



Toward an optimal observational array for improving two flavors of El Niño predictions in the whole Pacific

Meiyi Hou^{1,2} · Youmin Tang^{2,3} · Wansuo Duan^{4,5} · Zheqi Shen^{1,2}

Received: 6 January 2022 / Accepted: 6 May 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

This paper investigates the optimal observational array for improving the El Niño-Southern Oscillation (ENSO) prediction by exploring sensitive areas for target observations of two types of El Niño events in the whole Pacific. A target observation method based on the particle filter and pre-industrial control runs from six coupled model outputs in Coupled Model Intercomparison Project Phase 5 (CMIP5) experiments are used to quantify the relative importance of the initial accuracy of sea surface temperature (SST) in different Pacific areas. The initial accuracy of the tropical Pacific, subtropical Pacific, and extratropical Pacific can all exert influences on both types of El Niño predictions. The relative importance of different areas changes along with different lead times of predictions. Tropical Pacific observations are crucial for decreasing the root mean square error of predictions of all lead times. Subtropical and extratropical observations play an important role in decreasing the prediction uncertainty, especially when the prediction is made before and throughout the boreal spring. To consider different El Niño types and different start months for predictions, a quantitative frequency method based on frequency distribution is applied to determine the optimal observations of ENSO predictions. The final optimal observational array contains 31 grid points, including 21 grid points in the equatorial Pacific and 10 grid points in the North Pacific, suggesting the importance of the initial SST conditions for ENSO predictions not only in the tropical Pacific but also in the area outside the tropics. Furthermore, the predictions made by assimilating SST in sensitive areas have better prediction skills in the verification experiment, which can indicate the validity of the optimal observational array designed in this study.

Keywords ENSO · Initial error · Particle filter · Target observation

1 Introduction

The El Niño-Southern Oscillation (ENSO) is the dominant mode of interannual climate variability on Earth, alternating between warm (El Niño) and cold (La Niña) conditions, which are centered in the central and eastern equatorial

Pacific (Bjerknes 1969; Philander 1983; Webster and Yang 1992). It has long been a focus of exploration because of its profound influences on the tropical climate and even the global climate (Alexander et al. 2002; Andrews et al. 2004; Hoell et al. 2018; Wu et al. 2018; Zhang et al. 2016a). In theory, the self-sustaining nature of the ENSO is conducive to its potential predictability up to two years in advance (Liu 2021; Tang et al. 2018). However, most current real-time ENSO predictions, made by the existing dynamical and statistical models, can only provide useful skills two or three seasons in advance (Barnston et al. 2012; Jin et al. 2008; Liu 2021). More interestingly, even though the models are constantly improving, the El Niño prediction skill during 2002–2011 is the lowest for the 1981–2010 period (Barnston et al. 2012). In addition, most models substantially overestimated the amplitude of warming for the 2014/2015 event when the prediction was initialized around June 2014 (McPhaden 2015). It is notable that although much progress has been made in understanding the ENSO mechanism and

✉ Youmin Tang
ytang@unbc.ca

¹ Key Laboratory of Marine Hazards Forecasting, Ministry of Natural Resources, Hohai University, Nanjing, China

² College of Oceanography, Hohai University, Nanjing, China

³ Environmental Science and Engineering, University of Northern British Columbia, Prince George, BC, Canada

⁴ State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

⁵ University of Chinese Academy of Sciences, Beijing, China

improving the physical parameterizations in models in recent decades, El Niño prediction is only beyond a certain lead time (Tang et al. 2018).

ENSO diversity is certainly a crucial factor that hampers the skill of the ENSO prediction. Since the 1990s, a new flavor of El Niño (denoted as CP-El Niño) has occurred more frequently (Yu and Kao 2007). Different from the canonical El Niño (denoted as EP-El Niño), the largest warming SST anomalies (SSTA) center of CP-El Niño during peak time is located in the central tropical Pacific, instead of in the eastern tropical Pacific (Kao and Yu 2009; Kug et al. 2009). The spatial differences in the two types of El Niño events increase the difficulty of the ENSO prediction regarding both structure and intensity. Moreover, the difference between these two types of El Niño events is also reflected in other aspects, including the formation mechanism, evolutionary process, and climate influence (Ashok et al. 2007; Capotondi et al. 2015; Chen et al. 2015; Fang and Mu 2018; Guo et al. 2021; Ren and Jin 2013; Weng et al. 2007; Zhang et al. 2011, 2016b). Considering these differences, it is necessary to discern which type of El Niño will occur while making El Niño predictions. Indeed, machine learning (ML) techniques show remarkable progress in seasonal predictions, which can beat the traditional models (Dijkstra et al. 2019; Ham et al. 2019; Liu 2021). For example, Ham et al. 2019 found that the convolutional neural network (CNN) model can distinguish the types of El Niño 12 months in advance with high skill. Compared to ML models, the current classical statistical and dynamical models still have a large room for improvement in predicting the types of the El Niño events (Ren et al. 2018).

