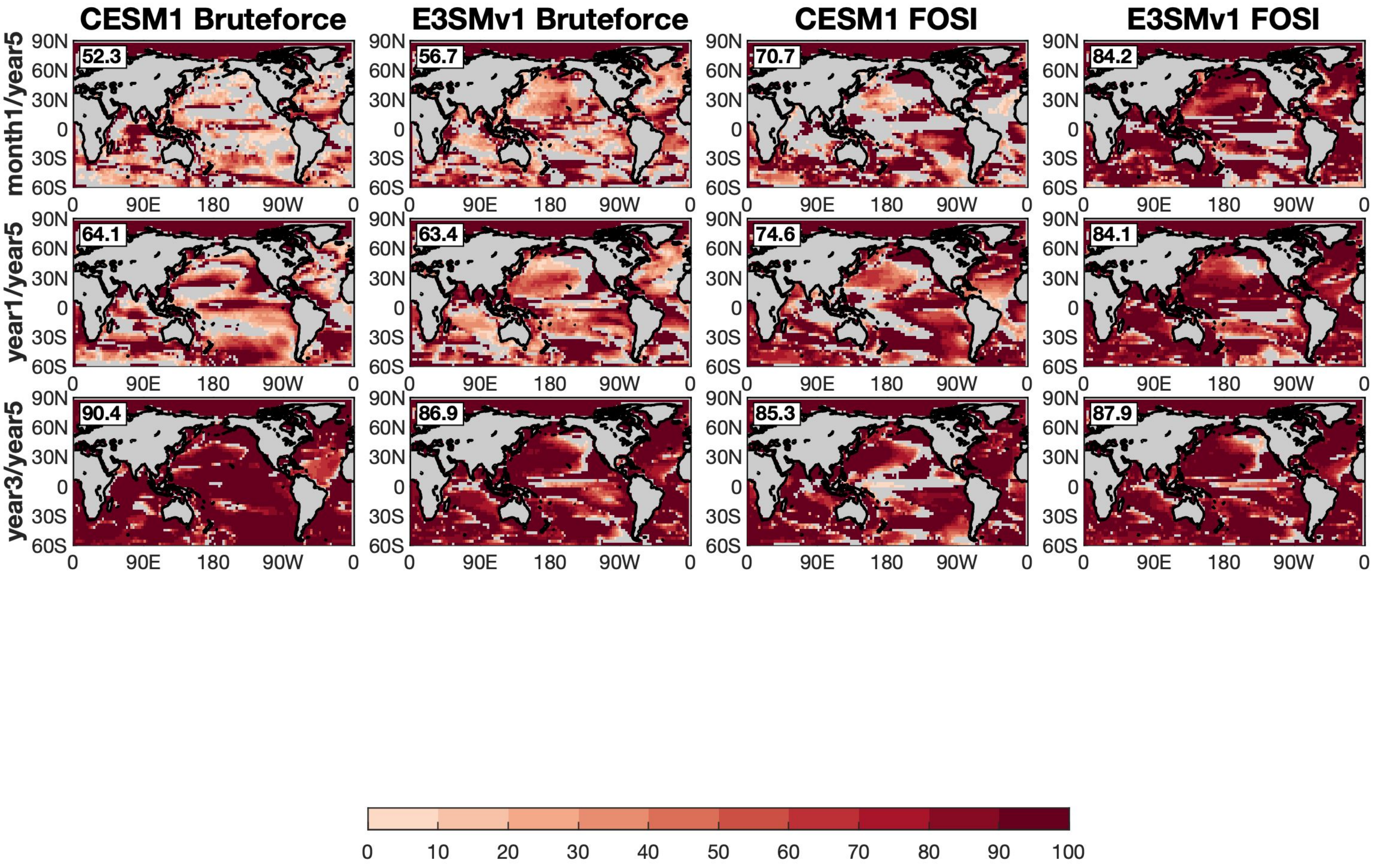
By year 1 the patterns are evident that then grow in amplitude through lead year 5;

The same error patterns are evident in the uninitialized historical large ensembles



Over 80% of the amplitude of the drift occurs by lead year 3 in most areas outside of tropical high interannual variability regions

