JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2023MS003653

Ensemble Spread Behavior in Coupled Climate Models: Insights From the Energy Exascale Earth System Model Version 1 Large Ensemble

Samantha Stevenson[1], Xingying Huang[2], Yingying Zhao[3], Emanuele Di Lorenzo[4], Matthew Newman[5,6], Luke van Roekel[7], Tongtong Xu[6], and Antonietta Capotondi[5,6]

[1]University of California, Santa Barbara, Santa Barbara, CA, USA, [2]National Center for Atmospheric Research, Boulder, CO, USA, [3]Pilot National Laboratory for Marine Science and Technology, Deep-Sea Research Center, Qingdao, China, [4]Georgia Institute of Technology, Atlanta, GA, USA, [5]University of Colorado, Boulder, CO, USA, [6]NOAA Physical Sciences Laboratory, Boulder, CO, USA, [7]Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract Assessing uncertainty in future climate projections requires understanding both internal climate variability and external forcing. For this reason, single-model initial condition large ensembles (SMILEs) run with Earth System Models (ESMs) have recently become popular. Here we present a new 20-member SMILE with the Energy Exascale Earth System Model version 1 (E3SMv1-LE), which uses a "macro" initialization strategy choosing coupled atmosphere/ocean states based on inter-basin contrasts in ocean heat content (OHC). The E3SMv1-LE simulates tropical climate variability well, albeit with a muted warming trend over the twentieth century due to overly strong aerosol forcing. The E3SMv1-LE's initial climate spread is comparable to other (larger) SMILEs, suggesting that maximizing inter-basin ocean heat contrasts may be an efficient method of generating ensemble spread. We also compare different ensemble spread across multiple SMILEs, using surface air temperature and OHC. The Community Earth system Model version 1, the only ensemble which utilizes a "micro" initialization approach perturbing only atmospheric initial conditions, yields lower spread in the first ~30 years. The E3SMv1-LE exhibits a relatively large spread, with some evidence for anthropogenic forcing influencing spread in the late twentieth century. However, systematic effects of differing "macro" initialization strategies are difficult to detect, possibly resulting from differing model physics or responses to external forcing. Notably, the method of standardizing results affects ensemble spread: control simulations for most models have either large background trends or multi-centennial variability in OHC. This spurious disequlibrium behavior is a substantial roadblock to understanding both internal climate variability and its response to forcing.

Plain Language Summary Large year-to-year variations in climate can arise due to chaos in the climate system, and as a result "large ensemble" sets of climate model simulations have become a popular way to estimate the magnitude of those variations. A new ensemble with the Energy Exascale Earth System Model version 1 is presented here, where the simulations have been started with as wide a range of amounts of heat contained in different oceans as possible. This allows a large spread across ensemble members, even with relatively few simulations. However, it is hard to say if this is the absolute "best" way of creating an ensemble, since many models have unrealistic trends in ocean heat content that lead to problems with calculating the differences between ensemble spreads.

Key Points:

- We present a new large ensemble with the Energy Exascale Earth System Model v1 (the E3SMv1-LE) initialized using ocean basin heat contrasts
- Ensemble spread in the 20-member E3SMv1-LE is comparable to that in other, larger model ensembles
- Model ensemble spread exhibits diverse behavior, resulting from initialization strategy, physical differences, and long-term climate drift

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Stevenson,
sstevenson@ucsb.edu

Citation:

Stevenson, S., Huang, X., Zhao, Y., Di Lorenzo, E., Newman, M., van Roekel, L., et al. (2023). Ensemble spread behavior in coupled climate models: Insights from the Energy Exascale Earth System Model version 1 large ensemble. *Journal of Advances in Modeling Earth Systems*, *15*, e2023MS003653. https://doi.org/10.1029/2023MS003653

Received 3 FEB 2023
Accepted 6 JUL 2023

Author Contributions:

Conceptualization: Samantha Stevenson, Xingying Huang, Emanuele Di Lorenzo, Matthew Newman, Luke van Roekel, Antonietta Capotondi

1. Introduction

Although current-generation climate models agree on many aspects of future climate change (Flato et al., 2014), the representation of climate variability and extreme events remains a challenge (e.g., Stevenson et al., 2021). This has implications for models' capacity to represent many quantities with significant impacts on human and natural systems: recent examples of significant events include the 1999-present drought in the western US (Williams et al., 2020), widespread wildfires over the West Coast during summer 2020 (Migliozzi et al., 2020), and marine ecosystem impacts from the 2015–2016 El Niño event (Brainard et al., 2018) and the 2014–2016 Northeast Pacific marine heat wave (Capotondi et al., 2022; Di Lorenzo & Mantua, 2016). Much of these observed variations in marine and terrestrial climatic conditions arises from "internal climate variability" (e.g., Deser et al., 2020); that is, variability arising from stochastic variations in climate in the absence of any external influence, including

Data curation: Samantha Stevenson, Xingying Huang, Matthew Newman, Luke van Roekel
Formal analysis: Samantha Stevenson, Xingying Huang, Yingying Zhao, Tongtong Xu
Funding acquisition: Xingying Huang, Emanuele Di Lorenzo, Matthew Newman
Investigation: Samantha Stevenson, Xingying Huang, Yingying Zhao, Emanuele Di Lorenzo, Matthew Newman, Antonietta Capotondi
Methodology: Samantha Stevenson, Xingying Huang
Project Administration: Samantha Stevenson, Emanuele Di Lorenzo, Matthew Newman
Resources: Samantha Stevenson, Xingying Huang, Emanuele Di Lorenzo
Software: Samantha Stevenson, Xingying Huang
Supervision: Samantha Stevenson, Emanuele Di Lorenzo, Matthew Newman
Validation: Samantha Stevenson, Xingying Huang, Matthew Newman
Visualization: Samantha Stevenson, Xingying Huang, Yingying Zhao, Tongtong Xu
Writing – original draft: Samantha Stevenson
Writing – review & editing: Samantha Stevenson, Xingying Huang, Yingying Zhao, Emanuele Di Lorenzo, Matthew Newman, Luke van Roekel, Tongtong Xu, Antonietta Capotondi

random variations in daily weather and large-scale, slowly evolving climate modes both within the atmosphere (e.g., the Pacific-North American pattern and North American Oscillation) and involving atmosphere-ocean coupled dynamics (e.g., the El Niño/Southern Oscillation, or ENSO).

Understanding the magnitude of both forced trends and internal variability is crucial for accurately quantifying the risks of climate-driven extremes (Deser et al., 2020; Huang & Stevenson, 2021; Xu et al., 2022). The two are additionally likely related: as anthropogenic climate impacts progress during the twenty-first century, modulations in background conditions are expected to modify both the processes operating during natural low-frequency climate variability (see examples in the Pacific by Joh & Di Lorenzo, 2017; Joh et al., 2021; Liguori & Di Lorenzo, 2018) and the associated impacts (Fasullo et al., 2018). This has motivated the rise in popularity of "large ensembles" in recent years: sets of multiple (typically 20 or more) simulations with a given climate model, varying only the initial conditions. An increasing number of modeling centers have chosen to run large ensembles (Kay et al., 2015; Maher et al., 2019; Otto-Bliesner et al., 2016; Rodgers et al., 2015, among others); intercomparisons of these so-called Single Model Initial-Condition Large Ensembles (SMILEs) might greatly enhance the accuracy of climate-driven risk quantification as well as advancing understanding of physical drivers of inter-model changes (Deser et al., 2020; Mankin et al., 2020).

Strategies for SMILE creation vary but generally fall into one of two categories: "micro" initialization, where each ensemble member starts with the same initial state apart from a small (typically roundoff-order) perturbation made to one model component (typically the atmosphere); and "macro" initialization, where each ensemble member is initialized with an entirely different atmosphere-land-ocean state (typically derived from the model's pre-industrial control simulation). The Community Earth System Model Version 1 (CESM1) Large Ensemble used micro initialization for most of its members (Deser, 2020; Kay et al., 2015), although a few were later run using varying ocean initial conditions (Kim et al., 2018). Most ensembles now rely on macro initialization, typically involving a quasi-random selection of initial states from an unforced control run of the corresponding model (e.g., Golaz et al., 2019; Maher et al., 2019). How a model's initialization strategy connects to the resulting spread across its ensemble members (analogous to internal variability), however, has not been well studied to date. Likewise, the possible role of temporally varying external influences from anthropogenic forcing on ensemble spread remains poorly understood. Here we present initial results from a new SMILE using the Energy Exascale Earth System Model version 1 (E3SMv1), and contrast with existing SMILE results to begin examining these questions.

2. E3SMv1 Model

In this study, we used the E3SMv1, the first version of the U.S. Department of Energy's flagship ESM (Golaz et al., 2019). Designed to make climate projections on scales relevant for energy applications, the E3SM has a higher resolution than many other models of its class, and a sophisticated representation of aerosol physical (MAM4 aerosols) and microphysical properties (CLUBB with MG2; Wang et al., 2020). The "standard resolution" E3SM version 1 (Golaz et al., 2019) used in the present study has a 1° resolution for the atmosphere and land, an 0.5° resolution river model, and a variable-resolution ocean/sea ice model ranging from 60 km (midlatitudes) to 30 km (equator/poles). Portions of the model are descended from the Community Earth System Model (CESM1), most notably the atmospheric component, which was branched from the Community Atmosphere Model version 5.3 (CAM5.3) but extensively modified thereafter. Changes from CAM include substantially finer vertical resolution and much more detailed aerosol microphysical schemes, which lead to large differences in cloud/aerosol interactions in both stratiform and shallow cumulus regimes (Rasch et al., 2019). The ocean, sea ice, and river components of E3SM are all new models, while the land component is based on the Community Land Model version 4.5 (CLM4.5) but with more complex hydrology and soil biogeochemistry parameterizations. A full description of E3SMv1 model physics is available in Golaz et al., 2019 and references therein describing the component models.

