Climate Dynamics
https://doi.org/10.1007/s00382-020-05621-8

Understanding the signal-to-noise paradox in decadal climate predictability from CMIP5 and an eddying global coupled model

Wei Zhang[1] · **Ben Kirtman**[1] · **Leo Siqueira**[1] · **Amy Clement**[1] · **Junfei Xia**[1]

Received: 7 June 2020 / Accepted: 28 December 2020

Abstract
Recent research suggests the widespread existence of the signal-to-noise paradox in seasonal-to-decadal climate predictions. The essence of the paradox is that the signal-to-noise ratio in models can be unrealistically small and models may make better predictions of the observations than they predict themselves. The paradox highlights a potentially serious issue with model predictions as previous studies may underestimate the limit of predictability. The focus of this paper is two-fold: the first objective is to re-examine decadal predictability from the lens of the signal-to-noise paradox in the context of CMIP5 models. We demonstrate that decadal predictability is generally underestimated in CMIP5 models possibly due to the existence of the signal-to-noise paradox. Models underestimate decadal predictability in regions where it is likely for the paradox to exist, especially over the Tropical Atlantic Ocean and Tropical Indian Ocean and eddy-rich regions, including the Gulf Stream, Kuroshio Current, and Southern Ocean. The second objective follows from the results of the first, attempting to determine if this underestimate of decadal predictability is, at least partially, due to missing ocean mesoscale processes and features in CMIP5 models. A suite of coupled model experiments is performed with eddying and eddy-parameterized ocean component. Compared with eddy-parameterized models, the paradox is less likely to exist in eddying models, particularly over eddy-rich regions. These also happen to be regions where increased decadal predictability is identified. We hypothesize that this enhanced predictability is due to the enhanced vertical connectivity in the ocean. The presence of mesoscale ocean features and associated vertical connectivity significantly influence decadal variability, predictability, and the signal-to-noise paradox.

Keywords Signal-to-noise paradox · Decadal predictability · CMIP5 · CCSM4 · Eddying model · Vertical connectivity

1 Introduction

There is a continuously growing demand for decadal climate predictions. Making skillful decadal predictions has potential benefits in terms of supporting decision-making processes in agriculture, energy and water management among other sectors (e.g., Kirtman et al. 2013; Kushnir et al. 2019; Merryfield et al. 2020). While seasonal climate prediction has matured into regular operational forecasts (e.g., Kirtman et al. 2014), forecasting the climate over decades has proven more challenging (Keenlyside et al. 2008; Meehl et al. 2014; Zhang and Kirtman 2019a).

One of the significant challenges in decadal prediction and often overlooked in previous studies is the so-called "signal-to-noise paradox" (e.g., Scaife et al. 2014; Siegert et al. 2016; Smith et al. 2019; Zhang and Kirtman 2019b). The essence of the paradox is that the signal-to-noise ratio estimated in climate models can be too small. Specifically, models seem to be better at predicting observations than predicting themselves as the model ensemble mean forecasts are better correlated with observations than with individual ensemble members. Scaife et al. (2014) first discussed the signal-to-noise paradox in seasonal prediction of the winter North Atlantic Oscillation (NAO) index, and subsequently, a growing list of examples in different atmospheric and climate models has emerged (Scaife and Smith 2018). For example, Zhang and Kirtman (2019b, hereafter ZK19) developed a simple Markov model framework and provided a comprehensive assessment of the NAO index indicating the widespread existence of the signal-to-noise paradox in coupled models from the fifth Coupled Model Intercomparison Project (CMIP5). The Markov model framework can easily reproduce the signal-to-noise paradox, which is

✉ Wei Zhang
wei.zhang@rsmas.miami.edu

[1] Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, USA

dependent on the magnitude of the persistence and noise variance. Smith et al. (2019) used multi-model decadal hindcasts from seven state-of-the-art coupled climate models with a total of 71 ensemble members suggesting the existence of the signal-to-noise paradox in decadal predictions. One of the key points highlighted in the Smith et al. (2019; 2020) papers was that model-based estimates of decadal predictability might actually be an underestimate, as previous studies could have mispresented the noise, or underestimated the magnitude of the predictable signal due to limited ensemble size.

The specific examples of the paradox and the associated model errors noted above suggest that model based estimates of climate predictability may seriously underestimate the limit of predictability. As we begin to understand the mechanisms for the paradox, predictability estimates also need to be revisited. Where and to what extent is the paradox leading to substantial underestimates of the limit of predictability? The first goal of the paper is to re-examine decadal predictability from the lens of the signal-to-noise paradox in the context of the CMIP5 models. We also attempt to understand if and how model initialization and external forcing contribute to the signal-to-noise paradox. Different from Smith et al. (2019), who investigated the signal-to-noise paradox from initialized decadal hindcasts from seven CMIP5 models, our study examines the paradox from a diagnostic perspective using thirty CMIP5 uninitialized historical and preindustrial control simulations. Instead of estimating decadal prediction skill, we examine decadal potential predictability (a diagnostic method to estimate predictability; Boer 2004) for both observations and CMIP5 models. Based on the Markov model framework in ZK19, we determine the widespread existence of the signal-to-noise paradox that applies to CMIP5 uninitialized simulations (focusing on decadal timescales), suggesting that the paradox is not due to problems with model initialization processes. Comparisons of the paradox between historical and preindustrial control simulations imply the potential impact of external forcing on the signal-to-noise paradox and decadal predictability.

The second goal of the paper follows from the results of the first. Essentially, the results from the first goal show that the CMIP5 models seriously underestimate the limit of decadal predictability, and models seem to underestimate decadal predictability in regions where it is likely for the paradox to exist. Specifically, we find that models underestimated decadal SST predictability, particularly in the Gulf Stream, Kuroshio Current, Southern Ocean and other eddy-rich regions, where the signal-to-noise paradox occurs. The underestimation of decadal predictability over eddy-rich regions in models suggests the importance of ocean model resolution or mesoscale ocean features in decadal SST predictability and the signal-to-noise paradox. Therefore, we hypothesize in this second goal that this underestimate of decadal SST predictability is, at least in part, due to missing ocean mesoscale processes and features in the CMIP5 models. Again, the results are presented in the context of the signal-to-noise paradox.

With the above in mind, there have been several studies examining the mechanisms for the paradox. For example, the signal-to-noise paradox has been attributed to a lack of persistence (Strommen and Palmer 2019; Zhang and Kirtman 2019b), weak extratropical air-sea coupling (Scaife and Smith 2018), stratospheric initialization (O'Reilly et al. 2019), and underestimated eddy feedbacks due to low atmospheric model resolution (Scaife et al. 2019). Little to no research, however, has asked how mesoscale ocean features affect the signal-to-noise paradox and associated estimate of decadal predictability. The role of ocean mesoscale processes is of particular interest since several previous studies have suggested that decadal SST variability in coupled models is improved when ocean mesoscale features and processes are correctly represented (e.g., He et al. 2018; Infanti and Kirtman 2019; Kim et al. 2018; Samanta et al. 2018; Kirtman et al. 2012, 2017; Siqueira and Kirtman 2016; Zhang and Kirtman 2019a; among others). As shown by Kirtman et al. (2017), for example, mesoscale ocean features can substantially influence large-scale climate variability, air-sea interactions, and predictability. Particularly in the North Atlantic region, a more realistic mean-state climate and improved representation of ocean–atmosphere coupling and decadal SST variability around the Gulf Stream region have been detected with eddying Global Coupled Models (GCMs; Siqueira and Kirtman 2016). Given the importance of eddies on low-frequency variability and ocean–atmosphere coupling, the lack of ocean eddy resolution in current coupled models (e.g., eddy-parameterized models used in CMIP5) can potentially affect the estimates of decadal climate predictability.

In addition to the overall representation of decadal variability, the second goal of this study is motivated by the hypothesis that low-resolution eddy-parameterized GCMs may misrepresent or even lack the vertical communication in the subsurface to the deeper ocean, contributing to a lack of persistence in models and thus the signal-to-noise paradox. The underestimated vertical communication between the deep ocean and surface processes in CMIP5 models compared to observations has been recently explored by Kravtsov (2020). Kravtsov (2020) introduced an updated linear energy-balance model considering the heat exchange between ocean mixed layer and thermocline in the Atlantic and Pacific oceans. By fitting the observed and CMIP5 model-simulated SST with the energy-balance model, Kravtsov (2020) identified stronger vertical communication between the deep ocean and surface processes in observations than CMIP5 models, contributing to a larger fraction of predictable variability at decadal timescales. This

significant difference in decadal potential predictability between observations and CMIP5 models, as suggested by Kravtsov (2020), may lead to the signal-to-noise paradox.

In this study, we first examine the decadal potential predictability in observations and CMIP5 models from a diagnostic perspective, i.e., the first goal. Again, through the lens of the signal-to-noise paradox we use the Markov model framework developed in ZK19 to diagnose predictability. In terms of the second goal, distinct from Kravtsov (2020) who estimated the coupling parameters between thermocline and mixed layer in the energy-balance model, we perform a suite of model experiments with and without eddying ocean component, again through the lens of the paradox. We argue that high-resolution models with better represented ocean mesoscale features have stronger vertical connectivity in the subsurface to the deeper ocean than low-resolution models, which may potentially, or at least partially eliminate the signal-to-noise issue and thus improve predictability over decadal timescales.

2 Data and method

2.1 Observations, reanalysis datasets, and CMIP5 models

Three observational monthly sea surface temperature (SST) datasets are used in this study; namely, the National Oceanic and Atmospheric Administration (NOAA) Extended Reconstructed SST version-5 (ERSST; Huang et al. 2017) on 2°×2° grids for 1854-present, the Hadley Center Global Sea Ice and SST data set (HadISST; Rayner et al. 2003) from 1870 to 2017 with a spatial resolution of 1° ×1°, and the Centennial Observation-Based Estimates of SST version-2 (COBE; Hirahara et al. 2014) from 1850 to 2017 on the same grid as HadISST data. Monthly mean sea level pressure (SLP) data are obtained from three resources, including two 20th century reanalysis datasets from the NOAA (20CR; 2°×2°; 1871–2012; Compo et al. 2011) and the European Centre for Medium-Range Weather Forecasts (ERA20C; 1× 1°; 1900–2010; Poli et al. 2016), as well as the Hadley Centre's Mean SLP data (HadSLP; 5×5°; 1850–2004; Allan and Ansell 2006). Here we simply consider reanalysis datasets as supplement to observations.

Both the historical (HIST, the first ensemble member) and the preindustrial control (PI) simulations of thirty CMIP5 models are used in this study to compare with observations (Table 1). We only use the first realization (r1i1p1) of each CMIP5 model to equally weight each model in the multi-model mean estimates. The HIST simulations are simulations of recent past climate (1850–2005) forced by changing conditions, while the PI simulations are preindustrial coupled ocean–atmosphere control simulations with non-evolving preindustrial conditions (Taylor et al. 2012). Variability in the PI simulations is generated only through interactions internal to the coupled system, while variability in the HIST simulations is also due to natural and anthropogenic forcing (Murphy et al. 2017).

2.2 CCSM4 model experiments

A suite of model experiments is performed with the National Center for Atmospheric Research Community Climate System Model Version 4 (CCSM4; see overview in Gent et al. 2011). CCSM4 is a state-of-the-art global coupled model composed of four component models, namely, the Community Atmosphere Model version 4 (CAM4), the Parallel Ocean Program version 2 (POP2), the Community Land Model version 4 (CLM4), the Community Ice Code version 4 (CICE4), and the coupling infrastructure version 7 (CPL7) (Gent et al., 2011). In the low-resolution eddy-parameterized experiment (hereafter referred to as LR), we use 1° CAM4 and CLM4 coupled to the POP2 and CICE4 with the nominal 1° horizonal resolution. The zonal resolution of the POP2 and CICE4 is 1.125°globally while the meridional resolution increases from 0.27° around the equator to 0.54° at 33°N/S and is constant at high latitudes. We conduct the LR CCSM4 experiment with present-day forcing (e.g., greenhouse gas concentrations for 1990) for 500 years. The LR simulation was initialized as a "cold-start", that is an ocean at rest and was spin-up for 200 years. Different from the LR experiment, we employ a 0.5° CAM4 and CLM4 coupled to the nominal 0.1° horizonal resolution of the POP2 and CICE4 in the high-resolution eddying experiment (hereafter referred as HR). We first perform a 155-year standard control simulation with the same greenhouse gas concentrations as in the LR experiment, and the first 100 years of the simulation are taken as spin-up and are discarded in the analysis. Restarting from the first experiment with small perturbations, we run two other experiments for 70 years, each with the first 20 years taken as spin-up periods (see details of HR initialization and configuration in Kirtman et al. 2012). In total, we analyze here 155 years of HR simulations and 300 years of LR simulations.

