https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

The moving target of simulating ENSO: A timeline of Community Earth System Model version 3 (CESM3) Development

Megan D. Fowler[1], Richard B. Neale[1], Cecile Hannay[1], David M. Lawrence[1], Adam S. Phillips[1], Isla R. Simpson[1], and Gustavo Marques[1]

[1]CGD Laboratory, NSF National Center for Atmospheric Research, Boulder, CO, 80305, USA

Correspondence: Megan D. Fowler (mdfowler@ucar.edu)

Abstract. The process of developing Earth System Models (ESMs) varies across modeling centers globally, but the overarching goal is largely the same - to improve the representation of physically-based processes such that biases in the mean climate state and its variability are minimized. Developers face a number of scientific and technical hurdles to ensure the best use of limited computational resources, storage space, and time. These challenges are particularly pronounced for modes of climate variability that are characterized by high internal variability, like the El Niño Southern Oscillation (ENSO). In this study, we leverage the development of the Community Earth System Model version 3 (CESM3) as a case study to illustrate some of these difficulties as well as to highlight new model developments and their impacts on ENSO.

ENSO is a complicated indicator of model performance in that it can be characterized by a wide number of metrics to assess how well (or poorly) it is simulated. This is at odds with the model development process as a whole, which requires that a manageably small number of metrics be selected for model analysis as teams conduct hundreds of simulations to arrive at a single "best" model configuration. Selecting too many metrics runs the risk of dramatically slowing progress in developing a coupled ESM. As a result, we discuss a minimal set of ENSO metrics here that we consider indicative of overall model performance. We find that biases in the spatial extent of ENSO events persisted throughout the development cycle, with sea surface temperature anomalies (SSTa) that extend too far into the West Pacific for all simulations. Other metrics are more sensitive to model changes, including ENSO amplitude and duration. Such metrics are, however, prone to significant internal variability. This is confirmed by temporally sub-sampling long pre-industrial control simulations of previous model versions (CESM1 and CESM2), which also adds critical context to the evaluation of CESM3; any new model version should ideally not be markedly worse than past iterations. Ultimately, CESM3 produces a reasonable ENSO in comparison to previous model versions and relative to observations, but it remains difficult to attribute changes in its representation to individual model changes due to the significant internal variability.

1 Introduction

The El Niño Southern Oscillation (ENSO) is the primary driver of weather and climate variability across the globe (Ropelewski and Halpert, 1987; Taschetto et al., 2020). It plays a central role in the interannual and decadal variability of precipitation and temperature (Diaz et al., 2001). ENSO variability often drives conditions such as persistent drought, precipitation extremes and

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

heat stress worldwide (Gershunov and Barnett, 1998; McPhaden et al., 2006; Williams and Funk, 2011; Power and Callaghan, 2016), with significant human impacts (Iizumi et al., 2014). An understanding of these impacts is necessary for successful earth system prediction capabilities, requiring the use of earth system models that capture the complex atmosphere ocean coupling mechanisms leading to ENSO. Correctly simulating ENSO has become a pre-requisite for model development success, and the gradual improvement over time has established a consistent reliability for models being used in Coupled Model Intercomparison Project (CMIP, Eyring et al., 2016) studies as well as more recently in seasonal prediction activities (Yeager et al., 2022). But questions as to how well a particular model is able to simulate ENSO compared to other models can be difficult to answer, and is further complicated by the challenge of understanding the physical mechanisms by which improvements occur (Wang, 2018).

Outside of model capabilities, even observations can be highly uncertain. The observed characteristics of ENSO, as reflected in the Niño 3.4 region sea surface temperature (SST) anomalies, are transitory over multi-decadal time periods (i.e., the observed characteristics of ENSO change over time). This is true for both the amplitude of events, where stronger SST anomalies exist after 1980, and the period, with the longest events in the 1950s and shortest events after the 2000s. It's clear that there is a distinct shift in the amplitude of Niño 3.4 SST anomalies at about the time where satellite products become available, such as for the Met Office Hadley Centre's sea ice and SST data set (HadISST, Rayner et al., 2003), where data was used from 1982 onwards. This limits the record we have available to validate simulations from models to such an extent that it may be insufficient (Wittenberg, 2009). To complicate the picture, studies have shown that the influence of climate change on the tropical atmosphere-ocean coupled system could also manifest as an increase in amplitude later in the observational period (Cai et al., 2023).

In assessing the fidelity of models in simulating ENSO, these uncertainties in the observational record are compounded by the difficulty in determining the inherent characteristics of a particular model configuration used in a particular forcing scenario. Improvements across model generations can be robust, primarily for SST (AchutaRao and Sperber, 2006), but this is not always the case for other local and remote impacts of ENSO (Alizadeh, 2024). In the past, improvements across model families has been revolutionary rather than evolutionary (Neale et al., 2008), resulting in a clear separation in multiple aspects of the phenomena (amplitude and phase, in particular). However, more recently model development has resulted in incremental improvements, which exist in the presence of large internal variability. Consequently, expected gradual changes in modeled ENSO behavior can be masked by abrupt changes in characteristics that can persist for decades even centuries in very long climate simulations (Wittenberg, 2009) .

In this study, we document the challenging process of understanding biases in the representation of ENSO across multiple simulations conducted while developing version 3 of the Community Earth System Model (CESM3), a final configuration of which will be used for CMIP7 (Dunne et al., 2025) submissions. This process entails understanding ENSO against the backdrop of simulations that are of different lengths, have different internal variability, are configured with different atmospheric model vertical extents, and/or are subject to different forcing scenarios (pre-industrial, historical, 4x CO_2). The process of finalizing CESM3 is, however, a lengthy one. Thus the final simulations shown in this study represent a near-final version of the model configuration but should not yet be taken as the official CESM3 release and ENSO performance may therefore continue to

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

evolve slightly from what is shown here. As the aim of this paper is not simply to assess ENSO in the next version of CESM, but to instead document the process by which we evaluate a complex mode of internal variability throughout the development cycle, we do not consider this a significant limitation.

The paper is laid out as follows. In section 2 we summarize the simulation set we use throughout the paper and how it relates to the development process for CESM3. Section 2 also summarizes a range of metrics used to quantify ENSO performance, based largely on SST. Section 3 details the metric performance as well as other measures of ENSO and ENSO feedbacks that can distinguish performance among the simulations. Conclusions and discussion are provided in 4.

2 Methods

We take a somewhat unique approach in this work by leveraging "simulations of opportunity" that have been conducted in the process of developing CESM3, to better understand how ENSO responds to a variety of model changes and provide context for current and future model development. The process of developing a skillful coupled earth system model takes a large team and many months or even years, as groups focused on individual earth system components come together with new or updated models, parameterizations, and tunings that were primarily developed independently. Once all of those components are brought together, the work begins to assess the performance of the fully coupled model and identify additional changes or tunings that are required for a reasonable representation of the Earth System. These changes are sometimes made in isolation (i.e., one change in one model component at a time), but more often are made in combination with others. This accelerates development, but complicates the process of identifying the unique impact of model changes on simulated weather and climate patterns. Nevertheless, various metrics of model performance are regularly tracked and drivers of changes can be attributed. Here, we attempt to bring at least part of that process to light.

2.1 Simulations

The simulations used in this analysis include active atmosphere, land, land-ice, sea-ice, and ocean components (i.e., "fully coupled"). Each run is initialized from spun-up initial conditions generated from an offline ocean and sea-ice model simulation, though the exact states vary throughout the development cycle as new offline spin-ups are performed when major model updates occur. Unless otherwise noted, all simulations are conducted using a pre-industrial (1850) forcing scenario due to computational constraints associated with longer transient historical simulations. Each simulation is tagged with an identification number, which we refer to here as its run ID. These are fairly straightforward; run 001 would refer to the first coupled simulation conducted during the development of CESM3; run 100 would refer to the hundredth run during the development process. At times, a letter is appended to note subtle changes in a series of concurrent experiments (changes in a single parameter setting or additional ensemble members, to explore the sensitivity of results, e.g., might be indicated as 020b, 020c, and so on). Additional appended identifiers indicate the height of the atmospheric model top ("MT" is a mid-top, mesopause top at 0.0008 hPA, which is distinct from the default low, upper stratosphere top at 2 Pa; see Simpson et al. (2025) for more details); different

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

forcing ("hist" indicates historically-evolving forcing instead of pre-industrial only); and ".002" indicates the second iteration of an existing simulation.

One of the challenges in using development simulations is that they are extended for various lengths of time, depending on the purpose of each run and a variety of performance factors. If the goal is to assess changes in low cloud radiative forcing, for instance, 20-30 years is likely sufficient to determine if a new model tuning has had the desired impact. But to assess a phenomena marked by high internal variability as in the case of ENSO, 100+ years would be much more suitable. In this analysis, we limit ourselves to simulations that have at least 40 years of monthly output, discarding the first five years to avoid the largest equilibration period in coupled simulations. One could imagine choosing a wide range of threshold values for this simulation length cutoff; we explore some of the impacts of that in Section 3. For simulations that have more than the minimum 40 years of output available, we include as much data as possible by not enforcing a cutoff. This means that the number of years used in each experiment is not constant, which in some cases may complicate comparison across model versions. That said, we explore the role of internal variability for each metric assessed and divide most 150+ year simulations into smaller pieces (denoted by "part1" and "part2" in their run ID). We view the different durations as an asset of the current study rather than a detriment, particularly as it reflects a challenge most modeling centers likely face when working through the model development process.

To assess the role of internal variability in our results, we leverage similar fully coupled pre-industrial control simulations from previous versions of CESM; CESM1 (Hurrell et al., 2013) and CESM2 (Danabasoglu et al., 2020). The CESM1 data was generated as part of the large ensemble project (Kay et al., 2015, 2021), and the CESM2 data as part of CMIP6 (Danabasoglu et al., 2019). For more detailed investigations of previous model versions, the ENSO representation in CESM2 is shown in Capotondi et al. (2020) and CESM1 in DiNezio et al. (2017). The control simulations we use here span 1200 years and are divided into non-overlapping 40 year periods. Each period is then treated as an individual ensemble member for the purposes of this analysis, resulting in 30 members per model version. We explore the impact of varying the length of each ensemble member below.

Importantly, the comparison of CESM3 development simulations to CESM1 and CESM2 again reflects a common practice used by CESM developers when producing a new model version. A first-order goal when creating an updated model is that it should not perform significantly *worse* than previous model versions. If, for instance, the spatial pattern of precipitation were to suddenly appear substantially more biased in a new model version, the development team would typically consider that unacceptable and return to the development cycle to identify causes of the bias and approaches to mitigate it. With regards to ENSO, this is more complex. As an emergent property of the coupled system, performance metrics related to ENSO are *influenced* by a large number of parameters but lack singular, identifiable parameters with known impacts on each measure of model performance.

2.2 Metrics

There are a wide variety of metrics that can be used to assess model simulations of ENSO (see for example, Planton et al. (2021); Capotondi et al. (2020)). We select a set of measures that describe the spatiotemporal characteristics of events in ways

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

that we have found especially useful for assessing model biases during the development of CESM3. Fig. 1 shows the behavior expected for each based on observations/reanalysis (HadISST and ERA5) over the period 1965-2002. Our goal in selecting this period is to minimize any anthropogenic signal (for better comparison to the pre-industrial CESM simulations) while also encompassing sufficient years to have confidence in the characteristics.

SST anomalies (SSTa) provide the basis of most ENSO metrics, including those used here. We compute these anomalies by removing the linear temporal trend at each point before removing the climatology of the annual cycle.