The overall performance of E3SMv1 has likewise been previously assessed (Golaz et al., 2019). This model reproduces historical climate features well, with mean-state biases comparable to other models of its class (e.g., equatorial Pacific cold biases, overly strong trade winds). The Atlantic Meridional Overturning Circulation is somewhat too weak in this model, as compared with other Coupled Model Intercomparison Project Phase 6 (CMIP6) simulations. Additionally, E3SM has a known bias toward overly cold global mean temperatures early in the twentieth century, with overly strong warming in the late twentieth/early twenty-first (Golaz et al., 2019). This is thought to result from a tendency for strong aerosol-driven radiative forcing.

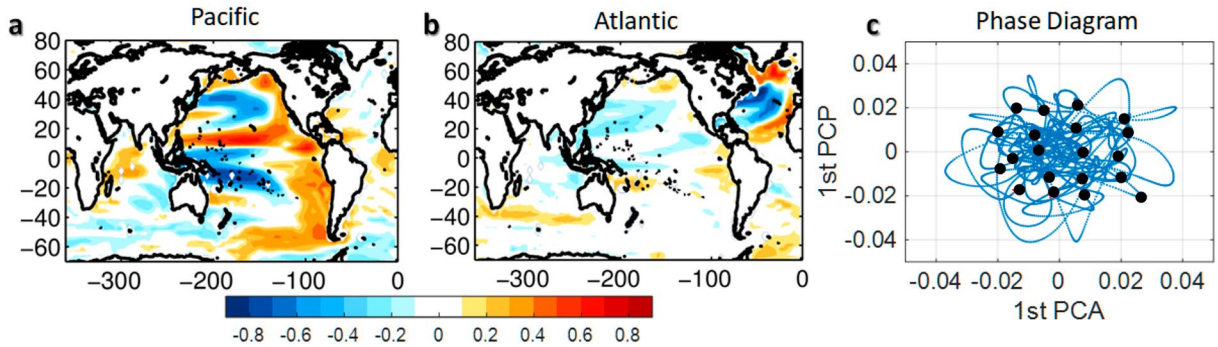

Figure 1. Analysis underlying the E3SMv1-LE initialization method. Correlation maps between 8-year lowpass OHC300 and (a) the leading principal component (PC) of 8-year lowpass OHC300 in the Pacific basin (first PCP), (b) the leading PC of 8-year lowpass OHC300 in the Atlantic basin (first PCA) for the 500-year E3SMv1 pre-industrial control. (c) Phase diagram of first PCA and first PCP. The blue dots show the positions within phase space as a function of time. The red stars show the locations of initial conditions used for ensemble generation in phase space.

The E3SMv1 Large Ensemble (E3SMv1-LE) follows a configuration that is similar to the historical E3SMv1 simulations run for CMIP6, apart from a few additional updates. These experiments are consistent with the CMIP6 Diagnosis, Evaluation, and Characterization of Klima (DECK) historical specifications for time-varying greenhouse gas concentrations, anthropogenic and volcanic aerosol emissions, and land use/land cover changes. The time period for all simulations is 1850–2015, consistent with simulations submitted as the DECK for CMIP6. E3SMv1-LE experiments differ from the CMIP6 E3SMv1 simulations only in the choice of timestep and the corresponding representation of horizontal mixing and the coefficients on momentum transport in the ocean. Additionally, many of the simulations also include code in the ocean component of the model allowing aggregation of terms in the ocean heat budget, such that regional closure of the budget is possible.

3. E3SMv1-LE Configuration: Inter-Basin Macro Initialization

A total of 20 simulations were run for the E3SMv1-LE. All simulations were branched off the 1850 pre-industrial control, providing an equilibrated set of initial states. We employed an initialization strategy based on upper-ocean heat content (OHC) to choose the appropriate initial state for each ensemble member, which is hereafter referred to as an "inter-basin" macro initialization due to its reliance on the relationship between ocean basin states. This approach bears some similarity to the strategy used by the HadGEM3-GC3.1 and UKESM1 ensembles run for CMIP6 (Sellar et al., 2020), but focuses on OHC rather than sea surface temperature. To maximize the intra-ensemble spread, we conducted a multi-basin analysis of OHC to identify those atmosphere/ocean states most distinct from one another, which were then selected as initial states. OHC integrated over the top 300 m (OHC300) was used to take advantage of the longer memory of OHC relative to sea surface temperature (SST) while remaining focused on the upper portion of the ocean.

Figures 1a and 1b show the correlation maps between 8-year lowpass OHC300 at each gridpoint and the leading principal component (PC) of 8-year lowpass OHC300, where the PCs have been determined separately for the Pacific Ocean (PCP hereafter) and the Atlantic Ocean PC1 (PCA hereafter), from the 500-year E3SMv1 pre-industrial (PI) control run. The patterns of PCP and PCA are spatially distinct, with PCP showing a larger loading in the Pacific and PCA a larger loading in the Atlantic, as expected. The independence of these modes is tested using the lead-lag correlation between PCA and PCP for lags of ±10 years, which shows insignificant correlations with maximum values of roughly 0.2 (not pictured). We note that the low correlation between Atlantic and Pacific OHC variability presents some contrast with previous work suggesting dynamical linkages between these basins (Dong et al., 2006; McGregor et al., 2014; Zanchettin et al., 2016; R. Zhang & Delworth, 2007). While it is most likely the case that there are indeed physical relationships between low-frequency variation in the Atlantic and the Pacific operative in E3SMv1, our results suggest that the amount of variance in upper-ocean heat content explained by inter-basin influences is relatively low on centennial timescales. Thus, the assumption of independence for PCA and PCP seems to be a reasonable one.

The PCP-PCA phase space was used to select specific restart years from the PI control (Figure 1c). Twenty points were chosen along the PI control's trajectory in phase space to be evenly distributed within the rectangle bounded by the 5th and 95th percentiles of PCA (x-axis) and PCP (y-axis). Additionally, the restart years were limited

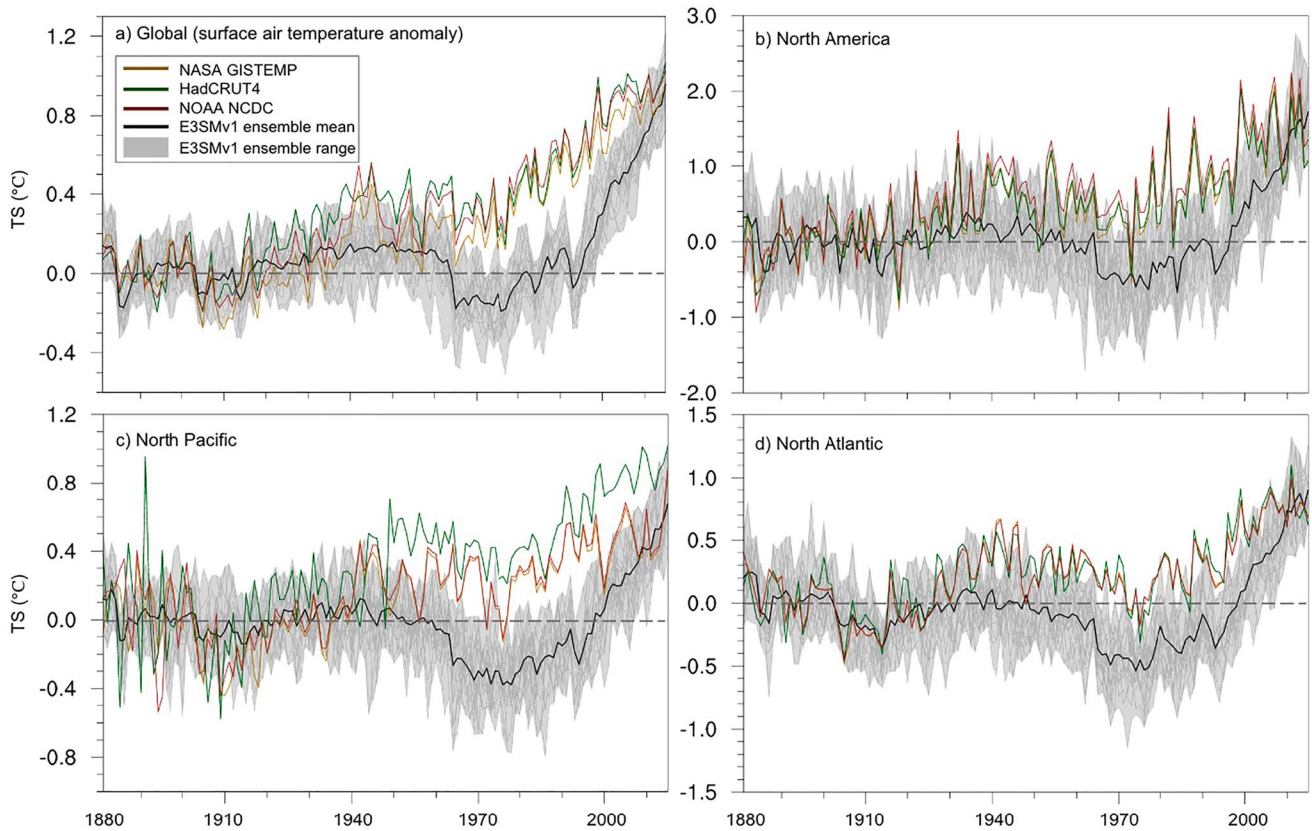

Figure 2. Area-weighted, spatially averaged time series of annual mean surface air temperature anomalies (with respect to 1880–1909) from the E3SMv1-LE (20 members) and observations. (a) Global average; (b) North America (regionally masked over the continental surface); (c) North Pacific (0–58N, 128–270E); (d) North Atlantic (0–68N, 277–357E).

since restart files were saved only every 5 years in the PI control run. Therefore, the phase space points corresponding to the restart file closest to the desired sample points were chosen as the initial conditions (Figure 1c).

4. E3SMv1-LE Performance

We first investigate the regional structure of surface temperature trends, as well as major modes of Pacific climate variability, as an assessment of overall model performance (other aspects of E3SMv1 performance have been presented by Golaz et al., 2019). The global surface temperature time series is shown in Figure 2a, referenced to the 1880–1909 time period and including comparison estimates from several observational datasets. E3SMv1 shows a relatively flat temperature trend between roughly 1850–1970, consistent with previous analyses and most likely due to the overly strong influence of anthropogenic aerosol forcing (Golaz et al., 2019; K. Zhang et al., 2022). After the 1970s, the global surface temperature recovers rapidly, and the warming trend from the mid-1970s to 2015 is even more rapid than in observations (Figure 2a).