2.3 Markov model framework

The Markov model framework is extensively described in ZK19 and has proven useful to determine the existence of the signal-to-noise paradox. The design of the Markov model framework starts from a linear signal-plus-noise model assuming that the future state forecasts depend linearly on the current state predictor and a stochastic noise term. The observations and models can be simulated with a statistical Markov chain model:

Table 1 CMIP5 models used in this study

Models	Atmospheric grid (Lat×Lon)	Ocean grid (Lat×Lon)	PI length (years)	Data/modeling institute
ACCESS1-0	1.3°×1.9°	1°×1°	500	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology (Australia)
ACCESS1-3				
BCC-CSM1-1	2.8°×2.8°	1°×1°	500	China Meteorological Administration
BCC-CSM1-1-m	1.1°×1.1°		400	
BNU-ESM	2.8°×2.8°	1°×1°	559	Beijing Normal University
CanESM2	2.8°×2.8°	0.9°×1.4°	996	Canadian Centre for Climate Modeling and Analysis
CCSM4	0.9°×1.3°	0.6°×1.1°	1051	National Center for Atmospheric Research
CESM1-BGC			500	
CESM1-CAM5			319	
CESM1-FASTCHEM			222	
CMCC-CM	0.8°×0.8°	1.9°×2.0°	330	Centro Euro-Mediterraneo per I Cambiamenti Climatici
CMCC-CMS	3.7°×3.7°		500	
CNRM-CM5	1.4°×1.4°	0.8°×1.0°	850	Centre National de Recherches Meteorologiques and Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique
CSIRO-Mk3-6-0	1.9°×1.9°	0.9°×1.9°	500	Australian Commonwealth Scientific and Industrial Research Organization and Queensland Climate Change Centre of Excellence
GFDL-CM3	2.0°×2.5°	1.0°×1.0°	500	Geophysical Fluid Dynamics Laboratory
GFDL-ESM2G				
GFDL-ESM2M				
GISS-E2-H-CC	2.0°×2.5°	1.0°×1.0°	251	Goddard Institute for Space Studies
GISS-E2-R-CC				
HadGEM2-CC	1.2°×1.9°	1.0°×1.0°	240	Met Office Hadley Centre
HadGEM2-ES			576	
inmcm4	1.5°×2.0°	0.4°×0.8°	500	Institute for Numerical Mathematics
IPSL-CM5A-LR	1.9°×3.8°	1.9°×2.0°	1000	Institut Pierre-Simon Laplace
IPSL-CM5B-LR				
IPSL-CM5A-MR	1.3°×2.5°	1.4°×1.6°	300	
MPI-ESM-LR	1.9°×1.9°	1.5°×1.5°	1000	Max Planck Institute for Meteorology
MPI-ESM-P				
MPI-ESM-MR		0.4°×0.4°	1156	
NorESM1-ME	1.9°×2.5°	0.6°×1.1°	252	Norwegian Climate Centre
NorESM1-M			501	

PI accounts for the preindustrial control simulations. We use historical simulations of CMIP5 models from 1870 to 2005. The CMIP5 model outputs and associated descriptions can be found in the CMIP5 archive (http://cmip-pcmdi.llnl.gov/cmip5). Atmospheric and ocean grids indicate horizonal resolution for each model. Ocean grids show approximate values as resolution may vary with different latitudes and longitudes. Note that in this study, we only use the last 200 years of PI for each model to estimate decadal predictability

$$O^{n+1} = \alpha O^n + N \tag{1}$$

$$M_i^{n+1} = \beta O^n + P_i \tag{2}$$

$$\bar{M}^{n+1} = \beta O^n + \bar{P} \tag{3}$$

where $\{O\}$ is the observation, and $\{M\}$ is the model forecasts initialized with observations and perturbed with different noise realizations $\{P_i\}$ ($i = 1, 2, 3, \ldots$). α and β are estimated as the lag-1 autocorrelation coefficients implying the persistence of the system. N and P are simply modeled as white-noise processes (see also Kirtman et al. 2005). $\{\overline{M}\}$ and $\{\overline{P}\}$ are model ensemble mean forecasts and ensemble mean noise. Following the procedure in ZK19, we can analytically derive the correlation between the model ensemble mean forecasts and observations ($corr(\overline{M}, O)$) and the correlation between the model ensemble mean forecasts and individual ensemble members ($corr(\overline{M}, M_i)$), and thus the ratio of squared correlation (RSC):

$$RSC = \frac{corr^2\left(\bar{M}, O\right)}{corr^2\left(\bar{M}, M_i\right)} = \frac{\alpha^2\beta^2\sigma_N^2 + \alpha^2\left(1-\alpha^2\right)\sigma_P^2}{\beta^2\sigma_N^2} \quad (4)$$

where σ_N^2 and σ_P^2 are noise variance for the observation and model ensemble members, which can be estimated as the total variance in observations and model forecasts multiplied by a factor of $1-\alpha^2$ and $1-\beta^2$, respectively. Based on Scaife and Smith (2018), there can be a signal-to-noise paradox when the *RSC* is greater than 1.0. So Eq. (4) becomes,

$$\frac{\alpha^2}{\sigma_N^2} > \frac{\beta^2}{\sigma_P^2} \quad (5)$$

Distinct from ZK19, who applied the Markov model framework to the monthly NAO index, this study further examines the signal-to-noise paradox, especially in SST fields based on CMIP5 HIST and PI simulations with a focus on decadal timescales.

3 Results and discussion

3.1 Underestimated decadal climate predictability in CMIP5 models

Decadal climate predictability is quantified using the potential predictability variance ratio (Boer 2004; Zhang et al. 2017), which indicates the relative intensity of decadal variability and provides an efficient approach to estimate decadal predictability from a diagnostic perspective. As suggested by Boer (2004), the total climate variability (σ_{Tot}^2) can be decomposed into a low-frequency component (σ_{Lf}^2) that may be potentially predictable and a high-frequency unpredictable noise component (σ_{Err}^2). This approach assumes that the slower potentially predictable component is independent from the noise term ($\sigma_{Lf}^2 = \sigma_{Tot}^2 - \sigma_{Err}^2$). Therefore, decadal potential predictability can be defined as the ratio of decadal-scale variability with respect to the total variability ($\sigma_{Lf}^2/\sigma_{Tot}^2 \times 100\%$). In this study, we extract decadal-scale climate variability by applying a 5-year low-pass Butterworth filter, after subtracting the annual cycle and linearly detrending the monthly time series from each observation and model simulation. The observational mean (based on ERSSTv5, COBE-SST2, and HadISST) and multi-model mean estimates (based on thirty CMIP5 HIST simulations) of decadal SST predictability are shown in Fig. 1a, b. Note that SST in model simulations is obtained by masking all the non-ocean regions of surface temperature, and for each observation and model simulation, SST is bilinearly interpolated to multiple $2^\circ \times 2^\circ$ grids before analysis. In observations, considerable ocean regions display high values of decadal SST predictability such as the North Atlantic, Western Pacific, Tropical Indian Ocean, and Southern Ocean, which is generally consistent with several earlier studies (e.g., Ding et al. 2016; Zhang and Kirtman 2019a). Meanwhile, low values of decadal SST predictability are detected over the Eastern Tropical Pacific, with values ranging from 25 to 60% gradually increasing westward, implying that decadal climate in this region may be still potentially predictable to some extent, though the Eastern Tropical Pacific is dominated by the El Niño-Southern Oscillation (ENSO) at interannual timescales. Uncertainty remains in the long-term predictability over the Eastern Tropical Pacific region (e.g., Gonzalez and Goddard 2016; Kirtman and Schopf 1998; Kravtsov 2012; Newman 2007), and as suggested by Wittenberg et al. (2014), for example, the ENSO potential predictability is lost after the 3–4 year range in the absence of external forcing. Compared with observational estimates, we find much lower values of decadal SST predictability for the CMIP5 multi-model mean estimates, except for some regions such as the Northeastern Pacific and subpolar North Atlantic (Fig. 1b). High decadal SST predictability in the subpolar North Atlantic is consistent with the results based on observations, where decadal SST variability is significantly impacted by internal atmospheric noise and subsurface ocean dynamics (e.g., Boer 2004; Robson et al. 2018; Wouters et al. 2013; Yan et al. 2018; Zhang and Kirtman 2019a; among others). Figure 1c shows the observational and CMIP5 multi-model estimates of decadal SST predictability averaged over six different ocean regions, namely, the North Atlantic (NA; 20°–65° N, 80°–0° W), North Pacific (NP; 20°–60° N, 120° E–120° W), Eastern Tropical Pacific (ETP; 10° S–10° N, 180°–85° W), Tropical Atlantic (TA; 15° S–15° N, 80°–0° W), Tropical Indian (TI; 15° S–25° N, 40°–100° E), and Southern Ocean (SO; 65°–40° S, 0°–360° E). The observational estimates show higher decadal SST predictability than most of the CMIP5 models except in the North Pacific; that is, decadal SST predictability is generally underestimated in CMIP5 models. The spatial distribution of decadal SST predictability based on CMIP5 models in the North Pacific is substantially different from those based on observational estimates, with larger values concentrated in the subpolar gyre, despite that the mean estimates in models and observations are somewhat comparable. Notably, there is significant inconsistency among different observational estimates in the Tropical Indian and the Southern Ocean; also, thirty CMIP5 models used here show better agreement with each other in the North Atlantic and North Pacific (based on larger percentage values) than other regions.

The impact of linear detrending is assessed in Fig. 2, which shows decadal SST predictability without detrending and with the regression-detrending method following Ting et al. (2009) by regressing out the global mean SST warming trend. Similar spatial patterns of decadal SST predictability

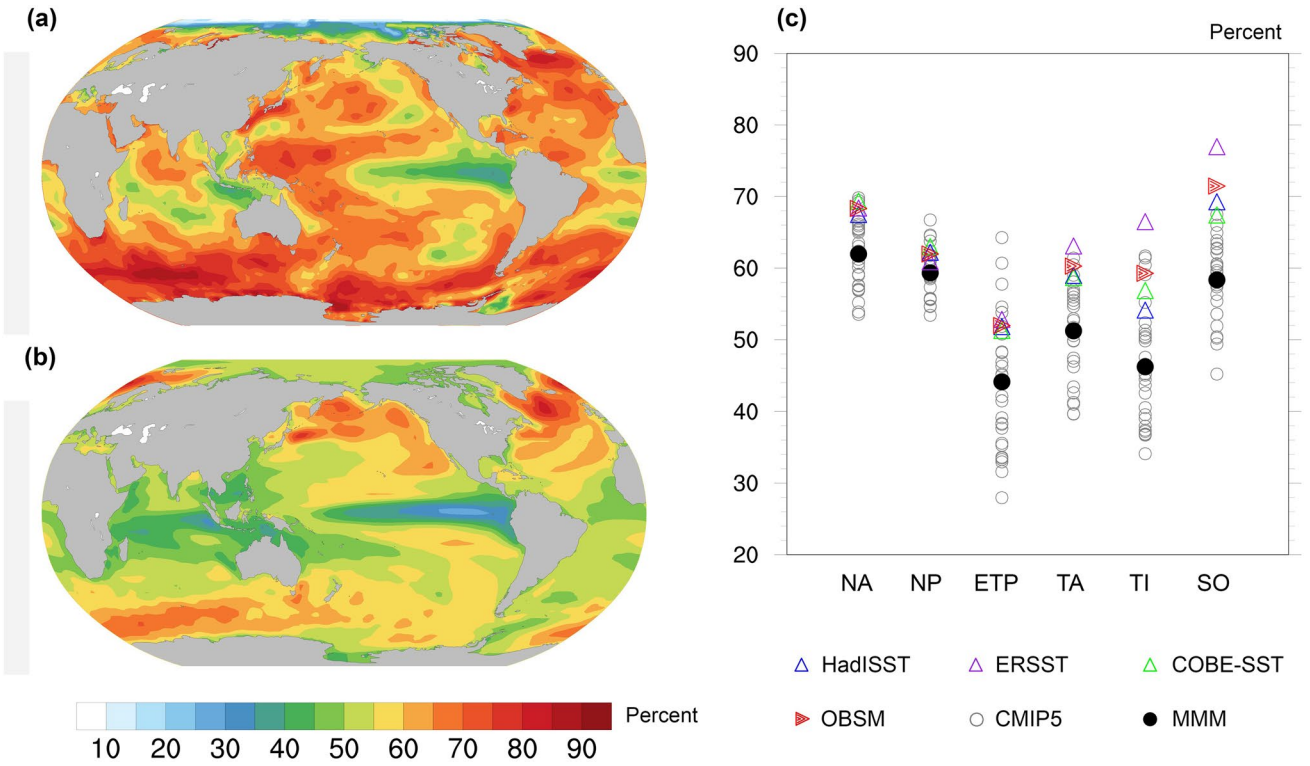

Fig. 1 Decadal SST predictability based on observations and CMIP5 models. **a** Observational mean estimates based on three observational SST datasets. We calculate decadal SST predictability for each observational dataset and then take the average as the observational mean estimates. **b** Multi-model mean estimates based on 30 CMIP5 historical simulations. **c** Estimates of Decadal SST predictability for each observational dataset and CMIP5 model averaged over six different ocean regions, namely, the North Atlantic (NA; 20°–65° N, 80°–0° W), North Pacific (NP; 20°–60° N, 120° E–120° W), Eastern Tropical Pacific (ETP; 10° S-10° N, 180–85° W), Tropical Atlantic (TA; 15° S–15° N, 80°–0° W), Tropical Indian (TI; 15°S-25°N, 40–100°E), and Southern Ocean (SO; 65°–40° S, 0°–360° E). HadISST = Hadley Centre Sea Ice and Sea Surface Temperature. ERSST = Extended Reconstructed Sea Surface Temperature. COBE-SST = Centennial in situ Observation-Based Estimates. OBSM = observational mean estimates. CMIP5 = Coupled Model Intercomparison Project 5. MMM = Multi-model Mean estimates

is found with the different detrending methods and even without detrending, supporting our argument of the underestimated decadal SST predictability in CMIP5 models.