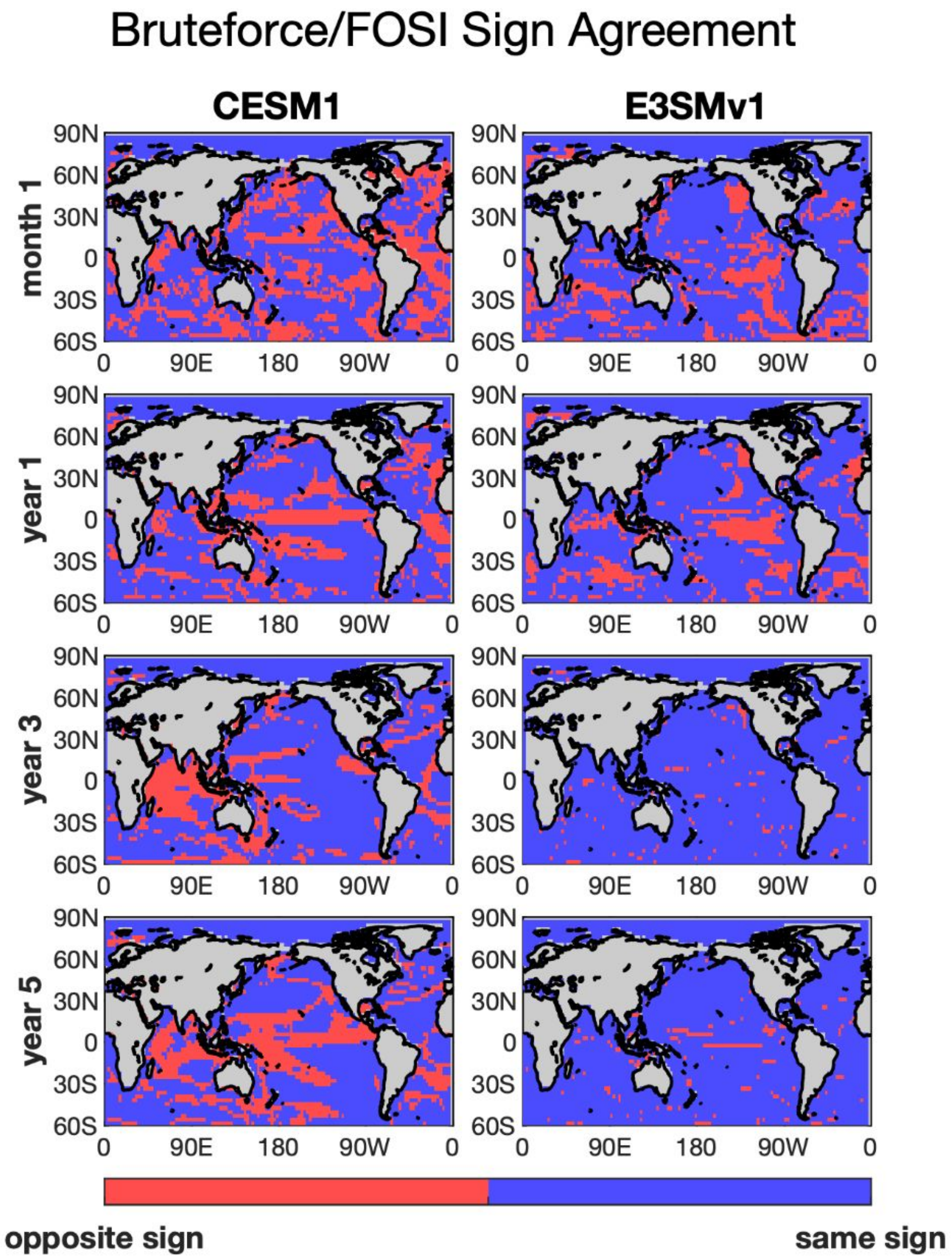
(bottom row, lead year 3 drift errors divided by lead year 5 drift errors)



There is a model dependence regarding relative agreement of drift errors between brute force and FOSI initialization schemes

By lead year 3, there is little difference between initialization schemes in E3SMv1, but areas of high interannual variability in the tropics in CESM1 show less agreement between the two schemes

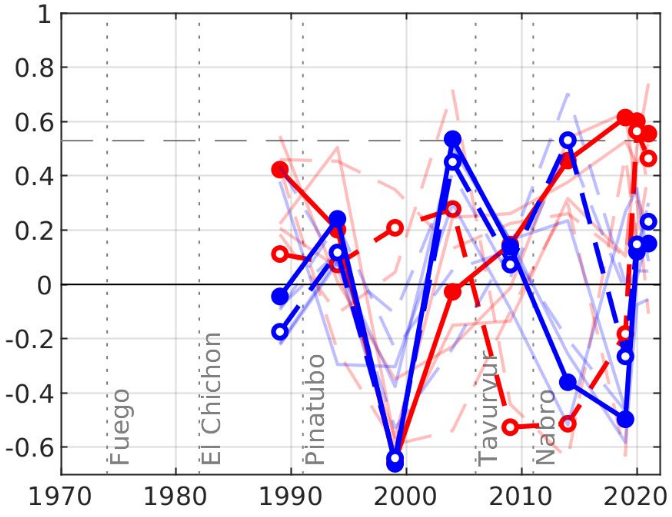
Right: sign of the drift between brute force and FOSI methods as the fraction of when brute and fosi are the same sign (blue = both negative or both positive) or opposite sign (red)



