ENSO Predictability over the Past 137 Years Based on a CESM Ensemble Prediction System

TING LIU,a,b XUNSHU SONG,a,b YOUMIN TANG,c,d ZHEQI SHEN,c,b and XIAOXIAO TANc,b

a State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China
b Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China
c College of Oceanography, Hohai University, Nanjing, China
d Environmental Science and Engineering, University of Northern British Columbia, Prince George, British Columbia, Canada

1. Introduction

El Niño–Southern Oscillation (ENSO) is a prominent interannual signal in the global climate system with widespread climatic influence. As one of the most predictable climatic modes with profound impacts, ENSO also is the strongest predictability source for global seasonal climate predictions. In recent decades, enormous strides have been made in improving ENSO prediction skill (Chen et al. 2004; Chen and Cane 2008; Barnston et al. 2012; Zhang et al. 2013; Kirtman et al. 2014; Saha et al. 2014; Luo et al. 2016; Takaya et al. 2017; Tang et al. 2018; Barnston et al. 2019; Lin et al. 2020; Zhang et al. 2020). However, uncertainties in initial conditions, imperfect model formulation, and inherent unpredictability limit the prediction skill to some extent. A promising strategy for addressing the above uncertainties is to establish an ensemble forecast by perturbing the initial conditions or model parameters. A suitable ensemble generation strategy can generate a marked improvement in the deterministic (ensemble mean) prediction skill as compared with any individual forecast (Jin et al. 2008). In addition, an ensemble forecast also provides an opportunity to estimate the potential predictability, which quantifies the upper limit of the prediction skill (Tang et al. 2018). Therefore, ensemble forecasts are a useful tool for studies of ENSO predictability both from the perspective of actual deterministic prediction skill and potential predictability. However, the potential predictability estimated from an ensemble prediction is model dependent and could differ from the predictability in nature.

Our current understanding of ENSO predictability is based mainly on retrospective/hindcast forecasts obtained from various ENSO models, for which the actual prediction skill and potential predictability are important factors. The former is a measure of model performance in forecasting ENSO events in advance, whereas the latter measures the intrinsic limit of
ENSO prediction skill under the assumption of a perfect model (Tang et al. 2018). Indeed, retrospective forecasts over long periods can undoubtedly yield statistically robust estimates of the overall ability of a model to predict ENSO, and they make it possible to explore the interdecadal variability of ENSO and the factors that control this variability. Such efforts have been based on intermediate complexity (Chen et al. 2004; Zheng et al. 2009a) and hybrid coupled models (Tang et al. 2008; Deng and Tang 2009; Tang and Deng 2011). These long-term retrospective ENSO predictions have advanced our knowledge of interdecadal variability and indicated that the variation in ENSO signal strength controls the decadal variability of ENSO deterministic prediction skill. Furthermore, Cheng et al. (2010c) undertook ensemble predictions with the model used by Chen et al. (2004) and showed that the probabilistic and deterministic prediction skills were not always consistent in an ENSO ensemble forecast. These ensemble prediction data also provide an opportunity to explore the relationship between the deterministic prediction skills and potential predictability (Cheng et al. 2011). Zheng and Yu (2017) further addressed the deterministic prediction skill for the two flavors of ENSO from 1950 to 2012 with an intermediate model, and showed a better deterministic skill for the eastern Pacific ENSO than for the central Pacific ENSO.

However, the aforementioned conclusions were derived solely from long-term hindcasts using intermediate complexity or hybrid coupled models. Compared with these models, coupled general circulation models (CGCMs) include more realistic physical processes and have the potential to reproduce the ENSO complexity. Retrospective forecasts are also a crucial component of a CGCM prediction system, which can examine the ENSO predictability and its variability. However, hindcast studies based on CGCMs have focused on just the last 20–60 years (Table 1), and only cover relatively few ENSO cycles, which hinders a statistically robust understanding of ENSO predictability and its interdecadal variations from CGCMs. Long-term hindcasts, especially ensemble retrospective forecasts using a CGCM, have not yet been widely undertaken. Such retrospective forecasts might provide an opportunity to comprehensively investigate the ENSO predictability in CGCMs, and examine new findings and compare these with the results of intermediate complexity or hybrid models. Recently, Weisheimer et al. (2020) introduced 110-yr-long (1901–2010) ensemble hindcasts [Coupled Seasonal Forecasts of the Twentieth Century (CFSv2)].