Figure 3 is the same as Fig. 1 but for decadal SLP predictability. The observations and reanalysis data (ERA20C, 20CR, and HadSLP) present higher values of decadal SLP predictability than almost all the CMIP5 models, indicating that CMIP5 models also underestimate the observed decadal SLP predictability. Figure 3c shows regionally averaged decadal SLP predictability for each observation and CMIP5 model in different ocean regions, namely, NA, NP, ETP, TA, TI, and SO. In all the ocean regions, CMIP5 models present substantially lower values of decadal SLP predictability than those based on observational SLP estimates. The underestimated decadal predictability in models is not merely due to the underestimate of decadal-scale SST and SLP variance in CMIP5 models. For example, the variance can be comparable or even higher than the observed variance estimate in substantial regions (Fig. 4). The results shown here suggest that the CMIP5 models largely underestimate the decadal predictability, but not necessarily the decadal variance. Even though the simulated decadal variance is large, its fractional contribution to the total variance is much smaller than in observations, meaning that the simulated non-decadal variance is much larger than in observations.

Wang et al. (2015) diagnose the leading Empirical Orthogonal Function (EOF) modes of SST on monthly and decadal timescales and argue that both CMIP3 and CMIP5 models fail to capture accurately the spatial structure compared with observational estimates of SST variability. The striking disagreement in SST variability and predictability among observational estimates and CMIP5 models may stem from the ocean–atmosphere coupling (Li et al. 2013; Sun et al. 2015), ocean dynamics (Kirtman et al. 2012), and intrinsic model errors (Gupta et al. 2013; Richter 2015), which requires further investigation.

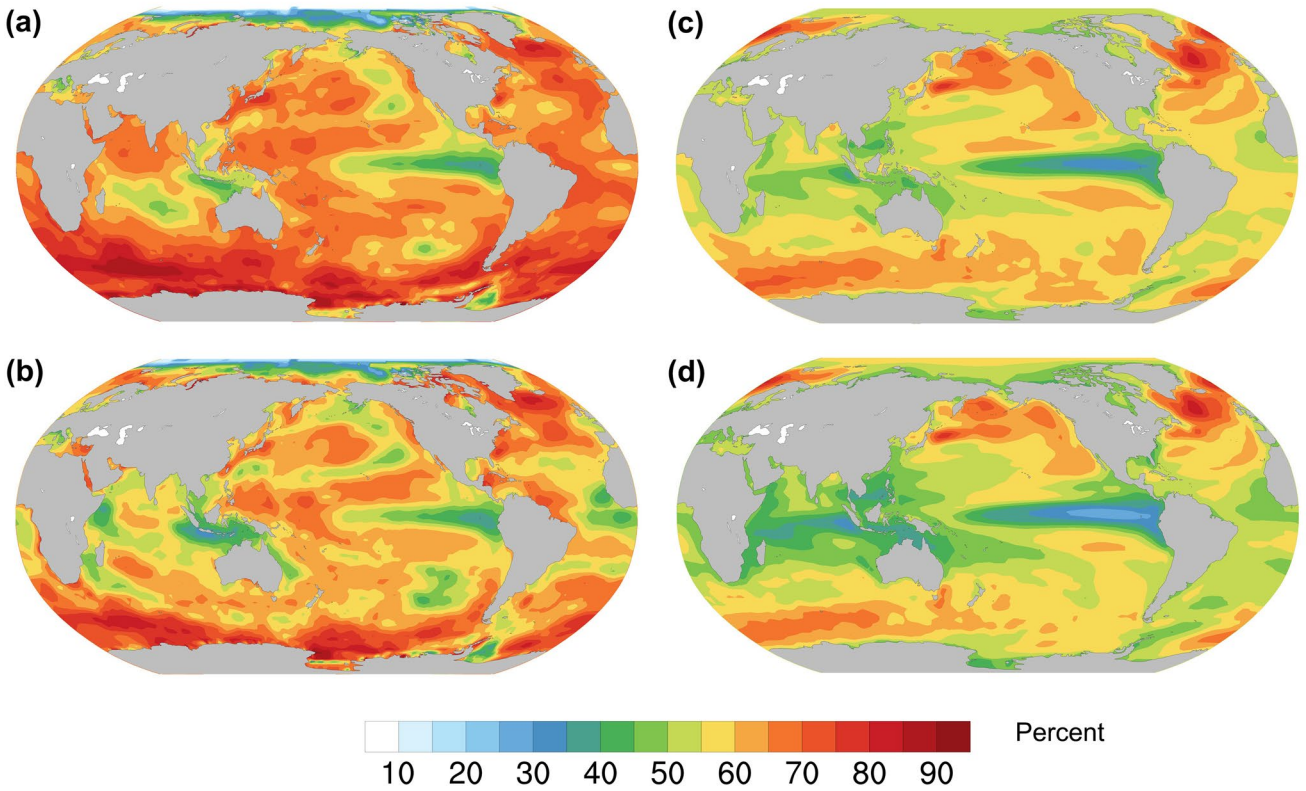

Fig. 2 Impact of linear detrending on decadal SST predictability. **a** Observational mean estimates of decadal SST predictability without any detrending. **b** Observational mean estimates of decadal SST predictability using the regression-detrending method (regressing out the global mean warming trend). **c** Multi-model mean estimates of decadal SST predictability without any detrending. **d** Multi-model mean estimates of decadal SST predictability using the regression-detrending method

3.2 Signal-to-noise paradox in decadal climate predictability

Recent research reveals the widespread existence of the signal-to-noise paradox in climate models. Here we ask is the underestimated decadal predictability in CMIP5 models related to the signal-to-noise paradox? Figure 5 encapsulates the relationship between the paradox and predictability. First, Fig. 5a shows the difference in decadal SST predictability between observations (Fig. 1a) and CMIP5 HIST simulations (Fig. 1b), which is indicative of the underestimation in CMIP5 models. Based on the Markov model framework (see Sect. 2), the chance of existence for the signal-to-noise paradox based on thirty CMIP5 HIST simulations is examined in the low-pass filtered SST field (Fig. 5b). Each SST observation and model simulation is detrended and normalized before analysis. We first compute the RSC values of global SST for each CMIP5 model based on the method introduced in Sect. 2. The paradox is present when the RSC values are greater than 1.0 at any spatial grid for each model. Here we apply the likelihood chart used in the Fifth Assessment Report of the United Nations Intergovernmental Panel on Climate Change (IPCC 2014): very likely 90–100%, likely 66–90%, about as likely as not 33–66%, unlikely 10–33%, and very unlikely 0–10%. The chance of existence for the paradox is then determined by considering all the thirty CMIP5 models; for example, if over 90% of thirty models present RSC values greater than 1.0 (or a paradox) in a region, this region will be considered "very likely" for the paradox to exist. The patterns in the possibility of the existence of the signal-to-noise paradox (Fig. 5b) are somewhat consistent with those in Fig. 5a, implying a possible relationship between the underestimated decadal SST predictability and the signal-to-noise paradox. Models are likely to underestimate decadal predictability in regions where it is likely to have a signal-to-noise paradox, especially around the Tropical Atlantic and the Tropical Indian Ocean and eddy-rich regions, including the Gulf Stream, the Kuroshio Current, and the Southern Ocean. We note that considerable regions show inconsistency between Fig. 5a and b such as the mid-latitude North Atlantic, subtropical Northwest Pacific, and eastern tropical Pacific. The signal-to-noise paradox (Fig. 5b) is determined by the Markov model framework, which is developed based on the hypothesis that the lack of persistence

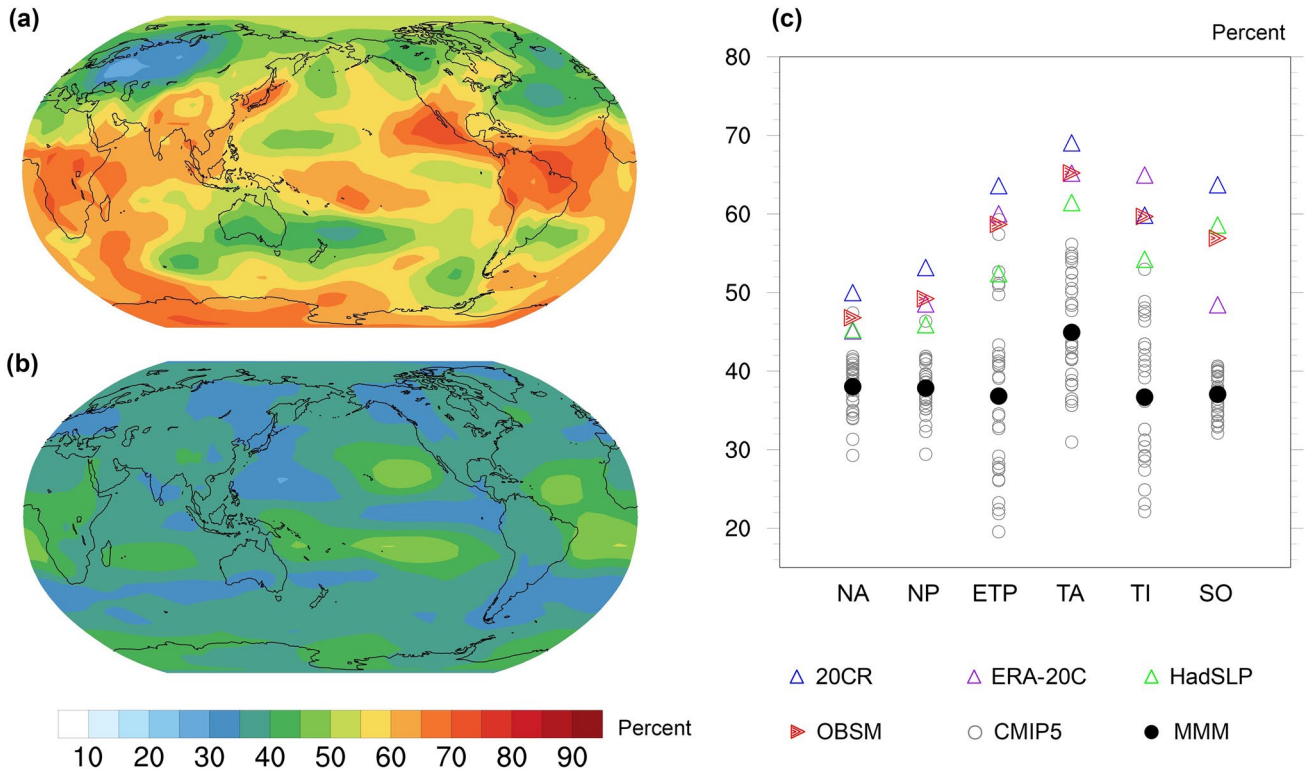

Fig. 3 Same as Fig. 1 but for decadal SLP predictability. **a** Observational mean estimates based on three observational SLP datasets. (b) Multi-model mean estimates based on 30 CMIP5 historical simulations. **c** Estimates of decadal SLP predictability for each observational dataset and CMIP5 model averaged over six different ocean regions (same as in Fig. 1c). 20CR = NOAA-CIRES-DOE Twentieth Century Reanalysis. ERA-20C = ECMWF's Astmospheric Reanalysis of the Twentieth Century. HadSLP = Hadley Centre's Mean Sea Level Pressure

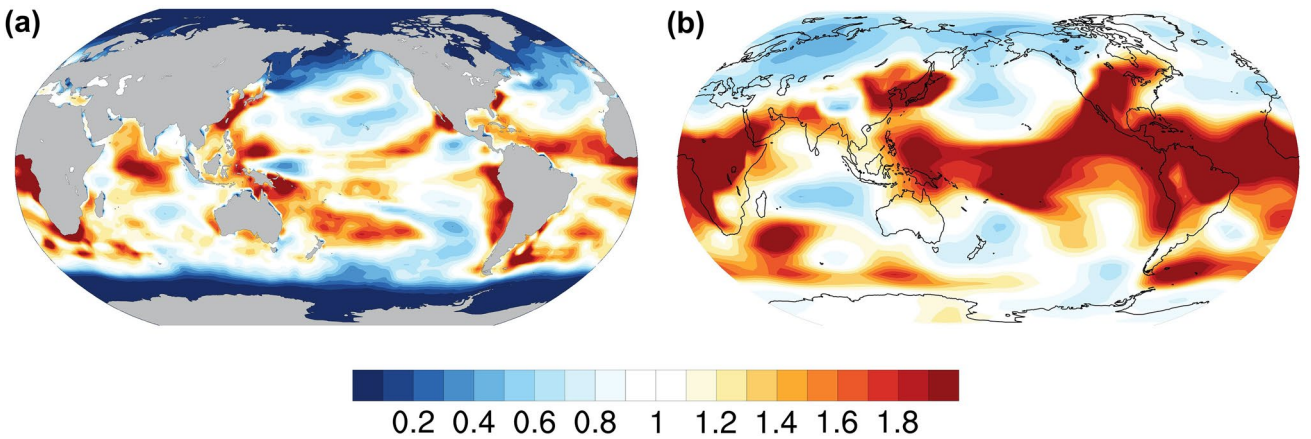

Fig. 4 Variance ratio between observations and CMIP5 models for **a** decadal SST variability and **b** decadal SLP variability. SST in models is obtained by masking all the non-ocean regions of surface temperature. Low ratio of decadal SST variance between models and observations in polar regions is significantly affected by sea ice temperature

leads to the paradox. Predictability can be highly related to persistence, but high persistence is not always indicative of larger predictability. Uncertainty remains in the simple Markov model as the lack of persistence may not be the only factor contributing to the signal-to-noise paradox (e.g., O'Reilly et al. 2019; Scaife et al. 2019) and the underestimated decadal predictability.