There appears to be substantial regional variation in the time series behavior of surface temperature. Figures 2b–2d show that North America, the North Pacific, and the North Atlantic have distinctly different surface temperature evolutions. While in all three cases, the overall temperature trend is insignificant through the mid-twentieth century, the degree of cooling from the 1960s–1980s is much more pronounced over the North Pacific (Figure 2c). The mechanisms for these differences are not investigated in detail here but may relate to spatially distinct patterns of aerosol advection, for instance from East Asian sources over the North Pacific (Golaz et al., 2019; Smith et al., 2016).

E3SMv1 also appears able to well capture the overall structure of Pacific climate variability. In addition to having an ENSO amplitude comparable to observations (Golaz et al., 2019), the behavior of "precursor" and "teleconnection" patterns are also investigated (Figure 3). This analysis follows the approach of Zhao et al. (2021), who examined Pacific decadal variability across the CMIP5 model suite. Here, the ENSO precursor is found as

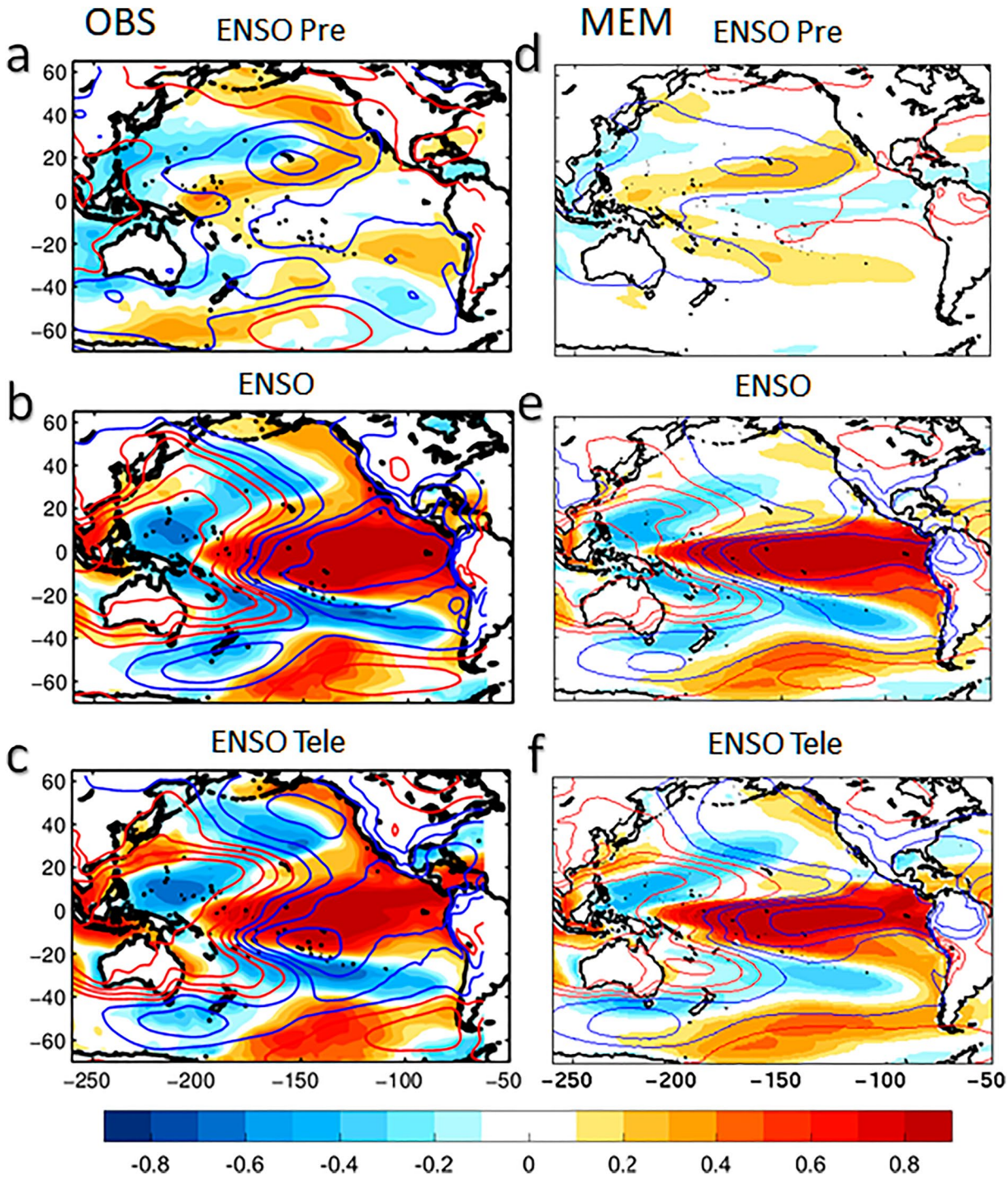

Figure 3. ENSO precursor, ENSO and ENSO teleconnection patterns. Left column: Shading: Observational oceanic ENSO precursor, ENSO and ENSO successor patterns (NOAA ERSST v3). Patterns are obtained by correlating the NDJ ENSO index with (a) SSTa in JFM, (b) SSTa in OND, and (c) SSTa in the following JFM. Contour: Observational atmospheric ENSO precursor, ENSO and ENSO successor patterns (NCEP SLP). Patterns are obtained by correlating NDJ ENSO index with (a) SLPa in JFM, (b) SLPa in OND, and (c) SLPa in the following JFM. The NDJ ENSO index is defined as the first principal component of the SST anomalies in November–December–January (NDJ) in the tropical Pacific (5°S–5°N). Right column: same with the left column but for MEM patterns.

the regression of the NDJ PC1 of Pacific SST over 5S–5N (hereafter the "NDJ ENSO index") onto gridpoint SST from the previous JFM period (Figures 3a and 3d); the ENSO pattern is the regression onto NDJ SST (Figures 3b and 3e); and the ENSO teleconnection pattern is the regression onto SST from the following JFM (Figures 3c and 3f). The spatial structures of both the precursor and teleconnection patterns are similar between E3SM and

Table 1
Large Ensembles Used in This Analysis

Model name	Initialization method	Ensemble size	Start date
E3SMv1	Macro-inter basin	20	1850
CESM1	Micro	40	1920
CSIRO Mk3.6	Macro	30	1850
IPSL-CM6A-LR	Macro	32	1850
CNRM-CM6-1	Macro	30	1850
EC-Earth3	Macro	19	1850
GFDL CM3	Macro/micro	20	1920

Note. Ensemble sizes indicate the number of members used for the present calculations; in most cases this is identical to that provided in the description paper, but in some cases online data availability increased following publication (e.g., for EC-Earth3).

observations, albeit with a tendency for E3SM to underestimate the strength of the midlatitude expression of both (Figures 3d and 3f). This is consistent with the overall behavior of CMIP-class models (Zhao et al., 2021).

5. Ensemble Spread Evolution Across SMILEs

A key goal of a large ensemble is to capture the internal variability of the climate system, which requires that the ensemble possess both a sufficient sample size and representative ensemble spread. Here, taking advantage of the recent proliferation of SMILEs, we consider the behaviors of ensemble spread calculated using several representative metrics:

- Surface air temperature
- OHC in the upper ocean (0–400 m)

These metrics provide contrasting information: OHC retains memory on decadal to multidecadal timescales, so its ensemble spread provides in some sense a "robust" estimate of memory communicated by initial condition spread, whereas surface air temperature has a shorter decorrelation timescale but is more directly related to societal impacts of climate variability.

The Multi-Model Large Ensemble Archive (MMLEA; Deser et al., 2020) was used as the primary data source for this analysis. All CMIP6-era large ensembles with available OHC output were included, and the CESM1 and GFDL CM3 large ensemble from the CMIP5 model generation were also included to provide comparison points with micro-initialized ensembles (Table 1). The ensemble initialization strategies for the macro ensembles differ somewhat, although all except E3SMv1 generally apply some form of semi-regular restart spacing stemming from the PI control run.

Details from the relevant description papers for the ensembles used are summarized as follows:

- *CSIRO Mk3.6*: restart years chosen at random intervals between 10 and 17 years in length (Jeffrey et al., 2013)
- *IPSL-CM6A-LR*: sampled "every 20 or 40 years" from the pre-industrial control beginning 20 years after the start of the simulation (Boucher et al., 2020)
- *CNRM-CM6-1*: restarts spaced roughly 10–40 years apart over the first 400 years of the equilibrated PI control run, method of choosing restart years not specified (Voldoire et al., 2019 Figure 3)
- *EC-Earth3*: restarts from PI control every 20 years (Döscher et al., 2021)
- *GFDL CM3*: restarts from five of the GFDL CM3 CMIP5 historical simulations, but initialized at four different starting years (five members each starting in 1907, 1908, 1909, and 1910), with micro-perturbations applied for each given starting year. Model years prior to 1920 were discarded as spinup (Sun et al., 2018; Sun, personal communication).

The starting month for the restarts is not explicitly specified in many of the papers cited above; however, common practice is for model restart files to begin in January (e.g., Kay et al., 2015; Otto-Bliesner et al., 2016) as is the case for the E3SMv1-LE.

Although the E3SMv1-LE is smaller than the other CMIP6-era ensembles (see Table 1), the choice to deliberately maximize the distance between initial states has the potential to make up for this smaller size by introducing spread within the ensemble more efficiently (e.g., Hawkins et al., 2016).

