Improve Climate Predictions by Reducing Initial Prediction Errors: A Benefit Estimate Using Multi-model ENSO Predictions

Gan Zhang[1]

[1]University of Illinois at Urbana Champaign

September 13, 2023

Abstract

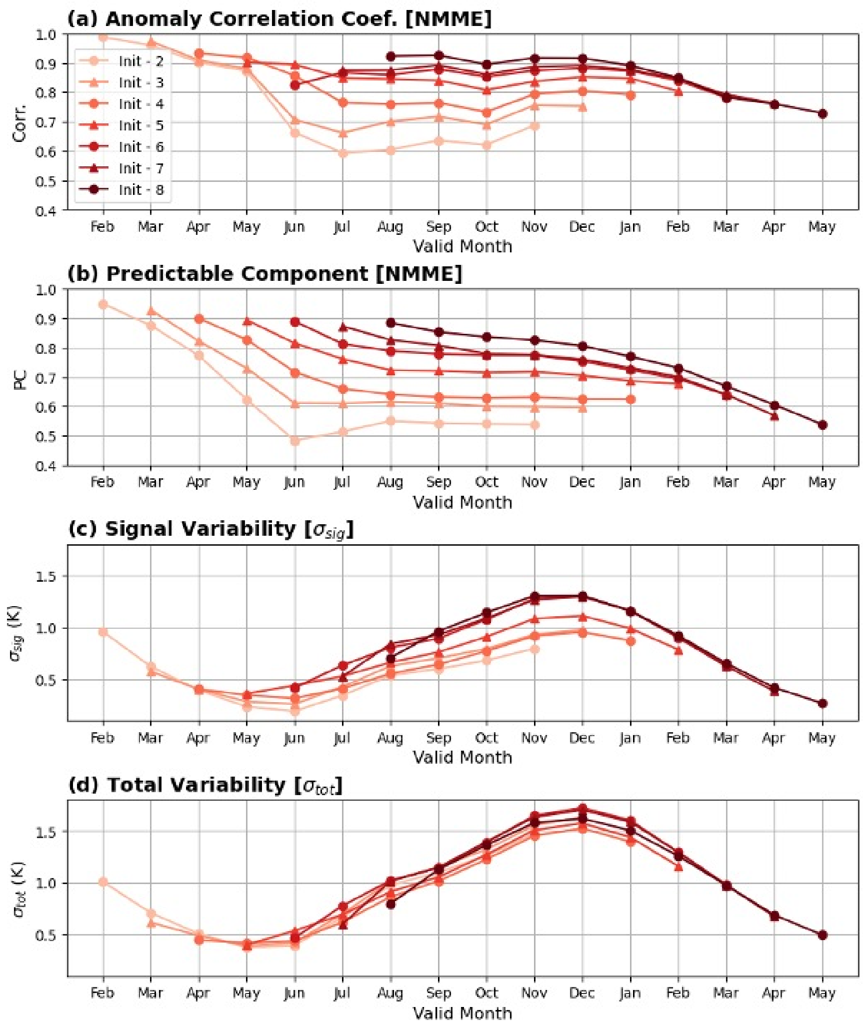
Climate risk management relies on accurate predictions of key climate variations such as El Niño-Southern Oscillation (ENSO), but the skill of ENSO predictions has recently plateaued or even degraded. Here we analyze the North American Multi-Model Ensemble (NMME) and estimate how the seasonal prediction of ENSO may benefit from reducing initial prediction errors. An analysis of predictable signals and system noises identifies a high-predictability regime and a low-predictability regime. The latter corresponds to the spring predictability barrier and is related to a rapid drop in the signal-to-noise ratio, which is caused by the comparably strong dampening of predictable signals. Reducing first-month prediction errors (FPEs) will likely reduce root-mean-square errors of the ENSO prediction. As a conservative estimate, halving the FPEs may extend the NMME's skill by one to two months. Importantly, this study identifies the regions where reducing FPE is the most effective. Unlike the predictions initialized after the boreal spring, the March-initialized predictions of the wintertime ENSO will likely benefit the most from FPE reductions in the tropical Northwest Pacific. An opportunistic thought experiment suggests the buoy observation changes during 1995–2020 may have contributed to FPEs associated with large cold biases (>1K) in some El Niño-year predictions. While data availability prevented in-depth analyses of physical processes, the findings suggest that prioritizing modeling and observation in certain regions can improve climate predictions cost-effectively. The analytical framework here is applicable to other climate processes, thus holding wide potential for benefiting climate predictions.

Hosted file

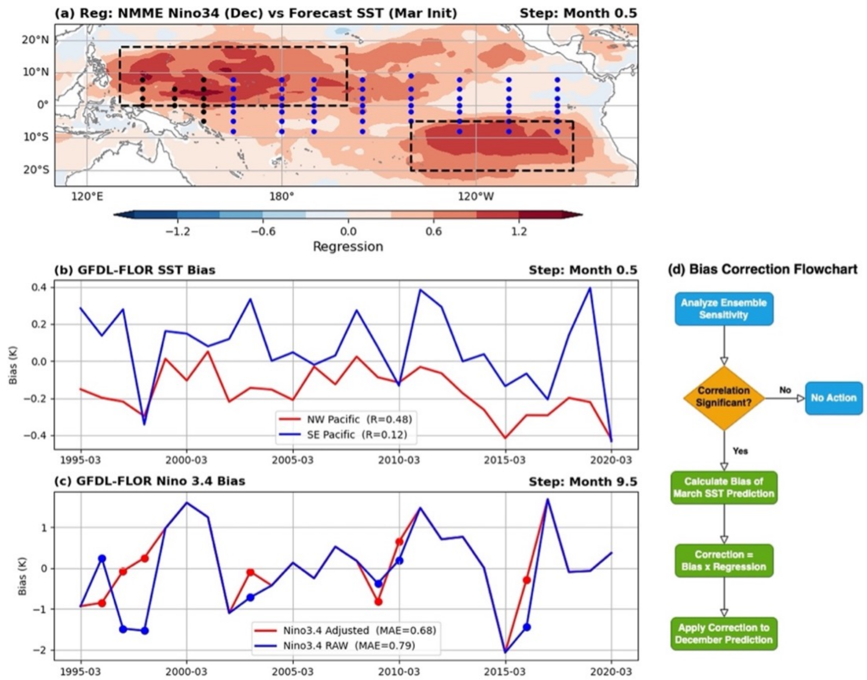
973609_0_art_file_11367316_s0qt3m.docx available at https://authorea.com/users/664189/articles/666017-improve-climate-predictions-by-reducing-initial-prediction-errors-a-benefit-estimate-using-multi-model-enso-predictions

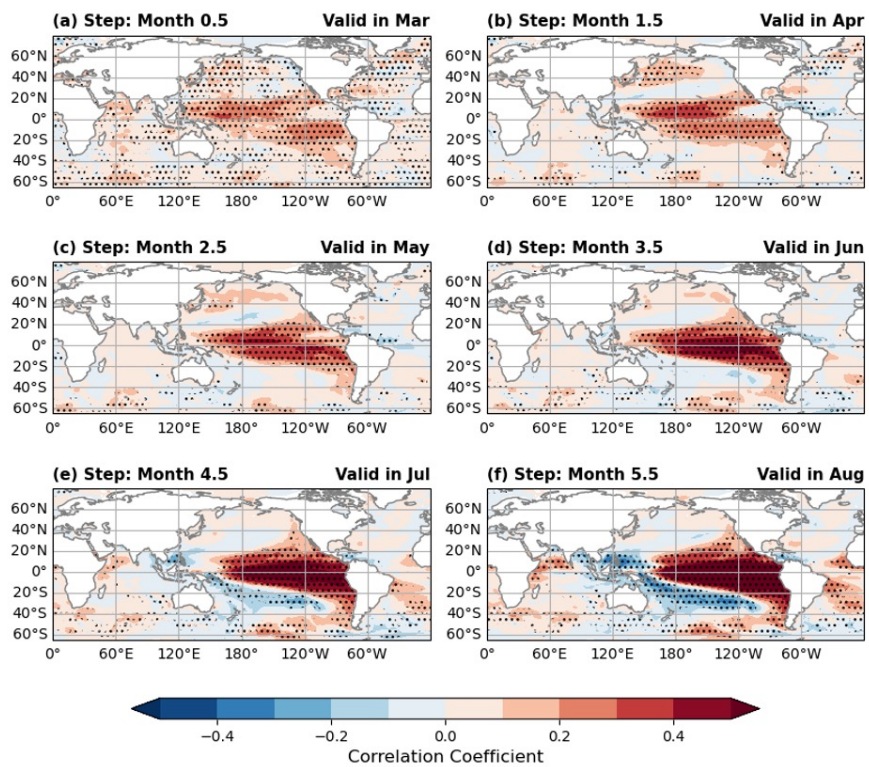
Hosted file

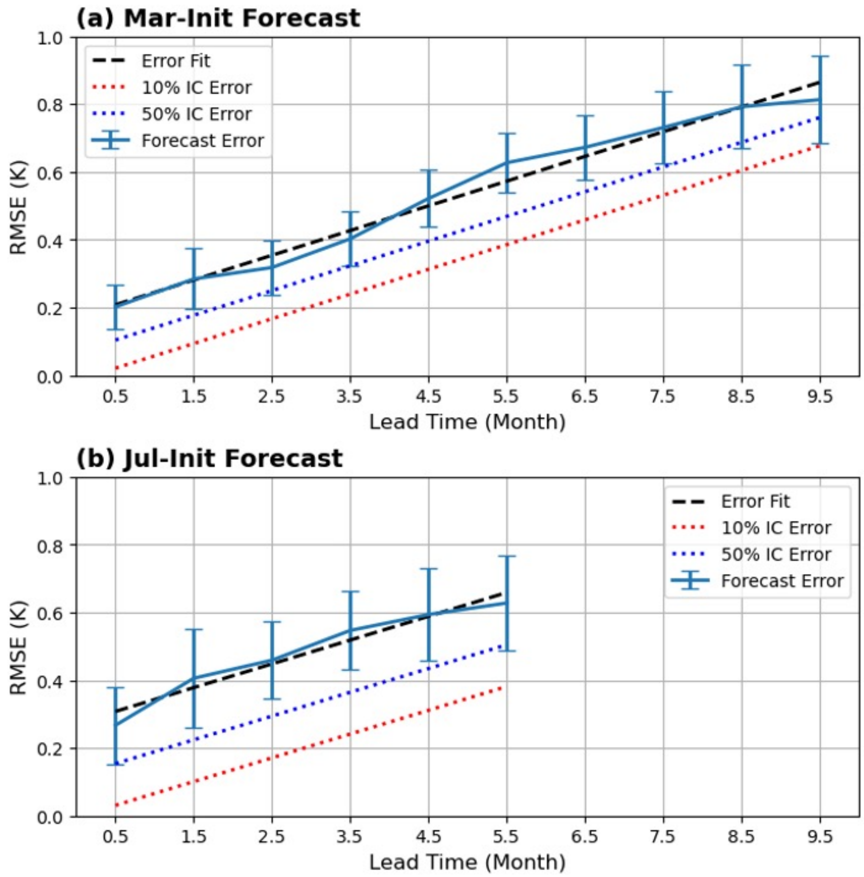
973609_0_supp_11367317_s0qt3m.docx available at https://authorea.com/users/664189/articles/666017-improve-climate-predictions-by-reducing-initial-prediction-errors-a-benefit-estimate-using-multi-model-enso-predictions



(a) Anomaly Correlation Coef. [NMME]
Corr.
Init - 2
Init - 3
Init - 4
Init - 5
Init - 6
Init - 7
Init - 8
Valid Month
(b) Predictable Component [NMME]
PC
Valid Month
(c) Signal Variability [σ_{sig}]
σ_{sig} (K)
Valid Month
(d) Total Variability [σ_{tot}]
σ_{tot} (K)
Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May
Valid Month