### Table 1. Information of the current ENSO hindcast.

<table>
<thead>
<tr>
<th>Model</th>
<th>Model type</th>
<th>Ensemble members</th>
<th>Max lead (months)</th>
<th>Hindcast period</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>24</td>
<td>1856–2004</td>
<td>Cheng et al. (2010c)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>24</td>
<td>1856–2016</td>
<td>Liu et al. (2018)</td>
</tr>
<tr>
<td>ICM</td>
<td></td>
<td>100</td>
<td>12</td>
<td>1886–2006</td>
<td>Zheng et al. (2009a)</td>
</tr>
<tr>
<td>HCM1</td>
<td>Hybrid</td>
<td>1</td>
<td>12</td>
<td>1881–2000</td>
<td>Tang et al. (2008)</td>
</tr>
<tr>
<td>HCM1</td>
<td></td>
<td>15</td>
<td>12</td>
<td>1881–2000</td>
<td>Tang and Deng (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>8</td>
<td>1982–2010</td>
<td>Zhu et al. (2017)</td>
</tr>
<tr>
<td>CMIP5</td>
<td></td>
<td>28</td>
<td>12</td>
<td>1982–2010</td>
<td>Ding et al. (2019)</td>
</tr>
<tr>
<td>CanSIPSv2</td>
<td></td>
<td>20</td>
<td>12</td>
<td>1980–2010</td>
<td>Lin et al. (2020)</td>
</tr>
<tr>
<td>ECMWF’s IFS</td>
<td></td>
<td>51</td>
<td>7</td>
<td>1981–2016</td>
<td>Johnson et al. (2019)</td>
</tr>
<tr>
<td>ECMWF’s IFS</td>
<td></td>
<td>51/25</td>
<td>4</td>
<td>1901–2010</td>
<td>Weisheimer et al. (2020)</td>
</tr>
<tr>
<td>FIO-ESM</td>
<td></td>
<td>10</td>
<td>6</td>
<td>1993–2013</td>
<td>Qiao et al. (2013)</td>
</tr>
<tr>
<td>NMME</td>
<td></td>
<td>6–24</td>
<td>8–12</td>
<td>1982–2010</td>
<td>Merryfield et al. (2013); Vernieres et al. (2012); Saha et al. (2014); Vecchi et al. (2014); Becker et al. (2014); Kirtman et al. (2014); Barnston et al. (2019)</td>
</tr>
</tbody>
</table>
Century (CSF-20C), based on the European Centre for Medium-Range Weather Forecasts (ECMWF)'s Integrated Forecasting System coupled model version cycle 41r1. The ENSO prediction skill exhibited remarkable decadal variation in the boreal winter (December–February). However, the CSF-20C data only run for four forecast months. The prediction skill variability at longer forecast lead times is unclear. Moreover, the variation of the potential predictability in a CGCM is also unclear. Therefore, it is also necessary to investigate whether the prediction skill variability is model dependent and to evaluate the performance of the predictability in the CGCMs.

In this study, we used the Community Earth System Model (CESM) to perform an ensemble hindcast of 12-month forecasts from 1881 to 2017. This paper addresses the following issues: 1) the consistency of the variation of the predictability estimated from the CGCM long-term retrospective ensemble forecast with that of simpler models and 2) the sources that dominate the variation in predictability on various scales. These issues have not yet been investigated in detail by previous retrospective ensemble forecasts. The remainder of this paper is organized as follows. A detailed description of the model, perturbation method, and measurement metrics are presented in section 2. The deterministic skill and its variability are explored in section 3. In section 4, we investigate the causes of the variations of the predictability on various scales. Finally, we provide a summary and discussion of our findings in section 5.