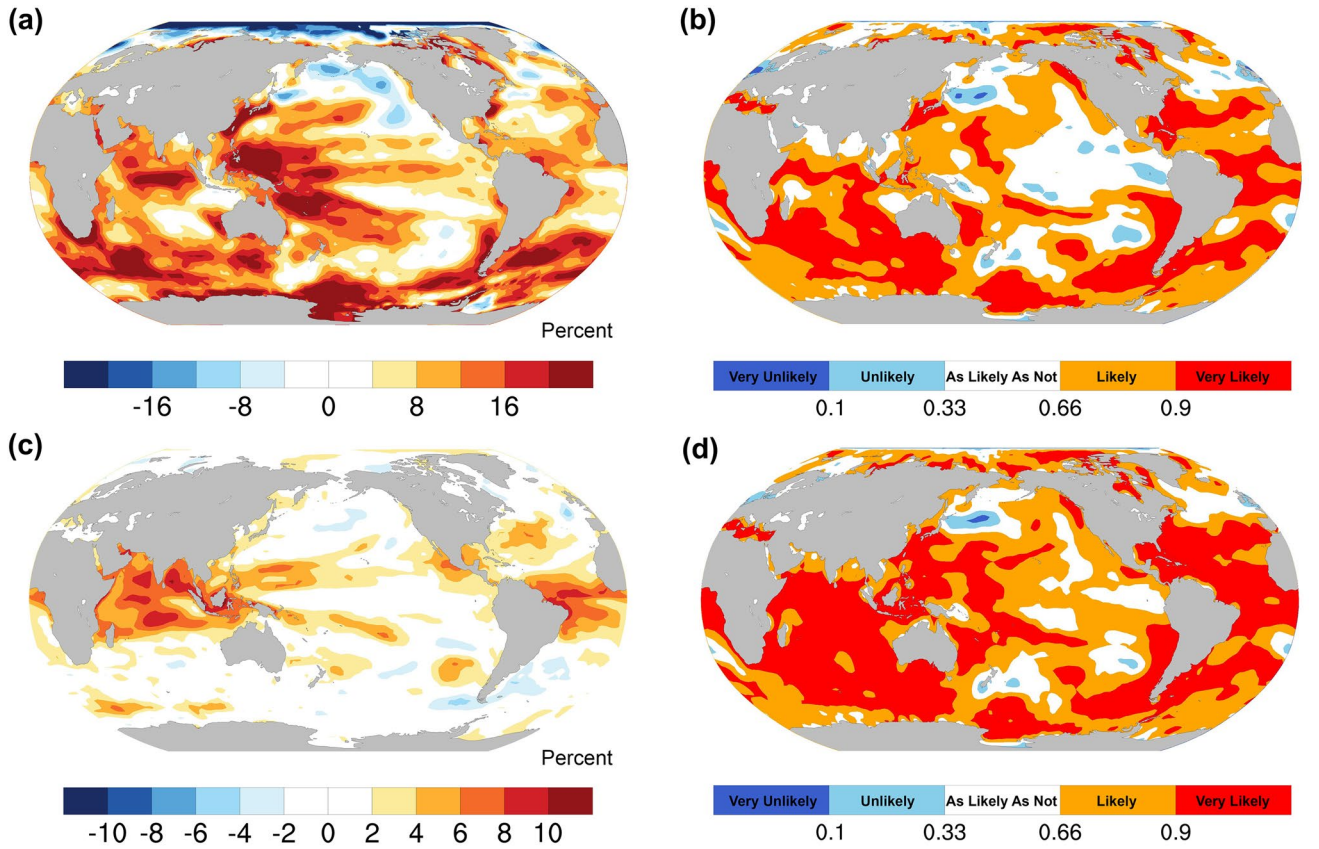

Fig. 5 Existence of the signal-to-noise paradox in CMIP5 models and the impact of external forcing. **a** Difference of decadal SST predictability between observations and CMIP5 historical simulations. **b** Chance of existence for the signal-to-noise paradox based on 30 CMIP5 historical simulations. Each SST simulation is detrended and normalized before analysis. The existence of the signal-to-noise paradox is estimated based on the Markov model framework. **c** Difference of decadal SST predictability between CMIP5 historical and pre-industrial control simulations, suggesting the impact of external forcing. **d** Chance of existence for the signal-to-noise paradox based on 30 CMIP5 pre-industrial control simulations

This relationship between the paradox and predictability is further supported by the ratio of squared correlation over six ocean areas (same as Fig. 1c) estimated for each model in CMIP5 (Fig. 6). The regional mean SST index for each ocean region is created and then assessed with the ratio of squared correlation for each model. There would be a signal-to-noise paradox when the ratio of squared correlation is higher than 1.0 (Eq. 4). Here we show that the signal-to-noise paradox is very likely to occur in extratropical regions (e.g., the North Atlantic), the Tropical Atlantic and the Southern Ocean; meanwhile, only about half of the CMIP5 models used in this study indicate a paradox in the North and Eastern Tropical Pacific. The results in Fig. 6 are in good agreement with decadal SST predictability estimates for regional averages based on three observational datasets and 30 CMIP5 models in Fig. 1c. For example, ratios of squared correlation for CMIP5 models are generally greater than 1.0 in the North Atlantic (Fig. 6a), where decadal predictability of SST is underestimated by most of CMIP5 models (Fig. 1c). The distributions for the chance of occurrence of the signal-to-noise paradox in this study show some consistency with several previous studies (Eade et al. 2014; Smith et al. 2019); for instance, Eade et al. (2014) show the distribution of the signal-to-noise paradox in surface air temperature by the ratio of predictable component based on a multi-model ensemble of decadal hindcasts from the Met Office decadal prediction system (DePreSys; Knight et al. 2014) and four CMIP5 models.

The impact of external forcing on decadal SST predictability and the signal-to-noise paradox is also examined here (Fig. 5c, d). Figure 5c displays the difference of decadal SST predictability between thirty CMIP5 HIST and PI simulations, with the latter having constant external forcing. Allowing for model drift, we only use the last 200 years of PI simulations to compute decadal SST predictability (Calculation 1). Some CMIP5 PI simulations may have short time records less than 500 years. Therefore, we recalculated decadal SST predictability based on selected models with simulation length over 500 years (Calculation 2). The results suggested no significant difference between Calculation 1

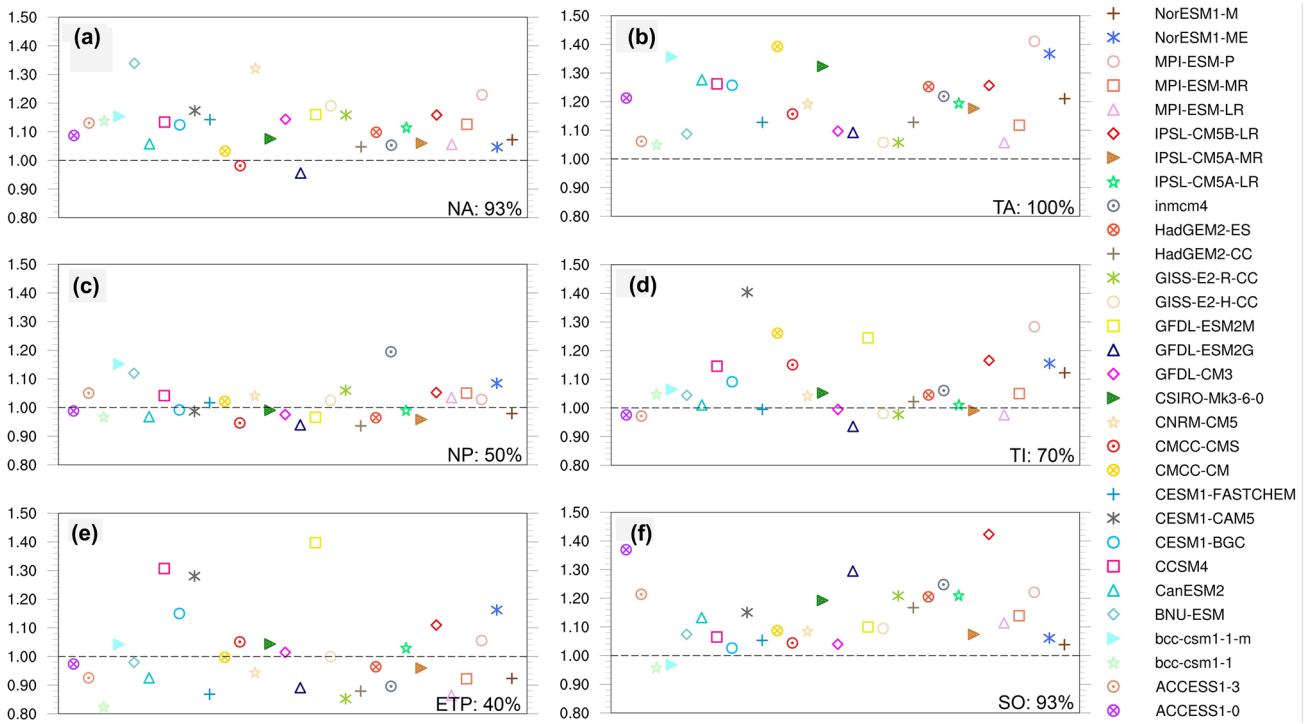

Fig. 6 Ratio of squared correlation estimated based on the Markov model framework for 30 CMIP5 models (historical simulations) in six different ocean regions: **a** NA: North Atlantic, **b** NP: North Pacific, **c** ETP: Eastern Tropical Pacific, **d** TA: Tropical Atlantic, **e** TI: Tropical Indian, and **f** SO: Southern Ocean. The percentage in each panel is computed as the number of models with RSC values greater than 1.0 divided by thirty

and 2. The most noticeable difference between CMIP5 HIST and PI simulations appears in the tropics such as the Tropical Atlantic and the Tropical Indian Ocean as larger decadal SST predictability is found in these regions, accompanied by a higher chance of the signal-to-noise paradox. Parts of western tropical Pacific and North Atlantic are also influenced by external forcing, where the chance of existence for the paradox is high for both HIST and PI simulations. Consistent with previous studies (e.g., Goddard et al. 2013; Guemas et al. 2013; Meehl et al. 2014), the Tropical Indian ocean stands out as the area significantly affected by the externally forced variability, which is shown to be much larger than the internally generated variability in both uninitialized simulations and initialized decadal hindcasts. We hypothesize that CMIP5 models underestimate the externally forced trend in the Tropical Indian ocean, though we cannot exclude the role of internal dynamics and any other associated factors in decadal predictability.

The Tropical Atlantic is another region of emerging interest in near-term climate predictability, where external forcing acts as an important factor driving decadal variability (Yeager and Robson 2017). The Tropical Atlantic has long been considered as a region with significant SST bias and poor upper ocean thermal structure and limited decadal predictability (Harlaß et al. 2018; Patricola et al. 2012; Xu et al. 2014). Shaffry et al. (2017) utilized a high-resolution eddying GCM (ocean model resolution $1/3° \times 1/3°$) and showed improved decadal prediction skills compared with low-resolution models, especially over the Tropical Atlantic region, pointing toward the importance of model resolution, in addition to external forcing.

Furthermore, the coexistence of the underestimated decadal SST predictability and the high chance of occurrence for the signal-to-noise paradox in eddy-rich regions is suggestive of the lack of ocean model resolution in CMIP5 models. This is because all the coupled models in CMIP5 use eddy parameterized ocean models that may have weak vertical connectivity between ocean mixed layer and thermocline (Kravtsov 2020). The role of mesoscale ocean eddies and fronts in climate variability, air-sea interaction, and predictability, particularly in the western boundary regions has been highlighted in previous work (Bryan et al. 2010; Kirtman et al. 2012; Minobe et al. 2008; Siqueira and Kirtman 2016), and is a potential source of decadal predictability that has not been fully accounted for or leveraged.