(a) Reg: NMME Nino34 (Dec) vs Forecast SST (Mar Init)
Step: Month 0.5
Regression
(b) GFDL-FLOR SST Bias
Step: Month 0.5
Bias (K)
NW Pacific (R=0.48)
SE Pacific (R=0.12)
(c) GFDL-FLOR Nino 3.4 Bias
Step: Month 9.5
Nino3.4 Adjusted (MAE=0.68)
Nino3.4 RAW (MAE=0.79)
(d) Bias Correction Flowchart
Analyze Ensemble Sensitivity
Correlation Significant?
No
No Action
Yes
Calculate Bias of March SST Prediction
Correction = Bias x Regression
Apply Correction to December Prediction


(a) Step: Month 0.5
Valid in Mar
(b) Step: Month 1.5
Valid in Apr
(c) Step: Month 2.5
Valid in May
(d) Step: Month 3.5
Valid in Jun
(e) Step: Month 4.5
Valid in Jul
(f) Step: Month 5.5
Valid in Aug
Correlation Coefficient


(a) Mar-Init Forecast
Error Fit
10% IC Error
50% IC Error
Forecast Error
RMSE (K)
Lead Time (Month)
(b) Jul-Init Forecast
Error Fit
10% IC Error
50% IC Error
Forecast Error
RMSE (K)
Lead Time (Month)

Improve Climate Predictions by Reducing Initial Prediction Errors: A Benefit Estimate Using Multi-model ENSO Predictions

Gan Zhang

Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign

1301 W Green Street, Urbana, IL 61801, United States

Corresponding Author:

Gan Zhang (gzhang13@illinois.edu)

Key Points

- The ratio of predictable signals and system noise explains the spring predictability barrier of ENSO and reveals biases in climate models.
- Reducing initial prediction errors will likely extend the ENSO prediction skill of a multi-model ensemble system by at least 1-2 months.
- Observations in the tropical Northwest Pacific may mitigate the predictability barrier and reduce prediction errors cost-effectively.

Abstract

Climate risk management relies on accurate predictions of key climate variations such as El Niño-Southern Oscillation (ENSO), but the skill of ENSO predictions has recently plateaued or even degraded. Here we analyze the North American Multi-Model Ensemble (NMME) and estimate how the seasonal prediction of ENSO may benefit from reducing initial prediction errors. An analysis of predictable signals and system noises identifies a high-predictability regime and a low-predictability regime. The latter corresponds to the spring predictability barrier and is related to a rapid drop in the signal-to-noise ratio, which is caused by the comparably strong dampening of predictable signals. Reducing first-month prediction errors (FPEs) will likely reduce root-mean-square errors of the ENSO prediction. As a conservative estimate, halving the FPEs may extend the NMME's skill by one to two months. Importantly, this study identifies the regions where reducing FPE is the most effective. Unlike the predictions initialized after the boreal spring, the March-initialized predictions of the wintertime ENSO will likely benefit the most from FPE reductions in the tropical Northwest Pacific. An opportunistic thought experiment suggests the buoy observation changes during 1995–2020 may have contributed to FPEs associated with large cold biases (>1K) in some El Niño-year predictions. While data availability prevented in-depth analyses of physical processes, the findings suggest that prioritizing modeling and observation in certain regions can improve climate predictions cost-effectively. The analytical framework here is applicable to other climate processes, thus holding wide potential for benefiting climate predictions.

Plain Language Summary

To manage climate risks effectively, accurate predictions of events like El Niño-Southern Oscillation (ENSO) are crucial. However, ENSO predictions have plateaued or worsened recently. Our study examined a collection of climate models called the North American Multi-Model Ensemble to find ways to improve seasonal ENSO predictions. Our analysis explores why the predictions are less skillful when they are made in the springtime of the northern hemisphere. The low skill is attributed to the relatively high sensitivity of the climate system to small perturbations in the input that help start model simulations ("butterfly effect"). The small perturbations can be reduced via science investments such as strengthening the observation of oceanic conditions. Following Dr. Edward Lorenz and others, we estimate such reductions may extend skillful ENSO predictions by at least 1-2 months. Crucially, our research pinpoints the specific regions where better observations may be particularly effective in reducing ENSO prediction errors. One of the regions is the Pacific waters near Guam, where budget-related changes in our observation capability allow a simple test of our idea. Despite some limitations, our findings suggest that modeling and observation improvements in particular regions could enhance climate predictions and benefit society in a cost-effective manner.

Keywords

Predictability, Climate, El Niño-Southern Oscillation, Ensemble Prediction, Signal-to-noise Ratio, Observation

1. Introduction

Thanks to continuous improvements in global climate models (GCMs) and observational networks, skillful predictions of climate anomalies on the seasonal to interannual scales have become increasingly accessible (e.g., Palmer & Anderson, 1994; Stern & Miyakoda, 1995; Shukla, 1998; Becker et al., 2022). Besides delivering predictions on their own, GCM simulations also helped machine learning and other methods push the envelope of predicting climate extremes (e.g., Murakami et al., 2016) and anomalies (e.g., Ham et al., 2019). Over the recent two decades, climate predictions have delivered significant societal benefits (Palmer et al., 2005; Becker et al., 2022) while co-benefiting the projection of anthropogenic climate change (e.g., Jain et al., 2023). As the need for climate risk management increases, improving predictions of societally important climate variations warrants continuous commitment. Yet surprisingly, the recent progress in theoretical understanding and model development has not led to skill gains in predicting the El Niño–Southern Oscillation (ENSO) (Chen & Cane, 2008; Barnston et al., 2012; Becker et al., 2020).

The skillfulness of modern climate predictions partly arises from the multi-model ensemble configuration. These ensembles often deliver superior skills in comparison to individual models (e.g., Doblas-Reyes et al., 2000; Palmer et al., 2004; Hagedorn et al., 2005; Weisheimer et al., 2009; Kirtman et al., 2014). Bundling multiple models affords an opportunity for diverse model representations of physical processes to mitigate the biases of individual models (Hagedorn et al., 2005). Such bundling also increases the ensemble size and helps sample the probability space, ultimately making predictions more reliable (e.g., Hagedorn et al., 2005; Tompkins et al., 2017). These findings motivated the implementation of operational multi-model prediction systems, including the European multi-model systems (Palmer et al., 2004; Weisheimer et al., 2009; Buontempo et al., 2022) and the North American Multi-Model Ensemble (NMME) (Kirtman et al., 2014; Becker et al., 2022). As summarized by Becker et al. (2022), numerous studies also explored how to post-process multi-model ensemble predictions to achieve better skill. Despite encouraging progress elsewhere, the skill of multi-model ensembles (at least the NMME) in the ENSO prediction has plateaued or even degraded in the past decade (Barnston et al., 2012; Becker et al., 2020).

On the seasonal to interannual scales, many societally valuable climate predictions are initial-value problems of predicting slowly varying climate processes (Palmer & Anderson, 1994). As suggested by the development of the weather forecasting (Lorenz, 1982; Bauer et al., 2015), an effective way to improve the skill of initial-value predictions is through high-quality initial conditions (ICs). Better ICs reduce initial prediction errors (IPEs) (Lorenz, 1982) and can be attained via improvements in observational networks, data assimilation, and model initialization. Nonetheless, implementing this strategy for climate predictions has proven more difficult. While weather forecasts are supported by atmospheric ICs from a vast network of observations (e.g., remote sensing), climate predictions heavily rely on oceanic ICs from sparse in-situ observations, especially in cloudy regions or beneath the ocean surface. Maintaining networks of in-situ oceanic observations can be challenging. For example, the in-situ buoy observations of the Tropical Pacific Observation System experienced a crisis and a subsequent decline in the 2010s (Ando et al., 2017; Fujii et al., 2015). Meanwhile, proper data assimilation and model initialization are complex and computationally intensive for the coupled climate system (Palmer & Zanna, 2013). The development of relevant model components often requires substantial effort (e.g., Goddard et al., 2001; Zhang et al., 2007; Lu et al., 2020) but does not always lead to skill gains (e.g., Chen & Cane, 2008). To effectively allocate resources, it is conducive to establish a priori knowledge about the potential skill gains from research activities, such as the circumstances where climate predictions may benefit from additional oceanic observations.

Motivated by these practical needs, this study explores the sensitivity of seasonal climate prediction to IPEs with a focus on the ENSO. Because of its global impacts, the ENSO has long been at the center of climate research and service (Latif et al., 1998; McPhaden et al., 2006). While modern models can deliver skillful predictions, the predictions initialized around the boreal spring often show lower prediction skills (Latif et al., 1994; Barnston et al., 2019; Tippett et al., 2019). Known as the spring predictability barrier, this low-skill regime received persistent research interest. Zebiak and Cane (1987) and Battisti (1988) first linked the predictability barrier to the weak instability growth rate of the equatorial Pacific during the springtime and argued that it reduces the memory of the coupled climate system. Using a conceptual recharge oscillator model, Levine and McPhaden (2015) showed that the annual cycle in the ENSO growth rate contributes to a spring predictability barrier. Other low-dimensional models also provided valuable insights into the ENSO predictability and its barrier (e.g., Newman & Sardeshmukh, 2017; Liu et al., 2019; Tippett & L'Heureux, 2020).

Another line of the spring predictability barrier research developed around error growth and stochastic forcings. Webster and Yang (1992) hypothesized that the predictability barrier is related to the faster springtime error growth in coupled forecasts and the interference from the Asian summer monsoon. Torrence and Webster (1998) suggested that the springtime climate system has a low signal-to-noise ratio (SNR) and is thus most susceptible to perturbations. The hypotheses were partly supported by GCM experiments. For example, Larson and Kirtman (2015) showed that the IPE growth is the largest in boreal spring, though the error growth rate in boreal summer appears comparable. Additionally, March was identified as a window of opportunity for stochastic perturbations of zonal wind to impose a long-lasting impact on the eastern Pacific (Larson & Kirtman, 2017). Intriguingly, the spring predictability barrier is less evident in models that emphasize the initialization procedure, have lower levels of system noise (Chen et al., 1995, 2004), or parameterize state-dependent atmospheric perturbations (Lopez & Kirtman, 2014).

To the best of our knowledge, recent climate studies paid little attention to the sensitivity of multi-model ensemble predictions to IPEs. Inspired by findings analyzing a prediction system that contributes to the NMME (G. Zhang et al., 2021), we posit that multi-model ensembles may offer insights into how reducing IPEs may improve predictions of climate anomalies such as the ENSO. Compared to a single-GCM study, multi-model ensembles—given their high prediction skills and large ensemble size—likely represent the climate system more reliably. This helps increase confidence in using the "perfect model" assumption (Lorenz, 1982) and generalizing conclusions from predictability analyses. Additionally, GCMs are often bounded by priori that differ from those in low-dimensional ENSO models. Therefore, multi-model ensembles may help validate past studies as independent sources or reveal novel findings. Accordingly, we address three research questions revolving around the ENSO prediction:

1. How may multi-model ensembles help understand the ENSO predictability, especially its spring barrier?
2. To what extent may reducing IPEs improve the skills of ENSO prediction?
3. What are the key oceanic observations that may drive such improvements?

Answering these questions will likely help with future development in climate predictions, especially when paired with strategies and tools that have been proven effective in improving weather forecasting.

2. Data and Methods

2.1 Prediction Data and Pre-processing

To study the spring barrier for predicting the wintertime ENSO, we look for multi-model ensembles that include predictions initialized in February–April and have a forecast lead of at least nine months. Additionally, we prioritize predictions with hindcasts for 1981–2010 over 1991–2020. The purpose is to account for more models (as in 2023), maximize the ensemble size, and bolster analyses related to the signal-to-noise ratio (SNR). These constraints exclude the European multi-model ensemble and several models in the NMME. The seasonal prediction systems used here are NCEP-CFSv2, NCAR-CESM1, GFDL-FLOR, COLA-RSMAS-CCSM4, and CanCM4i archived by IRI/LDEO (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/). The ensemble size of this group is seventy-eight, much larger than that of any individual prediction system.