2. Model and methodology

a. Model

We use CESM version 1.2, which is one of the most widely used fully coupled models with atmosphere, ocean, land, land-ice, and sea-ice components (Hurrell et al. 2013; Bellenger et al. 2014; Hu and Duan 2016; Bellomo et al. 2018; Hu et al. 2019; Xu et al. 2021). The atmospheric and oceanic components were the Community Atmosphere Model version 4 (CAM4; Neale et al. 2013), with a grid resolution of 0.9° × 1.25°, and the Los Alamos Parallel Ocean Program ocean model version 2 (POP2; Smith et al. 2010) with a Greenland pole and grid resolution of 1° × 1°. This configuration of CESM components can basically represent ENSO variability, including the amplitude, phase locking, and period (Yao et al. 2019). This model is also the operational model of the National Marine Environmental Forecasting Center in China (Li et al. 2015; S. W. Zhang et al. 2018; Zhang et al. 2019). We introduced a new nudging scheme to the CESM by adding wind data assimilation below 500 hPa and increasing the nudging weight of the subsurface ocean temperature at depths above 500 m (Song et al. 2021). Compared with the origin nudging scheme (Li et al. 2015), the new scheme improved the simulation and prediction skill for ENSO.

To initialize the long-term hindcast, the B20TR and B component sets were used before and after 1983 to formulate the CESM, respectively. The former represents 1850–2000 transient conditions and the latter represents the present-day situation. The model started with a 150-yr free run, and then we assimilated the upper ocean temperature and middle- to low-level wind reanalysis datasets from 1871 to 2017. The oceanic assimilation data were extracted from the monthly Simple Ocean Data Assimilation version 2.2.4 (SODA 2.2.4; Carton and Giese 2008) and the Global Ocean Data Assimilation System (GODAS; Behringer and Xue 2004) before and after 1983, respectively. The 6-hourly ERA-20C (Stickler et al. 2014) and ERA-Interim (Berrisford et al. 2011) were employed for the wind assimilation before and after 1983, respectively. The first 10 years are performed as a spinup and the ensemble retrospective forecasts started on 1 January, 1 April, 1 July, and 1 October of each year, and ran from 1881 to 2017 with 20 ensemble members. Each forecast continued for 12 months. The validation data for the sea surface temperatures (SSTs) were extracted from the SODA and GODAS datasets before and after 1983, respectively. Anomalies were identified by removing the climatology of the corresponding running 20-yr window both for the forecast and observational data as described by Deng and Tang (2009), which can exclude the effects of low-frequency variation.

b. Methodology

1) ENSEMBLE CONSTRUCTION STRATEGY

The pioneering work of Lorenz (1965) first introduced the singular vector (SV) method in studies of forecast uncertainty. Since the 1990s, the SV method has been employed to explore ENSO predictability and construct ENSO ensemble predictions with intermediate complexity and hybrid coupled models (Blumenthal 1991; Moore and Kleeman 1996; Chen et al. 1997; Xue et al. 1997a,b; Thompson 1998; Fan et al. 2000; Cai et al. 2003; Tang et al. 2006; Cheng et al. 2010a,b,c; Liu et al. 2018). The SV method aims to capture the rapid growth of initial errors, and has been widely used in studies of atmospheric and oceanic predictability. The traditional algorithm of the SV method requires the tangent linear and adjoint models of the original model, which is highly complicated. In addition, the traditional algorithm typically addresses weather-scale variability in fully coupled general circulation models (CGCMs), which has faster growth of errors than climate-relevant variability. As such, Kleeman et al. (2003) introduced the climatically relevant SV (CSV) method, whereby an ensemble approach is generated to replace the demand of tangent linear and adjoint models for calculating the SV. Also, weather-scale noise is filtered out by ensemble averaging and the climatically relevant rapid growth modes are remained. In this study, we employed the CSV method to obtain the optimum perturbation in the CESM. According to the CSV framework, we perturbed the sea temperatures at depths above 200 m to consider the effects of uncertainties in the subsurface ocean temperature on the prediction of SST. We used the leading three CSV modes to construct our ENSO ensemble forecast with 20 members. The first six ensemble members were obtained by perturbing the initial sea temperature conditions at depths above 200 m with positive and negative CSV1, CSV2, and CSV3,
which represent the orthogonality of the initial uncertainties. The last 14 ensemble members were obtained by perturbing the initial sea temperature conditions at depths above 200 m with a linear combination of a random number multiplied by one of the first three CSVs.