5.1. Surface Air Temperature

We first consider how the ensemble spread in surface air temperature, hereafter $\sigma_{\mathrm{TS-norm}}$, evolves over the course of each model's historical simulation. This quantity is calculated for each model by computing the standard deviation across all ensemble members as a function of time, for the surface air temperature averaged over different regions of interest:

$$\overline{\mathrm{TS}}(n,t) = \frac{1}{X}\frac{1}{Y}\sum_{x=1}^{X}\sum_{y=1}^{Y}\mathrm{TS}(n,x,y,t); \quad (1)$$

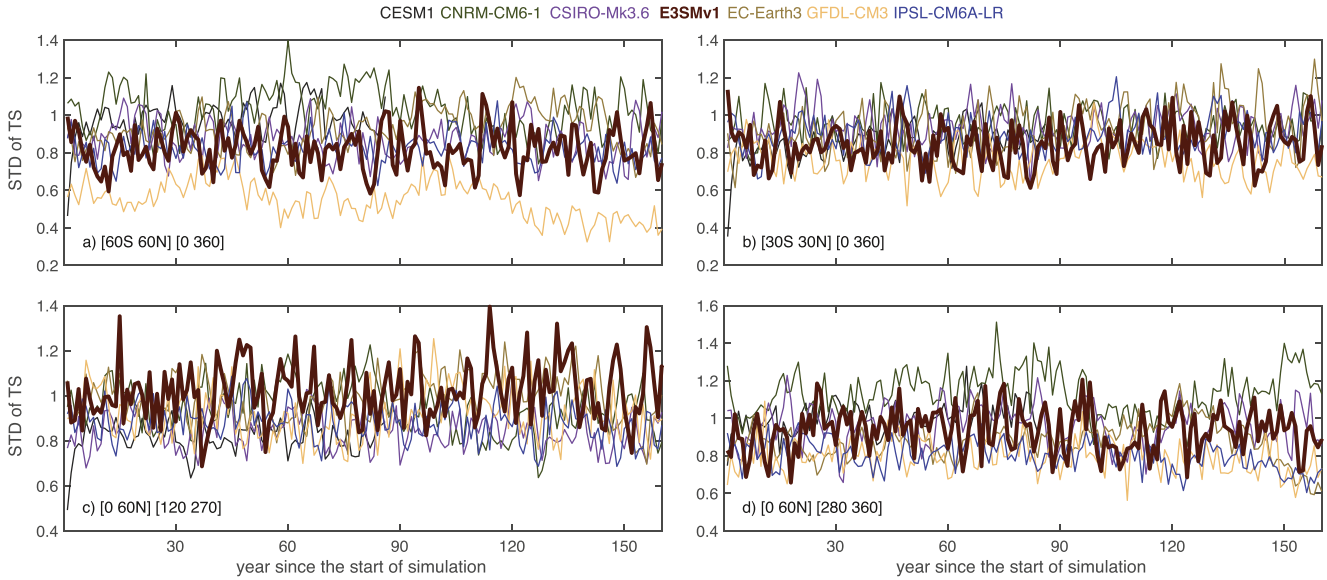

Figure 4. Time series of ensemble spread (standard deviation across all members as a function of time) for surface air temperature, averaged over (a) the entire globe excluding the polar regions; (b) the tropics; (c) the North Pacific; and (d) the North Atlantic. Values have been divided by the temporal standard deviation of the model's pre-industrial control simulation, and are therefore unitless.

$$\overline{\mathrm{TS}}(t) = \frac{1}{N}\sum_{n=1}^{N}\overline{\mathrm{TS}}(n,t); \tag{2}$$

$$\sigma(t)_{\mathrm{TS}} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\left(\overline{\mathrm{TS}}(n,t) - \overline{\mathrm{TS}}(t)\right)^2} \tag{3}$$

Equation 1 quantifies the region average surface temperature (TS) at each timestep for each ensemble member; Equation 2 calculates the ensemble mean TS at each timestep given Equation 1; Equation 3 is the expression for the standard deviation of TS at each timestep across the ensemble, or in other words, the ensemble spread. The output of Equations 1 and 2 is depicted in Figure 2.

To compare across models that have inherently different levels of internal variability, we divide the measure of ensemble spread in Equation 3 by the standard deviation of the regionally averaged surface air temperature time series derived from the corresponding pre-industrial simulation (referred to as $\sigma_{\mathrm{TS-PI}}$):

$$\sigma(t)_{\mathrm{TS-norm}} = \frac{\sigma(t)_{\mathrm{TS}}}{\sigma_{(t)\mathrm{TS-PI}}} \tag{4}$$

To ensure intercomparability between the standard deviation of the pre-industrial control and the ensemble spread, the PI control standard deviation $\sigma_{\mathrm{TS-PI}}$ is computed using deseasonalized data (see Section 5.2 for a more detailed discussion of seasonal effects). In all of the above equations, t is the monthly timestep, n refers to an individual ensemble member, N refers to the number of ensemble members, (x,y) spatial coordinates, and (X,Y) the total number of grid points in the x and y dimensions.

The resulting time series is shown in Figure 4. The largest differences between ensembles are seen at the very beginning of the simulation period (year ~0–10), where the micro-initialized CESM1 ensemble begins with a much smaller ensemble spread than any of the others. This behavior lasts only for the first few years, however, beyond which CESM1 reaches a roughly stable equilibrium spread on a similar order of magnitude as the other, macro-initialized ensembles.

The relative magnitudes of ensemble spread in surface air temperature appear to be highly regionally variable. Over the North Pacific (Figure 4c), the E3SMv1-LE exhibits one of the largest values of spread throughout the simulation period. However, in other regions such as the tropics (Figure 4b) and North Atlantic (Figure 4a),

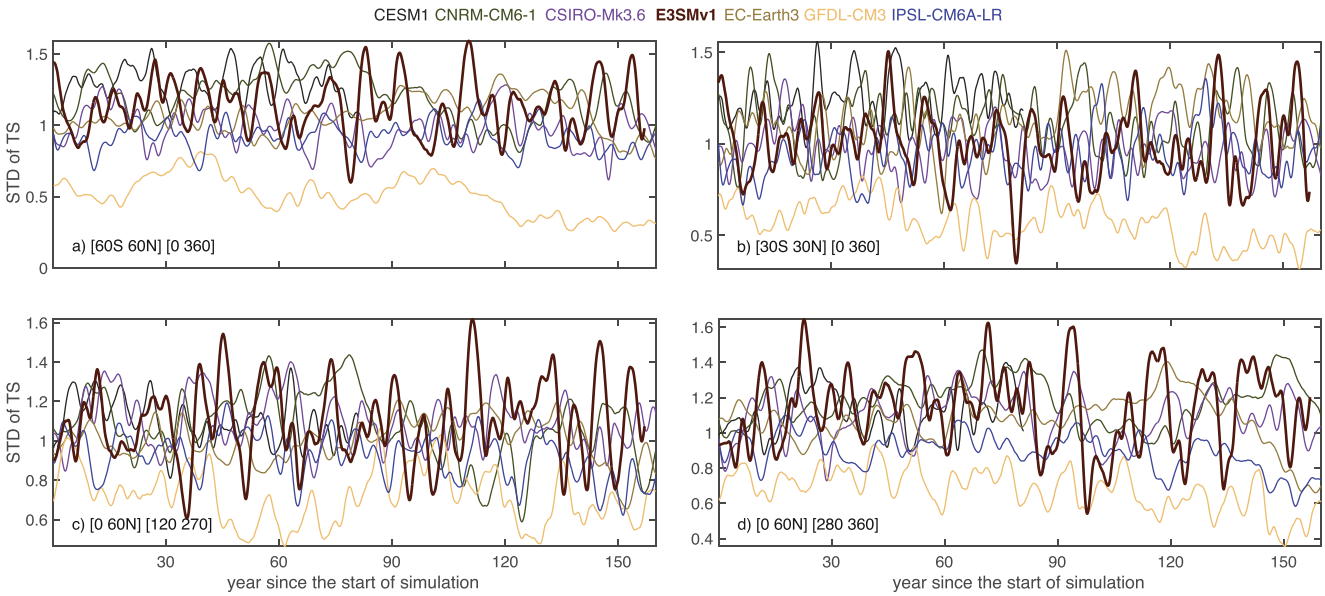

Figure 5. Same as Figure 4, but using 6-year lowpass-filtered values for historical surface air temperature. A lowpass filter is applied to the model's pre-industrial control simulation as well.

E3SMv1-LE falls roughly in the middle of the distribution of SMILEs. This suggests that the method of macro initialization employed in a given ensemble may not be predictive of the overall level of intra-ensemble spread, since there are no obvious systematic differences between particular macro-initialized ensembles which hold across all regions examined in Figure 4. There may be a role for the properties of internal modes of climate variability in setting these behaviors, however. For example, the ensemble ordering appears quite different in the North Atlantic/Pacific relative to the two tropical regions; in particular, CNRM-CM6-1 exhibits larger spread in the North Atlantic than other ensembles, and in the 60S–60N region, both EC-Earth3 and CNRM-CM6-1 show stronger spread than other models. A complete attribution of these behaviors is beyond the scope of this study, but should prove to be a fruitful area for future work.

Although the relative ordering of ensemble spread time series changes from region to region, it is also interesting that the regional spreads within a given ensemble tend to converge to fairly constant values. This convergence happens within a few years for both micro and macro ensembles, and the exact value appears to be a unique feature of each ensemble. Given the large number of physical differences among the models used to generate SMILEs, it is not possible to definitively attribute these lower spreads to the initialization strategy. However, it does appear that employing a macro initialization approach provides a greater chance of generating large ensemble spread, particularly within the first few years to decades of the simulation period.

We next examine lower-frequency variations in surface air temperature ensemble spread, which might conceivably be more strongly affected by ocean state given that high-frequency internal atmospheric variability is heavily influenced by synoptic influences and other sources of random noise. To that end, we apply a 6-year lowpass filter (Y. Zhang et al., 1997) to the surface air temperature time series and calculate ensemble spread as a function of time (Figure 5). The tendency for ensemble spread time series to remain roughly constant in most cases is apparent in the lowpass-filtered calculations, although the ordering of time series differs between Figures 4 and 5. Notable differences between the raw and lowpass-filtered spread include the larger separation of the GFDL-CM3 ensemble from the other SMILEs; GFDL-CM3 exhibits smaller lowpass-filtered spread than any other ensemble, across all regions considered in Figure 5. Also of interest is the fact that CESM1 no longer lies below other SMILEs, even in the first few years of the simulation period; this likely results from a combination of window effects from the bandpass filter masking the initialization "shock" at the beginning of the ensemble, and the large effect of ENSO on interannual climate variability in this model (e.g., Midhun et al., 2021).