3.3 Advancing decadal predictability from an eddying GCM

The following is based on the hypothesis that the presence of ocean mesoscale processes and features and the associated

vertical connectivity affects decadal variability, predictability, and the signal-to-noise paradox. Specifically, coupled models with an eddying ocean component may, at least partially, reduce the signal-to-noise issue and thus improve decadal-scale climate predictability. The enhanced predictability we argue, is in part, due to the enhanced vertical connectivity in the ocean. This enhanced vertical connectivity allows the deeper ocean to more efficiently communicate with the surface, which, given the slower sub-surface time-scales, leads to longer surface predictability. To test this argument, we perform a suite of model experiments using CCSM4 with HR (eddying; 0.1° ×0.1°) and LR (eddy-parameterized; 1° × 1°) ocean component models. The details of the CCSM4 model setup and experiment design have been provided in Sect. 2, which are generally consistent with Kirtman et al. (2017), but here we employ a much longer LR simulation.

Figure 7 encapsulates how better represented ocean mesoscales affect decadal predictability estimates. In particular, Fig. 7a, b shows the global distributions of decadal SST potential predictability based on CCSM4 HR and the ratio of decadal SST predictability between HR and LR simulations, respectively. In the HR simulations, we find a higher decadal SST predictability in eddy-rich regions such as the Gulf Stream and Kuroshio Current systems, Tropical Atlantic, and Southern Ocean, where decadal SST variability is also increased (Fig. 7c). Perhaps surprising is that we also detect higher decadal SST predictability in HR over the Tropical Indian region, suggesting the strong influence of mesoscale ocean features on decadal SST predictability in addition to the impact of external forcing (e.g., Guemas et al. 2013; Meehl et al. 2014). Another possible explanation is better represented ocean fronts and mesoscale eddies in the North Atlantic in HR are forcing basin-wide modes, enhancing decadal variability in the tropics. As suggested by Hameed et al. (2018), Gulf Stream variability in the decadal band is related to both the Atlantic Meridional Mode and North Atlantic Oscillation. Compared with HR, LR has substantially broader and more diffuse western boundary currents leading to underestimated covariability between Gulf Stream and Tropical Atlantic (Hameed et al. 2018). We also note that decadal SST variability is remarkably elevated with HR

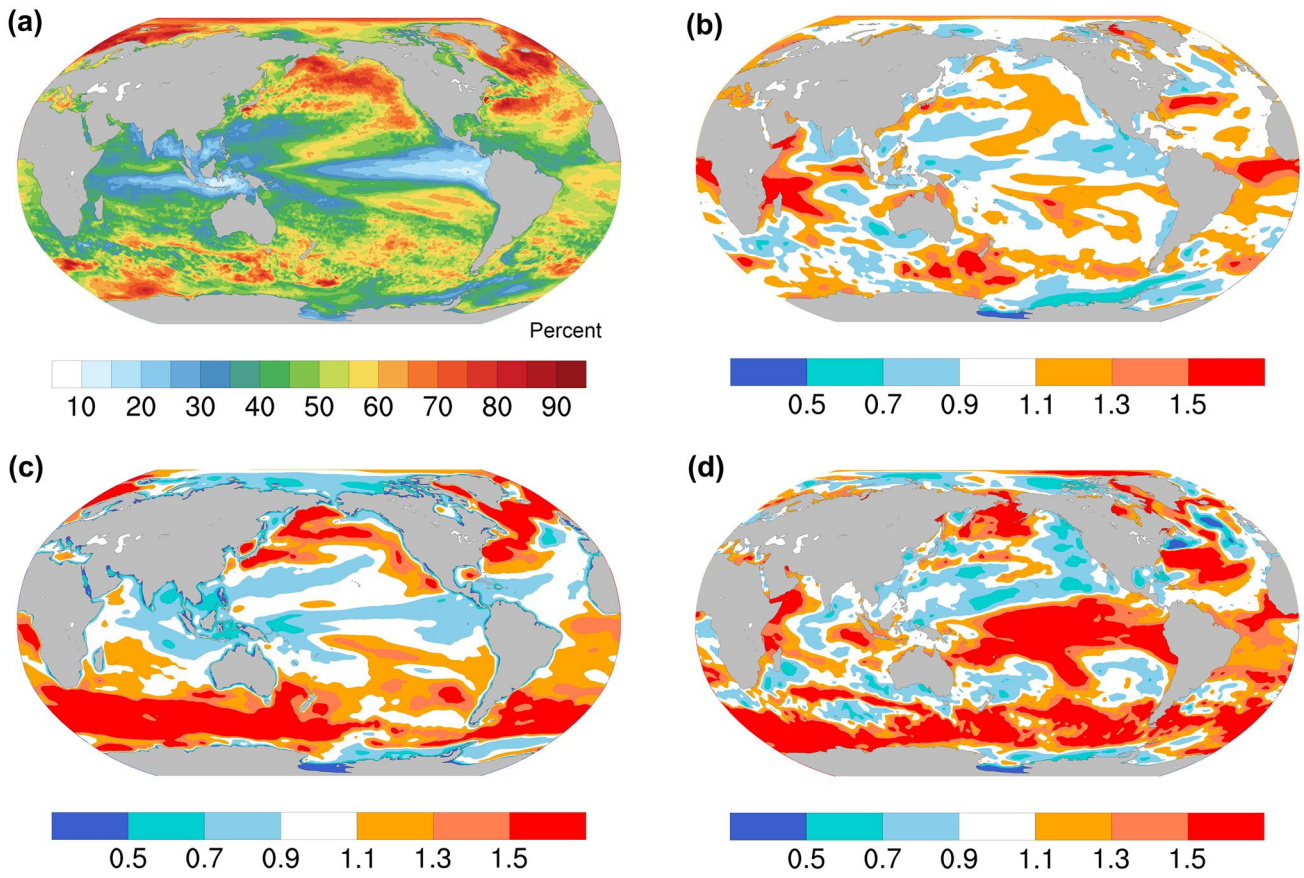

Fig. 7 Decadal SST predictability using CCSM4 models and the effect of ocean model resolutions. **a** Decadal SST predictability based on HR CCSM4. Ratio of HR and LR CCSM4 in terms of **b** decadal SST predictability, **c** decadal SST variability, and **d** the persistence of decadal SST variability. The persistence of the system is estimated as the lag-1 year autocorrelation of 5-year low-pass filtered SST data

over western and eastern boundary current regions in the extratropics (Fig. 7c).

Table 2 provides regionally averaged decadal SST predictability for CCSM4 LR and HR models in different ocean regions. In many ocean regions, we detect increased decadal SST predictability with the global mean HR/LR predictability ratio of 1.1. In the meanwhile, there are considerable regions such as the Eastern Tropical Pacific, Northern North Atlantic, as well as parts of the Southern Ocean showing reduced decadal predictability in HR. In the Northern North Atlantic, for example, both the LR and HR simulations show relatively higher decadal SST predictability than other ocean regions (Fig. 7a and Table 2), which is closely related with the persistence of ocean heat content (OHC) variability (Buckley et al. 2019; Foukal and Lozier 2018; Klavans et al. 2019; Robson et al. 2012; Yeager and Robson 2017), which some studies have related to the Atlantic Meridional Overturning Circulation (AMOC; Latif et al. 2006; Yan et al. 2018; Zhang 2017; Zhang and Zhang, 2015). The inclusion of mesoscale ocean features in HR results in greater decadal variability in the North Atlantic SST, but without a substantial increase in decadal predictability, except in the Gulf Stream and its extension. In particular, we find even decreased decadal predictability (or persistence) with HR in the Northern North Atlantic, with the HR/LR predictability (or persistence) ratio of 0.9 (Table 2).

The spatial patterns and area-averaged values of the HR/LR persistence ratio are included in Fig. 7d and Table 2. The persistence of the system is estimated as the lag-1 year autocorrelation of 5-year low-pass filtered SST data. We compute the HR/LR persistence ratio because of the lack of persistence in models can contribute to the signal-to-noise paradox (see also Eq. 5 in Sect. 2.3). A comparison of the persistence characteristics between HR and LR models indicates the possibility of the paradox's existence. Note that we cannot directly compare the persistence of CCSM4 model simulations (control simulations) with observational estimates. So the results solely based on the persistence characteristics can lead to considerable uncertainty. As shown in Table 2, we find that many regions with higher decadal predictability in HR than LR are accompanied with longer persistence and vice versa. We speculate that in these regions (larger predictability and persistence in HR than LR), there is a smaller chance for the paradox to exist in HR than LR. Uncertainty remains in regions such as parts of the Southern Ocean and Eastern Tropical Pacific, showing significant inconsistency between decadal predictability and persistence. For example, this predictability-persistence inconsistency in the Eastern Tropical Pacific is possibly due to different decadal ENSO variability in the simulations, as shown in Fig. 7c, decadal variance in HR is much lower than LR. Although HR shows longer persistence in the Eastern Tropical Pacific, the noise variance may be larger than LR, contributing to a higher chance for the paradox to exist in HR (Eq. 5 in Sect. 2.3).

Table 2 Regionally averaged decadal SST predictability for CCSM4 LR and HR models and HR/LR persistence ratio

Region	Decadal SST predictability (%)		HR/LR predictability ratio	HR/LR persistence ratio
	LR	HR		
NA	57.2	61.6	1.1	1.3
NNA	60.6	52.8	0.9	0.9
GS	48.9	63.6	1.3	2.3
NP	54.5	56.6	1.0	1.0
ETP	20.7	18.1	0.9	1.5
TA	38.5	47.5	1.2	1.3
TI	31.8	35.2	1.1	1.2
SO	48.5	53.0	1.1	1.7
Global Mean	47.7	51.0	1.1	1.3

HR/LR predictability (persistence) ratio is defined as HR predictability (persistence) divided by LR predictability (persistence)

NA North Atlantic (20°–65° N, 80°–0° W), *NNA* Northern North Atlantic (47°–52° N, 60°–45° W), *GS* Gulf Stream (32°–45° N; 80°–45° W), *NP* North Pacific (20°–60° N, 120° E–120° W), *ETP* Eastern Tropical Pacific (5° S–5° N, 120°–90° W), *TA* Tropical Atlantic (15° S–15° N, 80°–0° W), *TI* Tropical Indian (15° S–25° N, 40°–100° E), *SO* Southern Ocean (65°–40° S, 0°–360° E)

As noted earlier, we hypothesize on these decadal timescales (5–10 years) the enhanced vertical connectivity in the HR simulation compared to the LR simulation is a possible explanation for the longer limits of predictability (Buckley et al. 2019). We demonstrate this point by first taking a close look at the decadal predictability ratio of SST and subsurface ocean temperature averaged over 200–700 m depth (TEMP) in CCSM4 HR and LR models as well as the area-averaged values of SST/TEMP predictability ratio in different ocean regions (Fig. 8). In LR, decadal TEMP predictability is substantially higher than decadal SST predictability (except for parts of the Tropical Indian Ocean and Eastern Tropical Pacific; see Fig. 8b), while many regions in HR show comparable decadal SST and TEMP predictability (Fig. 8a). As shown in Fig. 8c, SST/TEMP decadal predictability ratio in HR is much closer to 1.0 than LR. We speculate that the difference between decadal SST and TEMP predictability in LR results from the lack of vertical connectivity. That is, a lack of vertical connectivity suggests that deeper ocean processes that tend to occur on longer timescales are not contributing to the SST variability, and thus the decadal predictability is lower.