2) ACTUAL PREDICTION SKILL MEASURES

We employ the following metrics to measure the deterministic prediction skill: the anomaly correlation coefficient (ACC), root-mean-square error (RMSE), and correlation of individual forecasts (CIP). These are defined as follows:

$$\text{ACC}(i) = \frac{\sum_{i=1}^{N} [y^i(t) - \bar{y}(t)][y^i(t) - \bar{y}(t)]}{\sqrt{\sum_{i=1}^{N} [y^i(t) - \bar{y}(t)]^2} \sqrt{\sum_{i=1}^{N} [y^i(t) - \bar{y}(t)]^2}},$$  \hspace{1cm} (1)

$$\text{RMSE}(i) = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} [y^i(t) - \bar{y}(t)]^2},$$  \hspace{1cm} (2)

$$\text{CIP}(i) = \frac{\sum_{i=1}^{L} [y^i(t) - \bar{y}^i(t)][y^i(t) - \bar{y}^i(t)]}{\sqrt{\sum_{i=1}^{L} [y^i(t) - \bar{y}^i(t)]^2} \sqrt{\sum_{i=1}^{L} [y^i(t) - \bar{y}^i(t)]^2}},$$  \hspace{1cm} (3)

where $y$ is the variable of interest, and $p$ and $o$ are the ensemble mean prediction and observation, respectively; also, $i$ and $t$ indicate the initial time and lead time of the forecast, respectively. $N$ denotes the total number of initial conditions (137 years $\times$ 4 months = 552), and $L$ denotes the length of the forecast in months (12 months). The overbar denotes the mean of the total initial conditions in Eqs. (1) and (2), and the mean of all lead times in Eq. (3). The ACC and RMSE measure the overall deterministic prediction skill of all the initial conditions, and are a function of lead time. The CIP is the correlation skill of an individual prediction, and varies with each initial condition.

3) POTENTIAL PREDICTABILITY MEASURES

For a variable $y$, an ensemble prediction can provide additional information about the characteristics of forecast distribution, such as the prediction mean and ensemble spread. Relative entropy (RE) is an information-based measure to quantify this additional information with the difference between the climatological distribution and predicted distributions (Cover and Thomas 1991), defined as follows:

$$\text{RE}(i) = \int p(y|i) \ln \frac{p(y|i)}{p(y)} \, dy,$$  \hspace{1cm} (4)

where $p(y)$ and $p(y|i)$ are the climatological and predicted conditional distributions at the given initial condition $i$ of the variable of interest, respectively. When the two distributions are Gaussian, which is usually the case for most seasonal variables (including ENSO prediction), RE is given by

$$\text{RE}(i) = \frac{1}{2} \left[ \ln \left( \frac{\sigma^2_i}{\sigma^2_{yi}} \right) + \frac{\sigma^2_{yi}}{\sigma^2_o} - 1 + \frac{(\mu_{yi} - \mu_o)^2}{\sigma^2_o} \right],$$  \hspace{1cm} (5)

where $\sigma^2_i$, $\sigma^2_{yi}$, $\sigma^2_o$, and $\mu_o$ are the ensemble variance, ensemble mean, climatological variance, and climatological mean, respectively. It is derived from the dispersion component (DC; the first three terms) and signal component (SC; the last term). The DC indicates a reduction in climatological uncertainty by prediction, and the SC represents extra information between the prediction and the climatological means.