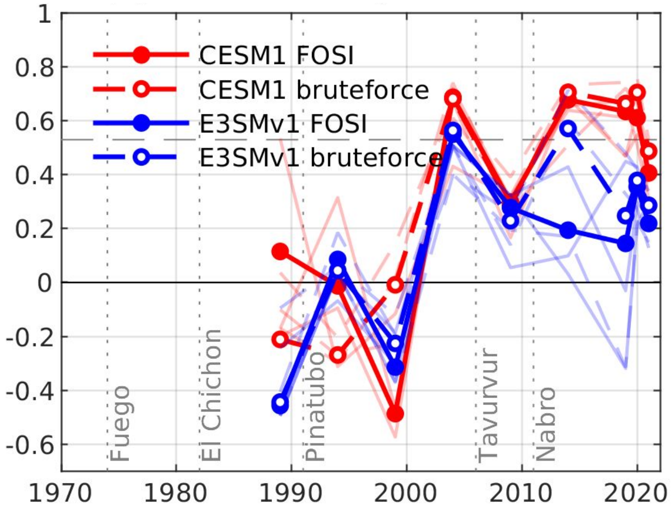
Limited start dates anomaly pattern correlations for IPO region in the Pacific for **lead years 3-5**

Negative impact of volcanic eruptions evident in loss of skill after eruptions (Meehl et al., 2015; Wu et al., 2023)