To evaluate the strength of ENSO events, we compute the annual cycle of SSTa variance within the Niño 3.4 region (Fig. 1a). The maximum, minimum, and mean of SSTa variances are then used as a metric to measure the seasonal behavior of ENSO amplitude in CESM. This will be discussed later. In observations, these variances peak during boreal winter (Nov-Jan), with a maximum variance in December of around 1.5 K^2. Values then weaken throughout spring, reaching a minimum near 0.4 K^2 in April. The difference between maximum and minimum SST variances indicates the overall strength of ENSO events relative to the background SST anomaly variance. In CESM cases with a wider than observed spread between maximum/minimum, SSTa undergoes too strong of annual cycle; in cases where the spread is smaller, the annual cycle is too muted and ENSO events are too weak relative to the mean.

In addition to the strength of ENSO events, it is important to consider their duration. Here, we use a simple metric for duration defined by the autocorrelation of Niño 3.4 SSTa out to lags of 24 months (Fig. 1b). The duration of a typical ENSO event in the dataset is then defined by the lag at which the autocorrelation changes sign (crosses the y-axis), determined by linear interpolation between the two consecutive points above and below the axis. Observations suggest a period of roughly 10 months for ENSO events before they transition to the opposite phase. This can, however, vary over time as we will discuss later.

We use two metrics to determine the spatial extent of ENSO. First, SSTa in each gridcell is correlated with the Niño 3.4 index in Fig. 1c. The western extent of positive SSTa correlations is defined by the westernmost point of the zero contour line, marked by a green circle in Fig. 1c. The western extent of "strong" SSTa correlations is similarly defined using the 0.5 contour line (marked by a green square). For reference, the Niño 3.4 region is indicated by a dashed-line rectangle. Within that region, correlations are maximized before they decay most rapidly towards the west.

To better assess the spatial characteristics of each event and feedbacks within the tropical Pacific, we assess the skewness of SSTa within the Niño 3 region (Fig. 1d). The PDF of SSTa exhibits a positive skew with a long tail towards warm events. The peak of the distribution lies on the cold side of the distribution. These characteristics reflect the non-linear atmosphere-ocean processes that promote the growth of El Nino events more favorably than La Nina events (An and Jin, 2004). Capturing this response has proven to be a challenge for large scale models, due to weak dynamical coupling and a muted precipitation response to El Nino SSTa (Zhao and Sun, 2022). ERA5 Reanalysis from ECMWF (Hersbach et al., 2020) and HadISST (Rayner et al., 2003) show a similar distribution, which is expected given that ERA5 SSTs are taken from the Hadley dataset prior to 2007. To quantify the tropical feedbacks at play during ENSO, we pair this distribution of SSTa with distributions of zonal surface stress anomalies (TAUXa) in the nearby Niño 4 region (Fig. 1e), and compute a linear regression between monthly values of each regional average, as discussed in Planton et al. (2021). We choose this metric since previous analysis

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

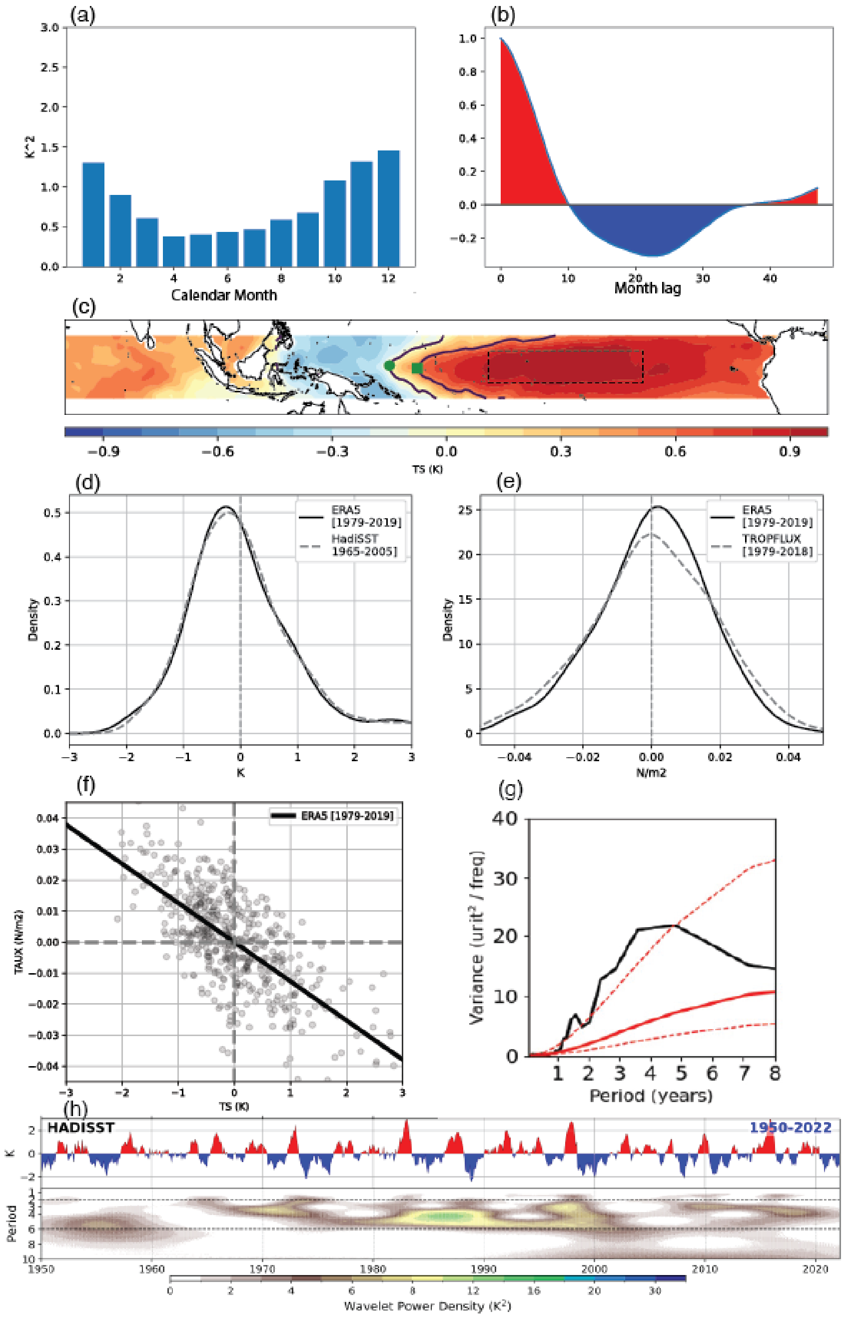

Figure 1. Characteristics of ENSO in observations. (a) Monthly Niño 3.4 SSTa variance from HadISST; (b) autocorrelation of Nino 3.4 SSTa, HadISST; (c) SSTa correlation with the Niño 3.4 index (shading), with black contour lines marking the 0.5 and 0 correlation contours and green points marking the western-most point of each contour; (d) probability density function of annual SSTa in the Niño 3 region; (e) probability density function of annual eastward-surface stress anomalies (TAUXa); (f) scatter plot and linear regression line between SSTa and TAUXa in (d) and (e); (g) power spectrum of SSTa variance in Niño 3.4 region, with red lines denoting the range of an AR1 red noise process. In d-f, observations are shown for ERA5 reanalysis (black) and additional observational products as available (dashed grey lines, with source noted in each legend); (h) nino3.4 observed complex Morlet wavelet spectrum (power density, K^2).

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

has identified common biases in a sub-set of CMIP6 models (Chen et al., 2021), and in particular the SSTa skewness has been shown to be deficient in previous versions of CESM (Zhang et al., 2017). The relationship between the two variables captures the strength of the feedback between surface temperature anomalies in the eastern Pacific and wind stress anomalies in the west (Fig. 1f), which are negatively correlated in reanalysis (i.e., negative slope in Fig. 1f).

The power spectrum (1g) shows that events have varying periodicity and although the maximum occurs at around four years, we can see periods ranging from three to five years and beyond. Multiple periodicity does not necessarily occur simultaneously, but rather evolves over multi-decadel time scales (1h). Consequentially a robust validation of periodicity may require many decades of simulation.

Biases in ENSO characteristics can in some instances be linked to biases in the mean state of models (e.g., cold SSTs, circulation biases, etc.). As appropriate, we compare the mean climatological state of CESM3 development simulations with the appropriate HadISST and ERA5 products and benchmarks as defined in Figure 1. As above, ERA5 and HadISST observations are primarily taken from the more modern record (1965-2002), though we note that this may be a source of disagreement when attempting to validate pre-industrial model behavior against modern climate. We thus also include a pre-industrial average (1870-1910) HadISST record when discussing biases in SST.

3 Results and Discussion

3.1 Internal Variability

Prior to investigating possible sensitivities of ENSO characteristics to developments made in CESM3, it is important to first assess internal variability of the various metrics. We do this by dividing 1200-year CESM1 and CESM2 pre-industrial simulations into non-overlapping 20-200 year segments, considering each time slice as an individual "simulation" on which ENSO metrics can be computed. These ensembles are thus comprised of simulations of varying length, all with identical model settings. The spread across ENSO metrics in each ensemble then yields an estimate of internal variability for each metric. In Figure 2, the sensitivity to sample length is calculated as the coefficient of variation (the ensemble standard deviation divided by the ensemble mean; CV hereafter). The closer a value is to zero (more blue in Fig. 2), the smaller the magnitude of spread is relative to the ensemble mean in the diagnosed metric, suggesting it can be more reliably computed than when the coefficient is larger.

For many metrics, it is striking how long of a simulation is required to achieve low CV. The Niño 3.4 annual maximum and minimum variances, as well as the skew of the Niño 3 SSTa distribution stand out as having very high CV when ensembles are relatively short (20-40 years), and may only reach <0.05 when two ensemble members of 500 years each are available, if then. Such a value suggests that the standard deviation across ensembles is 5 percent of the ensemble mean, at which point we suggest that the metric is likely to yield a reliable estimate that is unlikely to vary with longer simulation lengths.