The RE is a measure of the potential predictability of an individual prediction case (Kleeman 2002), which indicates the upper limit of the prediction skill under the assumption of a perfect model. The overall predictability of a prediction system can be estimated from the averaged REs over all prediction cases. If the prediction cases are sufficient, then the averaged REs is quantitatively equal to the mutual information (MI), which is an information-based theoretical metric of the potential predictability of a prediction system. Traditionally, the perfect correlation ($R$), defined by the correlation between an arbitrary ensemble member and the ensemble mean, is also a metric that describes the potential predictability of a prediction system. In this study, we used alternatively these three metrics to analyze the ENSO potential predictability. It has been proven that the three potential predictability measures for a Gaussian variable can be given by (DelSole 2004; DelSole and Tippett 2007; Tang et al. 2013)

$$\text{MI} = \frac{1}{2} \left[ \ln \frac{\sigma^2_o}{\sigma^2_{yi}} - 1 \frac{1}{N} \sum_{i=1}^{N} \ln \sigma^2_{yi} \right] = \text{RE} = - \frac{1}{2} \ln (1 - R^2),$$  \hspace{1cm} (6)

where $\sigma^2_{yi}$ and $\sigma^2_o$ are the ensemble and climatological variances, respectively, and $N$ denotes the total number of initial conditions (552). The overbar denotes the time mean.

3. Actual prediction skill

In this study, we employed the widely used Niño-3.4 index (averaged SST anomalies from 5°S–5°N, 190°–240°E) to depict ENSO variability. Figure 1 presents the predicted Niño 3.4 index (blue and gray lines indicate the ensemble mean and ensemble spread, respectively) at a lead time of 6 months, and the corresponding observations (red line). The model can predict the most prominent El Niño and La Niña events well at a lead time of 6 months, especially for recent years. Figure 2 shows the ACC and RMSE between the ensemble mean Niño-3.4 index and the corresponding observations. For all lead times, the prediction skill (blue line) is superior to the persistent skill (green line) for both the ACC (Fig. 2a) and RMSE (Fig. 2b) from 1881 to 2017, and high ACC corresponds to low RMSE. The ACC skill decreases with lead time, with a value of 0.71 at a 6-month lead, and is still larger than 0.5 at a 10-month lead. During the common period
(1982–2010, the current ENSO hindcasts based on various CGCMs are shown in Table 1), the ACC skill is 0.83 at a 6-month lead and 0.76 at a 9-month lead (red line). Compared with current state-of-the-art seasonal hindcasts, the ACC skill of our ensemble hindcast is competitive, and even superior to the performance of the averaged level of all North American Multimodel Ensemble (NMME) hindcast (0.75 and 0.63 at 6- and 9-month leads, respectively) and International Research Institute (IRI) hindcast (0.71 and 0.56 at 6- and 9-month leads, respectively) (Barnston et al. 2019). The prediction skill is also comparable with the multimodel ensemble mean (MME) of the NMME, with skills of 0.83 and 0.70 at 6- and 9-month leads, respectively. Notably, the skill of the MME is better than that of the any individual model (Barnston et al. 2019). Therefore, the prediction skill of our ensemble hindcast is comparable with that of current operational centers, which also gives us confidence to explore the ENSO predictability based on our ensemble hindcast.

To evaluate the seasonal variation of the prediction skill, Fig. 3 shows the ACC skill of the Niño-3.4 index as a function of the target month initiated on 1 January, 1 April, 1 July, and 1 October. There is a pronounced seasonal dependence of the prediction skill, referred to as the spring predictability barrier (SPB; Webster and Yang 1992), in this long-term ensemble ENSO hindcast, with the prediction skill decreasing significantly throughout the spring season regardless of the start month. The decrease of the prediction skill initiated in January and October is more rapid than that when initiated in April and July. For example, for the hindcasts started in January, the effective predictive skill (ACC > 0.5) is 5 months. This is much shorter than that of the hindcast started in April, which has a less rapidly decreasing prediction skill and an effective predictive skill of up to 11 months.