Drifted hindcast climatology



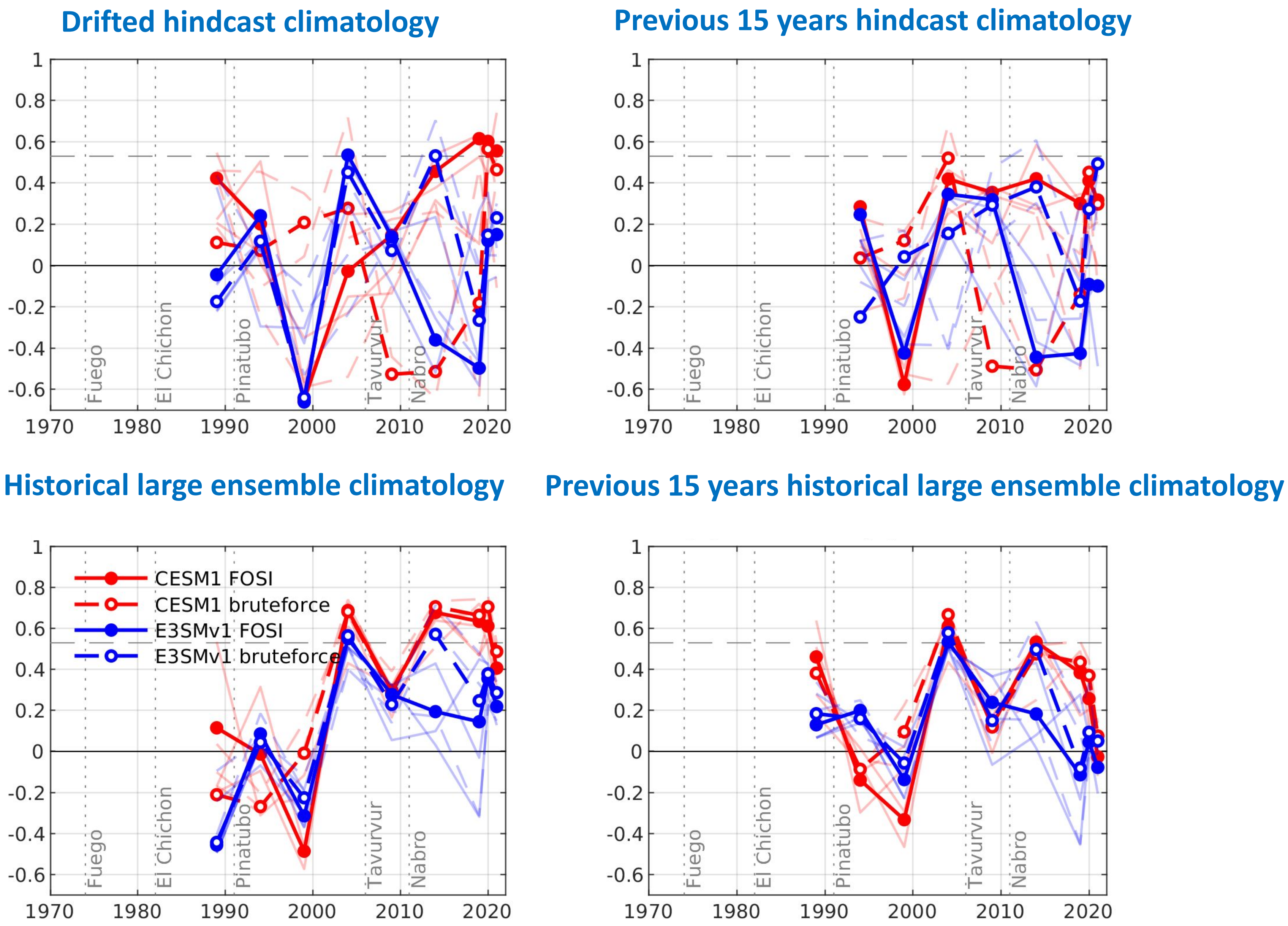
Historical large ensemble climatology



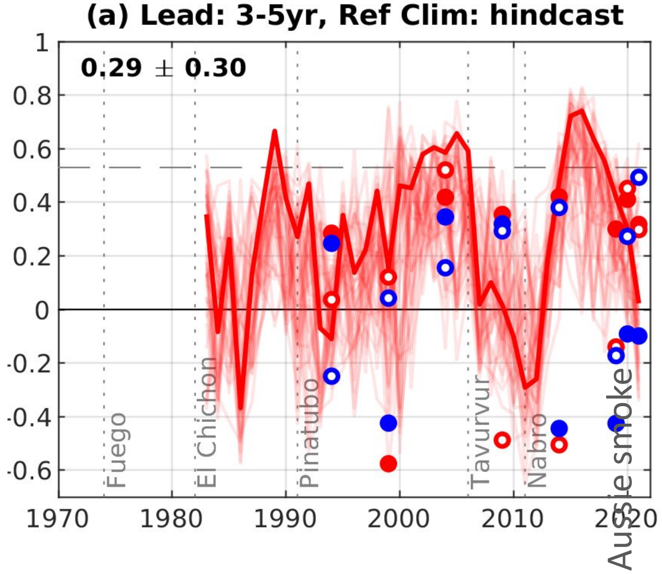
Limited start dates anomaly pattern correlations for IPO region in the Pacific for **lead years 3-5**

Negative impact of volcanic eruptions evident in loss of skill after eruptions (Meehl et al., 2015; Wu et al., 2023)

Highest and most consistent skill uses previous 15 years from historical large ensembles



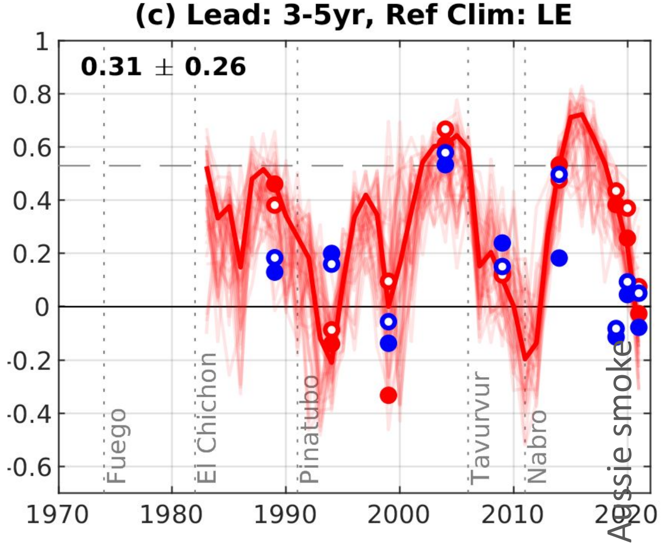
Drifted hindcast climatology (15 yr prior to initial year)



Anomaly pattern correlations for lead years 3-5 and 3-7 from full set of start dates from CESM1 DPLE using previous 15 years as reference

Negative impact of volcanic eruptions evident in loss of IPO skill after eruptions (Meehl et al., 2015; Wu et al., 2023) and Australian wildfire smoke (Meehl et al. yesterday's talk)