These five prediction systems use various ICs, with the exception that the NCEP-CFSv2 and COLA-RSMAS-CCSM4 share the ICs from the same reanalysis system. The NMME models use data assimilation procedures to various degrees, and simulations forced by boundary forcings are often used to generate ICs of certain model components (e.g., atmosphere). The sometimes ad-hoc approaches, together with the limited access to the actual IC data, made it challenging to directly analyze the impacts of ICs. Accordingly, we follow the approach of Lorenz (1982) and instead emphasize IPEs. These errors are closely associated with imperfections in the model initialization, including the observation uncertainty, data assimilation, and shocks in the spin-up of coupled GCMs. Interested readers can find more about the model initialization and other details in the references in Table 1.

This study focuses on analyzing the sea surface temperature (SST) since most of the public NMME data has only monthly aggregations of limited surface variables. This issue makes it challenging to conduct in-depth analyses of physical processes involving the subsurface ocean or high-frequency atmospheric perturbations. Due to model drifts, all the prediction systems of the NMME have their own model climatology that also depends on the initialization time. To calculate the SST anomalies, we remove the 30-year hindcast climatology independently for each model and each initialization time. The validation of SST predictions uses the monthly Optimum Interpolation SST (OISST) dataset (Reynolds et al., 2002). For brevity, the discussion of the ENSO prediction will be framed around the Niño 3.4 index, which is the area mean of SST anomalies in the equatorial Pacific (5°N-5°S, 170°W-120°W).

2.2 Predictability Analysis

Following the common practice of predictability analyses (e.g., Lorenz, 1982), we adopt the perfect model assumption. Specifically, we assume that the NMME ensembles realistically replicate the climate system and that the findings from studying the model systems are applicable to the real-world climate. While this assumption is likely reasonable given the NMME's skill in predicting the ENSO, the discussion will actively consider alternative possibilities, model biases, and their implications. With the perfect model assumption and a large ensemble, we consider the ensemble mean as the predictable signal. Accordingly, the error growth in predictions can be estimated using the root-mean-square error (RMSE) around the ensemble mean. With additional assumptions about the error growth, we estimate the potential impacts of reducing first-month prediction errors (FPEs) on the skill of ENSO prediction.

Our ENSO predictability analysis is also inspired by past SNR studies (e.g., Shukla, 1998; Torrence & Webster, 1998). To link the SNR to the skill metric, anomaly correlation coefficient (ACC), we follow Eade et al. (2014) and use a variation of SNR. Specifically, we define the strength of predictable signals as the temporal standard deviation of the ensemble mean (σ_{sig}). Similarly, the temporal standard deviations of individual ensemble members can be calculated. Their average is considered as an indicator of the total variability of the noisy model system (σ_{tot}). The $\sigma_{sig}/\sigma_{tot}$ is closely associated with the SNR and defined as the predictable component in the model hindcasts (PC) (Eade et al., 2014). For a perfectly reliable prediction system, its PC value is expected to be comparable to the ACC of the ensemble mean and the observation. Otherwise, $ACC/PC < 1$ indicates an overconfident, under-dispersive prediction system, whereas $ACC/PC > 1$ indicates an underconfident, over-dispersive prediction system.

To map the sensitivity of ENSO predictions to FPEs, we use the ensemble sensitivity analysis (Ancell & Hakim, 2007; Hakim & Torn, 2008; Torn & Hakim, 2008). The ensemble sensitivity is formally defined as the linear regression between a forecast response function and the ICs (Ancell & Hakim, 2007). Mathematically, the ensemble sensitivity is closely linked to the ensemble transform Kalman filter and the adjoint sensitivity analysis, yet the calculation is more straightforward and computationally inexpensive (Ancell & Hakim, 2007). Another advantage of this ensemble sensitivity analysis is its compatibility with statistical significance tests, which help deal with sampling errors. This technique was first developed to explore the relationship between the IC state variables and a 24-hour forecast of an extratropical cyclone and has later found wide applications in atmospheric predictability problems (see Ancell & Coleman, 2022 and references therein). But to the best of our knowledge, this technique has not been applied to climate predictions. Following Torn and Hakim (2008) and other studies, we will use the ensemble sensitivity analysis to evaluate the error growth and estimate the impact of missing observations in an opportunistic thought experiment (more details in Section 3.3).

3. Results

3.1 Predictability Analysis

We first examine the ACC (Figure 1a) and the PC (Figure 1b) of the predictions initialized between February and August. Based on the evolution of ACC and PC, the predictions can be categorized into two regimes: (I) initial ACC and PC values are relatively high but drop rapidly, which correspond to the predictions initialized in February-April, and (II) initial values are relatively low but decrease slowly, which correspond to the predictions initialized in June-August. With few exceptions, the ACC skill of the NMME is lower than the PC, suggesting the NMME is generally underconfident and over-dispersive. In other words, the error growth in the simulated ENSO may be faster than in the real world. Meanwhile, the rapid decay of PC in the springtime is consistent with the concurrent ACC drop. Therefore, the results from modern GCMs and simple statistical models (Torrence and Webster 1998) are qualitatively consistent, suggesting that the relatively low SNR in the springtime contributes to the ENSO predictability barrier.

To better understand the spring predictability barrier in GCMs, we analyze the σ_{sig} and σ_{tot} components of the PC. Figure 1c-d shows that Niño 3.4 anomalies decay rapidly in the springtime before growing in the summertime, consistent with the annual cycle of the ENSO. For predictions initialized in the springtime (Regime I), the annual cycle dampens SST anomalies immediately after model initialization. The dampening of predictable signals is relatively fast and persistent, contributing to the rapid PC drop. Taking the February-initialized prediction as an example, the signal strength decreases by about 0.8K (80%) over four months, while the total variability decreases by about 0.6 K (60%). This preferred dampening of predictable signals contributes to a rapid loss of predictability in the springtime. In comparison, the signal and noise components experience strong growth in the predictions initialized in the summertime (Regime II), with the signal dominating the overall variability till the seasonal peak of ENSO. These characteristics are consistent with the persistence of high prediction skills. Overall, the season-dependent evolution of the PC is consistent with the predictability barrier studies that emphasize the annual cycle of ENSO growth (Zebiak & Cane, 1987; Battisti, 1988; Levine & McPhaden, 2015; Liu et al., 2019).

Figure 1c-d also suggest intriguing inconsistencies of the model systems initialized in the boreal spring and summer. A close inspection of Figure 1c suggests the timing of the dampening-growth transition depends on the initialization month. Specifically, this transition occurs in June for the predictions initialized in February-April, but the May-initialized prediction suggests the transition occurs in May or even earlier. After this transition, the growth rate of Niño 3.4 anomalies in identical calendar months also shows a spurious dependence on the initialization month (Figure 1c). For example, the summer-initialized predictions suggest the signal growth during July-December is about 0.16 K month^{-1}. The value is notably higher than the July-December growth rate (0.12 K month^{-1}) indicated by the spring-initialized predictions. For a system that faithfully simulates the real-world system, such differences should be minimal and show little dependence on the initialization time. These inconsistent growth rates suggest GCM biases in simulating the stochastic forcing and signal growth of the ENSO system. Such biases likely interfere with the spring predictability barrier manifested in the NMME predictions, as predictability barriers are sensitive to the dampening-growth transition (Liu et al., 2019) and the signal growth rate (Levine & McPhaden, 2015; Liu et al., 2019).

Interestingly, some characteristics of NMME differ greatly from individual contributing models. While the NMME is generally over-dispersive and underconfident in the ENSO predictions, its members such as GFDL-FLOR and NCEP-CFSv2 are under-dispersive and overconfident (Figures S1 and S2). The two models also show large differences in the growth rate of predictable signals and system noises. While such differences are interesting, we will refrain from an extensive discussion of individual models. Overall, the ENSO prediction by individual models has room for improvement and will likely benefit from a well-calibrated representation of the signal growth and system noise.

We next follow Lorenz (1982) and turn to the IPEs to estimate the predictability limit. For brevity, we focus on the error growth in the March-initialized and July-initialized predictions (Figure 2). In the March-initialized prediction, the first-month RMSE is about 0.2 K, or ~0.1 K (30%) smaller than that in the July-initialized prediction. Nonetheless, the error growth rates in these predictions are comparable (0.08 K month^{-1}) afterward. The comparable growth rates in the springtime and the summertime are consistent with the COLA-RSMAS model (Larson and Kirtman 2015). This finding may appear to contrast the faster-error-growth hypothesis by Webster and Yang (1992), but the conclusion would depend on the choice of error growth metric. The error growth—if measured with the error doubling time—would be relatively fast for the spring-initialized predictions since they have smaller initial RMSE. Toward the end of the forecast range, the error growth shows signs of slow-down and saturation. Due to the relatively large uncertainty in estimating RMSE (Figure 2), it is not apparent whether the error growth needs to be fitted with the two-parameter empirical model by Lorenz (1982). For simplicity, we characterize the error growth with linear regressions instead.

As shown by Lorenz (1982) and ensuing studies (e.g., Simmons et al., 1995), an error growth model helps estimate the impact of reducing IPEs on the forecasts afterward. Here we assume the method is applicable to seasonal climate prediction and estimate the potential error reduction (or skill gain) in predicting the wintertime ENSO. We consider two scenarios of error reduction: 50% for an optimistic but technically plausible estimate, and 90% for an estimate of predictability bounds. In Figure 2, reducing the magnitude of first-month errors is equivalent to shifting the regression lines. The results suggest that halving errors may extend the skills of March-initialized predictions by one month or July-initialized predictions by two months. If the first-month errors decrease by 90%, the prediction skills may extend by two months and four months, respectively. While the potential skill gains with halving first-month errors may appear small, they are likely meaningful given that the improvement of seasonal ENSO predictions may have stalled or backtracked (Chen & Cane, 2008; Barnston et al., 2012; Becker et al., 2020). We speculate that such skill gains are potentially larger as the slope of error growth may be unrealistically large for the over-dispersive NMME ($ACC/PC > 1$; Figure 1a-b). Analysis of individual models indeed suggests the estimate of error growth rate has some uncertainty (Figures S3 and S4) and could be smaller than values indicated by Figure 2.

3.2 Ensemble Sensitivity Analysis

In practice, IPEs may be reduced with better observations, data assimilation, and model initialization. These error sources in principle can be disentangled with well-designed forecast experiments and the ensemble sensitivity analysis (e.g., Torn & Hakim, 2008). Due to data limitations, this study examines the SST only and traces error sources back to the first-month prediction. We analyzed predictions initialized in the springtime and the summertime (Figures 3, S5, and S6). The results suggest that the ENSO prediction errors generally originate from the equatorial Pacific, but the March-initialized prediction shows a distinct pattern of error growth. Given its relevance to the spring predictability barrier, the ensuing discussion will focus on the March-initialized prediction.

To highlight the most robust relationship, we first evaluate the relationship between December Niño 3.4 and the SST predictions in earlier months (Figure 3). Looking backward, the errors in predicting December Niño 3.4 can be traced back to a pattern resembling the ENSO pattern during the summertime (Figures 3d-f). This pattern also appears in the predictions initialized during the summertime and later months (Figures S5 and S6). The results suggest the error growth during the summertime primarily arises from the ENSO process itself. However, the sensitivity patterns in the springtime (Figure 3a-c) differ and show associations with off-equatorial regions. For example, the sensitivity pattern in April suggests that high Niño 3.4 values in December are associated with warmth in the off-equatorial regions of the tropical Central and Eastern Pacific. During March and April, the North Pacific pattern resembles the North Pacific meridional mode (Chiang & Vimont, 2004), which can lead to El Niño development in the observation (Chang et al., 2007) and GCMs (Larson & Kirtman, 2014; L. Zhang et al., 2009). The South Pacific pattern manifests in the tropics and thus differs from the South Pacific meridional mode (H. Zhang et al., 2014)

It is intriguing to compare the sensitivity patterns and the modern understanding of ENSO development. On one hand, the ensemble sensitivity analysis identifies plausible error sources that may feed into the ENSO growth (e.g., equatorial development and extratropical precursors).