Other ENSO metrics are surprisingly consistent even with short ensembles. For example, the western extent of positive SST anomalies associated with ENSO events has CV below 0.05, even with ensemble members that span just 20 years. Surface wind stress (TAUXa) skewness is prone to much less relative variability than SSTa, and as a result the slope of the regression

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

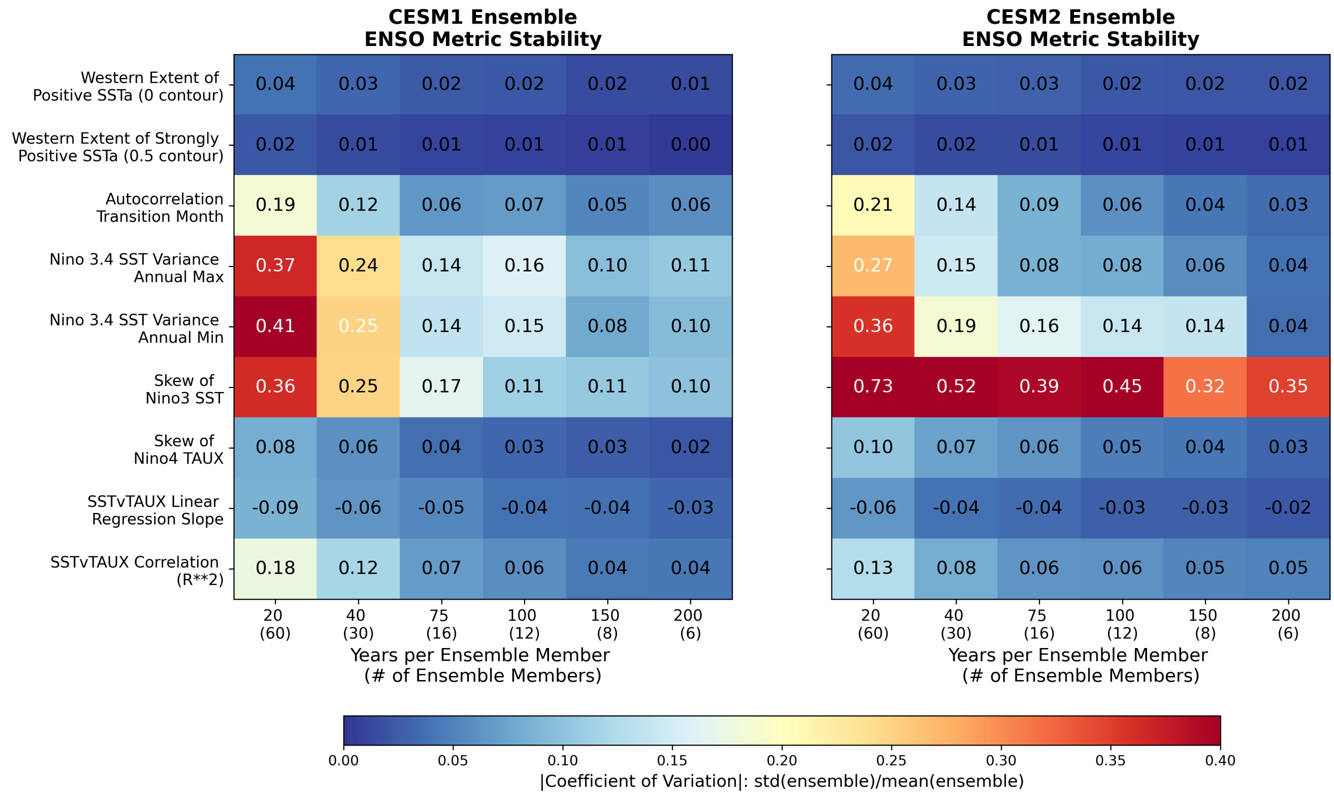

Figure 2. For a variety of ENSO metrics (y-axis) defined in the Methods section, values and shading represent the coefficient of variation across ensembles of CESM1 (left) and CESM2 (right), with ensemble members varying in the number of years that are included in each time-slice (x-axis)

line between the two variables has smaller CV as well. However, as we will discuss later, this measure does not indicate the *accuracy* of the metric relative to observations. CV only captures the spread within the model system. Based on this metric, samples should be on the order of 100+ years to limit uncertainties in the differences between simulations when assessing most characteristics of ENSO described here.

In most cases, CESM1 and CESM2 contain similar CV values (the shading and values in Figs. 2a and 2b are similar), though there are notable exceptions. The SSTa skew metric, for example, remains significantly more variable relative to the ensemble mean with increased time span in CESM2 (0.32 with samples of 150 years versus 0.11 in CESM1), suggesting more long term variability in upper-ocean temperatures. Conversely, the mean annual maximum of the Niño 3.4 SSTa metric yields a smaller CV in CESM2 than CESM1 (starting at a smaller value with 20-year ensemble members and reducing faster as ensemble member duration grows).

Figure 2 illustrates one of the largest difficulties in assessing ENSO during model development. Ideally, simulations should extend to at least 100 years to evaluate changes in the model's representation of ENSO with confidence. This is uncommon

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

for most development runs; in the course of CESM3 development, 300+ coupled simulations were conducted to test and tune the model with new physics options, resolutions, forcings, etc. Only a subset of these runs extended to the century mark to ensure that the overall climate and long term modes of variability, including ENSO, were behaving as expected. Through the development process, we recognized the need for increased frequency of these long simulations and it is now common to extend simulations for at least 70 years when diagnosing ENSO performance. Nevertheless, our goal here is to present a more comprehensive overview of model performance (from the beginning of the development cycle) and thus we will use the 40-year minimum criteria for further analysis. This gives us 64 CESM3 development simulations with a mean analysis duration (i.e., after subtracting the first 5 years of each simulation) of 76 years, a minimum of 36 years and maximum of 234 years.

At 40 years, we begin to see a slight reduction of CV in Figure 2 that is relatively unchanged from a higher threshold of 50 years (not shown, but which would limit the number of simulations included to 50). Only 10 development runs are conducted for 100+ years, which we consider too few to capture changes in ENSO across the development cycle of CESM3. We make the compromise of requiring simulations to run for 40 years, leaving 35+ years for analysis and provide perspective on the uncertainties involved using the CESM1 and CESM2 pre-industrial control sub-samples.

3.2 SST Sensitivity in CESM

Given that CESM development typically prioritizes the mean climate performance, we begin by assessing characteristics of the background ocean state in which ENSO evolves. We focus on equatorial SST magnitude and spatial pattern, which can exert a controlling influence on a number of ENSO characteristics.

We first consider the mean SST across the entirety of the equatorial Pacific (Fig. 3). For the DJF season, during which ENSO events tend to peak, CESM1 is biased cold by roughly 1 K, while CESM2 has a slight warm bias relative to present-day observations. These biases relative to the pre-industrial HadISST estimates are similar, though mean SST is 0.19 K warmer in present day as expected. Given the similarity between both periods, we show only the modern period HadISST average moving forward.

Taken as an ensemble, CESM3 development simulations exhibit more variations in mean SST than the sub-samples of the CESM1 and CESM2 pre-industrial controls. Despite this spread, almost all development simulations tend to show a significant cold bias that is more consistent with CESM1. The two most recent development simulations (blue points on the farthest right), which show the performance in near-final versions of CESM3, suggest a cold bias that may ultimately be even greater than seen in CESM1. We note, however, that additional development and tuning efforts may impact that bias prior to the official release of the model.

While SST bias in general is a useful summary metric of model performance, ENSO is sensitive to the gradient of temperatures across the basin (Heede and Fedorov, 2023; Maher et al., 2023). Figure 4 thus shows the longitudinal profile of the SST averages shown in Figure 3. While CESM1 is indeed biased cold relative to observations, this is especially true in the East and Central Pacific. Conversely, CESM2 is much closer to observations throughout the Central Pacific, though a slight cold bias is still present from 180-220E. Coupled with the strong warm bias at the western edge of the basin, this produces an amplified SST gradient compared to both observations and CESM1. The East Pacific cold bias in CESM1 is smaller than found in other

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

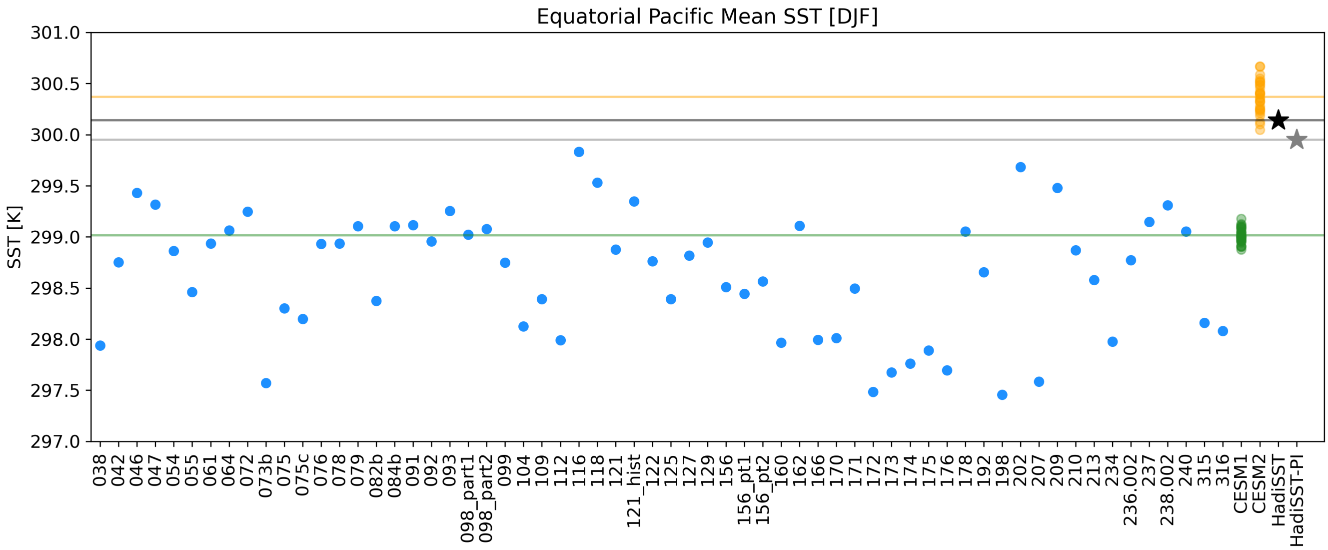

Figure 3. Average of equatorial Pacific (3S-3N, 140-270E) SST in boreal winter (DJF) for CESM3 development runs (blue), slices of a long-running CESM1 simulation (green) and slices of a long-running CESM2 simulation (orange). CESM1 and CESM2 averages are marked with horizontal lines of corresponding color. Observations from HadISST are marked with both a black star and horizontal black line for the modern record, and gray line/star for the pre-industrial period.

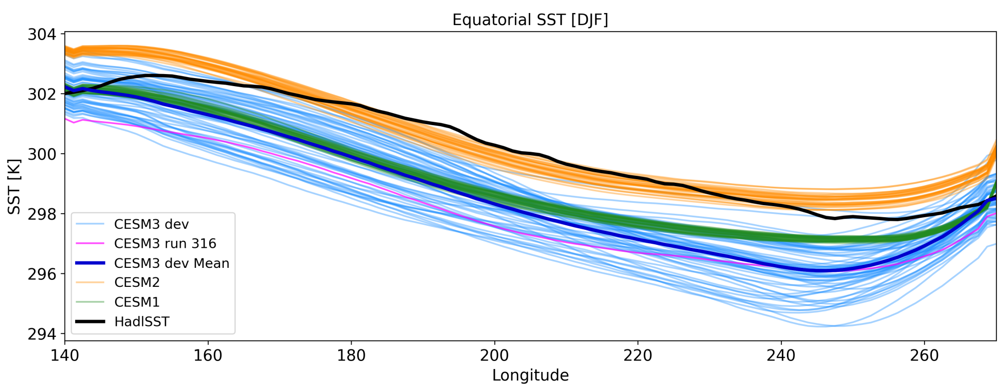

Figure 4. Equatorial average SST (3S-3N) in DJF across the Pacific Basin, as used to compute the averages in Fig 3.

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

parts of the basin, while the CESM2 warm bias is stronger. This again suggests a stronger gradient of SST in the eastern portion of the basin in CESM2 than in either observations or CESM1.

CESM3 development simulations exhibit a much wider spread than is found in either the CESM1 or CESM2 ensembles, especially in the eastern half of the basin. This suggests that there have indeed been substantive changes in the model representation of mean SST throughout development, compared to the final model versions otherwise shown. The most recent CESM3 simulation (pink line in Fig. 4) is cooler than the CESM1 ensemble members at all longitudes. Most strikingly, the problem of cold SSTs in the Niño 3.4 region, which has persisted through most of the development cycle, is reflected by nearly all simulations lying below the combined extent of both CESM1 and CESM2 ensembles. The along-basin SST gradient is more consistent with CESM1 than CESM2, though the decrease of SST across the West Pacific is still stronger in the model than in observations. Crucially, this establishes the SST gradient to be stronger than observations in almost the entire set of simulations shown. This has important consequences for the tropical climate and ENSO in particular, as the Bjerknes feedback suggests a strengthening of the trade winds in response and a corresponding deepening of the oceanic thermocline to the enhanced gradient (Bjerknes, 1969; Wang et al., 2016).