Previous studies using different models have reported that ENSO prediction skills have undergone interdecadal variability (Chen et al. 2004; Tang et al. 2008; Deng and Tang 2009; Zheng et al. 2009a; Cheng et al. 2011; Tang and Deng 2011; Weisheimer et al. 2020; Huang et al. 2021). We now examine whether the prediction skill of the CESM ensemble forecast system also exhibits interdecadal variability. Figure 4 presents the ACC (Fig. 4a) and RMSE (Fig. 4b) for seven 20-yr sub-periods. Higher prediction skill occurs during 1961–2017 for both the ACC and RMSE, with the highest prediction skill occurring during 1981–2000, and relatively low prediction skill prior to this period. Figure 5 presents the running ACC with a running window of 20 years, averaged over 1–12 lead months for the CESM (blue line) and an intermediate complexity coupled model (red line; the Lamont-Doherty Earth Observation model LDEO5; Chen et al. 2004; Liu et al. 2018). The variations of the ACC skill are generally consistent in the two models with relatively higher prediction skill in 1881–1910 and 1960–2000, and lower prediction skill from 1920 to 1950. This result also agrees well with previous results based on hybrid coupled models (Tang et al. 2008; Deng and Tang 2009; Tang and Deng 2011) and a CGCM (Weisheimer et al. 2020). These interdecadal variations are also evident in the running correlation of individual forecasts (CIP; Fig. 6). Figures 4–6 indicate that the ENSO prediction skill also exhibit distinct interdecadal variations in the CESM. This interdecadal variation is independent of the evaluation metrics and model complexity. The prediction skill was higher during 1881–1910, and it declines after 2000 in both models. Apparently, this feature cannot be simply explained by data quality or the number of ocean observations, which are actually well improved in recent years. One possible explanation may be due to the changes in ENSO signal intensity (Chen et al. 2004; Tang et al. 2008; Kumar et al. 2015). The high skill
periods are usually associated with strong ENSO events (green line in Fig. 5), which provide additional information as compared with climatological predictions, which corresponds to a high predictability. This is further discussed in the next section. In addition, the prediction skill is higher in the CESM than the LDEO5 model after the 1950s. There are two possible reasons for this. First, the CESM assimilates the atmospheric and subsurface reanalysis, apart from SST, which are of higher quality after 1950, and can help to improve the prediction skill of the CESM. Prior to 1950, the data quality of the atmospheric and subsurface reanalysis appears to make the prediction skill of the CESM worse than the LDEO5 model, which only assimilates the SST reanalysis. Second, the improvement of the prediction skill after 1950 may reflect the advantage of the CGCM in representing the diversity of ENSO, as it incorporates most realistic physical processes. For example, the CESM model can basically represent the variability of the two types of ENSO (Yao et al. 2019), whereas the LDEO5 model can only depict the traditional ENSO variability. This further demonstrates that the long-term ensemble hindcast based on a CGCM is an essential tool for investigating the predictability of ENSO.

4. Variation of predictability

As mentioned above, the prediction skill for ENSO in the CESM ensemble system exhibits striking seasonal and interdecadal variations. It is interesting to examine why this is the case. Previous research has reported that RE is a potential indicator of the actual prediction performance. A larger RE usually corresponds to good prediction, and vice versa (Tang et al. 2005; Cheng et al. 2011). Based on Eq. (5) in section 2, the ENSO predictability can be derived from the contribution of the DC and SC. We now investigate the relative importance of these two components on different time scales. Figure 7 presents the RE, DC, and SC for individual predictions as a function of lead month from 1881 to 2017. In general, the RE, SC, and DC all have significant interannual variations and decrease with lead time. Specifically, the SC is obviously related to the strength of ENSO events. A large SC typically corresponds to strong ENSO events, such as the 1982/83, 1997/98, and 2015/16 strong El Niño events, and 1973/74, 1988/89, 1999/00, and 2010/11 strong La Niña events. In contrast, the DC variation is smoother than the SC for different initial conditions and decreases abruptly above 4- or 5-month lead times. In general, the RE reflects the combined variation of the DC and SC, and follows the SC during ENSO events, especially for long lead times.