Considering the E3SMv1-LE in the context of other macro-initialized SMILEs, the effect of the inter-basin initialization strategy is once again unclear from Figure 5. E3SMv1 does appear to have a fairly large lowpass-filtered spread compared with other models, throughout the simulation period, but is generally not the largest of the

SMILEs with the possible exception of some excursions in the North Atlantic (Figure 5d). As the relative ordering of macro ensembles again does not exhibit systematic relationships with initialization method, we conclude that initialization plays only a limited role in setting the degree of spread.

5.2. Ocean Heat Content

Any potential influence of initialization strategy is expected to be largest for quantities with slower timescales of variability. We therefore next examine the ensemble spread in upper OHC across SMILE output. The depth range, 0–400 m, is chosen since the timescale of equilibration is on the order of several decades, providing sufficient "memory" to indicate disequilibrium due to ocean spinup; results are qualitatively similar if other upper ocean depth ranges are used (see Figures S1 and S2 in Supporting Information S1). Calculation of OHC is performed according to:

$$\mathrm{OHC}(z) = c_P \rho \int_0^Z T(z) dz \tag{5}$$

where T is the ocean potential temperature and z the depth below the ocean surface. c_p and ρ are the heat capacity and density of seawater, respectively. Here a constant ρ is assumed due to the incompressibility assumption for seawater, consistent with common usage for OHC calculations (e.g., review by Abraham et al., 2013). Values adopted for c_p and ρ are 3,990 J kg^{-1} K^{-1} and 1,026 kg m^{-3} respectively.

The pattern of OHC over the 1958–2005 period (Figure 6) reveals that all models contain regionally varying biases compared to observations. Models tend to simulate colder-than-observed conditions in the tropics and to exhibit warm biases in the subtropics and high latitudes, particularly in the Southern Hemisphere. E3SMv1 shares these tendencies, but has a larger cold bias in the tropical Pacific than most other models; the warm bias in the high-latitude Southern Hemisphere is also relatively large, although comparable to the IPSL-CM6A-LR and EC-Earth 3.

To examine the spatial patterns of initial ensemble spread at the start of each simulation, we compute the spatio-temporally varying ensemble mean (Figure 7) as:

$$\overline{\mathrm{OHC}}(x, y, t) = \frac{1}{N} \sum_{n=1}^{N} \mathrm{OHC}(n, x, y, t); \tag{6}$$

To isolate the time behavior of ensemble spread, an approach analogous to that applied for surface temperature is used, where the spatial averages are first computed over a region of interest for each ensemble member:

$$\overline{\mathrm{OHC}}(n, t) = \frac{1}{X} \frac{1}{Y} \sum_{x=1}^{X} \sum_{y=1}^{Y} \mathrm{OHC}(n, x, y, t); \tag{7}$$

$$\overline{\mathrm{OHC}}(t) = \frac{1}{N} \sum_{n=1}^{N} \overline{\mathrm{OHC}}(n, t); \tag{8}$$

These may then be used to calculate the standard deviation of OHC across the ensemble (i.e., the ensemble spread) either as a function of both space and time:

$$\sigma(x, y, t)_{\mathrm{OHC}} = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} \left(\mathrm{OHC}(n, x, y, t) - \overline{\mathrm{OHC}}(x, y, t) \right)^2} \tag{9}$$

or as a regionally-averaged (area-weighted) time series:

$$\sigma(t)_{\mathrm{OHC}} = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} \left(\overline{\mathrm{OHC}}(n, t) - \overline{\mathrm{OHC}}(t) \right)^2} \tag{10}$$

As for surface air temperature, to ensure that differences in the magnitude of internal variability across ensembles do not bias the results, the ensemble spread is finally divided by the deseasonalized standard deviation of the regionally averaged OHC in the PI control run from each model. This difference is crucial for ensuring a proper comparison; see Figure S3 in Supporting Information S1 for a comparison of the (spatially variable) ensemble

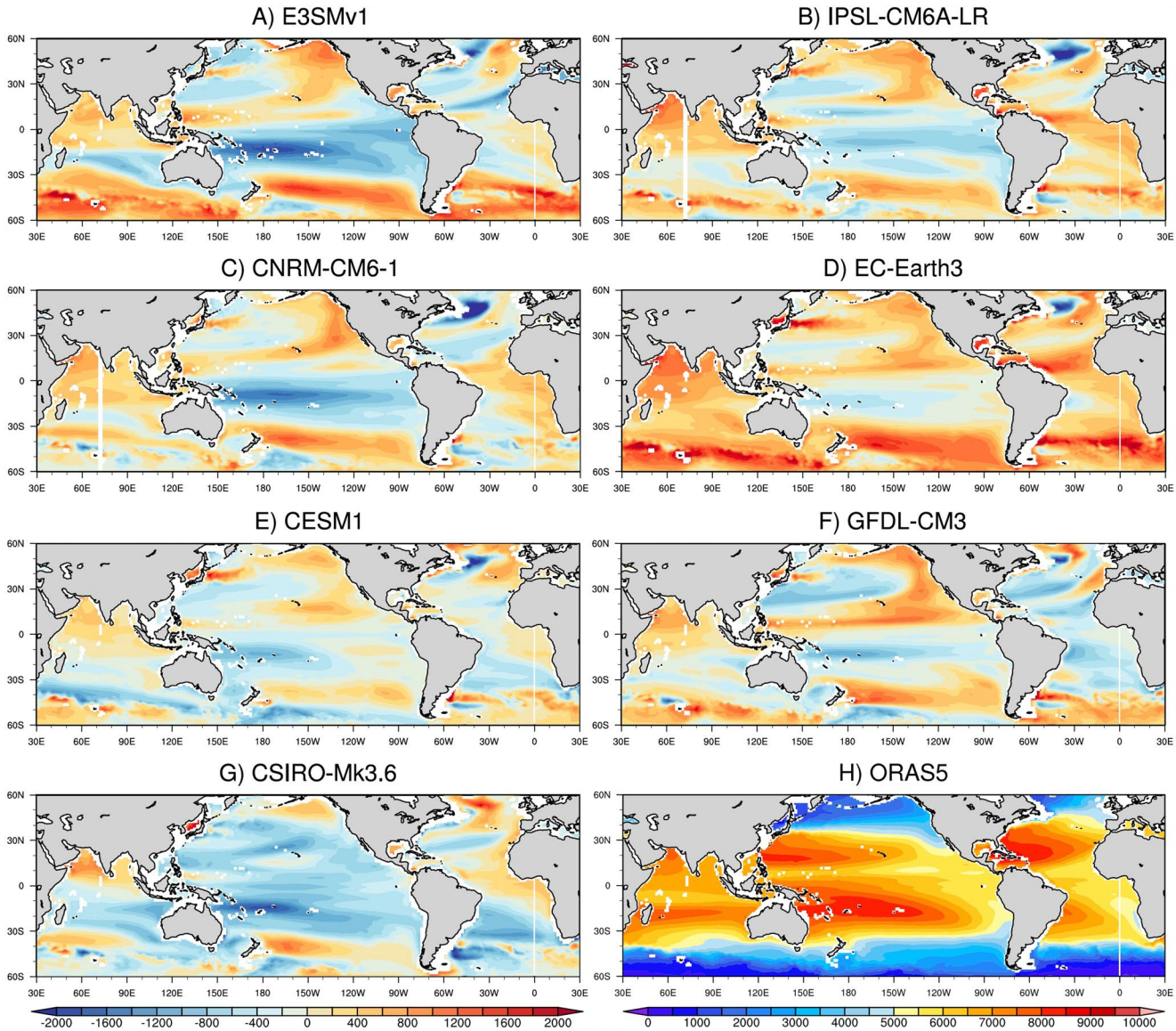

Figure 6. 0–400 m ocean heat content (OHC) difference (panels A–G) from the ORAS5 (panel H) reanalysis (Zuo et al., 2019) for each of the large ensembles (averaged over all members) (mean value during the 1958–2005 period). Unit for OHC is °C * m. All data is regridded to a resolution of 1°.

spread considered relative either to the "full" PI control standard deviation including the seasonal cycle, or one relative to the deseasonalized (January-only) standard deviation. The resulting metric can be expressed as:

$$\sigma(t)_{\mathrm{OHC-norm}} = \frac{\sigma(t)_{\mathrm{OHC}}}{\sigma_{(t)\mathrm{OHC-PI}}} \tag{11}$$

Using the above calculations, we examine σ_{OHC} during the first month of the simulations, as a measure of spread in the initial conditions (Figure 7). The ensembles show large initial OHC spreads, with one exception: the initial spread of the CESM1 ensemble is nearly zero, as is expected since this micro ensemble was constructed solely by varying atmospheric initial conditions. We also note that GFDL CM3 is not included since its various members began at different times (see above). Interestingly, the high-latitude North Atlantic appears as a region of strong intra-ensemble spread in several models, particularly the IPSL-CM6A-LR, CNRM-CM6-1, and EC-Earth 3. This likely relates to the magnitude of Atlantic Meridional Overturning Circulation (AMOC) variability in these ensembles, with a potential role for interaction with sea ice. The Kuroshio-Oyashio Extension region also appears as a region with large initial spread, although with smaller magnitude than in the North Atlantic, except for the E3SMv1 where spread appears large and comparable to EC-Earth3.