3.3 ENSO Magnitude

Building on the analysis of background SSTs, we begin to assess ENSO-specific model performance across a range of CESM3 development simulations. In Figure 5, we show the seasonal maximum, seasonal minimum and mean across months of Niño 3.4 SSTa variance, as would be derived from Figure 1a above. As before, the CESM1 and CESM2 experiments are 40-year slices of a single pre-industrial simulation, and HadISST is plotted for reference in black.

CESM2 shows slightly degraded model performance in this metric compared to CESM1; CESM2 variances are biased high relative to observations and are higher than CESM1. This is mostly a reflection of an inflated variance maximum in each slice. But the high internal variability means that a single time-slice of either model version could reasonably be expected to match observations fairly well or be strongly biased. It is difficult then to claim that changes in performance between many pairs of CESM3 development versions are due to individual model changes rather than a result of internal variability. There are, however, certain cases where the maximum and mean in particular fall outside the range of previous simulations. Runs 116, 122, and 171, for example, are characterized by unusually weak variability that is likely significant, while later simulations (i.e., runs 202 and 209) exhibit exceptional swings in monthly Niño 3.4 SSTa. These swings are, in part, to be expected. A number of parameterizations and tunings are tested during the development process, many of which target biases other than ENSO, and may be reverted/abandoned in subsequent simulations. It is thus important to note that although the simulations occur linearly in time, they are not necessarily linearly dependent on one another.

Across all simulations, we have not found that a clear, direct connection exists between background SST and ENSO magnitude (even though one might anticipate that there is a physical connection). Equatorial average SSTs between runs 116 and 202 for example are roughly identical at 299 K (Fig. 3) despite markedly different annual variance. In other words, there are additional complicating factors beyond SST gradient that influence the magnitude of simulated ENSO events. Yet there are indications that the spread between seasonal minimum and maximum Niño 3.4 SSTa variance is negatively correlated with

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

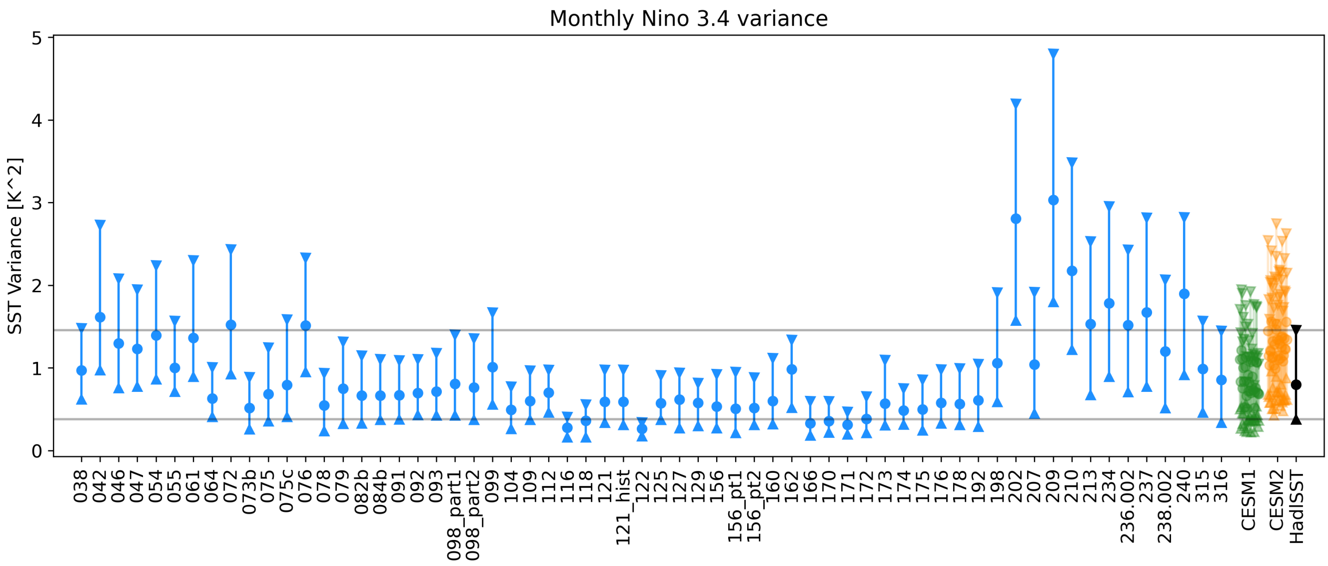

Figure 5. Mean (circle), maximum (downward triangle) and minimum (upward triangle) of monthly Niño 3.4 index variance across CESM3 development simulations (blue). 40 year periods are taken from the CESM1 (green) and CESM2 (gold) pre-industrial CMIP simulations to compute the Niño 3.4 index over as well. Observations from HadISST are shown in black, with the minimum and maximum values marked by horizontal grey lines.

equatorial SST gradient (Supplemental Fig. S1). Simulations with the weakest cross-Pacific SST gradient tend to exhibit the absolute smallest variability in Niño 3.4 SSTa seasonally, but the opposite does not hold true (i.e., simulations with the strongest gradient do not necessarily have the greatest variability).

Beyond individual stand-out experiments, there has been an overall trend in variances across the development cycle. In CESM2 and early development simulations, the SSTa variance was high, but that was reduced and even became biased low in subsequent development simulations. The overall spread went through another clear transition near the simulation 200 mark, with remarkably strong positive biases. The latest simulations, runs 315 and 316, have returned to within the range of CESM1 and agree well with observations.

We have attempted to link modifications made within the model to changes in ENSO performance where possible, though this is a somewhat complicated task. A notable change in ENSO behavior occurred in run 064, where SSTa variances decreased markedly compared to previous simulations. This simulation was the first to introduce a parameterization of convectively-driven surface wind gusts applied over oceans as in Redelsperger et al. (2000) and Harrop et al. (2018). The approach applies a "gustiness" factor to enhance 10-m winds in the presence of deep convection. The increase in surface winds leads to an enhancement of latent heat fluxes, which is especially prominent in the tropics where deep convection is common and surface winds are weak. This shift in the mean climate leads to a reduction in the strength of ENSO events. To confirm this impact, the parameterization was turned off in run 072, wherein the magnitude of SSTa variance increases to roughly the same level

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

as before the gustiness parameterization was implemented. During further development, a small bug was discovered in the implementation of the gustiness parameterization (fixed in 082b). To ensure that similar signals were still present using this corrected version we compare later runs 104 and 109. Though the impact is weaker with the updated formulation and in light of a range of other model changes, there remains a consistent signal that ENSO strength decreases when the convective gustiness parameterization is active. Variances also increased markedly near the simulation 200 mark, as noted above. This is largely related to experimental changes in the treatment of explicit diffusion within the atmospheric turbulence and boundary layer scheme, CLUBB (which was turned off or reduced in runs 202, 209, 210, and 213). The target of those changes was in fact not related to ENSO but to evaporation biases in the tropical oceans. Ultimately, explicit diffusion was returned to its previous magnitude (fully on), hence the subsequent reduction in maximum seasonal variance.

Relatively few of the model developments that have been implemented in CESM3 were designed with the explicit goal of reducing ENSO biases. Changes have included substantial shifts in not only parameterizations, but entire model components - CESM3 has a new ocean model (MOM6; Adcroft et al. (2019)) compared to CESM2 (POP; Smith et al. (2010)), for example. Given the high internal variability within this metric it is challenging to attribute substantive shifts in the strength of ENSO events to specific model changes. In fact, the spread across most development simulations lies within the spread generated by internal variability alone in the CESM1 and CESM2 ensembles, such that we likely require a long-running large ensemble to be confident that there has been a systematic change in this characteristic of ENSO by these metrics. The community would benefit from the creation of a CESM3 large ensemble with members that run for at least 100 years, or a long running PI control as is used for CESM1 and CESM2 here, to better assess the behavior of ENSO-related internal variability in isolation from changing model settings. Those types of simulations are expected after the finalization of CESM3.

3.4 ENSO Duration

Past versions of CESM, as well as most of the current CESM3 development simulations, are prone to overestimating the average duration of ENSO events (Fig. 6). Both CESM1 and CESM2 are biased slightly high relative to the 10.2 month event transition time observed in the HadISST record, though the bias is more pronounced in CESM1. The CESM3 development simulations span an especially wide range, from 9.8 to 46.7 months transition times, suggesting that some model changes may be significant (outside the range of internal variability). We have found, however, that this particular metric can at times be challenging to interpret. Autocorrelation curves may *approach* zero within a reasonable time (<12 months), but hover just above the inflection point without fully transitioning until 30+ months (Supplemental Fig. 2). In these cases, results should be interpreted with some caution but are retained here to illustrate the breadth of model behavior.

Throughout the development process, high internal variability has again played a prohibitive role in attributing changes. When separating run 156 into two 100-year time slices, for example, the mean event transition time changes from 10.7 months for years 6-106 to 19.5 months for years 107-207. Though not explicitly shown here, a handful of simulations initially suggested transition times of 30+ months despite being computed over at least 40 years. It was sometimes not until 70+ years that the duration would drop back into a more reasonable range. This was typically the result of a single anomalously long ENSO event that dominated the autocorrelation calculation in shorter simulations. In other simulations, discarding two or ten years

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

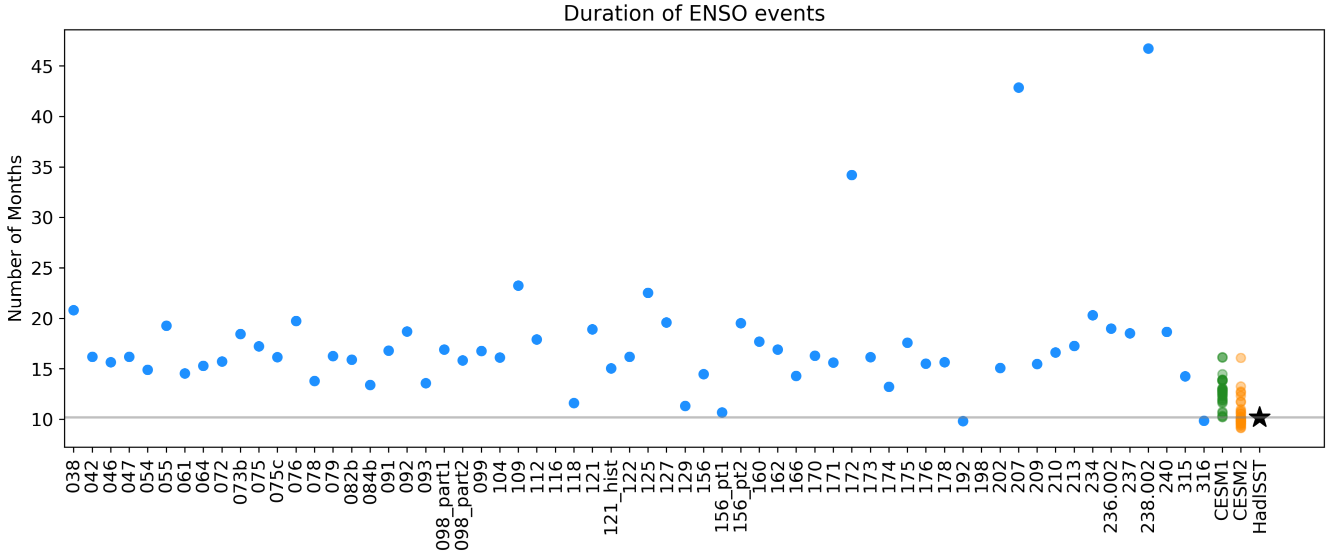

Figure 6. Transition time of ENSO events in each simulation based on the fractional month at which the lagged-autocorrelation of SSTa in the Niño 3.4 region transitions from positive to negative, as determined from plots similar to Fig 1
b.

at the beginning of the run had no real impact on the calculated duration of ENSO events, while it could create a fundamental difference in others. That said, there is at least some consistency across the development simulations as a whole and it suggests that the mean event transition time has increased in CESM3 relative to CESM1/2. This is one of the larger, persistent ENSO biases that later development runs have sought to mitigate. Indeed, the two most recent runs presented here (315 and 316) have event transition times that are closer to observations and within the range of previous model versions.