Figure 8 further presents scatterplots of SC and DC versus RE at lead times of 1–3 and 4–12 months, respectively. At shorter lead times, the DC has comparable or even a more important contribution to the RE than the SC, with correlation coefficients of 0.82 and 0.57, respectively. However, at
longer lead times, the RE is more significantly correlated with the SC, with a correlation coefficients of 0.9, whereas the correlation coefficient between the RE and DC is 0.56. As such, the DC and SC both make contributions to the interannual variation of the ENSO predictability in the CESM, while the DC is only significant at shorter lead times and in years with a weak ENSO signal years. In addition, the linear relationship between the SC and RE is generally consistent with, but weaker than, the results based on hybrid models during 1981–98 (Tang et al. 2005). This may reflect the different model complexity and hindcast period of the two models. Therefore, it is necessary to perform long-term hindcasts with variable complexity models to robustly understand ENSO predictability.

We further examined the seasonal variation of the R and corresponding DC and SC (Fig. 9). Similar to the ACC, the potential predictability (R) of ENSO also encounters the SPB, with a marked degradation in skill across spring, especially for predictions initiated in January and October (Fig. 9a). The SC also exhibits significant seasonal variations, and its strength rapidly decreases during spring. In contrast, the decay rate of the DC exhibits little difference among the seasons. It can be concluded that the seasonal variation of ENSO predictability is dominated by the SC, and the abrupt decrease of signal strength results in the SPB. Finally, we investigated the interdecadal variation of the potential predictability (Fig. 10). The RE also has interdecadal variations with relative lower predictability during 1920–50 and higher predictability during 1881–1910 and 1960–2000. This behavior is consistent with the interdecadal actual prediction skill (Figs. 5 and 6), and is mainly dominated by the SC, which has significant interdecadal variation with a relatively small signal during 1920–50 and large signals in other decades (Fig. 5, green line). However, the DC value is larger than the SC in all decades, but it has weak interdecadal variations. This indicates that the DC makes a greater contribution to the magnitude of ENSO predictability, while the variation of SC dominates its variability on an interdecadal time scale.

5. Summary and discussion

ENSO is one of the most predictable climate modes, and provides an important source of predictability for global seasonal climate predictions due to its worldwide climatic effects. Long-term ensemble retrospective forecasts are a useful strategy to comprehensively estimate ENSO predictability from
the perspective of both the actual prediction skill and potential predictability. In this study, we have employed the CSV method in the CESM to construct an ensemble forecast of ENSO from 1881 to 2017 in order to investigate ENSO predictability and the possible factors that control its variations. In general, the ensemble forecast has a comparable performance to other operational centers during the common period (1982–2010). It can provide a skillful deterministic prediction at a lead time of 10 months and capture the most prominent El Niño and La Niña events over the past 137 years. The actual deterministic prediction skill suffers from the SPB, with relatively high prediction skill for forecasts started in April and low prediction skill for forecasts started in January. In addition, the deterministic prediction skills exhibit a consistent interdecadal variability, with relatively higher prediction skill during 1881–1910 and 1960–2000. The highest skill occurs in 1981–2000, while relatively lower prediction skill occurs in 1920–50. This interdecadal variability of the deterministic prediction skills is scarcely model dependent and generally consistent with that obtained from intermediate
complexity coupled models (Chen et al. 2004), hybrid coupled models (Tang et al. 2008; Deng and Tang 2009; Tang and Deng 2011), and a CGCM (Weisheimer et al. 2020).