Historical large ensemble climatology (15 yr prior to initial year)



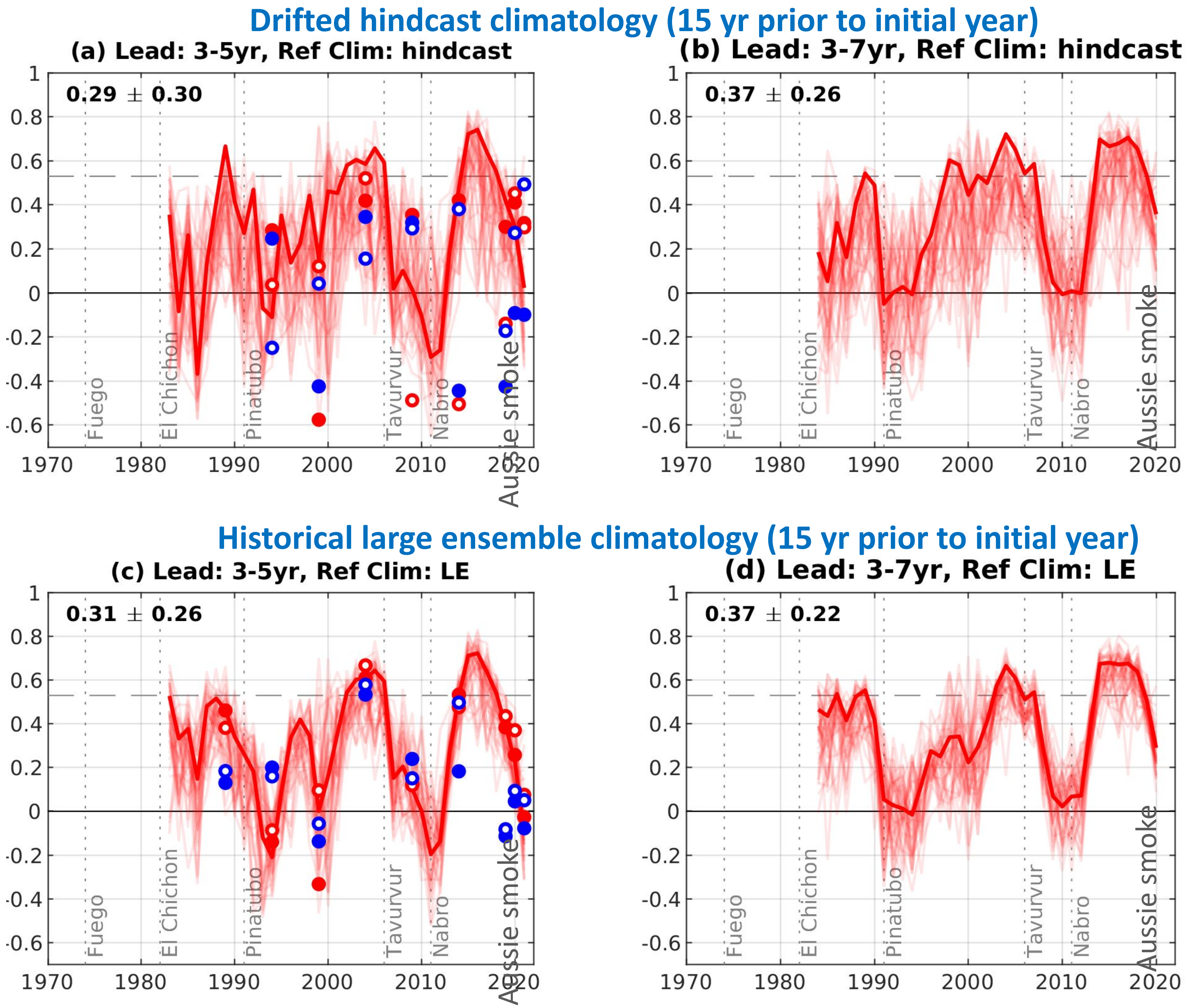
Highest and most consistent skill uses previous 15 years from historical large ensembles that have comparable skill to the drifted hindcasts

The closer the observations are to the ensemble mean, the higher the likelihood the ensemble mean will have the highest correlation

Anomaly pattern correlations for lead years 3-5 and 3-7 from full set of start dates from CESM1 DPLE using previous 15 years as reference