To understand possible drivers of the bias in ENSO event transition time, it is useful to return to the background climate state in each simulation. In Figure 7, we assess the relationship between ENSO duration (via autocorrelation transition time) and mean SST, but focus only on the East Pacific rather than the entire basin (assessments of the full basin and of the West Pacific did not demonstrate a clear pattern). In CESM3, there is a definitive relationship between the two fields, with a decrease in ENSO event transition time associated with warmer climatological SSTs. At colder SSTs, we hypothesize that La Niña events struggle to emerge; the background SST is simply so cold that the thermocline is already very near the sea surface and upwelling does not lead to even colder La Niña anomalies. Yet that relationship is not obviously present in either CESM1 or CESM2 pre-industrial controls. This suggests that variations in climatological SSTs from internal variability alone (as in the case of these ensembles) may not be related to event transition time, while structural model differences (as in CESM3 ensembles) can in fact drive variability in ENSO duration.

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

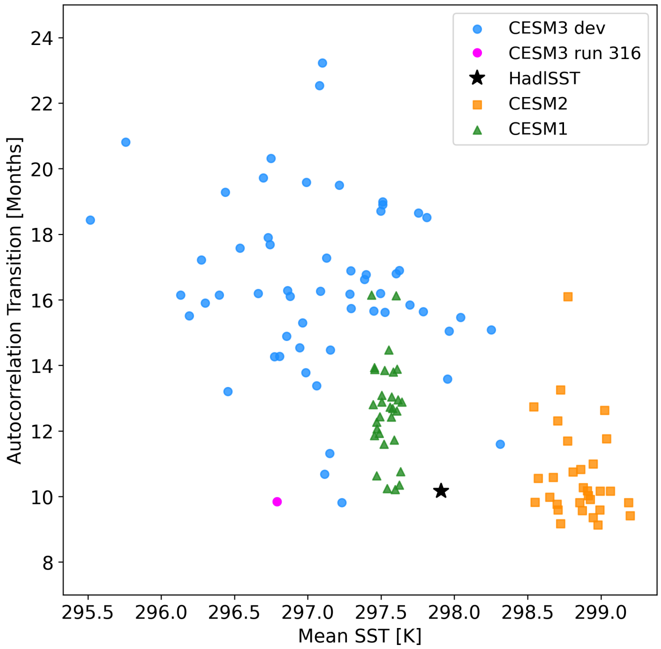

Figure 7. Mean DJF tropical East Pacific (-3S-3N, 220-280E) versus the month at which the Niño 3.4 SSTa autocorrelation reverses sign, in each CESM3 development simulation (blue), CESM2 (gold) and CESM1 (green) ensemble member, and HadSST observations (black star). The latest version of CESM3 development simulations is shown in pink for reference. Note that the y-axis is constrained to exclude outliers from Fig 6 and focus on the mean behavior

3.5 Spatial Extent

The western extent of ENSO-driven SSTa is persistently biased across many Earth System Models (Davey et al., 2002; Capotondi et al., 2015; Chen et al., 2021; Wu et al., 2022; Sandeep et al., 2026) and across generations of CESM (Capotondi et al., 2020). We confirm this here as well, measuered by the lag zero SSTa correlations with the Niño 3.4 index (Fig. 8). This bias was not specifically targeted during development of CESM3, but has remained difficult to improve upon all the same. The western extent remained nearly the same in early simulations (up to run 109 or so) and comparable to the extents in CESM1 and CESM2. The western extent of the zero contour, or the westernmost point of positive SSTa correlations with Niño 3.4, has been surprisingly insensitive to a range of model variations. The extent of "strongly" correlated SSTa values (indicated by the 0.5 contour, squares in Fig. 8) exhibits a bit more variability, but remains consistently biased relative to observations. This reflects the pervasive extension of Niño 3.4 SSTa into the central and western Pacific.

Starting in runs 112-118, both the extent of positive SSTa correlations with the Niño 3.4 index and especially the extent of *strongly* positive correlations become more variable compared to earlier experiments. Though no single change seems to drive the behavior, the largest responses in western extent have come as a result of changes in the ocean model, MOM6. For example, one difference between runs 116 and 118 is the addition of a parameterization for mixed layer eddies, which coincides with a

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

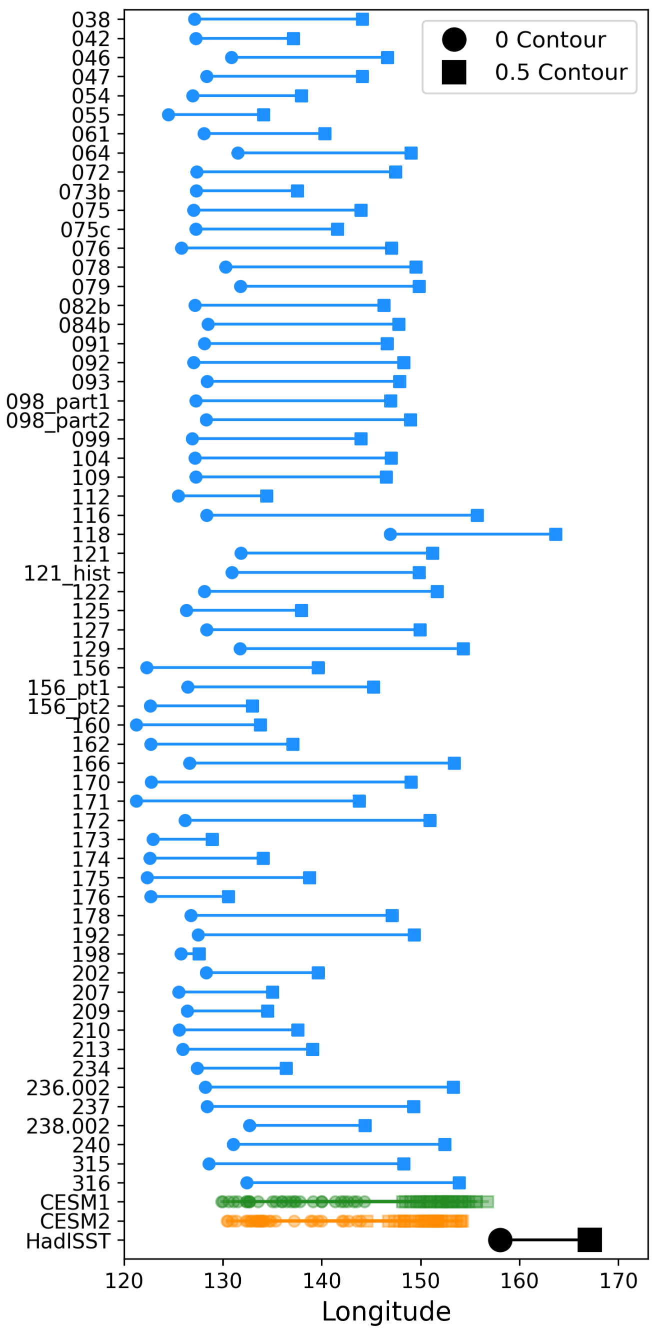

Figure 8. Western-most longitude of the 0 (circle) and 0.5 (square) contours of SSTa correlation with the Niño 3.4 Index, determined as in 1c

.

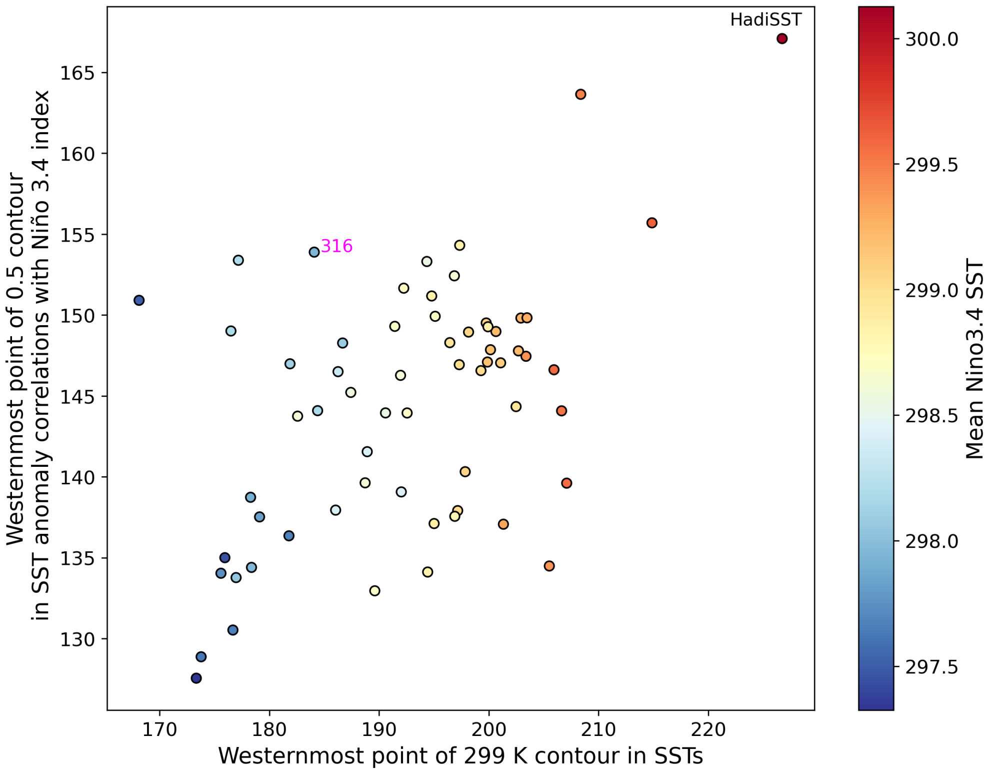

Figure 9. Scatter plot of CESM3 spatial extents, with shading denoting the mean SST in the Niño 3.4 region. The westernmost point of the 299 K contour in SSTs is used to mark the western extent of the Pacific cold tongue, while the 0.5 contour of SSTa autocorrelations with the Niño 3.4 index (squares in Fig 8) is used to mark the western extent of ENSO anomalies.

sizable reduction in the western extent of positive correlations. Runs 170 and 171 differ only in the application of an energy backscatter scheme in MOM6 in 171, but again the runs exhibit sizable differences in their western extent.

As mentioned briefly above, it's important to note that the CESM development process is far from linear. Changes in the model are made for a variety of reasons, including sensitivity tests, attempts to better understand drivers of biases, remove bugs in the code, etc. That is, not every simulation marks a step forward with which we find a positive result on CESM3 performance and keep the settings, building on them iteratively in every subsequent simulation. Adjustments and decisions to keep or abandon settings are made using a variety of criteria. It is thus difficult to assess performance in a single metric (in this case, ENSO western extent) and determine whether or not a certain model configuration should be retained. While some simulations do show improved performance in this bias, as in the case of run 118, this alone is not a sign that the model settings in that case were the "best" overall.