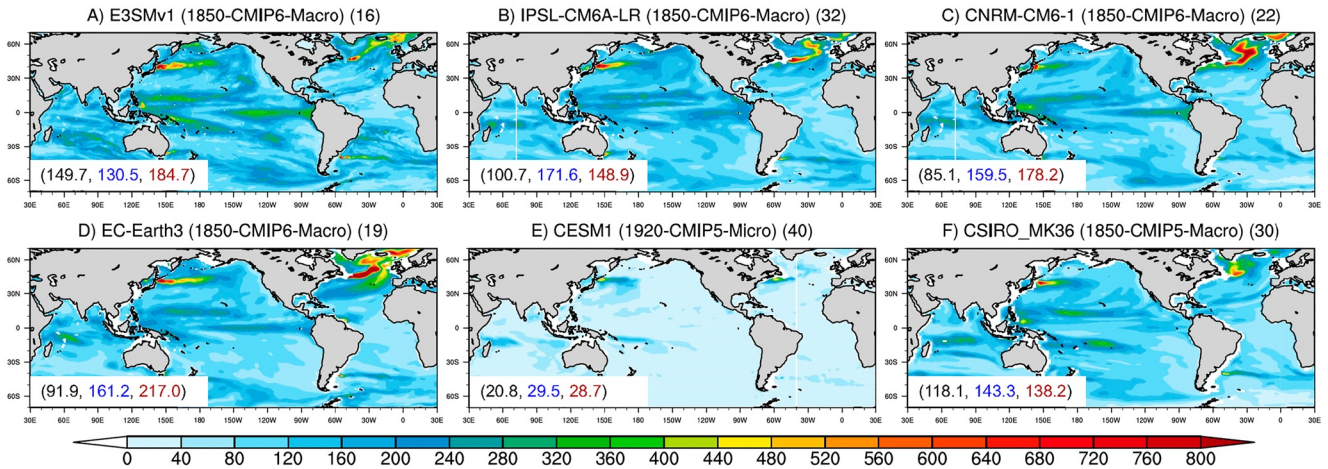

Figure 7. σ_{OHC} in initial month of simulation (unit: °C * m), using ocean heat content (OHC) averaged within the top 400 m for all ensembles in Table 1. The ensemble starting year, CMIP version, initialization method, and number of ensemble members are denoted in the figure labels. Numbers in each panel indicate the area-average of the heat content over different regions: black denotes the region [−60S, 60N] [0, 360]; blue is for the region [0, 60N] [120, 270]; brown for the [0, 60N] [280, 360].

To examine ensemble spread throughout the simulation period, we next plot $\sigma_{OHC-norm}$ as a function of time. This is first evaluated over the global oceans (Figure 8a), where the regions poleward of 60° have been excluded to prevent complications due to sea ice interactions. As noted above, CESM1 begins with the smallest initial spread; interestingly, throughout the simulation period this spread remains relatively low, suggesting that the initial low spread has temporal persistence beyond the first few years. This result is consistent with previous analyses of

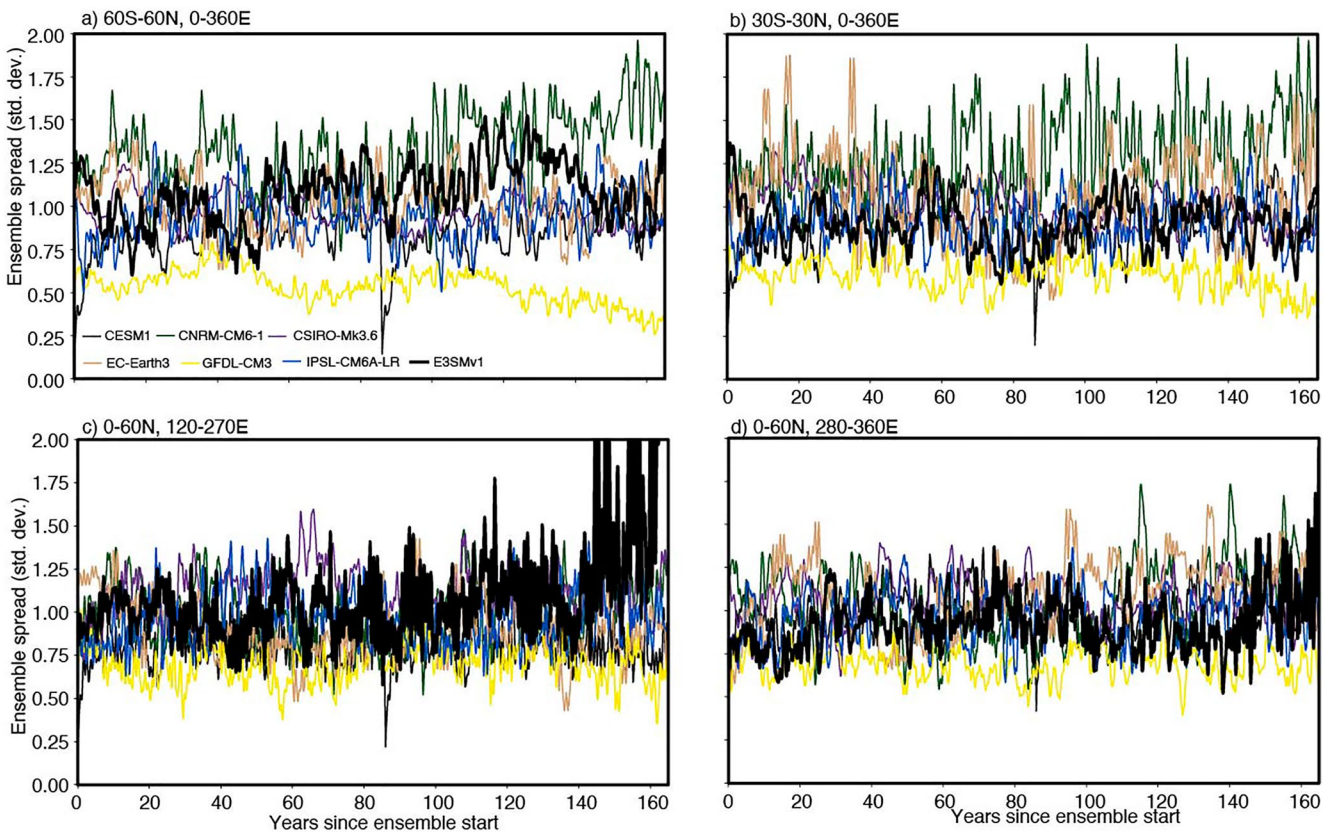

Figure 8. $\sigma_{OHC-norm}$ calculated using 0–400 m ocean heat content (OHC) ensemble spread divided by the temporal standard deviation of the model's pre-industrial control. (a) Global, excluding high latitudes; (b) Tropics; (c) North Pacific; (d) North Atlantic.

CESM ensembles suggesting a significant contribution of ocean initial state to ensemble spread (Kim et al., 2018). The GFDL-CM3 is consistently lower in overall ensemble spread even than CESM1 after the first few simulation years, a result which is true across averaging regions and consistent with the lowpass-filtered spread in surface temperature shown in Figure 5. The other (macro) ensembles show relatively stable degrees of spread, but at different overall levels. The E3SMv1-LE appears on the larger end of the spread estimates over 60S–60N, and especially in the latter half of the simulation period is the ensemble with the largest degree of spread.

The regional dependence of the OHC ensemble spread is shown in the remaining panels of Figure 8. In the tropical oceans (Figure 8b) the ensemble spread is generally comparable to or slightly larger overall than the global mean values for most models (see Figure 8a). The causes for this are as yet unclear, but may relate to coupled modes of tropical climate variability such as ENSO or the Pacific Decadal Oscillation. The tropics also appear to be the region in which micro initialization's effects manifest more strongly, as evidenced by the systematic lower ensemble spread present in CESM1 and GFDL-CM3 in Figures 8a and 8b. However, in general it is difficult to identify systematic patterns in ensemble spread as a function of ensemble initialization strategy, as the various macro ensembles' ordering shifts from region to region. That said, the fact that the E3SMv1-LE simulates the same magnitude of spread as other SMILEs despite its overall smaller size again points to the utility of an inter-basin initialization approach in efficiently generating spread.

The results of Figure 8 are suggestive of a possible role for external forcing in altering the temporal evolution of ensemble spread. All ensembles exhibit values of spread which are centered around 1, suggesting that ensemble spread is comparable to the degree of variability generated by coupled dynamics in the corresponding pre-industrial control simulation. However, in some cases, the magnitude of spread changes with time. This is particularly apparent in the North Pacific and North Atlantic in E3SMv1-LE (Figures 8c and 8d), where near year 140 (corresponding to simulated calendar year 1990) there is an abrupt increase in the average ensemble spread. This increase is so large that by the end of the simulation period, E3SMv1-LE is the model with the largest spread in both regions. The onset of this increase in spread coincides temporally with the start of rapid warming in E3SMv1 (Figures 2c and 2d), which has a large signature in the North Pacific and North Atlantic (Figures 2c and 2d), and suggests that greenhouse warming-driven changes in variability in these basins may be responsible. Interestingly, the post-1970s warming trend does not appear to be accompanied by a rise in the ensemble spread in surface temperature (Figure 4); external forcing, if responsible, appears to be more strongly expressed in ensemble spread as a change in the accumulation of OHC.

5.3. Pre-Industrial Control OHC Behavior

The use of the pre-industrial control simulations for calibration of ensemble OHC raises the question of how OHC behaves within those control simulations themselves—a question which has as yet received relatively little attention. The time series of OHC is therefore shown in Figure 9 for each of the PI control simulations. Several of these simulations appear to be out of equilibrium through the early portions of the run (CNRM-CM6-1, E3SMv1) or through the entire run (IPSL-CM6A-LR, GFDL-CM3). In some cases, the later portions of the control simulation do appear to reach a relatively stable equilibrium; this is the case for E3SMv1, CNRM-CM6-1, and CSIRO-Mk3.6.

Interestingly, centennial-scale oscillations are present in several of the PI control simulations, and it is unclear whether these are true features of the underlying model or are an artifact of numerical drift. Some contribution from coupled internal variability is likely at play, as mentioned above; centennial-scale oscillations in OHC as a result of AMO and/or AMOC variability have been identified in previous work as well (Meccia et al., 2022). Attempts at recalibration using either “stable” subsets of the PI control simulation or using highpass filtering to remove spurious trends reorder the ensemble spread (not pictured); we therefore conclude that a more rigorous examination of numerical drift is required to fully characterize the effects of ocean initial conditions on resulting spread across coupled ensembles. Additionally, given the magnitude of centennial variability in some models (e.g., EC-Earth3, GFDL-CM3), investigations of the possible contribution of such low-frequency variability to model behavior over the historical period are likely necessary.