Without a priori about the growth mechanisms, such consistency is encouraging for the application of the ensemble sensitivity analysis. On the other hand, the sensitivity patterns suggest additional regions of error sources, such as the West Pacific and the South Pacific in March. The differences from meridional modes could arise from possible biases of the NMME models in simulating the meridional modes. Yet an alternative explanation is possible: the key regions for error growth do not necessarily have to be regions with the maximum SST variance (i.e., meridional modes). If one holds the perfect model assumption, the additional high-sensitivity regions in GCM simulations could instead indicate a new opportunity to help models mitigate the spring predictability barrier.

3.3 Impacts of First-Month Prediction Error

Focusing on the March-initialized prediction by the NMME, we hypothesize that the ENSO prediction in some high-sensitivity years may benefit from reducing first-month prediction errors (FPEs). Figure 4a shows the regression of December Niño 3.4 predictions onto the March SST predictions. The regression coefficients indicate the sensitivity of December predictions to the FPEs in March. The coefficients are evaluated for individual years separately and averaged over 1982-2010. Despite large year-to-year variations (not shown), the greatest regression coefficients tend to appear in the tropical Northwest Pacific and the Southeast Pacific (Figure 4a). While the average regression coefficients are about one, values in individual years can reach four or higher (not shown). Noting that the uncertainty among SST observational datasets often exceeds 0.1 K (Yang et al., 2021), first-month prediction errors caused by March observational biases could be associated with a non-trivial error ($\geq$0.4K) in predicting the December Niño 3.4.

What could contribute to FPEs and their changes in practice? We conduct an opportunistic thought experiment about increasing IC errors in the high-sensitivity regions. For the tropical Northwest Pacific, a main source of oceanic observation data is the Triangle Trans-Ocean Buoy Network (TRITON) (Ando et al., 2017; black dots in Figure 4a). The TRITON array was deployed in 1998 and gradually expanded in later years. But starting in the early 2010s, the off-equator buoys were decommissioned. Fujii et al. (2015) showed that the loss of buoy data is impactful on local SST errors in several ocean data assimilation systems. For the GFDL assimilation system, the RMSE against the observed SST in the tropical Northwest Pacific may increase by >0.8 K (see their Figure 10).

Consistent with Fujii et al. (2015), Figure 4b suggests the absence of TRITON data may contribute to a cold bias in March SST predictions. We define the bais as the difference between predictions and observation (OISST). Coincident with the absence of TRITON data (pre-1998 and post-2012), March-initialized GFDL-FLOR predictions show first-month cold biases of ~0.3 K in the tropical Northwest Pacific (Figure 4b). Such biases are not apparent in other regions (e.g., the Southeast Pacific; Figure 4b), suggesting the cold biases are related to region-specific changes. Since the sensitivity regression coefficient can reach four (e.g., 2016) in the tropical Northwest Pacific, the initial cold biases may lead to a 1.2 K cold bias in the prediction of December Niño 3.4.

Accordingly, we formulate a simple linear model to estimate the impacts of reducing the IPEs in the tropical Northwest Pacific (Figure 4c-d). The linear model can also serve as a bias correction, and its workflow is outlined in Figure 4d. When selecting the region of SST input, we examine the correlations between the March SST and December Niño 3.4 predictions. For March-initialized GFDL-FLOR, the Niño 3.4 prediction is more sensitive to the SST of the

Northwest Pacific (Figure 4b). Focusing on this region, we define a correction term for predicting December Niño 3.4 as the product of the March SST biases and the sensitivity regression coefficients. The application of this correction is conditional: it is only activated when the ensemble sensitivity analysis suggests that March SST errors and December Niño 3.4 errors are correlated at the 90% confidence level.

Figure 4c shows the conditionally corrected predictions and the original predictions. By replacing the predicted SST in the Northwest Pacific with the observation, the simple conditional correction reduces the mean average error (MAE) by 0.11 K (14%). The most substantial improvements exceed 1 K and appear in 1997, 1998, and 2016, when the TRITON observations were absent. While such bias corrections are not equivalent to reducing global FPEs (Figure 2), the results are consistent with the expectation that reducing IPEs helps improve the ENSO prediction. Meanwhile, the three years with the most substantial improvements are around extreme El Niño conditions. The underlying physical processes and whether the high sensitivity depends on a specific climate state warrant future study.

4. Summary and Discussion

This study was partly inspired by early pioneering research in predicting the Earth system (e.g., Lorenz 1982). To explore potential avenues to improve climate prediction, we analyze the NMME predictions and use the ENSO prediction as a testing ground. The analyses focus on the ENSO's seasonal predictability and sensitivity to first-month prediction errors. The key findings are summarized as follows:

- The predictions initialized in the springtime experience a rapid drop in the predictable component, consistent with the SNR interpretation of the spring predictability barrier

(Torrence and Webster 1998). Furthermore, predictable signals in the equatorial Pacific SST are preferentially dampened relative to noises.

- Contributors of the ENSO predictability barrier, such as the dampening-growth transition and the post-spring growth rate, are found to depend on the initialization time of the NMME models. These model biases may have interfered with the simulated spring predictability barrier and suggest inconsistency between the NMME and the real world.
- Reducing IPEs will likely extend the skill of predicting the wintertime ENSO. Halving the errors may extend the prediction skills by one to two months. The practical limit of such skill gains may be up to four months. These estimates with the NMME could be conservative since the error growth might be too fast in this over-dispersive model group.
- Where reducing IPEs may be effective varies across seasons and years. For example, the March-initialized predictions will likely benefit from error reductions in the tropical Northwest Pacific rather than the equatorial Pacific or the off-equatorial East Pacific. The ensemble sensitivity analysis also suggests much higher sensitivity in some El Niño-year predictions (e.g., 1997, 1998, and 2016).
- Using changes in the TRITON array as an opportunistic test, we found that reducing March SST prediction errors may reduce the MAE of December Niño 3.4 prediction by about 14%. For the GFDL-FLOR, the potential improvements may exceed 1 K in some high-sensitivity years.

Admittedly, the seasonal prediction of the ENSO may have been near the predictability limit (Chen & Cane, 2008; Newman & Sardeshmukh, 2017), and the prediction skill is subject to impacts of the ENSO decadal variability (Balmaseda et al., 1995; Barnston et al., 2012; Weisheimer et al., 2022). Nonetheless, avenues toward better climate predictions are worth pursuing. Such effort is critical when climate risk management increasingly needs reliable predictions, and improving predictions in a cost-effective way is particularly valuable. Our predictability analysis suggests such improvements for the ENSO prediction are possible by reducing IPEs. Importantly, our ensemble sensitivity analysis highlighted the tropical Northwest Pacific as a region where better ICs may be particularly effective.

The discussion of this study focuses on the NMME group instead of individual models. We caution that the conclusions about the NMME group should not be extended to individual models without scrutiny. For example, the NMME is over-dispersive in simulating the ENSO system, even though the contributing models suffer from overconfidence issues. As suggested by Figures S1-S4, the characteristics of the ENSO system (e.g., signal growth and system noise) simulated by individual models are highly diverse. To address the spring predictability barrier, future research should consider constraining the uncertainty of the related system parameters. We also caution that many NMME models show negative prediction skills in the tropical Pacific (Newman & Sardeshmukh, 2017), so model biases may also need to be addressed to fully materialize the benefits of reducing IPEs.

Although the findings based on the ensemble sensitivity analysis are promising, a notable caveat of this study is the lack of independent verification by numeric experiments. While complex and expensive to implement, such experiments may eliminate ambiguities related to changes in models or their input, such as the 2011 change of the SST input for GFDL FLOR (Bushuk et al., 2019). By making additional variables available, numeric experiments can facilitate the analysis of physical processes (e.g., Bushuk et al., 2019). For the ENSO prediction, this will enable analysis of the subsurface ocean and atmosphere-ocean anomalies, which may be important sources of predictability (e.g., Li et al., 2023). One may also follow Fujii et al. (2015) or Torn and Hakim (2008) and quantify the impacts of observational errors using data assimilation systems. Together with the ensemble sensitivity analysis, these experiments may assist the management and improvements of the climate observation system.

We hope our findings may motivate future improvements in climate prediction. The analytical framework used here has found success in improving weather forecasting and showed promises with the ENSO prediction. In principle, the framework is applicable to other climate variations or anomalies, including droughts, heatwaves, cyclone activity, and renewable energy resources. Systematic applications to societally important phenomena may help maximize the societal benefits of research and development. A well-coordinated effort might lead to a revolution like what the weather community has celebrated (Bauer et al., 2015) and bolster climate risk management.

Acknowledgment

The author thanks the Massachusetts Institute of Technology (MIT) and the European Centre for Medium-Range Weather Forecasting (ECMWF) for granting the public access to numerous historical and technical documents, including many valuable manuscripts by Dr. Edward Lorenz. The author also thanks Dr. Mitch Bushuk for his technical insights about the GFDL-FLOR model and Dr. Zhuo Wang for suggestions that improved the clarity of result presentation. The work is supported by the faculty start-up fund of the University of Illinois at Urbana-Champaign.

Conflict of Interest Statement

The author has no conflicts of interest to declare.

Data Availability Statement

The NMME dataset and the OISST dataset are available at IRI/LDEO (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/). The analysis code is available to reviewers upon request. The author will deposit the code at Zenodo before the article is accepted for publication.

References

Ancell, B., & Coleman, A. A. (2022). New Perspectives on Ensemble Sensitivity Analysis with Applications to a Climatology of Severe Convection. *Bulletin of the American Meteorological Society*, *103*(2), E511–E530. https://doi.org/10.1175/BAMS-D-20-0321.1

Ancell, B., & Hakim, G. J. (2007). Comparing Adjoint- and Ensemble-Sensitivity Analysis with Applications to Observation Targeting. *Monthly Weather Review*, *135*(12), 4117–4134. https://doi.org/10.1175/2007MWR1904.1

Ando, K., Kuroda, Y., Fujii, Y., Fukuda, T., Hasegawa, T., Horii, T., et al. (2017). Fifteen years progress of the TRITON array in the Western Pacific and Eastern Indian Oceans. *Journal of Oceanography*, *73*(4), 403–426. https://doi.org/10.1007/s10872-017-0414-4

Balmaseda, M. A., Davey, M. K., & Anderson, D. L. T. (1995). Decadal and Seasonal Dependence of ENSO Prediction Skill. *Journal of Climate*, *8*(11), 2705–2715. https://doi.org/10.1175/1520-0442(1995)008<2705:DASDOE>2.0.CO;2

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., & DeWitt, D. G. (2012). Skill of Real-Time Seasonal ENSO Model Predictions during 2002–11: Is Our Capability Increasing? *Bulletin of the American Meteorological Society*, *93*(5), 631–651. https://doi.org/10.1175/BAMS-D-11-00111.1

Barnston, A. G., Tippett, M. K., Ranganathan, M., & L'Heureux, M. L. (2019). Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dynamics*, *53*(12), 7215–7234. https://doi.org/10.1007/s00382-017-3603-3

Battisti, D. S. (1988). Dynamics and Thermodynamics of a Warming Event in a Coupled Tropical Atmosphere–Ocean Model. *Journal of the Atmospheric Sciences*, *45*(20), 2889–2919. https://doi.org/10.1175/1520-0469(1988)045<2889:DATOAW>2.0.CO;2

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55. https://doi.org/10.1038/nature14956

Becker, E. J., Kirtman, B. P., & Pegion, K. (2020). Evolution of the North American Multi-Model Ensemble. *Geophysical Research Letters*, *47*(9). https://doi.org/10.1029/2020GL087408