As before, we look for relationships that might exist between the background SST and the western extent of ENSO temperature anomalies (Fig. 9). The extent of the East Pacific cold tongue, indicated here by the westernmost point of the 299 K SST contour, is closely correlated with the westernmost extent of strongly positive ENSO SSTa and with Niño 3.4 SST (shading

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

in 9), in agreement with our previous findings and with past work (e.g., Wu et al. (2022); Capotondi et al. (2020)). These relationships between regional SST and ENSO duration/extent are not proof positive that the background SST climatology is the exclusive driver of biases in ENSO characteristics, but both speak to a measurable relationship between a key mode of variability and mean climate in a coupled model system.

3.6 Feedbacks, Asymmetry and Intermittency

Anomalies in equatorial SST and surface wind stress are key components for representing the impacts of ENSO on the atmosphere. Planton et al. (2021) highlight the feedback between East Pacific (Niño 3) SSTa and West Pacific (Niño 4) TAUXa as an important mechanism in the Bjerknes feedback that drives ENSO intensification. We evaluate the performance of CESM on this particular piece of the feedback loop to continue our emphasis on atmospheric-markers of ENSO performance, while noting that the evolution of the ocean mixed layer and the transit of oceanic Kelvin waves are also critical but outside the scope of this study.

Figure 10a and b show the relationship between SSTa and TAUXa in two ways: first by the Pearson correlation coefficient between monthly anomalies of each, and second by the slope of a linear regression line drawn through those monthly points. In both cases, CESM3 exhibits clearly wider spread compared to the CESM1 and CESM2 pre-industrial control samples. CESM3 primarily shows a weaker relationship between SSTa and TAUXa, suggesting that for a given SSTa it is less predictable how TAUXa will respond. In the case of correlation coefficient (Fig. 10a), the difference is enough to completely reverse the sign of the bias in the CESM3 ensemble relative to CESM2 and CESM1. In fact, it is only simulations made after run 192 that have a positive correlation coefficient, more in line with previous CESM versions (not shown). The most recent development run (marked in pink in Fig. 10a, b) is in fact similar to observations and in range of the CESM1 ensemble.

The mean slope of the regression line between SSTa and TAUXa is generally more comparable to previous CESM versions but again marked by significantly larger spread across simulations (Fig. 10b). For many CESM3 cases, this slope is much closer zero, suggesting no discernible linear relationships between the two fields despite a slight slope in observations. But as in the case of other metrics, the latest model version is one of the better performers. It's difficult, however, to rule out the possibility that other characteristics within the development simulations offset some of the expected behavior in this one feedback.

The individual probability distributions of SSTa and TAUXa (Fig. 10c, d) vary markedly between generations of CESM. CESM2, for example, has denser tails in its SSTa distribution compared to observations and other CESM versions. The stronger anomalies, on top of an overall warmer tropical Pacific (Fig. 4), likely contributed to more realistic patterns of teleconnections to the midlatitudes from the tropics (Fasullo et al., 2024). Distributions of TAUXa are less sensitive to model version, especially in the tails of the distribution. Yet all are markedly more peaked compared to observations, with smaller anomalies more common than larger ones.

CESM3 development simulations that exhibit a weaker relationship between SSTa and TAUXa (Fig. 10b), their wind stress distributions tend to be significantly more peaked than in runs that have a stronger relationship (Fig. 10e); that is, they have a larger kurtosis in TAUXa distributions. This is in the background of an overall bias, however, wherein all CESM versions suggest TAUXa Niño 4 distributions that continue to fail to capture the tails of the observed distribution. The SSTa distributions

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

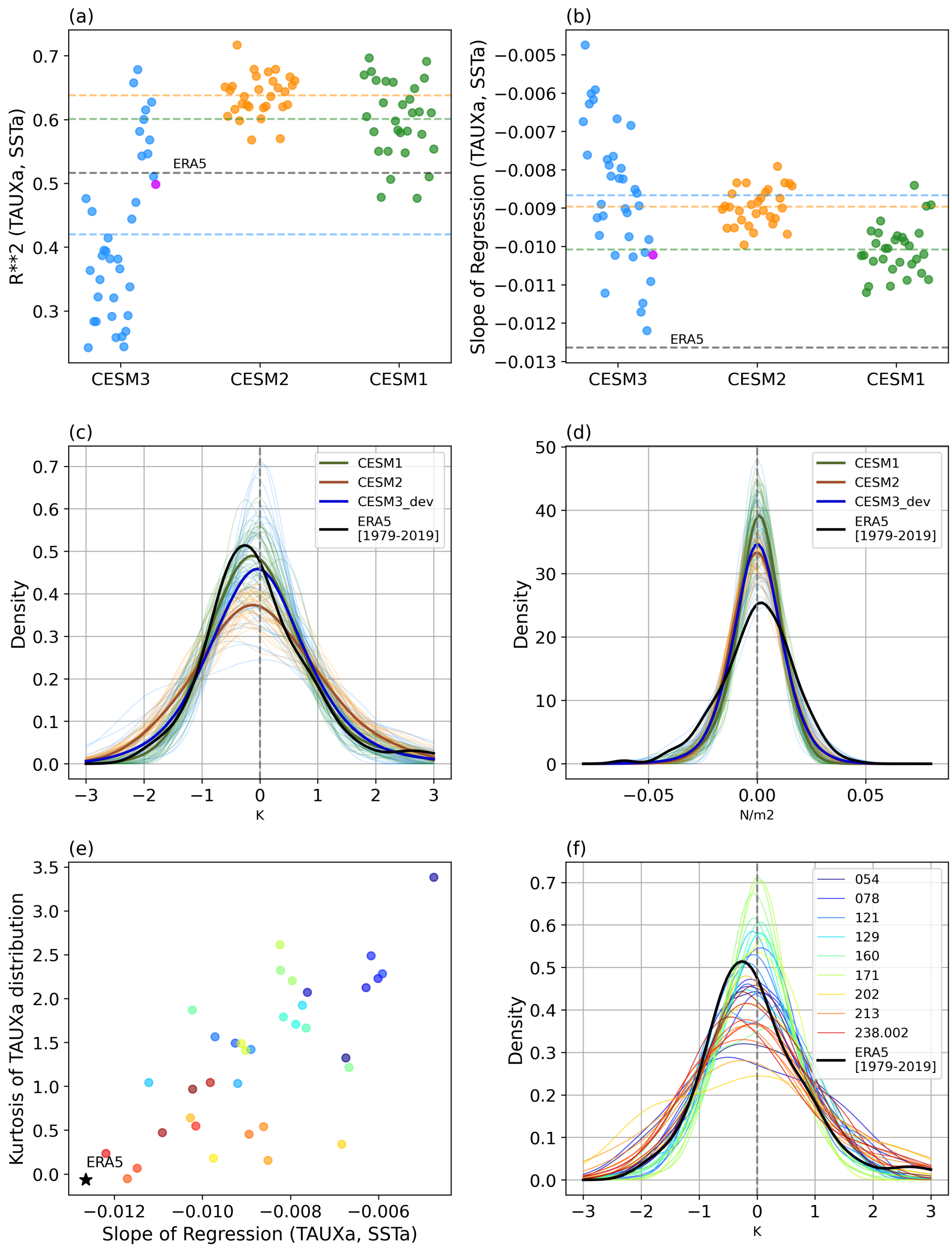

Figure 10. Tropical feedbacks between SSTa and TAUXa are assessed as in Fig 1d-f. Each dot in (a) and (b) represent one simulation or ensemble member, which are also represented by light transparent lines in (c) and (d). Means across model generations are marked by dashed lines (a,b), or bold lines (c, d) and the latest development run of CESM3 is marked in pink for reference (a-d). The relationship between the slope of the regression in (b) is plotted against a measure of TAUXa peaked-ness (kurtosis) in (e) for all available CESM3 experiments, shaded based on simulation number (with early runs in blue and late runs in red). To similarly assess evolution through development, (f) shows CESM3 SSTa distributions as in (c) but with shading as in (e). All panels include ERA5 reanalysis for reference (black).

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

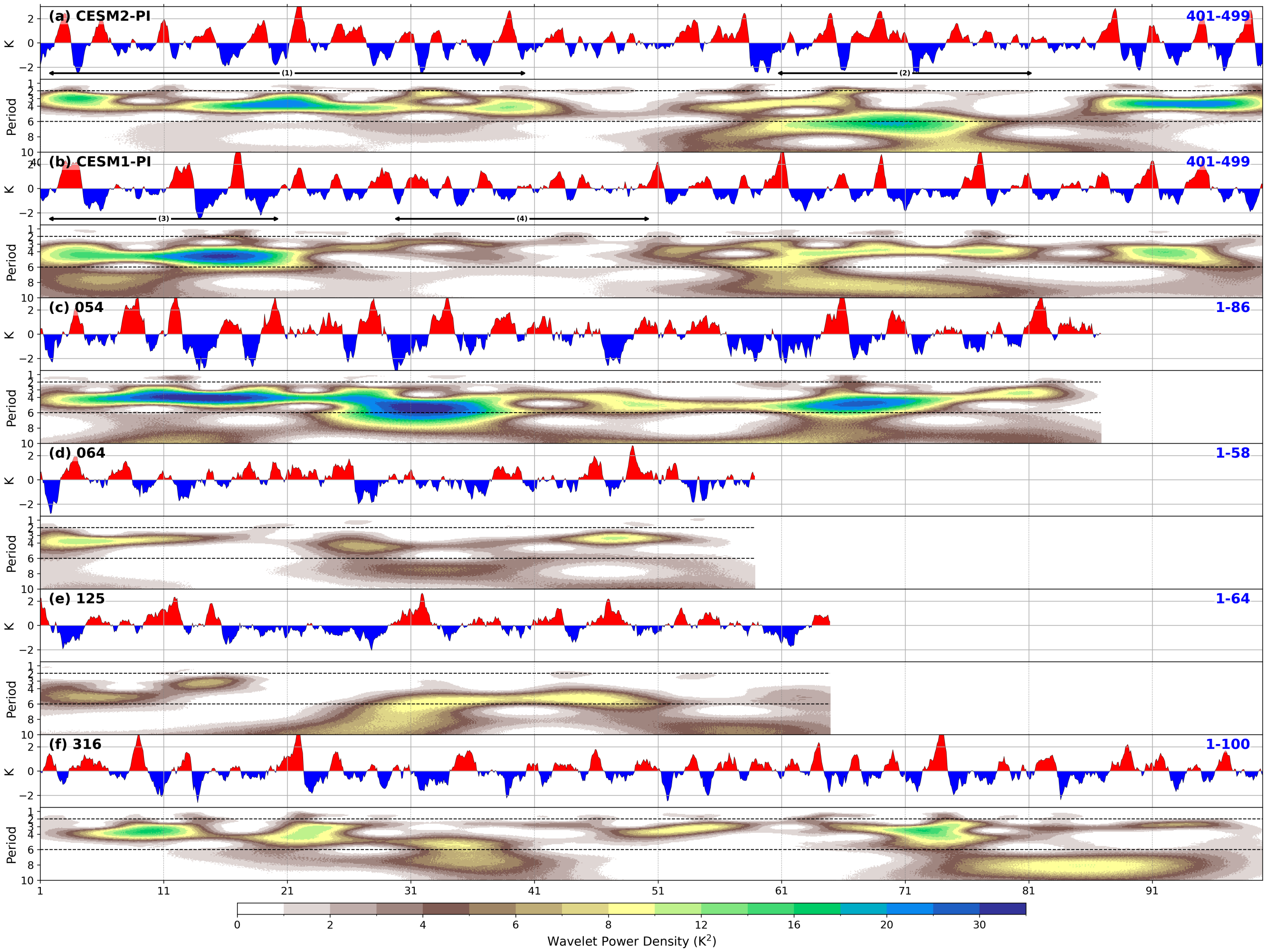

Figure 11. Timeseries and wavelet decomposition (Morlet method) of Niño 3.4 SSTa from (a) CESM2, (b) CESM1 and (c) through (g) a subset of CESM3 development simulations.