Figure 9 shows the lead-lag correlations of SST anomalies (SSTA) and OHC at different depths over the Northern North Atlantic, Gulf Stream, Eastern Tropical Atlantic, and Eastern Tropical Pacific, where large differences between LR and HR decadal SST predictability are situated (see

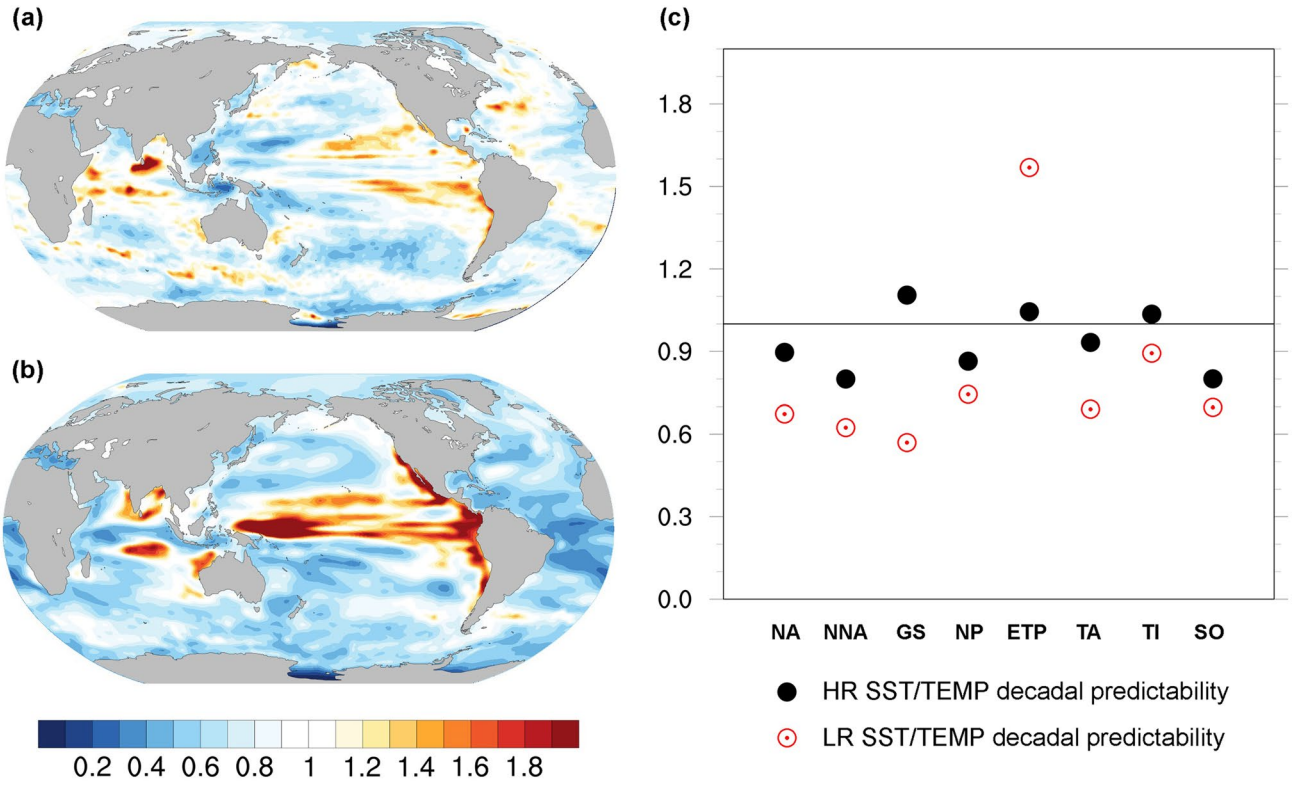

Fig. 8 SST/TEMP decadal predictability ratio for **a** CCSM4 HR and **b** LR models. **c** Regionally averaged values in different ocean regions, including NA, NNA, GS, NP, ETP, TA, TI, and SO (same as regions in Fig. 7). TEMP decadal predictability is computed as decadal predictability of subsurface ocean temperature averaged over 200–700 m depth. The black reference line at 1.0 in **c** implies comparable values of decadal SST and TEMP predictability

Fig. 7b). For shallow depth integrals of 100 m and 200 m, both LR and HR models show OHC anomalies highly correlated with SSTA, especially in the Gulf Stream (Fig. 9c, d). However, for deeper depth integrals, such as to 400 and 700 m, the OHC anomalies get progressively less correlated with the SSTA in the LR model, especially in the Northern North Atlantic (Fig. 9a, b) and Eastern Tropical Atlantic (Fig. 9e, f), indicating evident differences with HR, which maintains a consistent vertical structure. Notably, LR has OHC leading SST up to a year over the Eastern Tropical Pacific (Fig. 9g, h), which can contribute to a higher decadal SST predictability in LR than HR (Fig. 7b). Of particular interest is the Northern North Atlantic, where HR seems to have more vertical connectivity (Fig. 9a, b) but shorter decadal SST predictability (Table 2). So we speculate that the reduced predictability in HR cannot be simply explained by vertical connectivity, other factors such as different mean state model bias (e.g., Kirtman et al., 2017), and teleconnection patterns (i.e., the effect of ENSO on mid-latitude climate) between HR and LR may contribute to the predictability reduction.

To underscore the results in Fig. 9, i,e., the enhanced vertical connectivity associated with ocean mesoscale features and processes, we show the vertical correlation between SSTA and subsurface ocean temperature anomalies (Fig. 10) and the instantaneous correlation (zero-lag cross-correlation) between the 400 m OHC anomalies and surface heat flux anomalies (Fig. 11). Figure 10 shows the vertical correlation between SSTA and thermocline ocean temperature averaged over 200–500 m (upper thermocline; Fig. 10a, b) and 700–1000 m (deeper thermocline; Fig. 10c, d), as well as the difference of vertical correlation coefficients (absolute values) between HR and LR (Fig. 10e, f). We note that there is a strong vertical connection between SST and upper thermocline in many regions in both HR (Fig. 10a) and LR (Fig. 10b), such as the Northern North Atlantic, Gulf Stream, Eastern Tropical Pacific, and the Southern Ocean. For example, the vertical correlation coefficients in the Gulf Stream and its extension regions reach the maximum values of about 0.8 for both HR and LR simulations. Note that HR shows stronger vertical connections in part of the Northern North Atlantic compared with LR, with an area-averaged difference around 0.21. Different from LR, this strong vertical connectivity in HR persists in the deeper thermocline (Fig. 10c) albeit weaker than in the upper thermocline

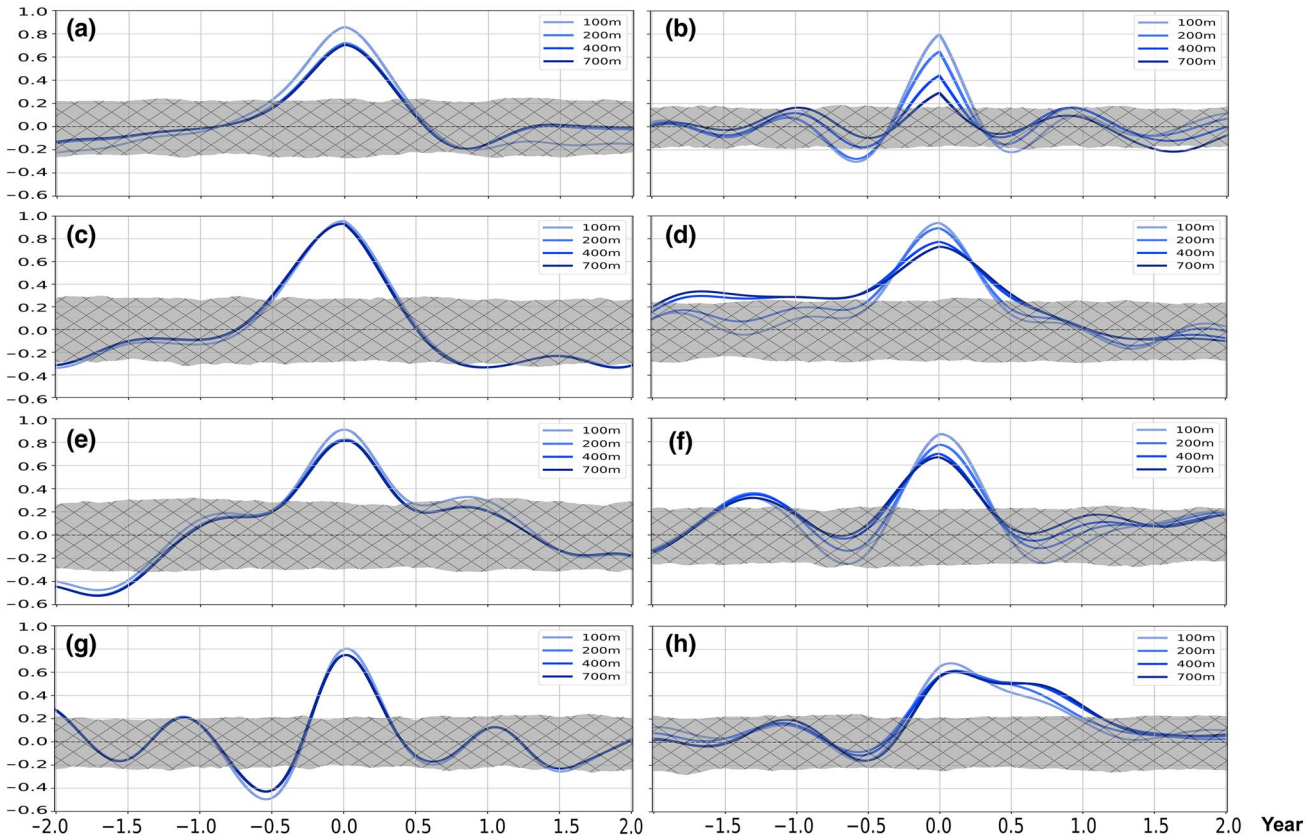

Fig. 9 Lagged cross-correlation between SSTA and OHC anomalies for depths of integration to 100, 200, 400 and 700 m in (**a**, **b**) Northern North Atlantic (47°–52° N, 60°–45° W), **c**, **d** Gulf Stream (32°–45° N; 80°–45° W), **e**, **f** Eastern Tropical Atlantic (20° S–10° N; 15° W–15° E), and **g**, **h** Eastern Tropical Pacific (5° S–5° N, 120°–90° W). The left panels (**a**, **c**, **e**, **g**) are based on HR while the right panels are based on LR. Negative (positive) years indicate the SSTA leading (lagging) the OHC anomalies for lags between −2 and 2 years. All variables are 5-year low-pass filtered. The one-tailed (95%) significance threshold for the cross-correlation is depicted by the hatched area and estimated using the non-parametric random phase method (Ebisuzaki 1997)

(Fig. 10a), except for the Eastern Tropical Pacific where both HR and LR show very weak or no vertical connectivity in the deeper thermocline (Fig. 10c, d). In the North Atlantic, for instance, the area-averaged vertical correlations in the upper thermocline in HR and LR are 0.38 and 0.32, respectively. The North Atlantic area-averaged vertical correlation coefficient in the lower thermocline slightly decreases to 0.22 in HR, but in LR, the vertical correlation is almost missing. The difference of vertical correlation coefficients (absolute values) between HR and LR (Fig. 10e, f) suggests that HR has more robust vertical connectivity, especially over eddy-rich regions (i.e., Gulf Stream, Kuroshio Current, and the Southern Ocean) that may contribute to larger decadal SST predictability and the smaller chance for the existence of the paradox than LR (Fig. 7b, d). HR has reduced vertical connectivity in the upper thermocline of Northern North Atlantic, which may provide some hints for the decreased decadal SST predictability (Table 2). However, larger vertical correlation values are detected in HR in the deeper thermocline (Fig. 10e, f).

Figure 11 shows the correlation between the surface net heat flux anomalies and OHC anomalies for HR and LR, as well as the difference of correlation coefficients (absolute values) between HR and LR. The results are generally consistent with Buckley et al. (2019). The correlations in frontal ocean zones and eddy rich regions in HR (Fig. 11a) are stronger than LR (Fig. 11a). Further, in LR, there are widespread regions of negative correlations in the Gulf Stream and Kuroshio subtropical recirculation gyres, subtropical North and South Pacific, and Southern Ocean (Fig. 11b). This reversal is difficult to detect or appears only in isolated regions of HR, notably in the Pacific warm pool and the sub-Antarctic zone of deep mixed layers. Though uncertainty remains, HR has higher correlation values between ocean heat content and net heat flux than LR, particularly over the Gulf Stream, Kuroshio Current,