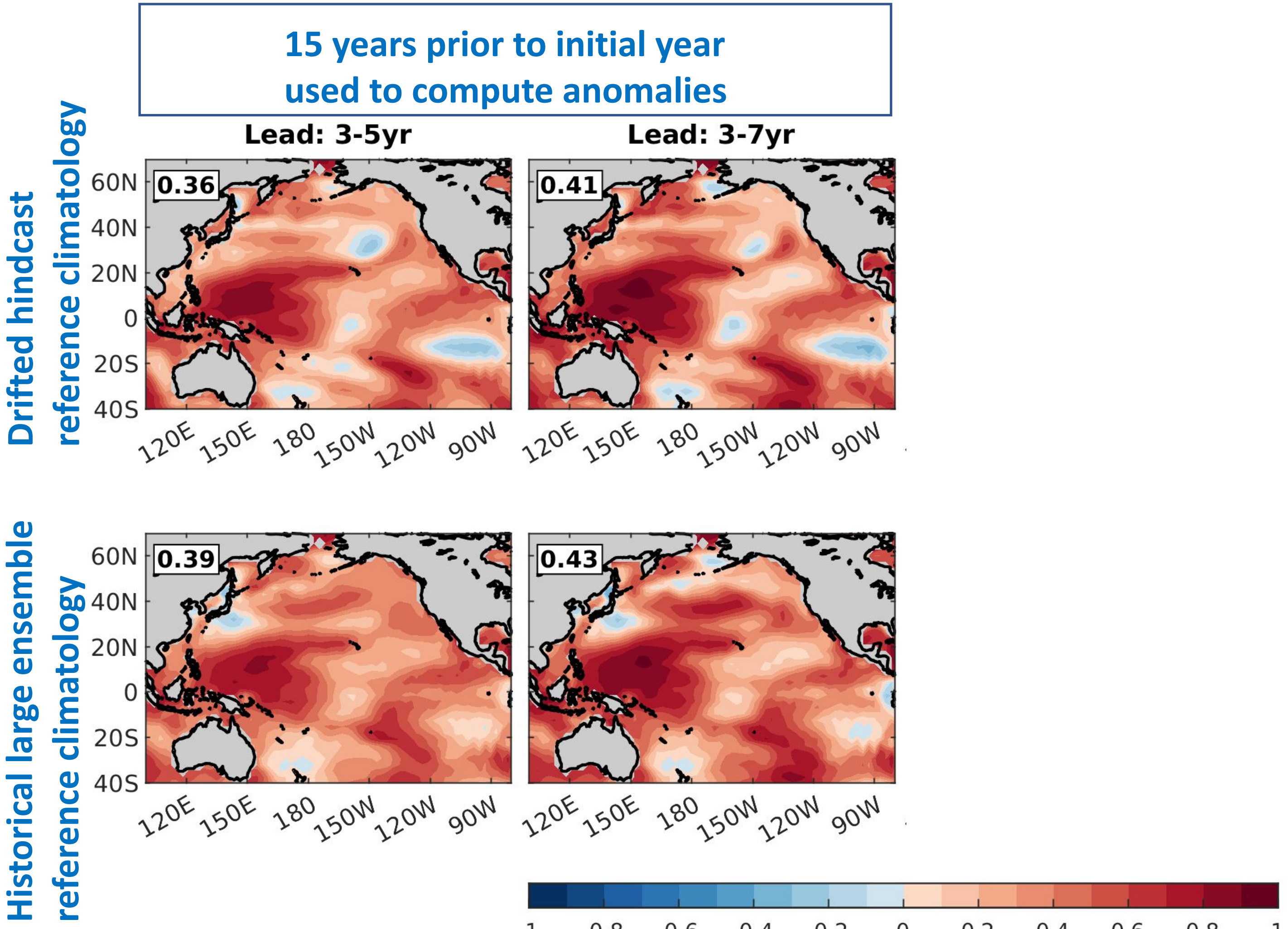
Negative impact of volcanic eruptions evident in loss of IPO skill after eruptions (Meehl et al., 2015; Wu et al., 2023) and Australian wildfire smoke (Meehl et al. yesterday's talk)

Highest and most consistent skill uses previous 15 years from historical large ensembles that have comparable skill to the drifted hindcasts