in the East Pacific also vary markedly across simulations in CESM3, though these differences have evolved fairly steadily throughout development (Fig. 10f). Early simulations (marked by blue/green colors in Fig. 10e, f) were characterized by SSTa distributions too centered around zero, with overly strong kurtosis values of the TAUXa distribution. Later simulations (marked by orange/red colors) have a more realistic TAUXa distribution as well as a regression slope between TAUXa and SSTa that is more similar to ERA5 reanalysis (Fig 10e). That said, the SSTa distribution has developed denser tails than observations in the most recent simulations (Fig 10f)

The multi-decadal internal variability of ENSO signals in models can be diagnosed from the perspectives of time variations in amplitude and period. Figure 11 shows the timeseries and wavelet decomposition of Niño 3.4 SSTa. Comparing to the observed

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

equivalent in Fig. 1g, we can see behavior that shows little stationarity. Observations exhibit periods of weak variability prior to 1980 and stronger variability through the 1980s and 1990s. In contrast, previous versions of CESM exhibit persistent variations in both amplitude and period. CESM2 maintains a consistent pattern of 2-5 year ENSO events for a 40 year period (marked with arrows spanning (1) in Fig. 11a), but transitions to a period where events become 6-8 years in length for 20 years (2), before returning to shorter periods. CESM1 shows similar behavior, but with multi-decadal time periods varying from strong amplitudes (3) to a near absence of events (4). Many aspects of these behaviors can be reproduced in the sub-set of CESM3 simulations shown. Consistent with the metrics in previous figures, the mean amplitude variations can be determined, but the inherent decadal variations make it challenging to categorically determine real behavior changes without 100+ year simulations. Finally, these spectra reveal that much of the model power seen in periodicities longer than 6 years is absent in the observed spectra. This may be reflected in the extended event transition time/long duration characteristics seen in most of the CESM3 simulations (Fig. 6)

4 Conclusions

The development of the latest generation of a world-class Earth System Model, CESM3, provides a unique opportunity to investigate how various decisions made throughout the development cycle impact the model's representation of ENSO, a key mode of climate variability. This process is complicated by a number of factors, including a changing model baseline (i.e., there is no one defined "CESM3 configuration" for the experiments studied here), varying simulation lengths, and high internal variability. The latter can be somewhat constrained by assessing variability in previous versions of the model (CESM1 and CESM2), which reveals that certain ENSO metrics are more variable than others. Figure 2 suggests that measures of spatial extent, for example, are more reliable with shorter ensemble members than those of ENSO amplitude. Lee et al. (2021) similarly suggest that most measures of ENSO characteristics require large sample sizes, with measures of amplitude and seasonality necessitating upwards of 40 ensemble members to come within 10% of a given large-ensemble mean, while metrics associated with spatial patterns may only require half that number (see their Fig. 4). This is an ambitious target to reach, but serves as additional evidence that ENSO metrics require a large number of years to reasonably assess.

Overall, CESM3 development has produced mixed results regarding the representation of ENSO. Our results are consistent with well-documented biases of CMIP models that simulate too wide of a westward SSTa extent, events that last too long and/or are overly intense (Bellenger et al., 2013; Capotondi et al., 2015; Ayar et al., 2023). ENSO seasonal amplitude (Fig. 5) varied considerably over the development of CESM3, but has stabilized in recent simulations near the observed values and within the range of previous CESM versions. Similarly, the duration of ENSO events (Fig. 6) has varied considerably, though typically CESM3 events are longer than in observations and often longer than what is captured in the envelope of CESM1 and CESM2 time-slices. But again, the latest simulations have settled both closer to observations and within the range of previous CESM performance.

All CESM3 simulations are marked by an overly-extended East Pacific cold tongue, which results in ENSO-related SSTa extending too far into the West Pacific for all simulations (Fig. 9). The bias is consistent with previous studies across multiple

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

Earth System Models (Davey et al., 2002; Capotondi et al., 2015; Chen et al., 2021; Wu et al., 2022; Sandeep et al., 2026). That spatial bias in Figure 8 is resistant to even large changes in the structure and parameterizations of CESM (i.e., the substitution of an entirely new ocean model occurs in CESM3 while producing little sensitivity versus CESM2). The current bias is within the range simulated by CESM1 and CESM2, but likely warrants dedicated model development effort given its persistence across models.

From a physical standpoint, the biases in ENSO characteristics are perhaps unsurprising given the biases governing tropical feedbacks between the West and East Pacific (Fig. 10). Recent studies have highlighted the critical importance of accurate Bjerknes feedback representation for both regional and remote ENSO impacts (McPhaden et al., 2006; Wang, 2018; Ferrett and Collins, 2016; Sullivan et al., 2025). Any weakening of this relationship in CESM3, as indicated especially in early model development versions, could thus have additional effects on seasonal prediction skill, though an exploration of such is outside the scope of the current study. It is reassuring to see that later versions of CESM3 have improved what was initially a large bias.

The goal of the work presented here is only partially to highlight changes in the representation of ENSO that can be expected once a final version of CESM3 is released. In fact, it is hard to conclude with any certainty that the representation in CESM3 is "better" or "worse" than CESM2 without the final configuration yet determined. The larger thrust of this work is aimed at shedding light on how one model development center (the National Science Foundation's National Center for Atmospheric Research, NSF NCAR) evaluates the simulation of a realistic ENSO while balancing competing priorities in developing a fully coupled Earth System Model.

Our approach thus makes use of “simulations of opportunity” rather than carefully designed experiments that test the impact of one model change at a time, reflecting the practicalities of the development exercise. The high internal variability inherent in ENSO metrics demands simulation lengths of 100+ years for robust assessment, yet most development runs are necessarily shorter to maintain development momentum. This tension between scientific rigor and practical constraints represents a fundamental challenge for model development groups, and one that may be handled differently across centers. This study represents a first step in establishing an open and transparent process for assessing just one mode of variability, but could reasonably be extended to additional metrics, expanded upon with input from other modeling groups, and serve as a foundation for a shared approach to enhancing model performance. Those tasks are more than can be accomplished in a single study and are reserved for future work.

Code and data availability. The CESM3 model output and analysis scripts used to create the figures in this study are archived on Zenodo (https://doi.org/10.5281/zenodo.21658223, Fowler et al. (2026)), along with code to reproduce the simulations. CESM1 and CESM2 data are publicly available at the data citations included in the text, with a subset of relevant data also available in the Zenodo archive.

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

Author contributions. MDF and RBN led the analysis and development of this study. MDF and RBN led the writing of the manuscript. CH and GM conducted the CESM3 model simulations included in this study, and contributed significant expertise along with AP, IS, and DML. DML led the overall CESM3 development effort. All authors contributed to the development of this manuscript through their shared insights and by review of the paper.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Geoscientific Model Development.

Acknowledgements. This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. Additional support was provided by NSF Award 1755088. Computing resources were provided by the Climate Simulation Laboratory at NCAR's Computational and Information Systems Laboratory (CISL) on Derecho (doi:10.5065/qx9a-pg09).

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

References

AchutaRao, K. and Sperber, K. R.: ENSO simulation in coupled ocean-atmosphere models: are the current models better?, Climate Dynamics, 27, 1–15, https://doi.org/10.1007/s00382-006-0119-7, 2006.

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., Dunne, J. P., Griffies, S. M., Hallberg, R., Harrison, M. J., Held, I. M., Jansen, M. F., John, J. G., Krasting, J. P., Langenhorst, A. R., Legg, S., Liang, Z., McHugh, C., Radhakrishnan, A., Reichl, B. G., Rosati, T., Samuels, B. L., Shao, A., Stouffer, R., Winton, M., Wittenberg, A. T., Xiang, B., Zadeh, N., and Zhang, R.: The GFDL Global Ocean and Sea Ice Model OM4.0: Model Description and Simulation Features, Journal of Advances in Modeling Earth Systems, 11, 3167–3211, https://doi.org/10.1029/2019MS001726, 2019.

Alizadeh, O.: A review of ENSO teleconnections at present and under future global warming, WIREs Climate Change, 15, https://doi.org/10.1002/wcc.861, 2024.

An, S.-I. and Jin, F.-F.: Nonlinearity and Asymmetry of ENSO, Journal of Climate, 17, 2399–2412, https://doi.org/10.1175/1520-0442(2004)017<2399:NAAOE>2.0.CO;2, 2004.

Ayar, P. V., Battisti, D. S., Li, C., King, M., Vrac, M., and Tjiputra, J.: A Regime View of ENSO Flavors Through Clustering in CMIP6 Models, Earth's Future, 11, e2022EF003 460, https://doi.org/10.1029/2022EF003460, 2023.

Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M., and Vialard, J.: ENSO representation in climate models: from CMIP3 to CMIP5, Climate Dynamics 2013 42:7, 42, 1999–2018, https://doi.org/10.1007/S00382-013-1783-Z, 2013.

Bjerknes, J.: ATMOSPHERIC TELECONNECTIONS FROM THE EQUATORIAL PACIFIC, Monthly Weather Review, 97, 163–172, https://doi.org/10.1175/1520-0493(1969)097, 1969.

Cai, W., Ng, B., Geng, T., Jia, F., Wu, L., Wang, G., Liu, Y., Gan, B., Yang, K., Santoso, A., Lin, X., Li, Z., Liu, Y., Yang, Y., Jin, F.-F., Collins, M., and McPhaden, M. J.: Anthropogenic impacts on twentieth-century ENSO variability changes, Nature Reviews Earth & Environment, 4, 407–418, https://doi.org/10.1038/s43017-023-00427-8, 2023.

Capotondi, A., Ham, Y.-G., Wittenberg, A., and Kug, J.-S.: Climate model biases and El Niño Southern Oscillation (ENSO) simulation, https://repository.library.noaa.gov/view/noaa/31041, 2015.

Capotondi, A., Deser, C., Phillips, A. S., Okumura, Y., and Larson, S. M.: ENSO and Pacific Decadal Variability in the Community Earth System Model Version 2, Journal of Advances in Modeling Earth Systems, 12, e2019MS002 022, https://doi.org/10.1029/2019MS002022, 2020.

Chen, H.-C., Jin, F.-F., Zhao, S., Wittenberg, A. T., and Xie, S.: ENSO Dynamics in the E3SM-1-0, CESM2, and GFDL-CM4 Climate Models, Journal of Climate, pp. 1–59, https://doi.org/10.1175/JCLI-D-21-0355.1, 2021.

Danabasoglu, G., Lawrence, D., Lindsay, K., Lipscomb, W., and Strand, G.: NCAR CESM2 model output prepared for CMIP6 CMIP piControl, https://doi.org/10.22033/ESGF/CMIP6.7733, 2019.

Danabasoglu, G., Lamarque, J., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, https://doi.org/10.1029/2019MS001916, 2020.