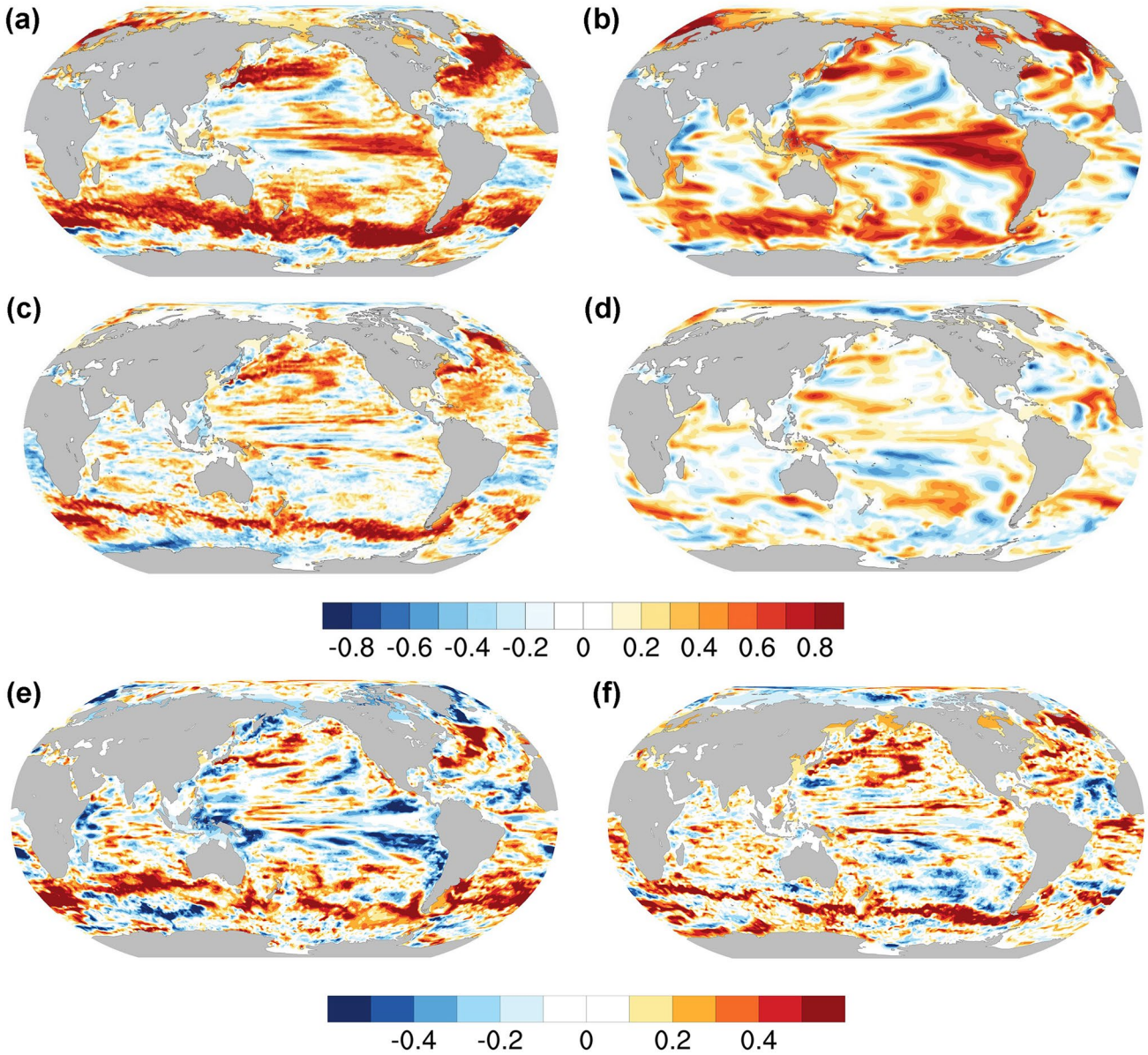

Fig. 10 Vertical correlation between SSTA and subsurface ocean temperature anomalies averaged over (**a**, **b**) the upper 200–500 m and **c**, **d** the upper 700–1000 m. **e**, **f** Difference of vertical correlation coefficients (absolute values) between HR and LR models: **e** difference between **a** and **b**; **f** difference between **c** and **d**. **a** and **c** are for HR and **b** and **d** are for LR. Vertical correlation coefficients are estimated as simultaneous correlation between SSTA and subsurface ocean temperature anomalies. All the data are linearly detrended and 5-year low-pass filtered

and the Southern Ocean, consistent with results based on SST-TEMP relationships. Only a slight increase of vertical correlation in HR (about 0.1), however, is detected over the Tropical Indian and Tropical Atlantic Ocean, as shown in Fig. 11c. It is found that HR generally displays a higher correlation of OHC and net heat flux in eddy-rich regions, where higher decadal SST predictability and reduced signal-to-noise paradox are also shown (Fig. 7b).

4 Summary

To understand the underlying mechanisms for the signal-to-noise paradox, we focus on two main questions: (i) where and to what extent is the paradox leading to substantial underestimates of the limit of predictability? (ii) Is this underestimate of predictability and associated signal-to-noise paradox related to the representation of ocean

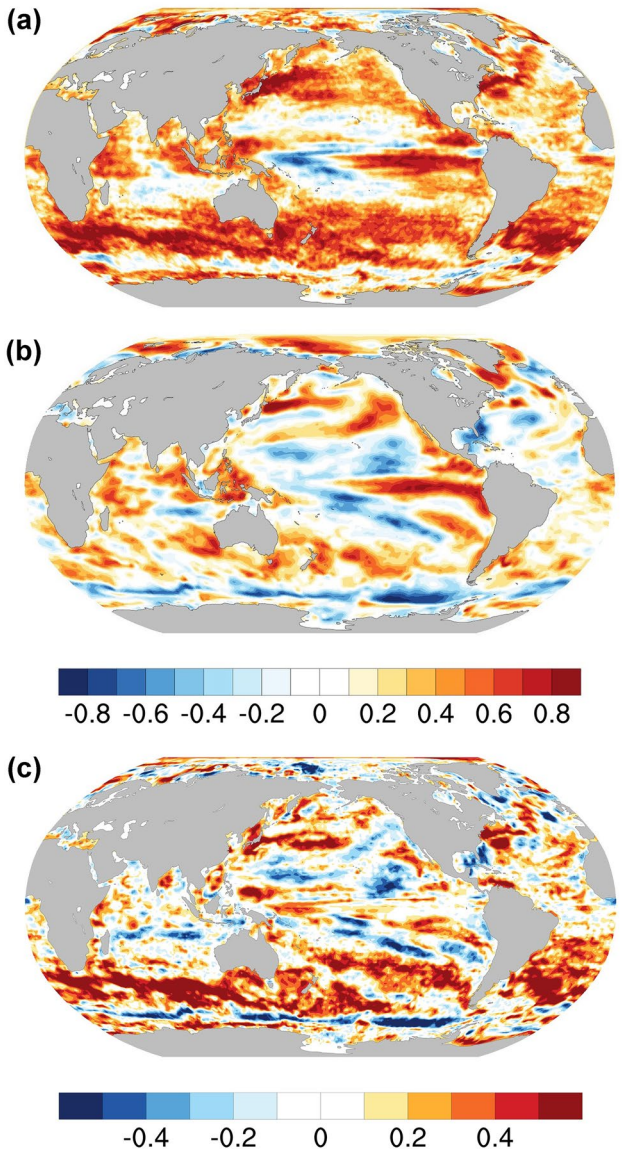

Fig. 11 Correlation between 5 year low-pass filtered anomalies of 0–400 m ocean heat content and net heat flux for **a** HR, **b** LR, and **c** difference of correlation coefficients (absolute values) between HR and LR. The sign convention in **a** and **b** indicates positive heat flux warms the atmosphere

mesoscale processes and features? To address the first question, we re-examine decadal predictability in CMIP5 models from the signal-to-noise paradox's perspective. We first compare decadal predictability of SST and SLP in observations and CMIP5 models, showing that decadal predictability estimates based on models are generally underestimated, particularly in the Tropical Atlantic, Tropical Indian ocean, and many eddy-rich regions. The distribution of the signal-to-noise paradox in the SST fields of CMIP5 models is presented following the Markov model framework in Zhang and Kirtman (2019b). The difference between observed and model-simulated decadal predictability is closely associated with the signal-to-noise paradox. Models are likely to underestimate decadal predictability in regions where it is likely for the signal-to-noise paradox to exist. Note that some regions fail to satisfy these predictability-paradox relationships, which requires further investigation. We also examined this question in the context of so-called historical climate simulations and pre-industrial control runs. For example, the Tropical Indian and Tropical Atlantic oceans are two distinct regions significantly influenced by external forcing, where we detect a lower chance of existence for the signal-to-noise paradox in CMIP5 HIST simulations compared with PI simulations. Many regions in the North Atlantic are also impacted by external forcing in terms of decadal SST predictability and the signal-to-noise paradox. The signal-to-noise paradox applies to CMIP5 uninitialized simulations (i.e., HIST and PI), implying that the paradox is not due to model initialization processes' problems.

To address the second question, we perform CCSM4 HR model experiments with eddying ocean component compared to the LR model, and the results are also discussed through the lens of the signal-to-noise paradox. The model experiments' design is based on the hypothesis that the presence of ocean mesoscale processes and features and the associated vertical connectivity impact decadal variability, predictability, and the signal-to-noise paradox. For the first time, we have addressed the signal-to-noise paradox using eddying GCMs. Our argument here is generally consistent with Strommen and Palmer (2019) and Zhang and Kirtman (2019b) in that we attribute the low signal-to-noise ratio to the lack of persistence, which can be seen in decadal predictability estimates in Fig. 7. We further argue that the lack of persistence in climate models stems from the lack of vertical connectivity in the subsurface ocean between ocean mixed layer and thermocline. As shown in SST/TEMP decadal predictability ratio (Fig. 8), there is a significant difference in decadal SST and TEMP predictability in LR resulting from the lack of vertical connectivity, which implies close relationships between vertical connectivity and decadal SST predictability. The inclusion of a high-resolution ocean component shows comparable decadal predictability of ocean temperature in the upper ocean in many regions. We compare HR and LR models in terms of the relationships among SST, surface heat flux, OHC, and subsurface ocean temperature, suggesting that HR and LR models have substantially different subsurface vertical structure (Figs. 9, 10, 11). In the HR model with better represented ocean mesoscales, we have demonstrated that there is consistent upper ocean vertical structure and strong vertical connection in the subsurface ocean that is weaker or even absent in the LR model. The differences in the HR and LR models' vertical connectivity can thus contribute to the differences in the persistence of

decadal SST variability and the decadal SST predictability. Compared with LR, HR suggests more vertical connectivity, particularly over eddy-rich regions, such as the Gulf Stream, Kuroshio Current, and the Southern Ocean, where increased decadal predictability and longer persistence (or reduced signal-to-noise paradox) are detected. Note that there are some exceptions, such as the Northern North Atlantic, where we detect increased vertical connectivity but decreased decadal SST predictability in HR compared with LR. This reduced decadal SST predictability in HR cannot be simply explained by vertical connectivity, which requires further investigation. Overall, we argue that the better represented mesoscale ocean features may (at least partially) eliminate the signal-to-noise issue and improve decadal-scale climate predictability in comparison with observational estimates.

Acknowledgements The research is supported by NOAA NA18OAR4310293, NA15OAR4320064, NSF OCE1419569, OCE1559151, and DOE DE-SC0019433. All the observations and CMIP5 model data used in this study are publicly available from the links and citations provided in the manuscript. CCSM4 model codes, experiments, and outputs were performed and archived at the University of Miami Center for Computational Science.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Allan R, Ansell T (2006) A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. J Clim 19(22):5816–5842. https://doi.org/10.1175/JCLI3937.1

Boer GJ (2004) Long time-scale potential predictability in an ensemble of coupled climate models. Clim Dyn 23(1):29–44. https://doi.org/10.1007/s00382-004-0419-8

Bryan FO, Tomas R, Dennis JM, Chelton DB, Loeb NG, McClean JL (2010) Frontal scale air–sea interaction in high-resolution coupled climate models. J Clim 23(23):6277–6291. https://doi.org/10.1175/2010JCLI3665.1

Buckley MW, DelSole T, Lozier MS, Li L (2019) Predictability of North Atlantic Sea Surface Temperature and Upper-Ocean Heat Content. J Clim 32(10):3005–3023. https://doi.org/10.1175/JCLI-D-18-0509.1

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Brönnimann S (2011) The twentieth century reanalysis project. Q J R Meteorol Soc 137(654):1–28. https://doi.org/10.1002/qj.776

Ding R, Li J, Zheng F, Feng J, Liu D (2016) Estimating the limit of decadal-scale climate predictability using observational data. Clim Dyn 46(5–6):1563–1580. https://doi.org/10.1007/s00382-015-2662-6

Eade R, Smith D, Scaife A, Wallace E, Dunstone N, Hermanson L, Robinson N (2014) Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? Geophys Res Lett 41(15):5620–5628. https://doi.org/10.1002/2014GL061146

Ebisuzaki W (1997) A method to estimate the statistical significance of a correlation when the data are serially correlated. J Clim 10(9):2147–2153. https://doi.org/10.1175/1520-0442(1997)010%3c2147:AMTETS%3e2.0.CO;2

Foukal NP, Lozier MS (2018) Examining the origins of ocean heat content variability in the eastern North Atlantic subpolar gyre. Geophys Res Lett 45(20):11–275. https://doi.org/10.1029/2018GL079122

Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR et al (2011) The community climate system model version 4. J Clim 24(19):4973–4991. https://doi.org/10.1175/2011JCLI4083.1

Goddard L, Kumar A, Solomon A, Smith D, Boer G, Gonzalez P, Kirtman BP (2013) A verification framework for interannual-to-decadal predictions experiments. Clim Dyn 40(1–2):245–272. https://doi.org/10.1007/s00382-012-1481-2

Gonzalez PL, Goddard L (2016) Long-lead ENSO predictability from CMIP5 decadal hindcasts. Clim Dyn 46(9–10):3127–3147. https://doi.org/10.1007/s00382-015-2757-0

Guemas V, Corti S, García-Serrano J, Doblas-Reyes FJ, Balmaseda M, Magnusson L (2013) The Indian Ocean: the region of highest skill worldwide in decadal climate prediction. J Clim 26(3):726–739. https://doi.org/10.1175/JCLI-D-12-00049.1

Gupta AS, Jourdain NC, Brown JN, Monselesan D (2013) Climate drift in the CMIP5 models. J Clim 26(21):8597–8615. https://doi.org/10.1175/JCLI-D-12-00521.1

Hameed S, Wolfe CL, Chi L (2018) Impact of the Atlantic meridional mode on Gulf Stream North Wall position. J Clim 31(21):8875–8894. https://doi.org/10.1175/JCLI-D-18-0098.1

Harlaß J, Latif M, Park W (2018) Alleviating tropical Atlantic sector biases in the Kiel climate model by enhancing horizontal and vertical atmosphere model resolution: climatology and interannual variability. Clim Dyn 50(7–8):2605–2635. https://doi.org/10.1007/s00382-017-3760-4