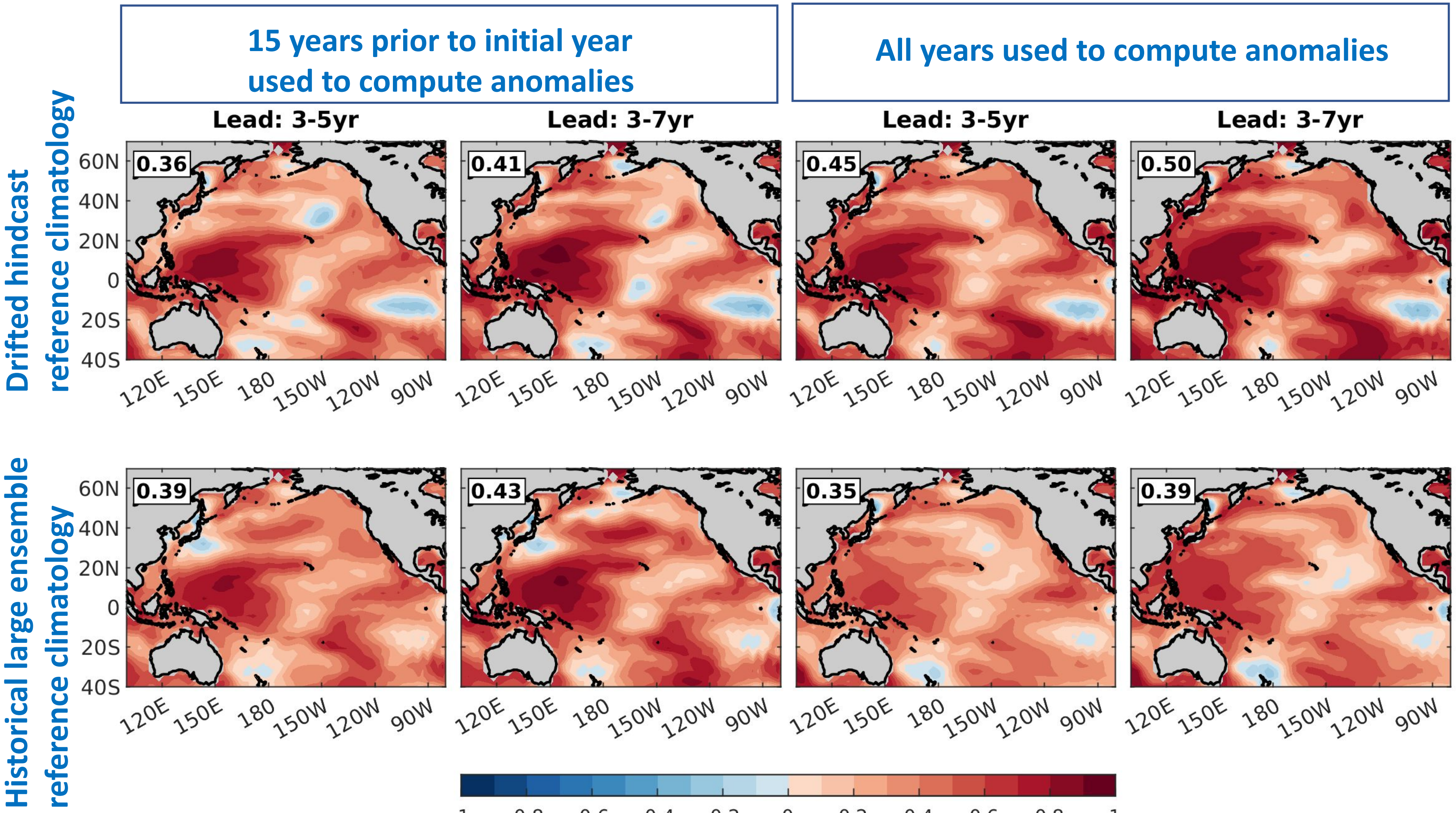


The closer the observations are to the ensemble mean, the higher the likelihood the ensemble mean will have the highest correlation

Anomaly correlations for the full set of initial years (each of 64 initial years from 1954 to 2017) from the CESM1 DPLE, lead years 3-5 and 3-7)