Becker, E. J., Kirtman, B. P., L'Heureux, M., Muñoz, Á. G., & Pegion, K. (2022). A Decade of the North American Multimodel Ensemble (NMME): Research, Application, and Future Directions. *Bulletin of the American Meteorological Society*, *103*(3), E973–E995. https://doi.org/10.1175/BAMS-D-20-0327.1

Buontempo, C., Burgess, S. N., Dee, D., Pinty, B., Thépaut, J.-N., Rixen, M., et al. (2022). The Copernicus Climate Change Service: Climate Science in Action. *Bulletin of the American Meteorological Society*, *103*(12), E2669–E2687. https://doi.org/10.1175/BAMS-D-21-0315.1

Bushuk, M., Yang, X., Winton, M., Msadek, R., Harrison, M., Rosati, A., & Gudgel, R. (2019). The Value of Sustained Ocean Observations for Sea Ice Predictions in the Barents Sea. *Journal of Climate*, *32*(20), 7017–7035. https://doi.org/10.1175/JCLI-D-19-0179.1

Chang, P., Zhang, L., Saravanan, R., Vimont, D. J., Chiang, J. C. H., Ji, L., et al. (2007). Pacific meridional mode and El Niño-Southern Oscillation: PACIFIC MERIDIONAL MODE AND ENSO. *Geophysical Research Letters*, *34*(16). https://doi.org/10.1029/2007GL030302

Chen, D., & Cane, M. A. (2008). El Niño prediction and predictability. *Journal of Computational Physics*, *227*(7), 3625–3640. https://doi.org/10.1016/j.jcp.2007.05.014

Chen, D., Zebiak, S. E., Busalacchi, A. J., & Cane, M. A. (1995). An Improved Procedure for EI Niño Forecasting: Implications for Predictability. *Science*, *269*(5231), 1699–1702. https://doi.org/10.1126/science.269.5231.1699

Chen, D., Cane, M. A., Kaplan, A., Zebiak, S. E., & Huang, D. (2004). Predictability of El Niño over the past 148 years. *Nature*, *428*(6984), 733–736. https://doi.org/10.1038/nature02439

Chiang, J. C. H., & Vimont, D. J. (2004). Analogous Pacific and Atlantic Meridional Modes of Tropical Atmosphere–Ocean Variability*. *Journal of Climate*, *17*(21), 4143–4158. https://doi.org/10.1175/JCLI4953.1

Doblas-Reyes, F. J., Déqué, M., & Piedelievre, J.-P. (2000). Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Quarterly Journal of the Royal Meteorological Society*, *126*(567), 2069–2087. https://doi.org/10.1002/qj.49712656705

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?: Seasonal to decadal predictability. *Geophysical Research Letters*, *41*(15), 5620–5628. https://doi.org/10.1002/2014GL061146

Fujii, Y., Cummings, J., Xue, Y., Schiller, A., Lee, T., Balmaseda, M. A., et al. (2015). Evaluation of the Tropical Pacific Observing System from the ocean data assimilation perspective. *Quarterly Journal of the Royal Meteorological Society*, *141*(692), 2481–2496. https://doi.org/10.1002/qj.2579

Goddard, L., Mason, S. J., Zebiak, S. E., Ropelewski, C. F., Basher, R., & Cane, M. A. (2001). Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology*, *21*(9), 1111–1152. https://doi.org/10.1002/joc.636

Hagedorn, R., Doblas-Reyes, F. J., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus A: Dynamic Meteorology and Oceanography*, *57*(3), 219. https://doi.org/10.3402/tellusa.v57i3.14657

Hakim, G. J., & Torn, R. D. (2008). Ensemble Synoptic Analysis. In L. F. Bosart & H. B. Bluestein (Eds.), *Synoptic—Dynamic Meteorology and Weather Analysis and Forecasting* (pp. 147–161). Boston, MA: American Meteorological Society. https://doi.org/10.1007/978-0-933876-68-2_7

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7

Infanti, J. M., & Kirtman, B. P. (2016). Prediction and predictability of land and atmosphere initialized CCSM4 climate forecasts over North America. *Journal of Geophysical Research: Atmospheres*, *121*(21), 12,690-12,701. https://doi.org/10.1002/2016JD024932

Jain, S., Scaife, A. A., Shepherd, T. G., Deser, C., Dunstone, N., Schmidt, G. A., et al. (2023). Importance of internal variability for climate model assessment. *Npj Climate and Atmospheric Science*, *6*(1), 68. https://doi.org/10.1038/s41612-023-00389-0

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., et al. (2014). The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Larson, S. M., & Kirtman, B. P. (2014). The Pacific Meridional Mode as an ENSO Precursor and Predictor in the North American Multimodel Ensemble. *Journal of Climate*, *27*(18), 7018–7032. https://doi.org/10.1175/JCLI-D-14-00055.1

Larson, S. M., & Kirtman, B. P. (2015). Revisiting ENSO Coupled Instability Theory and SST Error Growth in a Fully Coupled Model. *Journal of Climate*, *28*(12), 4724–4742. https://doi.org/10.1175/JCLI-D-14-00731.1

Larson, S. M., & Kirtman, B. P. (2017). Drivers of coupled model ENSO error dynamics and the spring predictability barrier. *Climate Dynamics*, *48*(11–12), 3631–3644. https://doi.org/10.1007/s00382-016-3290-5

Latif, M., Barnett, T. P., Cane, M. A., Flügel, M., Graham, N. E., Von Storch, H., et al. (1994). A review of ENSO prediction studies. *Climate Dynamics*, *9*(4–5), 167–179. https://doi.org/10.1007/BF00208250

Latif, M., Anderson, D., Barnett, T., Cane, M., Kleeman, R., Leetmaa, A., et al. (1998). A review of the predictability and prediction of ENSO. *Journal of Geophysical Research: Oceans*, *103*(C7), 14375–14393. https://doi.org/10.1029/97JC03413

Levine, A. F. Z., & McPhaden, M. J. (2015). The annual cycle in ENSO growth rate as a cause of the spring predictability barrier. *Geophysical Research Letters*, *42*(12), 5034–5041. https://doi.org/10.1002/2015GL064309

Li, Y., Xie, S., Lian, T., Zhang, G., Feng, J., Ma, J., et al. (2023). Interannual Variability of Regional Hadley Circulation and El Niño Interaction. *Geophysical Research Letters*, *50*(4), e2022GL102016. https://doi.org/10.1029/2022GL102016

Liu, Z., Jin, Y., & Rong, X. (2019). A Theory for the Seasonal Predictability Barrier: Threshold, Timing, and Intensity. *Journal of Climate*, *32*(2), 423–443. https://doi.org/10.1175/JCLI-D-18-0383.1

Lopez, H., & Kirtman, B. P. (2014). WWBs, ENSO predictability, the spring barrier and extreme events: WWBs and ENSO Predictability. *Journal of Geophysical Research: Atmospheres*, *119*(17), 10,114-10,138. https://doi.org/10.1002/2014JD021908

Lorenz, E. N. (1982). Atmospheric predictability experiments with a large numerical model. *Tellus*, *34*(6), 505–513. https://doi.org/10.1111/j.2153-3490.1982.tb01839.x

Lu, F., Harrison, M. J., Rosati, A., Delworth, T. L., Yang, X., Cooke, W. F., et al. (2020). GFDL's SPEAR Seasonal Prediction System: Initialization and Ocean Tendency Adjustment (OTA) for Coupled Model Predictions. *Journal of Advances in Modeling Earth Systems*, *12*(12). https://doi.org/10.1029/2020MS002149

McPhaden, M. J., Zebiak, S. E., & Glantz, M. H. (2006). ENSO as an Integrating Concept in Earth Science. *Science*, *314*(5806), 1740–1745. https://doi.org/10.1126/science.1132588

Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., et al. (2013). The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization. *Monthly Weather Review*, *141*(8), 2910–2945. https://doi.org/10.1175/MWR-D-12-00216.1

Murakami, H., Villarini, G., Vecchi, G. A., Zhang, W., & Gudgel, R. (2016). Statistical–Dynamical Seasonal Forecast of North Atlantic and U.S. Landfalling Tropical Cyclones Using the High-Resolution GFDL FLOR Coupled Model. *Monthly Weather Review*, *144*(6), 2101–2123. https://doi.org/10.1175/MWR-D-15-0308.1

Newman, M., & Sardeshmukh, P. D. (2017). Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures?: Seasonal Predictability of Tropical SSTs. *Geophysical Research Letters*, *44*(16), 8520–8529. https://doi.org/10.1002/2017GL074088

Palmer, T. N., & Anderson, D. L. T. (1994). The prospects for seasonal forecasting—A review paper. *Quarterly Journal of the Royal Meteorological Society*, *120*(518), 755–793. https://doi.org/10.1002/qj.49712051802

Palmer, T. N., & Zanna, L. (2013). Singular vectors, predictability and ensemble forecasting for weather and climate. *Journal of Physics A: Mathematical and Theoretical*, *46*(25), 254018. https://doi.org/10.1088/1751-8113/46/25/254018

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délécluse, P., et al. (2004). DEVELOPMENT OF A EUROPEAN MULTIMODEL ENSEMBLE SYSTEM FOR SEASONAL-TO-INTERANNUAL PREDICTION (DEMETER). *Bulletin of the American Meteorological Society*, *85*(6), 853–872. https://doi.org/10.1175/BAMS-85-6-853

Palmer, T. N., Doblas-Reyes, F. J., Hagedorn, R., & Weisheimer, A. (2005). Probabilistic prediction of climate using multi-model ensembles: from basics to applications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1463), 1991–1998. https://doi.org/10.1098/rstb.2005.1750

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An Improved In Situ and Satellite SST Analysis for Climate. *Journal of Climate*, *15*(13), 1609–1625. https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP Climate Forecast System Version 2. *Journal of Climate*, *27*(6), 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Shukla, J. (1998). Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting. *Science*, *282*(5389), 728–731. https://doi.org/10.1126/science.282.5389.728

Simmons, A. J., Mureau, R., & Petroliagis, T. (1995). Error growth and estimates of predictability from the ECMWF forecasting system. *Quarterly Journal of the Royal Meteorological Society*, *121*(527), 1739–1771. https://doi.org/10.1002/qj.49712152711

Small, R. J., Bacmeister, J., Bailey, D., Baker, A., Bishop, S., Bryan, F., et al. (2014). A new synoptic scale resolving global climate simulation using the Community Earth System Model. *Journal of Advances in Modeling Earth Systems*, *6*(4), 1065–1094. https://doi.org/10.1002/2014MS000363

Stern, W., & Miyakoda, K. (1995). Feasibility of Seasonal Forecasts Inferred from Multiple GCM Simulations. *Journal of Climate*, *8*(5), 1071–1085. https://doi.org/10.1175/1520-0442(1995)008<1071:FOSFIF>2.0.CO;2

Tippett, M. K., & L'Heureux, M. L. (2020). Low-dimensional representations of Niño 3.4 evolution and the spring persistence barrier. *Npj Climate and Atmospheric Science*, *3*(1), 24. https://doi.org/10.1038/s41612-020-0128-y

Tippett, M. K., Ranganathan, M., L'Heureux, M., Barnston, A. G., & DelSole, T. (2019). Assessing probabilistic predictions of ENSO phase and intensity from the North American Multimodel Ensemble. *Climate Dynamics*, *53*(12), 7497–7518. https://doi.org/10.1007/s00382-017-3721-y

Tompkins, A. M., Ortiz De Zárate, M. I., Saurral, R. I., Vera, C., Saulo, C., Merryfield, W. J., et al. (2017). The Climate-System Historical Forecast Project: Providing Open Access to Seasonal

Forecast Ensembles from Centers around the Globe. *Bulletin of the American Meteorological Society*, *98*(11), 2293–2301. https://doi.org/10.1175/BAMS-D-16-0209.1