He J, Kirtman B, Soden BJ, Vecchi GA, Zhang H, Winton M (2018) Impact of ocean eddy resolution on the sensitivity of precipitation to CO2 increase. Geophys Res Lett 45(14):7194–7203. https://doi.org/10.1029/2018GL078235

Hirahara S, Ishii M, Fukuda Y (2014) Centennial-scale sea surface temperature analysis and its uncertainty. J Clim. https://doi.org/10.1175/JCLI-D-12-00837.1

Huang B, Thorne PW, Banzon VF, Boyer T, Chepurin G, Lawrimore JH, Zhang HM (2017) Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. J Clim 30(20):8179–8205. https://doi.org/10.1175/JCLI-D-16-0836.1

Infanti JM, Kirtman BP (2019) A comparison of CCSM4 high-resolution and low-resolution predictions for south Florida and southeast United States drought. Clim Dyn 52(11):6877–6892. https://doi.org/10.1007/s00382-018-4553-0

IPCC (2014) Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change. In: Pachauri (RK), Meyer (LA) (eds) Core writing team. IPCC, Geneva, Switzerland, 151 pp

Keenlyside NS, Latif M, Jungclaus J, Kornblueh L, Roeckner E (2008) Advancing decadal-scale climate prediction in the North Atlantic sector. Nature 453(7191):84. https://doi.org/10.1038/nature06921

Kim WM, Yeager S, Chang P, Danabasoglu G (2018) Low-frequency North Atlantic climate variability in the community earth system model large ensemble. J Clim 31(2):787–813. https://doi.org/10.1175/JCLI-D-17-0193.1

Kirtman BP, Schopf PS (1998) Decadal variability in ENSO predictability and prediction. J Clim 11(11):2804–2822. https://doi.org/10.1175/1520-0442(1998)011%3c2804:DVIEPA%3e2.0.CO;2

Kirtman BP, Pegion K, Kinter SM (2005) Internal atmospheric dynamics and tropical Indo-Pacific climate variability. J Atmos Sci 62(7):2220–2233. https://doi.org/10.1175/JAS3449.1

Kirtman BP, Bitz C, Bryan F, Collins W, Dennis J, Hearn N, Stan C (2012) Impact of ocean model resolution on CCSM climate simulations. Clim Dyn 39(6):1303–1328. https://doi.org/10.1007/s00382-012-1500-3

Kirtman B, Power SB, Adedoyin AJ, Boer GJ, Bojariu R, Camilloni I, Prather M (2013) Near-term climate change: projections and predictability. In: Stocker TF et al (eds) Climate change 2013: the physical science basis. Contribution of Working Group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge

Kirtman BP, Min D, Infanti JM, Kinter JL III, Paolino DA, Zhang Q, Peng P (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. Bull Am Meteorol Soc 95(4):585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Kirtman BP, Perlin N, Siqueira L (2017) Ocean eddies and climate predictability. Chaos 27(12):126902. https://doi.org/10.1063/1.4990034

Klavans JM, Clement AC, Cane MA (2019) Variable external forcing obscures the weak relationship between the NAO and North Atlantic Multidecadal SST variability. J Clim 32(13):3847–3864. https://doi.org/10.1175/JCLI-D-18-0409.1

Knight JR, Andrews MB, Smith DM, Arribas A, Colman AW, Dunstone NJ, Scaife AA (2014) Predictions of climate several years ahead using an improved decadal prediction system. J Clim 27(20):7550–7567. https://doi.org/10.1175/JCLI-D-14-00069.1

Kravtsov S (2012) An empirical model of decadal ENSO variability. Clim Dyn 39(9–10):2377–2391. https://doi.org/10.1007/s00382-012-1424-y

Kravtsov S (2020) Dynamics and predictability of hemispheric-scale multidecadal climate variability in an observationally constrained mechanistic model. J Clim. https://doi.org/10.1175/JCLI-D-19-0778.1

Kushnir Y, Scaife AA, Arritt R, Balsamo G, Boer G, Doblas-Reyes F, Matei D (2019) Towards operational predictions of the near-term climate. Nat Clim Change 9(2):94–101. https://doi.org/10.1038/s41558-018-0359-7

Latif M, Collins M, Pohlmann H, Keenlyside N (2006) A review of predictability studies of Atlantic sector climate on decadal time scales. J Clim 19(23):5971–5987. https://doi.org/10.1175/JCLI3945.1

Li J, Sun C, Jin FF (2013) NAO implicated as a predictor of Northern Hemisphere mean temperature multidecadal variability. Geophys Res Lett 40(20):5497–5502. https://doi.org/10.1002/2013GL057877

Marzocchi A, Hirschi JJM, Holliday NP, Cunningham SA, Blaker AT, Coward AC (2015) The North Atlantic subpolar circulation in an eddy-resolving global ocean model. J Mar Syst 142:126–143. https://doi.org/10.1016/j.jmarsys.2014.10.007

Meehl GA, Goddard L, Boer G, Burgman R, Branstator G, Cassou C, Karspeck A (2014) Decadal climate prediction: an update from the trenches. Bull Am Meteorol Soc 95(2):243–267. https://doi.org/10.1175/BAMS-D-12-00241.1

Merryfield WJ, Baehr J, Batté L, Becker EJ, Butler AH, Coelho CA, Ferranti L (2020) Current and emerging developments in subseasonal to decadal prediction. Bull Am Meteorol Soc. https://doi.org/10.1175/BAMS-D-19-0037.1

Minobe S, Kuwano-Yoshida A, Komori N, Xie SP, Small RJ (2008) Influence of the Gulf Stream on the troposphere. Nature 452(7184):206–209. https://doi.org/10.1038/nature06690

Murphy LN, Bellomo K, Cane M, Clement A (2017) The role of historical forcings in simulating the observed Atlantic multidecadal oscillation. Geophys Res Lett 44(5):2472–2480. https://doi.org/10.1002/2016GL071337

Newman M (2007) Interannual to decadal predictability of tropical and North Pacific sea surface temperatures. J Clim 20(11):2333–2356. https://doi.org/10.1175/JCLI4165.1

O'Reilly CH, Weisheimer A, Woollings T, Gray LJ, MacLeod D (2019) The importance of stratospheric initial conditions for winter North Atlantic Oscillation predictability and implications for the signal-to-noise paradox. Q J R Meteorol Soc 145(718):131–146. https://doi.org/10.1002/qj.3413

Patricola CM, Li M, Xu Z, Chang P, Saravanan R, Hsieh JS (2012) An investigation of tropical Atlantic bias in a high-resolution coupled regional climate model. Clim Dyn 39(9–10):2443–2463. https://doi.org/10.1007/s00382-012-1320-5

Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vitart F, Trémolet Y (2016) ERA-20C: An atmospheric reanalysis of the twentieth century. J Clim 29(11):4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1

Rayner NAA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res Atmos. https://doi.org/10.1029/2002JD002670

Richter I (2015) Climate model biases in the eastern tropical oceans: causes, impacts and ways forward. Wiley Interdiscip Rev Clim Change 6(3):345–358. https://doi.org/10.1002/wcc.338

Robson JI, Sutton RT, Smith DM (2012) Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s. Geophys Res Lett. https://doi.org/10.1002/2016GL070559

Robson J, Polo I, Hodson DL, Stevens DP, Shaffrey LC (2018) Decadal prediction of the North Atlantic subpolar gyre in the HiGEM high-resolution climate model. Clim Dyn 50(3–4):921–937. https://doi.org/10.1007/s00382-017-3649-2

Samanta D, Karnauskas KB, Goodkin NF, Coats S, Smerdon JE, Zhang L (2018) Coupled model biases breed spurious low-frequency variability in the tropical Pacific Ocean. Geophys Res Lett 45(19):10–609. https://doi.org/10.1029/2012GL053370

Scaife AA, Smith D (2018) A signal-to-noise paradox in climate science. npj Clim Atmos Sci 1(1):28. https://doi.org/10.1038/s41612-018-0038-4

Scaife AA, Arribas A, Blockley E, Brookshaw A, Clark RT, Dunstone N, Hermanson L (2014) Skillful long-range prediction of European and North American winters. Geophys Res Lett 41(7):2514–2519. https://doi.org/10.1002/2014GL059637

Scaife AA, Camp J, Comer R, Davis P, Dunstone N, Gordon M, Roberts M (2019) Does increased atmospheric resolution improve seasonal climate predictions? Atmos Sci Lett. https://doi.org/10.1002/asl.922

Shaffrey LC, Hodson D, Robson J, Stevens DP, Hawkins E, Polo I, Smith D (2017) Decadal predictions with the HiGEM high resolution global coupled climate model: description and basic evaluation. Clim Dyn 48(1–2):297–311. https://doi.org/10.1007/s00382-016-3075-x

Siegert S, Stephenson DB, Sansom PG, Scaife AA, Eade R, Arribas A (2016) A Bayesian framework for verification and recalibration of ensemble forecasts: how uncertain is NAO predictability? J Clim 29(3):995–1012. https://doi.org/10.1175/JCLI-D-15-0196.1

Siqueira L, Kirtman BP (2016) Atlantic near-term climate variability and the role of a resolved Gulf Stream. Geophys Res Lett 43(8):3964–3972. https://doi.org/10.1002/2016GL068694

Smith DM, Eade R, Scaife AA, Caron LP, Danabasoglu G, DelSole TM, Kharin V (2019) Robust skill of decadal climate predictions. npj Clim Atmos Sci 2(1):13. https://doi.org/10.1038/s41612-019-0071-y

Smith DM, Scaife AA, Eade R, Athanasiadis P, Bellucci A, Bethke I, Danabasoglu G (2020) North Atlantic climate far more predictable than models imply. Nature 583(7818):796–800. https://doi.org/10.1038/s41586-020-2525-0

Strommen K, Palmer TN (2019) Signal and noise in regime systems: a hypothesis on the predictability of the North Atlantic Oscillation. Q J R Meteorol Soc 145(718):147–163. https://doi.org/10.1002/qj.3414

Sun C, Li J, Jin FF (2015) A delayed oscillator model for the quasi-periodic multidecadal variability of the NAO. Clim Dyn 45(7–8):2083–2099. https://doi.org/10.1007/s00382-014-2459-z

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteor Soc 93(4):485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Ting M, Kushnir Y, Seager R, Li C (2009) Forced and internal twentieth-century SST trends in the North Atlantic. J Clim 22(6):1469–1481. https://doi.org/10.1175/2008JCLI2561.1

Wang G, Dommenget D, Frauen C (2015) An evaluation of the CMIP3 and CMIP5 simulations in their skill of simulating the spatial structure of SST variability. Clim Dyn 44(1–2):95–114. https://doi.org/10.1007/s00382-014-2154-0

Wittenberg AT, Rosati A, Delworth TL, Vecchi GA, Zeng F (2014) ENSO modulation: is it decadally predictable? J Clim 27(7):2667–2681. https://doi.org/10.1175/JCLI-D-13-00577.1

Wouters B, Hazeleger W, Drijfhout S, Van Oldenborgh GJ, Guemas V (2013) Multiyear predictability of the North Atlantic subpolar gyre. Geophys Res Lett 40(12):3080–3084. https://doi.org/10.1002/grl.50585

Xu Z, Chang P, Richter I, Tang G (2014) Diagnosing southeast tropical Atlantic SST and ocean circulation biases in the CMIP5 ensemble. Clim Dyn 43(11):3123–3145. https://doi.org/10.1007/s00382-014-2247-9

Yan X, Zhang R, Knutson TR (2018) Underestimated AMOC variability and implications for AMV and predictability in CMIP models. Geophys Res Lett 45(9):4319–4328. https://doi.org/10.1029/2018GL077378

Yeager SG, Robson JI (2017) Recent progress in understanding and predicting Atlantic decadal climate variability. Curr Clim Change Rep 3(2):112–127. https://doi.org/10.1007/s40641-017-0064-z

Zhang R (2017) On the persistence and coherence of subpolar sea surface temperature and salinity anomalies associated with the Atlantic multidecadal variability. Geophys Res Lett 44(15):7865–7875. https://doi.org/10.1002/2017GL074342

Zhang W, Kirtman B (2019a) Estimates of decadal climate predictability from an interactive ensemble model. Geophys Res Lett 46(6):3387–3397. https://doi.org/10.1029/2018GL081307

Zhang W, Kirtman B (2019b) Understanding the signal-to-noise paradox with a simple Markov model. Geophys Res Lett 46(22):13308–13317. https://doi.org/10.1029/2019GL085159

Zhang J, Zhang R (2015) On the evolution of Atlantic meridional overturning circulation fingerprint and implications for decadal predictability in the North Atlantic. Geophys Res Lett 42(13):5419–5426. https://doi.org/10.1002/2015GL064596

Zhang L, Delworth TL, Jia L (2017) Diagnosis of decadal predictability of Southern Ocean sea surface temperature in the GFDL CM2 1 model. J Climate 30(16):6309–6328. https://doi.org/10.1175/JCLI-D-16-0537.1