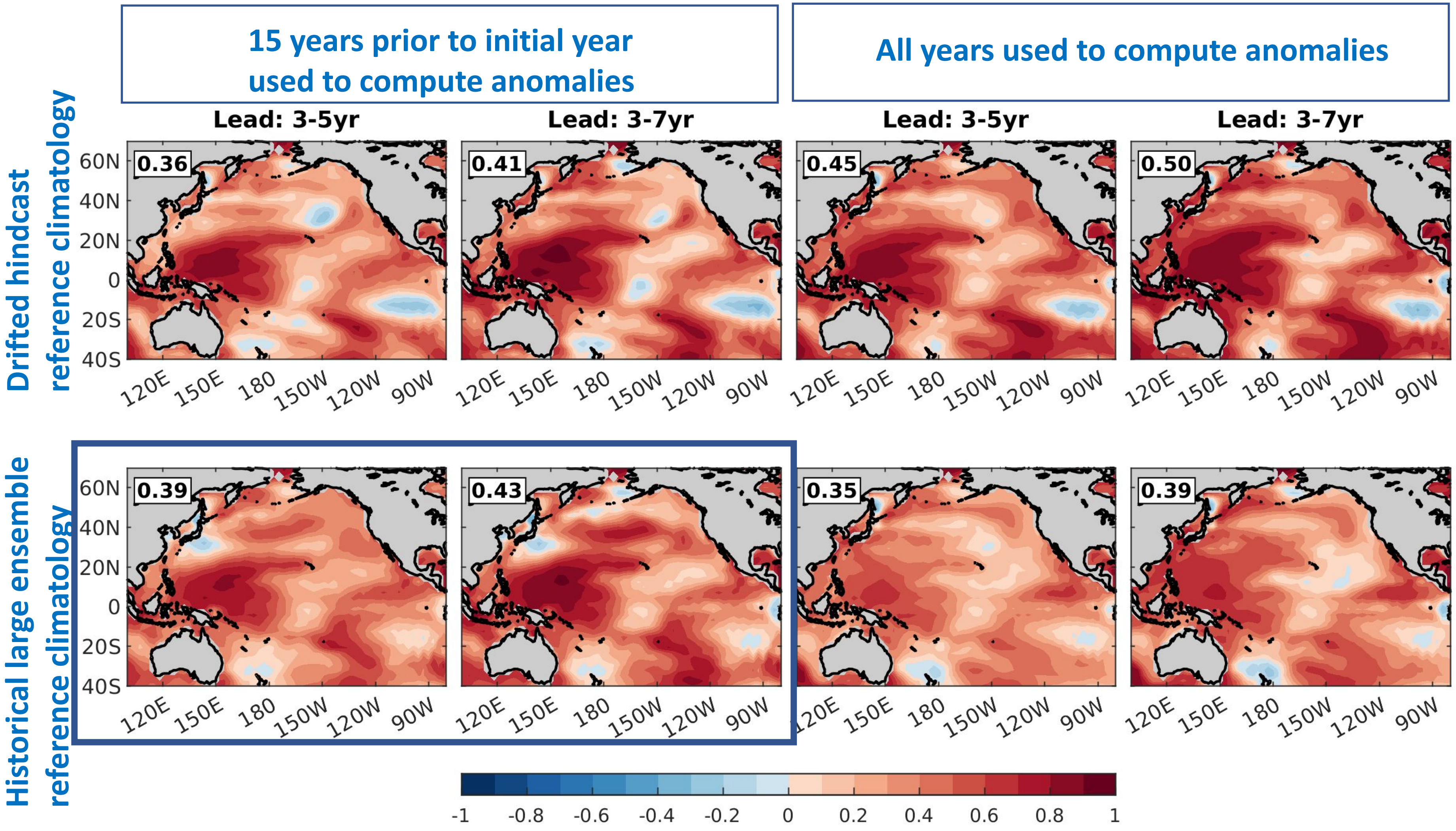


Anomaly correlations for the full set of initial years (each of 64 initial years from 1954 to 2017) from the CESM1 DPLE, lead years 3-5 and 3-7)



Anomaly correlations for the full set of initial years (each of 64 initial years from 1954 to 2017) from the CESM1 DPLE, lead years 3-5 and 3-7)

Highest and most consistent skill uses previous 15 years from historical large ensembles for lead ensembles for lead years 3-5 and 3-7 (two lower left panels)



Summary

For IPO predictions in the Pacific, it is possible to run a smaller number of start years for hindcasts and use the uninitialized free-running historical simulations to form the model climatologies to compute anomalies and remove drifts after about lead year 3

There is a model dependence regarding relative agreement of the pattern of drift errors between brute force and FOSI initialization schemes initially, but there are similar drifts after about lead year 2

Over 80% of the amplitude and pattern of the drift occurs by lead year 3 in most areas outside of tropics where there is high interannual variability

In DPLE, highest and most consistent skill uses previous 15 years from historical large ensemble to assess IPO prediction skill for lead years 3-5 and 3-7

caveat: there is a loss of skill for IPO predictions after major volcanic eruptions
(e.g. Meehl et al., 2015; Wu et al., 2023)
and after the Australian wildfire smoke event of 2019-2020 (Meehl et al., CVCWG talk yesterday)