Torn, R. D., & Hakim, G. J. (2008). Ensemble-Based Sensitivity Analysis. *Monthly Weather Review*, *136*(2), 663–677. https://doi.org/10.1175/2007MWR2132.1

Torrence, C., & Webster, P. J. (1998). The annual cycle of persistence in the El Nño/Southern Oscillation. *Quarterly Journal of the Royal Meteorological Society*, *124*(550), 1985–2004. https://doi.org/10.1002/qj.49712455010

Vecchi, G. A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg, A. T., et al. (2014). On the Seasonal Forecasting of Regional Tropical Cyclone Activity. *Journal of Climate*, *27*(21), 7994–8016. https://doi.org/10.1175/JCLI-D-14-00158.1

Webster, P. J., & Yang, S. (1992). Monsoon and Enso: Selectively Interactive Systems. *Quarterly Journal of the Royal Meteorological Society*, *118*(507), 877–926. https://doi.org/10.1002/qj.49711850705

Weisheimer, A., Doblas-Reyes, F. J., Palmer, T. N., Alessandri, A., Arribas, A., Déqué, M., et al. (2009). ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophysical Research Letters*, *36*(21), L21711. https://doi.org/10.1029/2009GL040896

Weisheimer, A., Balmaseda, M. A., Stockdale, T. N., Mayer, M., Sharmila, S., Hendon, H., & Alves, O. (2022). Variability of ENSO Forecast Skill in 2-Year Global Reforecasts Over the 20th Century. *Geophysical Research Letters*, *49*(10). https://doi.org/10.1029/2022GL097885

Yang, C., Leonelli, F. E., Marullo, S., Artale, V., Beggs, H., Nardelli, B. B., et al. (2021). Sea Surface Temperature Intercomparison in the Framework of the Copernicus Climate Change Service (C3S). *Journal of Climate*, *34*(13), 5257–5283. https://doi.org/10.1175/JCLI-D-20-0793.1

Zebiak, S. E., & Cane, M. A. (1987). A Model El Ni&ntilde–Southern Oscillation. *Monthly Weather Review*, *115*(10), 2262–2278. https://doi.org/10.1175/1520-0493(1987)115<2262:AMENO>2.0.CO;2

Zhang, G., Murakami, H., Yang, X., Findell, K. L., Wittenberg, A. T., & Jia, L. (2021). Dynamical Seasonal Predictions of Tropical Cyclone Activity: Roles of Sea Surface Temperature Errors and Atmosphere–Land Initialization. *Journal of Climate*, *34*(5), 1743–1766. https://doi.org/10.1175/JCLI-D-20-0215.1

Zhang, H., Clement, A., & Di Nezio, P. (2014). The South Pacific Meridional Mode: A Mechanism for ENSO-like Variability. *Journal of Climate*, *27*(2), 769–783. https://doi.org/10.1175/JCLI-D-13-00082.1

Zhang, L., Chang, P., & Ji, L. (2009). Linking the Pacific Meridional Mode to ENSO: Coupled Model Analysis. *Journal of Climate*, *22*(12), 3488–3505. https://doi.org/10.1175/2008JCLI2473.1

Zhang, S., Harrison, M. J., Rosati, A., & Wittenberg, A. (2007). System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies. *Monthly Weather Review*, *135*(10), 3541–3564. https://doi.org/10.1175/MWR3466.1

Table 1 Summary of the Analyzed NMME Models

Model Name	NCEP-CFSv2	NCAR-CESM1	GFDL-CM25-FLOR	COLA-RSMAS-CCSM4	CanCM4i
Ensemble Size	24	10	24	10	10
Availability (IRI/LDEO)	1982-2010	1980-2010, 2016-2017	1980-2020	1982-present	1981-2010, 2016-present
Reference	Saha et al., (2014)	Small et al., (2014)	Vecchi et al., (2014)	Infanti & Kirtman, (2016)	Merryfield et al., (2013)

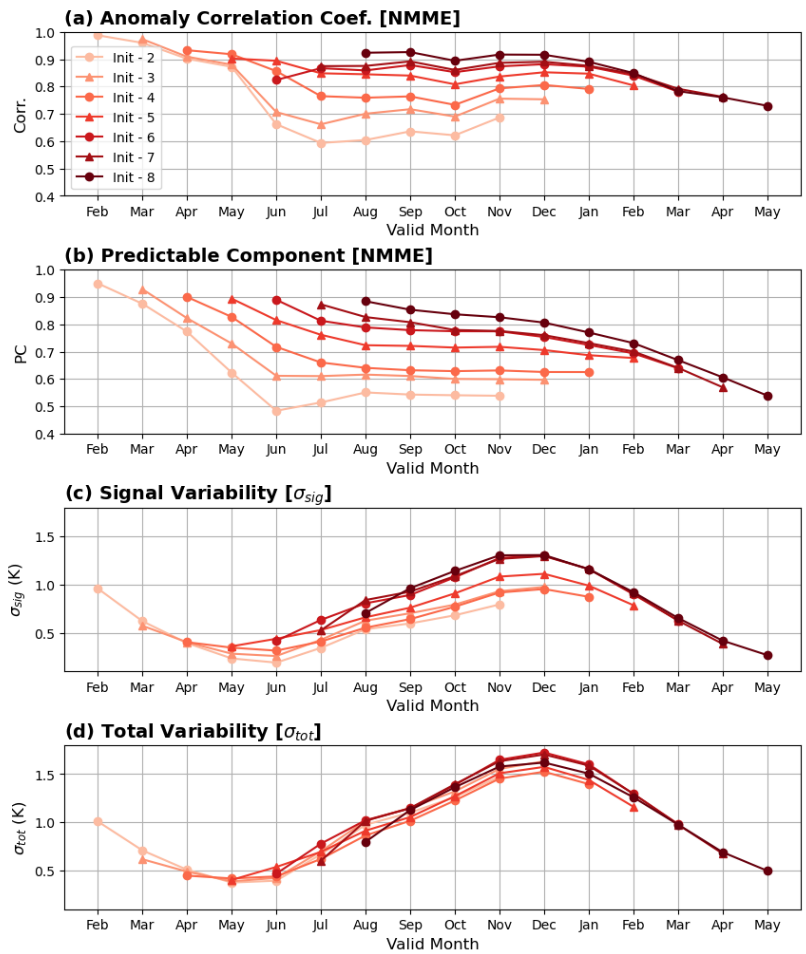


Figure 1 The skill and predictable component of the Niño 3.4 predictions by the NMME models initialized in February–August. (a) The anomaly correlation coefficient (unitless) between the multi-model means and the observation. (b) The predictable component ($\sigma_{sig}/\sigma_{tot}$; unitless) indicated by the NMME models. (c) The signal variability part (σ_{sig}; unit: K) of the predictable component. (d) The total variability part (σ_{tot}; unit: K) of the predictable component. All the examined predictions are grouped based on the initialization time during 1982-2010. For each group, we evaluate ten monthly steps of predictions, with the first step corresponding to the month when the predictions are initialized. The horizontal axis indicates the valid time of monthly predictions.

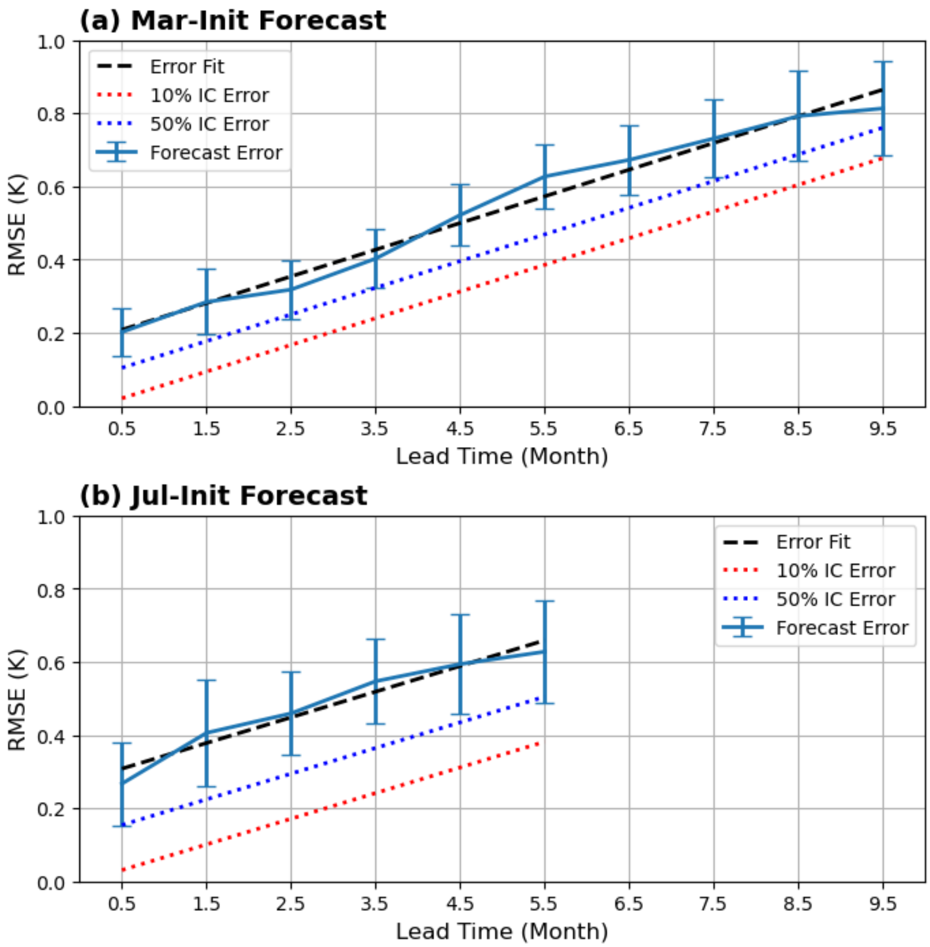


Figure 2 Error growth in the December Niño 3.4 predictions by the NMME models initialized in March and July. Blue solid lines show the RMSE around the NMME ensemble mean. Error bars indicate the ±1 standard deviation range of the RMSE in individual years between 1982 and 2010. Black dashed lines show the linear regressions of error growth. The dotted lines show the potential error growth if first-month prediction errors can be reduced by 50% and 90%. The horizontal axis shows the prediction lead time. Following the convention of the data source, the prediction lead time is denoted with the middle point of monthly mean windows. For example, a 0.5-month lead corresponds to the first-month prediction immediately after model initialization.