6. Conclusions

This study presents the first large ensemble performed with the E3SMv1, using the low-resolution configuration (nominal 1 resolution). To date, 20 simulations have been completed covering the 1850–2015 period, with

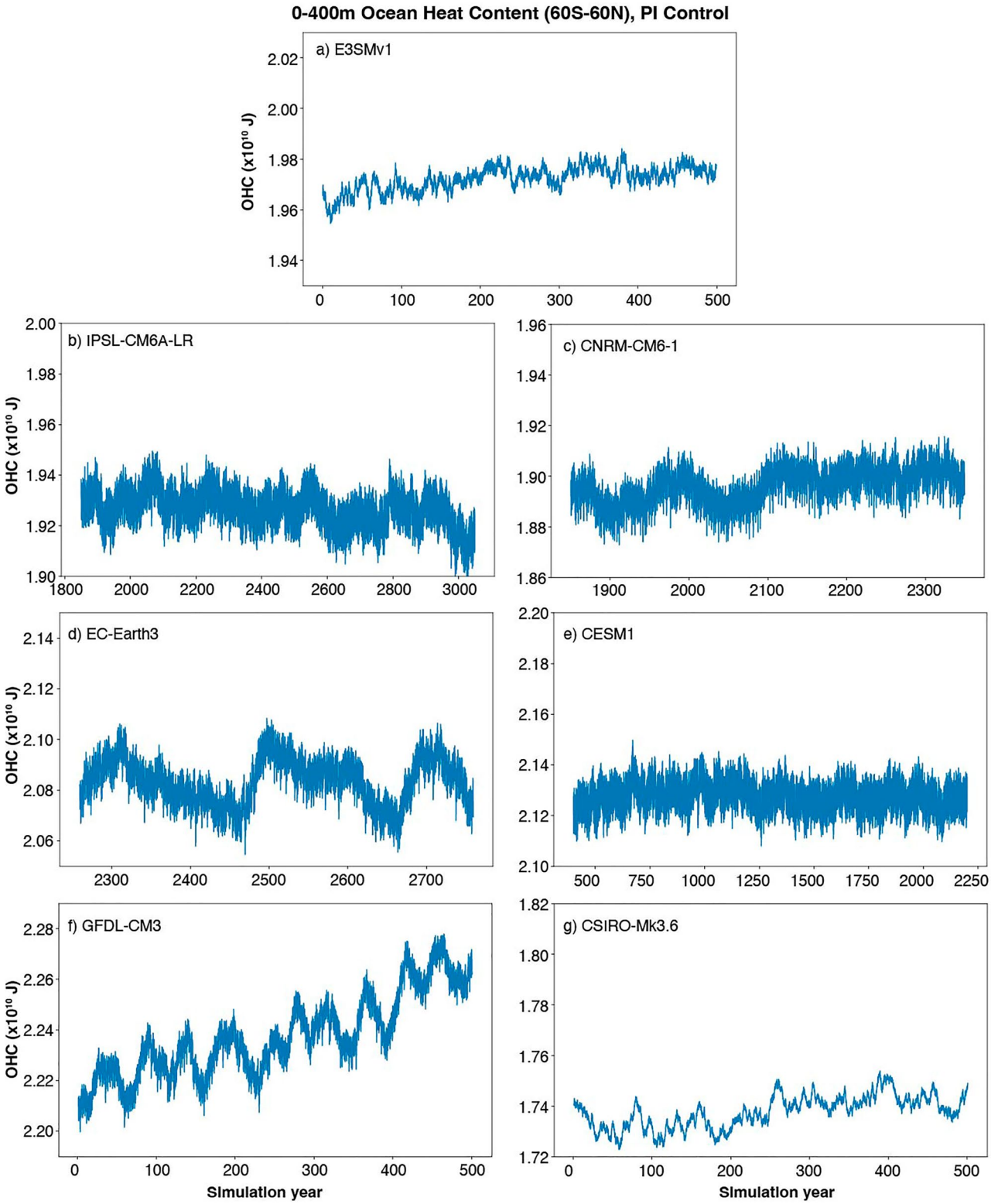

Figure 9. Time series of 0–400 m ocean heat content averaged over 60S–60N, 0–360E, for the pre-industrial control simulations used to initialize each ensemble. Units are 10^{10} J, and the (arbitrarily defined) time coordinate for each simulation is used on the *x*-axis. The difference between the max and min values on the *y*-axis are identical across simulations (0.1×10^{10} J) to facilitate visual intercomparison.

extensions to 2100 underway. The E3SMv1 Large Ensemble employs a macro strategy for ensemble initialization, designed to maximize the intra-ensemble spread; the dominant mode of variability in upper OHC is computed for the Atlantic and Pacific Oceans in the 1850 pre-industrial control simulation, and restart years spanning the Atlantic:Pacific phase space as regularly as possible selected for ensemble initial conditions. This strategy allows the initial states to attain a wider degree of spread for a relatively small ensemble size, compared to the commonly employed approach of randomly or regularly choosing start years from the relevant control simulation.

The overall performance of the E3SMv1-LE relative to observations is similar to the E3SMv1 CMIP6 contribution, although the two model configurations differ slightly in timestep and representation of ocean mixing. As for the E3SM CMIP6 simulations (Golaz et al., 2019), the trend in global surface temperature in E3SMv1 is weaker than observed, with the time series remaining nearly flat until the late 1970s, after which warming proceeds more rapidly than in observations. This is most likely related to the known strong sensitivity of E3SMv1 climate to anthropogenic aerosol emissions, and the spatial structure of temperature behavior confirms: strong cooling is seen in the mid-twentieth century in the North Pacific, consistent with the regional distribution of aerosols over that period.

Ensemble spread behavior in E3SMv1-LE is contrasted with other CMIP5 and CMIP6-era large ensembles, taken from the MMLEA. When surface temperature is considered, the majority of models exhibit a spread similar to their corresponding PI control runs, suggesting that they are all sampling the internal variability of the model relatively well. Over the first few simulation years, the micro-initialized CESM1 ensemble shows by far the smallest spread, but this effect generally disappears later in the simulation period. Additionally, CESM1 exhibits comparable spread to other macro-initialized ensembles when low-frequency surface temperature variability is considered. Overall, it is difficult to identify clear signatures of a specific macro initialization strategy, as the relative magnitudes of ensemble spread vary from region to region. However, the E3SMv1-LE generally shows spread comparable to (and regionally sometimes larger than) many other macro ensembles with considerably more ensemble members, suggesting that choosing "maximally independent" ocean initial states may lead to more efficient strategies for generating large spread, even in cases where computational resource limits dictate that the ensemble size must remain smaller than is otherwise desirable.

Ensemble spread in upper-ocean heat content shows behaviors broadly similar to surface air temperature across SMILE simulations. The difference between micro and macro initialization approaches is more obvious, with CESM1 and especially the hybrid micro-macro GFDL-CM3 ensemble showing lower spread in globally averaged OHC. However, regional differences in ensemble spread ordering are substantial, especially between macro ensembles, and obscure the potential influence of ensemble initialization strategy. The substantial differences in model physics are an additional complicating factor, which could not be mitigated given the relatively limited availability of SMILE data. There also appears to be a possible role for external forcing in altering ensemble spread: in E3SMv1-LE, ensemble spread in the North Atlantic and Pacific is amplified after the 1980s, which may be an effect of the rapid onset of warming in this model due to greenhouse gas increases and reductions in anthropogenic aerosol emissions. Influences of external forcing on internal modes of variability such as the AMO and AMOC, some of which have previously been identified in coupled model simulations (Menary & Scaife, 2014; Otterå et al., 2010), may also be playing a role.

This work may serve as a starting point for further investigation of the optimal strategies for construction of large ensembles that have a representative spread across their members. Through the selection of more independent initial ocean states, it may be possible to construct effective ensembles using relatively fewer members. However, the limited number of models used here and the lack of capacity to make "clean" comparisons between micro and macro initializations indicates that further work is needed to definitively identify best practices for ensemble creation. This work also demonstrates that conclusive evidence on the generation of ensemble spread will also requite a more thorough understanding of spurious model drift and multi-centennial variability, as both of these significantly alter the apparent relative degree of spread across ensemble members in SMILEs.

Data Availability Statement

Data from the MMLEA is publicly accessible via the NCAR Climate Data Gateway (2023; https://www.earthsystemgrid.org/dataset/ucar.cgd.ccsm4.CLIVAR_LE.html). CMIP5 and CMIP6 data, as well as the E3SMv1 Large Ensemble output, are available via the Earth System Grid (2023; http://esgf-node.llnl.gov).

Acknowledgments
We acknowledge the Multi-Model Large Ensemble Archive (MMLEA) at the National Center for Atmospheric Research for making multiple climate models' ensemble members readily accessible. We also acknowledge the CMIP5 and CMIP6 data portal provided by ESGF enterprise system (http://esgf-node.llnl.gov). XH, EDL, MN, and SS were supported by the U.S. Department of Energy, DE-SC0019418. SS was also supported by NSF CAREER, OCE-2142953. TX was funded by the NOAA Physical Sciences Laboratory through National Research Council (NRC) Research Associateship Programs. AC was supported by the NOAA Climate Program Office CVP and MAPP programs. We thank Lantao Sun for discussion of the GFDL CM3 ensemble configuration. LVR was supported as part of the Energy Exascale ESM (E3SM) project, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

References

Abraham, J. P., Baringer, M., Bindoff, N. L., Boyer, T., Cheng, L. J., Church, J. A., et al. (2013). A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Reviews of Geophysics*, *51*(3), 450–483. https://doi.org/10.1002/rog.20022

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., et al. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010. https://doi.org/10.1029/2019ms002010

Brainard, R. E., Oliver, T., McPhaden, M. J., Cohen, A., Venegas, R., Heenan, A., et al. (2018). Ecological impacts of the 2015/16 El Niño in the central equatorial Pacific. *Bulletin of the American Meteorological Society*, *99*(1), S21–S26. https://doi.org/10.1175/bams-d-17-0128.1

Capotondi, A., Newman, M., Xu, T., & Di Lorenzo, E. (2022). An optimal precursor of northeast Pacific marine heatwaves and central Pacific El Niño events. *Geophysical Research Letters*, *49*(5), e2021GL097350. https://doi.org/10.1029/2021gl097350

Deser, C. (2020). Certain uncertainty: The role of internal climate variability in projections of regional climate change and risk management. *Earth's Future*, *8*(12), e2020EF001854. https://doi.org/10.1029/2020EF001854

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., et al. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, *10*(4), 277–286. https://doi.org/10.1038/s41558-020-0731-2

Di Lorenzo, E., & Mantua, N. (2016). Multi-year persistence of the 2014/15 North Pacific marine heatwave. *Nature Climate Change*, *6*(11), 1042–1047. https://doi.org/10.1038/nclimate3082

Dong, B., Sutton, R. T., & Scaife, A. A. (2006). Multidecadal modulation of El Niño–Southern Oscillation (ENSO) variance by Atlantic Ocean sea surface temperatures. *Geophysical Research Letters*, *33*(8), L08705. https://doi.org/10.1029/2006gl025766

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., et al. (2021). The EC-Earth3 Earth system model for the climate model intercomparison project 6. *Geoscientific Model Development Discussions*, 1–90.