Davey, M., Huddleston, M., Sperber, K., Braconnot, P., Bryan, F., Chen, D., Colman, R., Cooper, C., Cubasch, U., Delecluse, P., DeWitt, D., Fairhead, L., Flato, G., Gordon, C., Hogan, T., Ji, M., Kimoto, M., Kitoh, A., Knutson, T., Latif, M., Treut, H. L., Li, T., Manabe, S., Mechoso, C., Meehl, G., Power, S., Roeckner, E., Terray, L., Vintzileos, A., Voss, R., Wang, B., Washington, W., Yoshikawa, I., Yu, J., Yukimoto, S., and Zebiak, S.: STOIC: A study of coupled model climatology and variability in tropical ocean regions, Climate Dynamics, 18, 403–420, https://doi.org/10.1007/s00382-001-0188-6, 2002.

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

DiNezio, P. N., Deser, C., Okumura, Y., and Karspeck, A.: Predictability of 2-year La Niña events in a coupled general circulation model, Climate Dynamics 2017 49:11, 49, 4237–4261, https://doi.org/10.1007/S00382-017-3575-3, 2017.

Dunne, J. P., Hewitt, H. T., Arblaster, J. M., Bonou, F., Boucher, O., Cavazos, T., Dingley, B., Durack, P. J., Hassler, B., Juckes, M., Miyakawa, T., Mizielinski, M., Naik, V., Nicholls, Z., O'Rourke, E., Pincus, R., Sanderson, B. M., Simpson, I. R., and Taylor, K. E.: An evolving Coupled Model Intercomparison Project phase 7 (CMIP7) and Fast Track in support of future climate assessment, Geoscientific Model Development, 18, 6671–6700, https://doi.org/10.5194/gmd-18-6671-2025, 2025.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fasullo, J. T., Caron, J. M., Phillips, A., Li, H., Richter, J. H., Neale, R. B., Rosenbloom, N., Strand, G., Glanville, S., Li, Y., Lehner, F., Meehl, G., Golaz, J.-C., Ullrich, P., Lee, J., Arblaster, J., Fasullo, J. T., Caron, J. M., Phillips, A., Li, H., Richter, J. H., Neale, R. B., Rosenbloom, N., Strand, G., Glanville, S., Li, Y., Lehner, F., Meehl, G., Golaz, J.-C., Ullrich, P., Lee, J., and Arblaster, J.: Modes of Variability in E3SM and CESM Large Ensembles, Journal of Climate, 37, 2629–2653, https://doi.org/10.1175/JCLI-D-23-0454.1, 2024.

Ferrett, S. and Collins, M.: ENSO feedbacks and their relationships with the mean state in a flux adjusted ensemble, Climate Dynamics 2016 52:12, 52, 7189–7208, https://doi.org/10.1007/S00382-016-3270-9, 2016.

Fowler, M., Neale, R. B., Hannay, C., Lawrence, D., Phillips, A., Simpson, I., and Marques, G.: Data & scripts for "The moving target of simulating ENSO: A timeline of Community Earth System Model version 3 (CESM3) Development", https://doi.org/10.5281/zenodo.21658223, 2026.

Gershunov, A. and Barnett, T. P.: ENSO Influence on Intraseasonal Extreme Rainfall and Temperature Frequencies in the Contiguous United States: Observations and Model Results, Journal of Climate, 11, 1575–1586, https://doi.org/https://doi.org/10.1175/1520-0442(1998)011<1575:EIOIER>2.0.CO;2, 1998.

Harrop, B. E., Ma, P. L., Rasch, P. J., Neale, R. B., and Hannay, C.: The Role of Convective Gustiness in Reducing Seasonal Precipitation Biases in the Tropical West Pacific, Journal of Advances in Modeling Earth Systems, 10, 961–970, https://doi.org/10.1002/2017MS001157, 2018.

Heede, U. K. and Fedorov, A. V.: Towards understanding the robust strengthening of ENSO and more frequent extreme El Niño events in CMIP6 global warming simulations, Climate Dynamics 2023 61:5, 61, 3047–3060, https://doi.org/10.1007/S00382-023-06856-X, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/QJ.3803;ISSUE:ISSUE:DOI, 2020.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A Framework for Collaborative Research, Bulletin of the American Meteorological Society, 94, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1, 2013.

Iizumi, T., Luo, J.-J., Challinor, A. J., Sakurai, G., Yokozawa, M., Sakuma, H., Brown, M. E., and Yamagata, T.: Impacts of El Niño Southern Oscillation on the global yields of major crops, Nature Communications, 5, 3712, https://doi.org/10.1038/ncomms4712, 2014.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., Vertenstein, M., Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., and Vertenstein, M.: The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability, Bulletin of the American Meteorological Society, 96, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1, 2015.

Kay, J. E., Deser, C., Phillips, A. S., and Simpson, I. R.: CESM1 Large Ensemble Community Project, https://doi.org/10.5065/D6J101D1, 2021.

Lee, J., Planton, Y. Y., Gleckler, P. J., Sperber, K. R., Guilyardi, E., Wittenberg, A. T., McPhaden, M. J., and Pallotta, G.: Robust Evaluation of ENSO in Climate Models: How Many Ensemble Members Are Needed?, Geophysical Research Letters, 48, e2021GL095 041, https://doi.org/10.1029/2021GL095041, 2021.

Maher, N., Wills, R. C., Dinezio, P., Klavans, J., Milinski, S., Sanchez, S. C., Stevenson, S., Stuecker, M. F., and Wu, X.: The future of the El Niño-Southern Oscillation: using large ensembles to illuminate time-varying responses and inter-model differences, Earth System Dynamics, 14, 413–431, https://doi.org/10.5194/ESD-14-413-2023, 2023.

McPhaden, M. J., Zebiak, S. E., and Glantz, M. H.: ENSO as an Integrating Concept in Earth Science, Science, 314, 1740–1745, https://doi.org/10.1126/SCIENCE.1132588, 2006.

Neale, R. B., Richter, J. H., and Jochum, M.: The impact of convection on ENSO: From a delayed oscillator to a series of events, Journal of Climate, https://doi.org/10.1175/2008JCLI2244.1, 2008.

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., and Voldoire, A.: Evaluating Climate Models with the CLIVAR 2020 ENSO Metrics Package, Bulletin of the American Meteorological Society, 102, E193–E217, https://doi.org/10.1175/BAMS-D-19-0337.1, 2021.

Power, S. B. and Callaghan, J.: Variability in Severe Coastal Flooding, Associated Storms, and Death Tolls in Southeastern Australia since the Mid–Nineteenth Century, Journal of Applied Meteorology and Climatology, 55, 1139–1149, https://doi.org/10.1175/JAMC-D-15-0146.1, 2016.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD002670, 2003.

Redelsperger, J. L., Guichard, F., and Mondon, S.: A parameterization of mesoscale enhancement of surface fluxes for large-scale models, Journal of Climate, 13, 402–421, https://doi.org/10.1175/1520-0442(2000)013<0402:APOMEO>2.0.CO;2, 2000.

https://doi.org/10.5194/egusphere-2026-4558
Preprint. Discussion started: 9 September 2026

Ropelewski, C. F. and Halpert, M. S.: Global and Regional Scale Precipitation Patterns Associated with the El Niño/Southern Oscillation, Monthly Weather Review, 115, 1606–1626, https://doi.org/10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2, 1987.

Sandeep, S., Chen, N., Manna, G., and Sandeep, S.: Linking mean-state biases to ENSO diversity in climate models: the key role of westerly wind burst suppression, Climate Dynamics 2026 64:6, 64, 256–, https://doi.org/10.1007/S00382-026-08227-8, 2026.

Simpson, I. R., Garcia, R. R., Bacmeister, J. T., Lauritzen, P. H., Hannay, C., Medeiros, B., Caron, J., Danabasoglu, G., Herrington, A., Jablonowski, C., Marsh, D., Neale, R. B., Polvani, L. M., Richter, J. H., Rosenbloom, N., and Tilmes, S.: The Path Toward Vertical Grid Options for the Community Atmosphere Model Version 7: The Impact of Vertical Resolution on the QBO and Tropical Waves, Journal of Advances in Modeling Earth Systems, 17, e2025MS004 957, https://doi.org/10.1029/2025MS004957, 2025.

Smith, R., Jones, P., Briegleb, B., Bryan, F., Danabasoglu, G., Dennis, J., Dukowicz, J., Eden, C., Fox-Kemper, B., Gent, P., Hecht, M., Jayne, S., Jochum, M., Large, W., Lindsay, K., Maltrud, M., Norton, N., Peacock, S., Vertenstein, M., and Yeager, S.: The Parallel Ocean Program (POP) Reference Manual Ocean Component of the Community Climate System Model (CCSM), Tech. rep., https://n2t.org/ark:/85065/d70g3j4h, 2010.

Sullivan, A., McGregor, S., Planton, Y., Zhong, W., Marsland, S., Rashid, H. A., Li, Z., and Geng, T.: The Biennial ENSO in the ACCESS-CM2 Due to a Weakened Atmospheric Bjerknes Feedback in May–June, Journal of Climate, 38, 1005–1020, https://doi.org/10.1175/JCLI-D-23-0743.1, 2025.

Taschetto, A. S., Ummenhofer, C. C., Stuecker, M. F., Dommenget, D., Ashok, K., Rodrigues, R. R., and Yeh, S.: ENSO Atmospheric Teleconnections, pp. 309–335, https://doi.org/10.1002/9781119548164.ch14, 2020.

Wang, C.: A review of ENSO theories, National Science Review, 5, 813–825, https://doi.org/10.1093/nsr/nwy104, 2018.

Wang, C., Deser, C., Yu, J.-Y., Dinezio, P., Clement, A., Wang, C., Deser, C., Yu, J.-Y., Dinezio, P., and Clement, A.: El Niño and Southern Oscillation (ENSO): A Review, Coral Reefs of the Eastern Pacific, 8, 85–106, https://doi.org/10.1007/978-94-017-7499-4_4, 2016.

Williams, A. P. and Funk, C.: A westward extension of the warm pool leads to a westward extension of the Walker circulation, drying eastern Africa, Climate Dynamics, 37, 2417–2435, https://doi.org/10.1007/S00382-010-0984-Y, 2011.

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

Wu, X., Okumura, Y. M., DiNezio, P. N., Yeager, S. G., Deser, C., Wu, X., Okumura, Y. M., DiNezio, P. N., Yeager, S. G., and Deser, C.: The Equatorial Pacific Cold Tongue Bias in CESM1 and Its Influence on ENSO Forecasts, Journal of Climate, 35, 3261–3277, https://doi.org/10.1175/JCLI-D-21-0470.1, 2022.

Yeager, S. G., Rosenbloom, N., Glanville, A. A., Wu, X., Simpson, I., Li, H., Molina, M. J., Krumhardt, K., Mogen, S., Lindsay, K., Lombardozzi, D., Wieder, W., Kim, W. M., Richter, J. H., Long, M., Danabasoglu, G., Bailey, D., Holland, M., Lovenduski, N., Strand, W. G., and King, T.: The Seasonal-to-Multiyear Large Ensemble (SMYLE) prediction system using the Community Earth System Model version 2, Geoscientific Model Development, 15, 6451–6493, https://doi.org/10.5194/gmd-15-6451-2022, 2022.

Zhang, T., Shao, X., and Li, S.: Impacts of Atmospheric Processes on ENSO Asymmetry: A Comparison between CESM1 and CCSM4, Journal of Climate, 30, 9743–9762, https://doi.org/10.1175/JCLI-D-17-0360.1, 2017.

Zhao, Y. and Sun, D.-Z.: ENSO Asymmetry in CMIP6 Models, Journal of Climate, 35, 5555–5572, https://doi.org/10.1175/JCLI-D-21-0835.1, 2022.