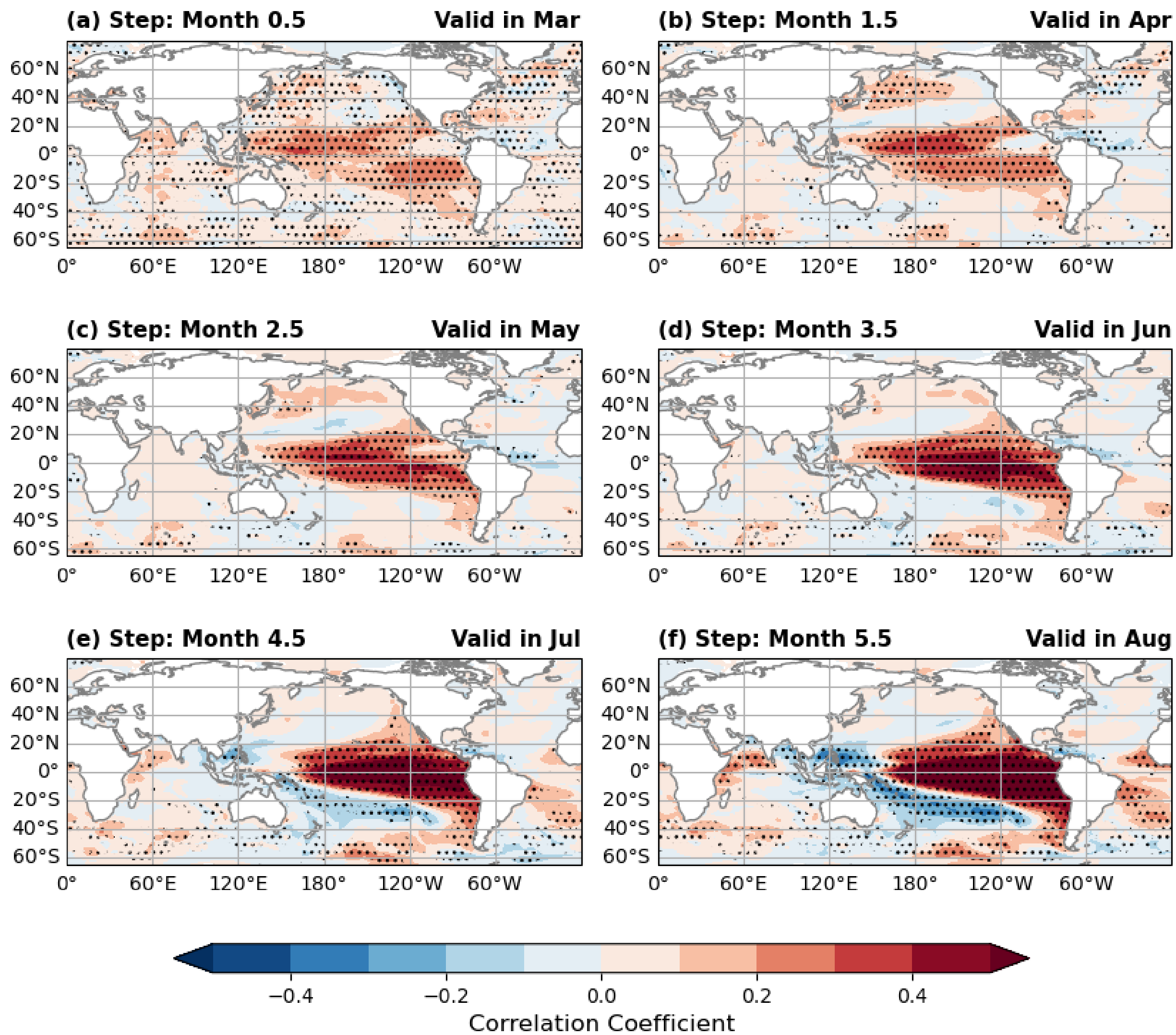


Figure 3 The correlation between the ensemble spreads of the predicted December Niño 3.4 and the SST at the earlier steps. The analyzed NMME predictions are initialized in March 1982-2010. The correlation coefficients are calculated for each individual year and then averaged to highlight the regions with the most robust relationship. We evaluated the statistical significance of correlation coefficients for each year, and the regions where 95% confidence level correlations appear in at least 9 years are denoted with the stippling. Subplots (a)-(f) show prediction steps of Months 0.5 to 5.5, which correspond to the monthly means valid from March to August.

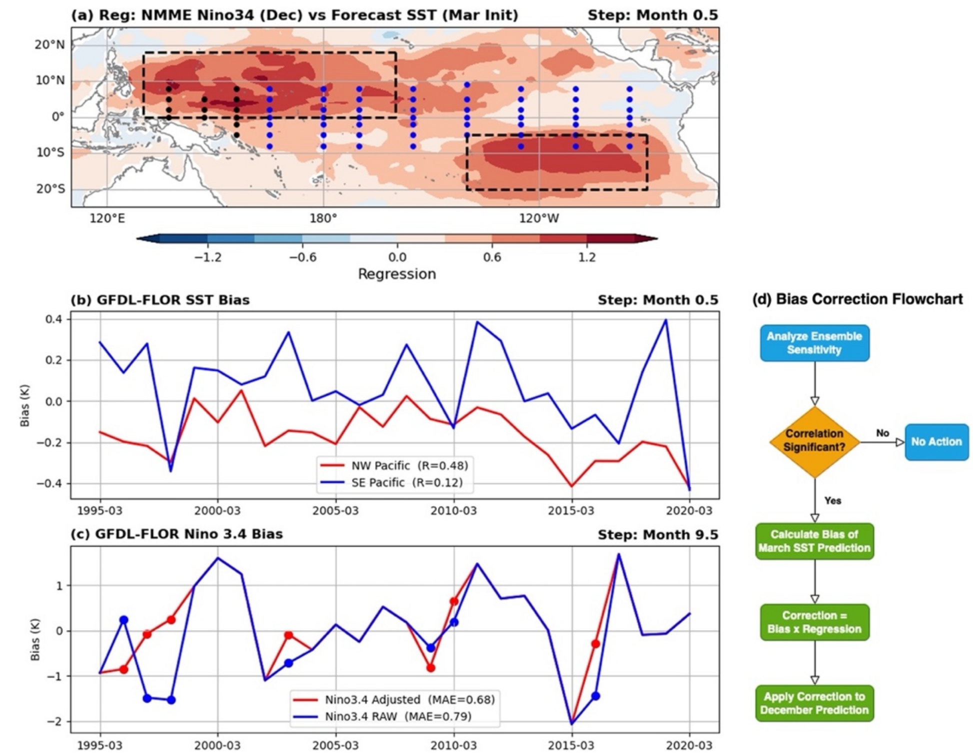
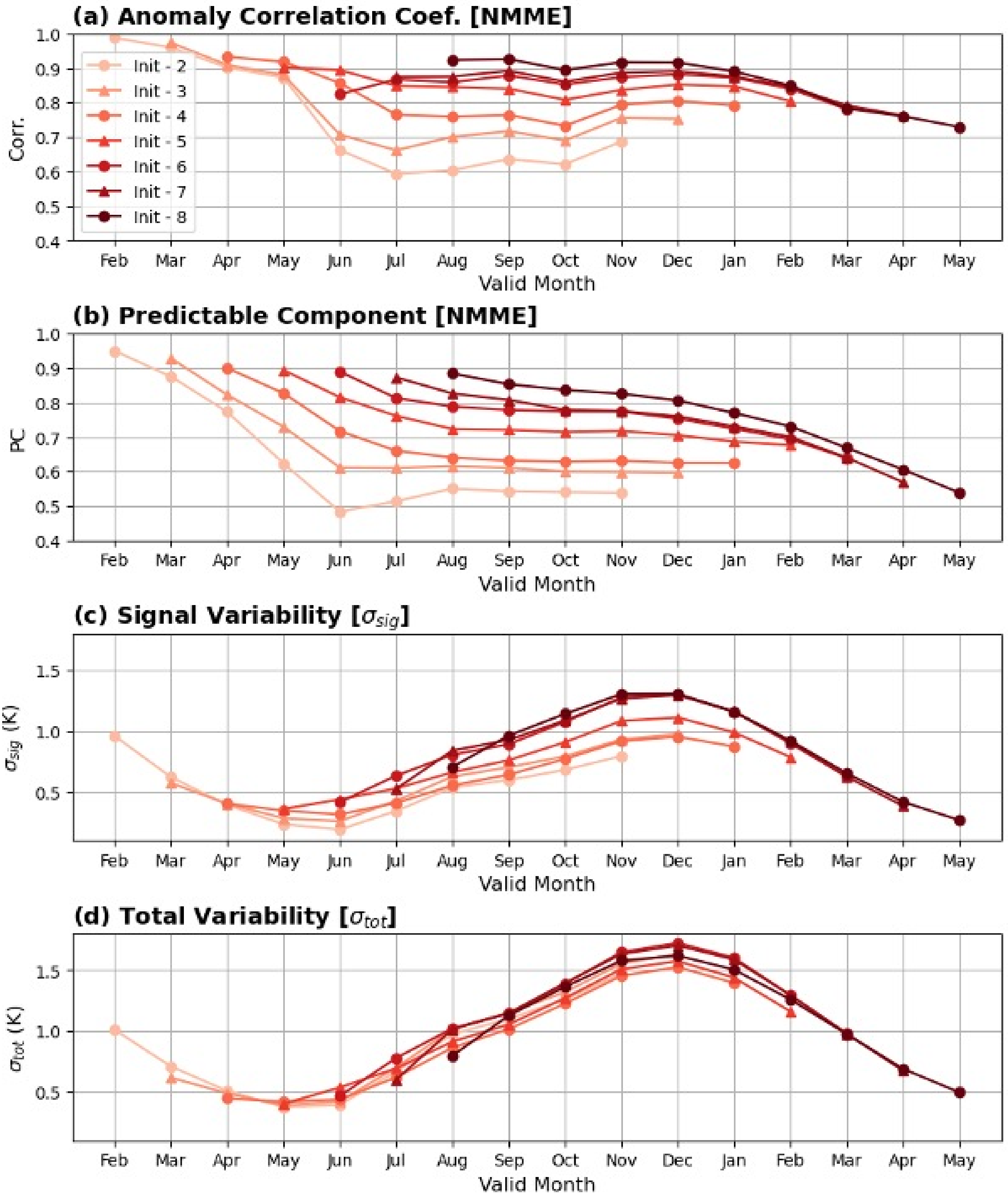
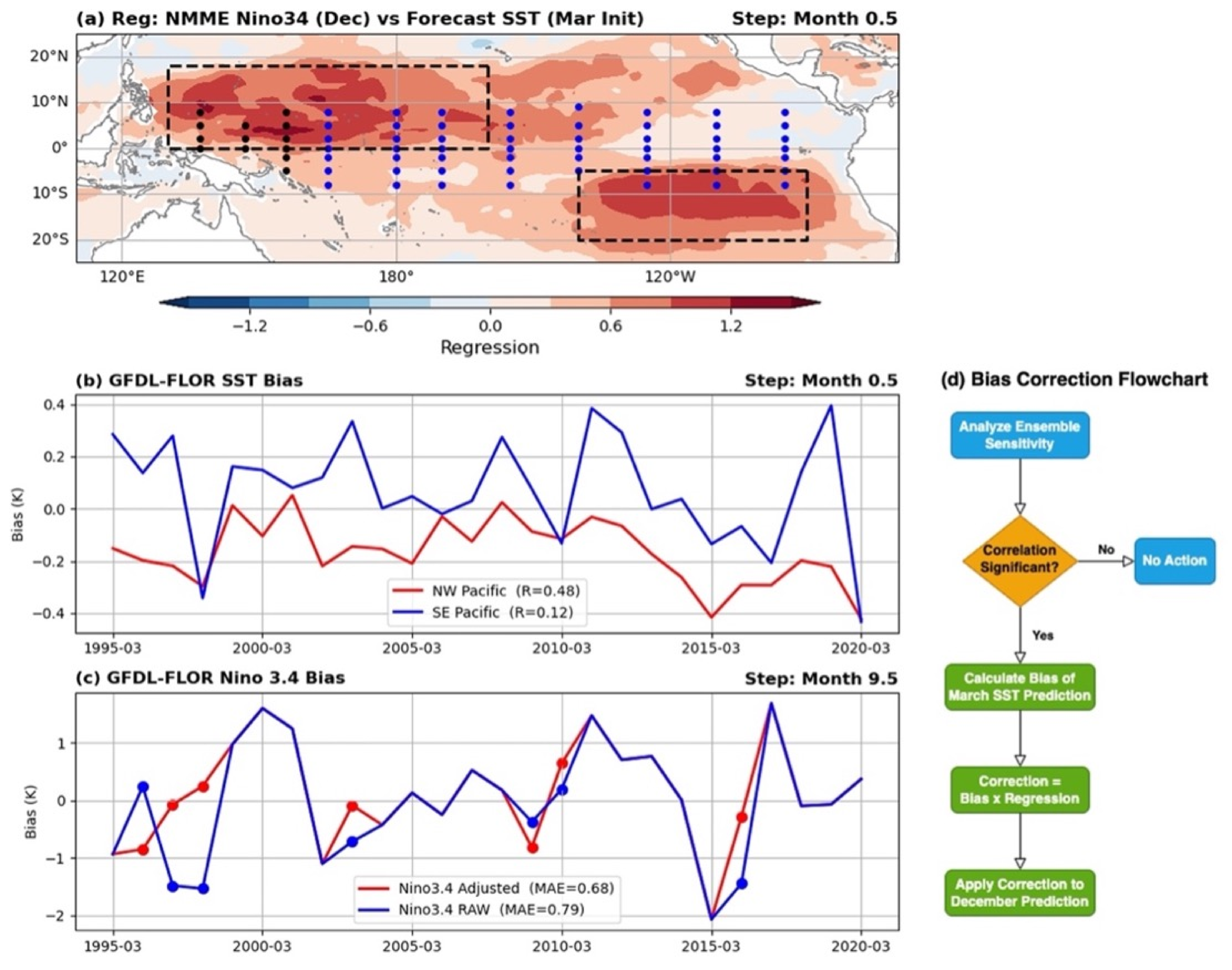