Earth System Grid Federation. (2023). [Dataset]. Retrieved from https://esgf-node.llnl.gov/

Fasullo, J. T., Otto-Bliesner, B. L., & Stevenson, S. (2018). ENSO's changing influence on temperature, precipitation, and wildfire in a warming climate. *Geophysical Research Letters*, *45*(17), 9216–9225. https://doi.org/10.1029/2018gl079022

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., et al. (2014). Evaluation of climate models. In *Climate change 2013: The physical science basis. Contribution of Working Group I to the fifth assessment report of the Intergovernmental Panel on Climate Change* (pp. 741–866). Cambridge University Press.

Golaz, J. C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089–2129. https://doi.org/10.1029/2018ms001603

Hawkins, E., Smith, R. S., Gregory, J. M., & Stainforth, D. A. (2016). Irreducible uncertainty in near-term climate projections. *Climate Dynamics*, *46*(11), 3807–3819. https://doi.org/10.1007/s00382-015-2806-8

Huang, X., & Stevenson, S. (2021). Connections between mean North Pacific circulation and western US precipitation extremes in a warming climate. *Earth's Future*, *9*(6), e2020EF001944. https://doi.org/10.1029/2020ef001944

Jeffrey, S., Rotstayn, L., Collier, M., Dravitzki, S., Hamalainen, C., Moeseneder, C., et al. (2013). Australia's CMIP5 submission using the CSIRO-Mk3. 6 model. *Australian Meteorological and Oceanographic Journal*, *63*(1), 1–14. https://doi.org/10.22499/2.6301.001

Joh, Y., & Di Lorenzo, E. (2017). Increasing coupling between NPGO and PDO leads to prolonged marine heatwaves in the northeast Pacific. *Geophysical Research Letters*, *44*(22), 11663–11671. https://doi.org/10.1002/2017gl075930

Joh, Y., Di Lorenzo, E., Siqueira, L., & Kirtman, B. (2021). Enhanced interactions of Kuroshio Extension with tropical Pacific in a changing climate. *Nature Scientific Reports*, *11*(6247), 6247. https://doi.org/10.1038/s41598-021-85582-y

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/bams-d-13-00255.1

Kim, W. M., Yeager, S., Chang, P., & Danabasoglu, G. (2018). Low-frequency North Atlantic climate variability in the Community Earth System Model large ensemble. *Journal of Climate*, *31*(2), 787–813. https://doi.org/10.1175/jcli-d-17-0193.1

Liguori, G., & Di Lorenzo, E. (2018). Meridional modes and increasing Pacific Decadal variability under anthropogenic forcing. *Geophysical Research Letters*, *45*(2), 983–991. https://doi.org/10.1002/2017gl076548

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., et al. (2019). The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2050–2069. https://doi.org/10.1029/2019ms001639

Mankin, J. S., Lehner, F., Coats, S., & McKinnon, K. A. (2020). The value of initial condition large ensembles to robust adaptation decision-making. *Earth's Future*, *8*(10), e2012EF001610. https://doi.org/10.1029/2020ef001610

McGregor, S., Timmermann, A., Stuecker, M. F., England, M. H., Merrifield, M., Jin, F. F., & Chikamoto, Y. (2014). Recent Walker circulation strengthening and Pacific cooling amplified by Atlantic warming. *Nature Climate Change*, *4*(10), 888–892. https://doi.org/10.1038/nclimate2330

Meccia, V. L., Fuentes-Franco, R., Davini, P., Bellomo, K., Fabiano, F., Yang, S., & von Hardenberg, J. (2022). Internal multi-centennial variability of the Atlantic Meridional Overturning Circulation simulated by EC-Earth3. *Climate Dynamics*, *60*(11–12), 1–18. https://doi.org/10.1007/s00382-022-06534-4

Menary, M. B., & Scaife, A. A. (2014). Naturally forced multidecadal variability of the Atlantic meridional overturning circulation. *Climate Dynamics*, *42*(5–6), 1347–1362. https://doi.org/10.1007/s00382-013-2028-x

Midhun, M., Stevenson, S., & Cole, J. E. (2021). Oxygen isotopic signatures of major climate modes and implications for detectability in speleothems. *Geophysical Research Letters*, *48*(1), e2020GL089515. https://doi.org/10.1029/2020gl089515

Migliozzi, B., Reinhard, S., Popovich, N., Wallace, T., & McCannSept, A. (September 24, 2020). Record wildfires on the west coast are capping a disastrous decade. *New York Times*. Retrieved from https://www.nytimes.com/interactive/2020/09/24/climate/fires-worst-year-california-oregon-washington.html

National Center for Atmospheric Research Climate Data Gateway. (2023). [Dataset]. Retrieved from https://www.earthsystemgrid.org/dataset/ucar.cgd.ccsm4.CLIVAR_LE.html

Otterå, O. H., Bentsen, M., Drange, H., & Suo, L. (2010). External forcing as a metronome for Atlantic multidecadal variability. *Nature Geoscience*, *3*(10), 688–694. https://doi.org/10.1038/ngeo955

Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., et al. (2016). Climate variability and change since 850 CE: An ensemble approach with the Community Earth System Model. *Bulletin of the American Meteorological Society*, *97*(5), 735–754. https://doi.org/10.1175/bams-d-14-00233.1

Rasch, P. J., Xie, S., Ma, P. L., Lin, W., Wang, H., Tang, Q., et al. (2019). An overview of the atmospheric component of the Energy Exascale Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2377–2411. https://doi.org/10.1029/2019ms001629

Rodgers, K. B., Lin, J., & Frölicher, T. L. (2015). Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model. *Biogeosciences*, *12*(11), 3301–3320. https://doi.org/10.5194/bg-12-3301-2015

Sellar, A. A., Walton, J., Jones, C. G., Wood, R., Abraham, N. L., Andrejczuk, M., et al. (2020). Implementation of UK Earth system models for CMIP6. *Journal of Advances in Modeling Earth Systems*, *12*(4), e2019MS001946. https://doi.org/10.1029/2019ms001946

Smith, D. M., Booth, B. B., Dunstone, N. J., Eade, R., Hermanson, L., Jones, G. S., et al. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, *6*(10), 936–940. https://doi.org/10.1038/nclimate3058

Stevenson, S., Wittenberg, A. T., Fasullo, J., Coats, S., & Otto-Bliesner, B. (2021). Understanding diverse model projections of future extreme El Niño. *Journal of Climate*, *34*(2), 449–464. https://doi.org/10.1175/jcli-d-19-0969.1

Sun, L., Alexander, M., & Deser, C. (2018). Evolution of the global coupled climate response to Arctic sea ice loss during 1990–2090 and its contribution to climate change. *Journal of Climate*, *31*(19), 7823–7843. https://doi.org/10.1175/jcli-d-18-0134.1

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al. (2019). Evaluation of CMIP6 deck experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2177–2213. https://doi.org/10.1029/2019ms001683

Wang, H., Easter, R. C., Zhang, R., Ma, P. L., Singh, B., Zhang, K., et al. (2020). Aerosols in the E3SM Version 1: New developments and their impacts on radiative forcing. *Journal of Advances in Modeling Earth Systems*, *12*(1), e2019MS001851. https://doi.org/10.1029/2019ms001851

Williams, A. P., Cook, E. R., Smerdon, J. E., Cook, B. I., Abatzoglou, J. T., Bolles, K., et al. (2020). Large contribution from anthropogenic warming to an emerging North American megadrought. *Science*, *368*(6488), 314–318. https://doi.org/10.1126/science.aaz9600

Xu, T., Newman, M., Capotondi, A., Stevenson, S., Di Lorenzo, E., & Alexander, M. A. (2022). An increase in marine heatwaves without significant changes in surface ocean temperature variability. *Nature Communications*, *13*(1), 1–12. https://doi.org/10.1038/s41467-022-34934-x

Zanchettin, D., Bothe, O., Graf, H. F., Omrani, N. E., Rubino, A., & Jungclaus, J. H. (2016). A decadally delayed response of the tropical Pacific to Atlantic multidecadal variability. *Geophysical Research Letters*, *43*(2), 784–792. https://doi.org/10.1002/2015gl067284

Zhang, K., Zhang, W., Wan, H., Rasch, P. J., Ghan, S. J., Easter, R. C., et al. (2022). Effective radiative forcing of anthropogenic aerosols in E3SM version 1: Historical changes, causality, decomposition, and parameterization sensitivities. *Atmospheric Chemistry and Physics*, *22*(13), 9129–9160. https://doi.org/10.5194/acp-22-9129-2022

Zhang, R., & Delworth, T. L. (2007). Impact of the Atlantic multidecadal oscillation on North Pacific climate variability. *Geophysical Research Letters*, *34*(23), L23708. https://doi.org/10.1029/2007gl031601

Zhang, Y., Wallace, J. M., & Battisti, D. S. (1997). ENSO-like interdecadal variability: 1900-93. *Journal of Climate*, *10*(5), 1004–1020. https://doi.org/10.1175/1520-0442(1997)010<1004:eliv>2.0.co;2

Zhao, Y., Di Lorenzo, E., Sun, D., & Stevenson, S. (2021). Tropical Pacific decadal variability and ENSO precursor in CMIP5 models. *Journal of Climate*, *34*(3), 1023–1045. https://doi.org/10.1175/jcli-d-20-0158.1

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Science*, *15*(3), 779–808. https://doi.org/10.5194/os-15-779-2019