To our knowledge, this is the first study to examine ENSO predictability and its controlling factors on different time scales, based on a long-term ensemble hindcast performed with a complex CGCM. The DC and SC both make contributions to the interannual variability of ENSO predictability, while the SC is more important during ENSO events and the DC is of equal importance at short lead times and weak ENSO signal years. The seasonal variability of the ENSO predictability is controlled by the SC, and the marked degradation of its strength results in the SPB feature of the ENSO. At an interdecadal scale, the DC determines the magnitude of ENSO predictability, while the SC dominates its variability. This finding is somewhat different with the results derived from intermediate complexity (Cheng et al. 2011) and hybrid models (Tang et al. 2005), which have shown that the SC is more important to the variation of ENSO predictability in these models. This may be because the CGCM contains more

Fig. 8. Plots of the SC and DC vs RE at (a),(b) 1–3- and (c),(d) 4–12-month leads for the ensemble Niño-3.4 index over the past 137 years.
realistic high-frequency “noise” processes (e.g., westerly wind bursts and the MJO) and external forcing mechanisms (e.g., aerosol and volcanic eruption), while the intermediate complexity and hybrid coupled models mainly represent the ENSO signal dynamic process. Therefore, it is unsurprising that the SC has a greater contribution to the predictability in the intermediate complexity and hybrid coupled models.

These high-frequency noise and external forcing mechanisms can also provide additional information on the ENSO predictability and improve the prediction skill (Hirono 1988; Perigaud and Cassou 2000; Vitart et al. 2003; Church et al. 2005; Mann et al. 2005; Gleckler et al. 2006; Gebbie et al. 2007; Tan et al. 2009, 2020; McGregor and Timmermann 2011; Lau 2012; Ohba et al. 2013; Zhang et al. 2013; Lian et al. 2014;
Chen et al. 2015; Maher et al. 2015; McPhaden et al. 2015; Thual et al. 2016; Lian et al. 2018) of CGCMs. Its contribution is contained in the DC, thereby increasing the importance of the DC in ENSO predictability in a CGCM. It should also be noted that ENSO predictability is affected by both uncertainties in the initial conditions and model biases (Zheng et al. 2009b; Kang et al. 2017; Gao et al. 2018; Tao et al. 2018; R. H. Zhang et al. 2018), and only the uncertainties in the initial conditions were considered in this study. Therefore, future studies should explore the possible influence of model biases, or the combined effects of uncertainties in the initial conditions and model biases on ENSO predictability. In addition, some studies have noted that the actual skill is higher than the perfect skill (Mehta et al. 2000; Kumar et al. 2014; Jin et al. 2018). A possible reason is that the ensemble prediction system underestimated the observational SST persistence (Jin et al. 2001).

Fig. 10. Running (a) RE, (b) SC, and (c) DC for the ensemble Niño-3.4 index with a 20-yr running window.
et al. 2018). This phenomenon does not characterize our ensemble forecast of the Niño-3.4 region, where the CESM shows a reasonable performance in capturing the observational persistence. It would be interesting to explore the performance of the CESM in other regions in future studies.

Acknowledgments. This study was jointly supported by the National Natural Science Foundation of China (Grant 41690124), the Scientific Research Fund of the Second Institute of Oceanography, Ministry of Natural Resources (Grant QNYC2101), the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (Grant SML2021SP310), and the National Key Research and Development Program (2017YFA0604202) and Predictable Pattern of Precipitation over Asian Summer Monsoon Regions (41690120).

Data availability statement. The SODA 2.2.4 data were provided by the SODA group at http://apdrc.soest.hawaii.edu/. The GODAS data were provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, at https://psl.noaa.gov/data/gridded/data.godas.html. The ERA-20C and ERA-interim datasets were provided by the ECMWF at https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets.

REFERENCES