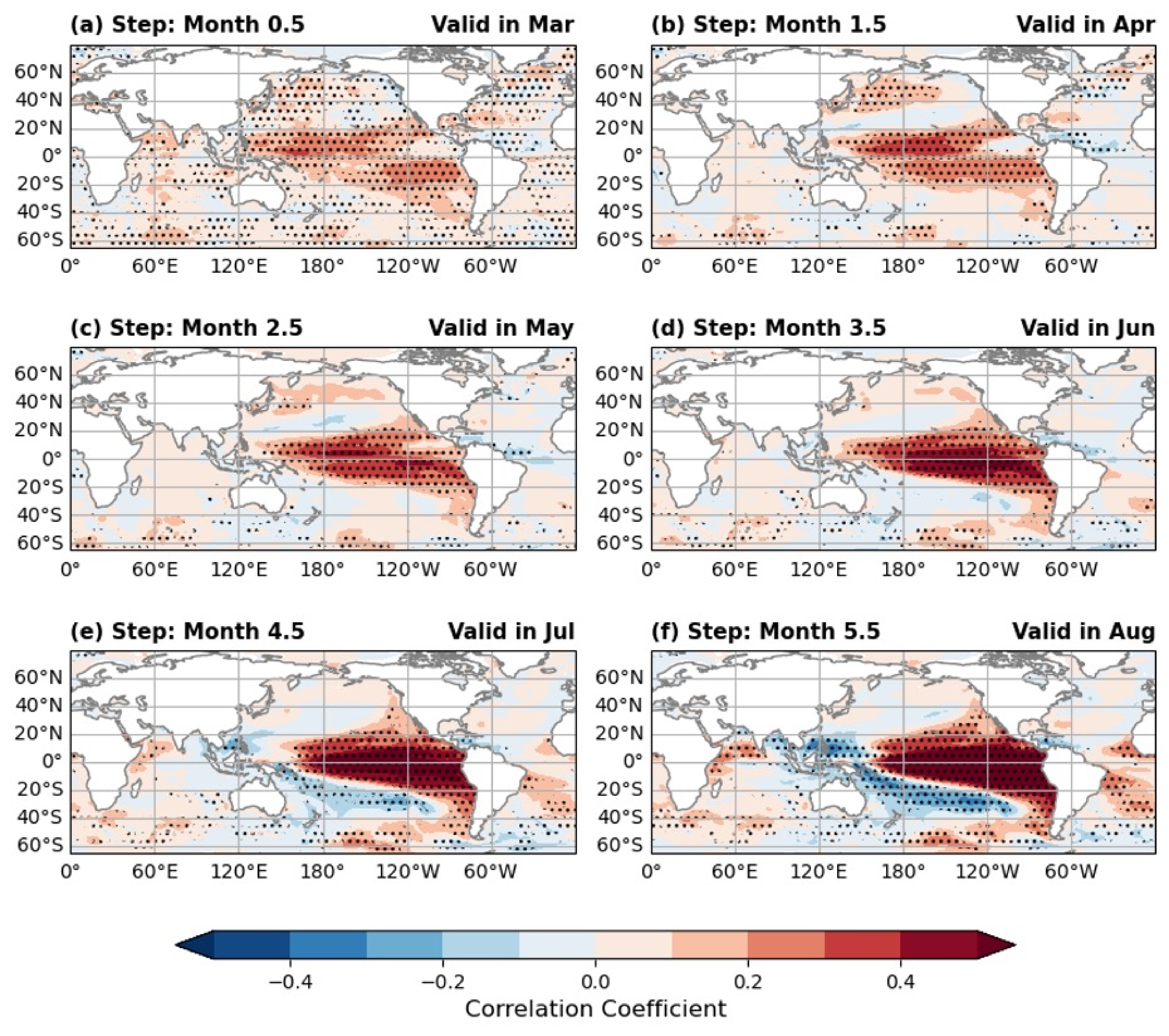
Figure 4 The ensemble sensitivity analysis and the estimated impacts of reducing model biases. The input data are the March SST predictions and the December Niño 3.4 predictions by (a) the NMME models and (b-c) the GFDL-FLOR initialized in March. (a) Regression coefficients of the December Niño 3.4 predictions onto the March SST predictions. The dots indicate the location of the buoy observations by the TRITON (black) and the Tropical Atmosphere Ocean (TAO; blue) projects. The black dash lines highlight two high-sensitivity regions, the tropical Northwest Pacific (0-18ºN, 130ºE-160ºW) and the tropical Southeast Pacific (5ºS-20ºS, 90ºW-140ºW). (b) The prediction biases of the March SST averaged in the tropical Northwest and Southeast Pacific. The line legends indicate the correlation coefficients between the March SST averages and the December Niño 3.4 predictions. (c) The raw prediction and the corrected prediction of December Niño 3.4. The line legends show the mean absolute errors evaluated against the observation. (d) The workflow of generating the corrected prediction in (c).



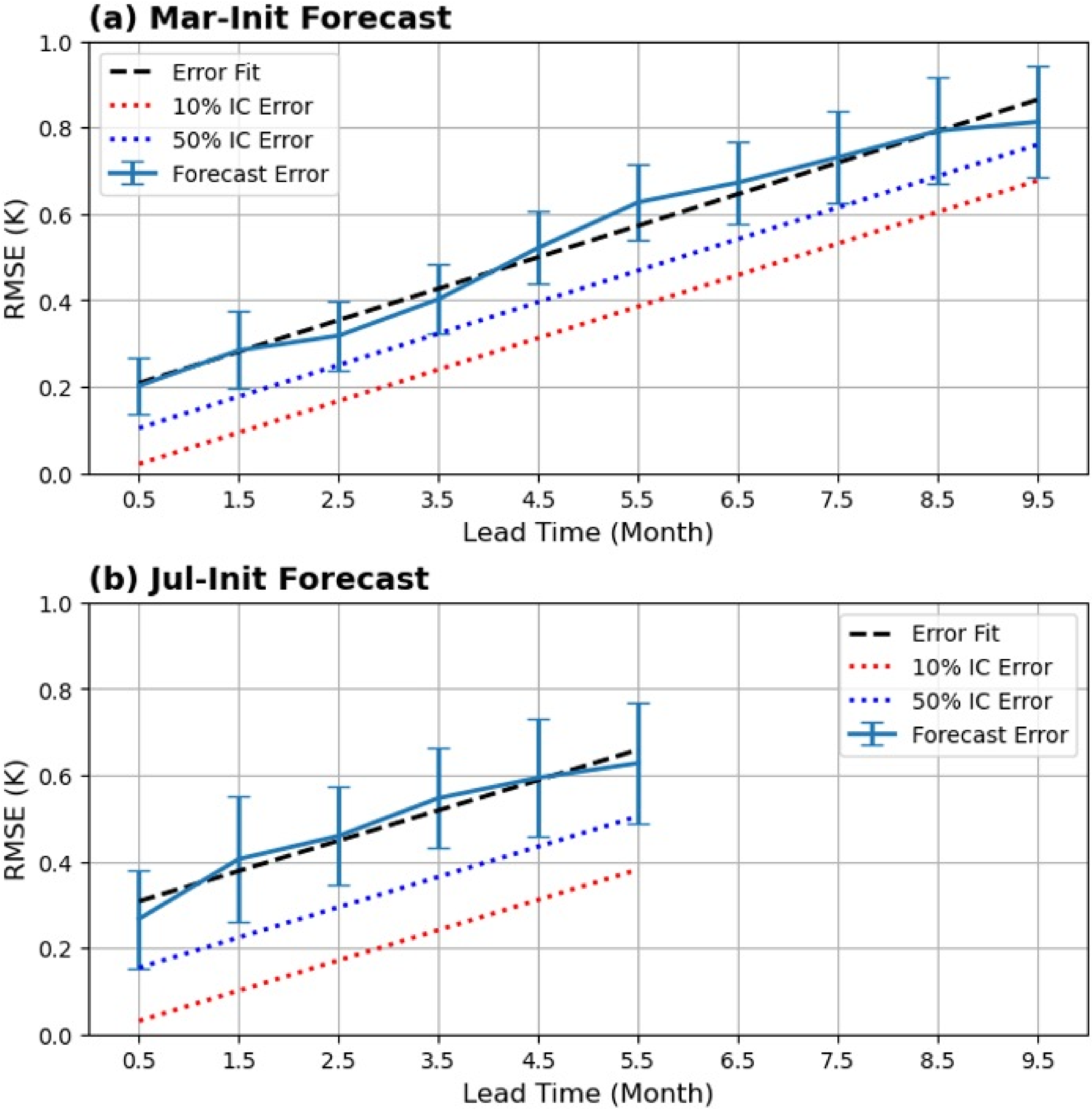
(a) Anomaly Correlation Coef. [NMME]
Corr.
Init - 2
Init - 3
Init - 4
Init - 5
Init - 6
Init - 7
Init - 8
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Valid Month
(b) Predictable Component [NMME]
PC
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Valid Month
(c) Signal Variability [σ_{sig}]
σ_{sig} (K)
1.5
1.0
0.5
Valid Month
(d) Total Variability [σ_{tot}]
σ_{tot} (K)
1.5
1.0
0.5
Valid Month



(a) Reg: NMME Nino34 (Dec) vs Forecast SST (Mar Init)
Step: Month 0.5
20°N
10°N
0°
10°S
20°S
120°E
180°
120°W
−1.2
−0.6
0.0
0.6
1.2
Regression
(b) GFDL-FLOR SST Bias
Step: Month 0.5
Bias (K)
NW Pacific (R=0.48)
SE Pacific (R=0.12)
1995-03
2000-03
2005-03
2010-03
2015-03
2020-03
(c) GFDL-FLOR Nino 3.4 Bias
Step: Month 9.5
Bias (K)
Nino3.4 Adjusted (MAE=0.68)
Nino3.4 RAW (MAE=0.79)
(d) Bias Correction Flowchart
Analyze Ensemble Sensitivity
Correlation Significant?
No
No Action
Yes
Calculate Bias of March SST Prediction
Correction = Bias x Regression
Apply Correction to December Prediction



(a) Step: Month 0.5
Valid in Mar
(b) Step: Month 1.5
Valid in Apr
(c) Step: Month 2.5
Valid in May
(d) Step: Month 3.5
Valid in Jun
(e) Step: Month 4.5
Valid in Jul
(f) Step: Month 5.5
Valid in Aug
60°N
40°N
20°N
0°
20°S
40°S
60°S
0°
60°E
120°E
180°
120°W
60°W
−0.4
−0.2
0.0
0.2
0.4
Correlation Coefficient


(a) Mar-Init Forecast
Error Fit
10% IC Error
50% IC Error
Forecast Error
RMSE (K)
Lead Time (Month)
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.5
2.5
3.5
4.5
5.5
6.5
7.5
8.5
9.5
(b) Jul-Init Forecast
Error Fit
10% IC Error
50% IC Error
Forecast Error
RMSE (K)
Lead Time (Month)