Great efforts have been made to obtain skillful predictions of two types of El Niño events. From the perspective of ENSO predictability dynamics, one effective way to improve ENSO prediction is to decrease the error of the initial condition, which is related to the first kind of predictability problem (Lorenz 1975). It has been acknowledged that the accuracy of the initial condition is of great importance to ENSO prediction (Chen et al. 1995, 2004; Duan and Hu 2016; Gao et al. 2016; Moore and Kleeman 1996; Tao et al. 2017, 2018). Thus, observations are crucial not only for understanding the ENSO mechanism but also for improving prediction skills. However, launching a large and intensive observation network is costly. Therefore, the best approach is to employ optimal observations in some “key areas” or “sensitive areas”, which exert the largest influences on ENSO prediction (Mu et al. 2015). The question is how to detect and locate these sensitive areas.

Usually, two approaches are utilized to determine the optimal observation locations. One approach seeks the largest growth of initial errors, which are mostly assumed to affect the prediction. This approach includes methods such as singular vector (SV; Palmer et al. 1998; Tang

et al. 2006), breeding vector (BV; Toth and Kalnay 1997), adjoint sensitivity (Bergot 1999), conditional nonlinear optimal perturbation (CNOP; Mu et al. 2003; Duan et al. 2018b) and other uncertainty analysis approaches of prediction (Hou et al. 2019; Zhang et al. 2015). By focusing on the error growth, these methods help locate the most sensitive area where the initial errors grow the most dramatically and unavoidably interfere with the prediction. Most research on detecting ENSO optimal observational arrays uses this kind of method. In particular, Duan et al. (2018b) designed an optimal observational array for ENSO prediction in the tropical Pacific by using the CNOP method. The other kind of approach is based on assimilation methods, including the ensemble transform (Bishop and Toth 1999), ensemble transform Kalman filter (ETKF; Bishop et al. 2001), and ensemble Kalman filter (EnKF; Liu and Kalnay 2008; Wu et al. 2020). However, the EnKF and its variants assume that both the model errors and observation errors are Gaussian. The Kalman filter, which is the fundamental basis of EnKF, is only applied to linear state-space systems. As such, a new assimilation method referred to as the particle filter (PF; Gordon et al. 1993; Van Leeuwen 2009; Shen et al. 2017) has recently attracted broad attention and is appropriate for any non-Gaussian and nonlinear system.

By using an offline numerically efficient method, Kramer and Dijkstra (2013; hereafter referred to as KD13) applied the PF to explore the predictability barrier for two types of El Niño events in the tropical Pacific domain (also see Duan et al. 2018a). By performing an identical twin approach, they proposed an offline approach without model forward integration to update the weights of particles (ensemble members). In this way, they discovered that the initial accuracy of the SST in the tropical Pacific near the Niño3 and Niño4 areas is very significant for ENSO predictions.

The aforementioned studies on the target observations for ENSO predictions applied the PF but were limited within the tropical Pacific area. However, numerous recent studies have indicated that the subtropical Pacific is also important to ENSO formation and its predictability (Chang et al. 2007; Lin et al. 2015; Lu et al. 2017; Zhang et al. 1998). Specifically, the North Pacific Meridional Mode (NPMM) is more closely related to the formation of CP-El Niño events while the South Pacific Meridional Mode (SPMM) has a greater effect on EP-El Niño events (Ding et al. 2015, 2017; Min et al. 2017; Vimont et al. 2014; Yu et al. 2010). Furthermore, Hou et al. (2019) and Qi et al. (2021) investigated the impact of the initial accuracy of the tropical and extratropical ocean temperatures in the Pacific on ENSO predictions from the perspective of error growth. They showed that the accuracy of the extratropical Pacific temperature also exerts large influences on the ENSO prediction, especially on the prediction of El Niño types.

As previously discussed, the most effective way to improve the ENSO prediction is to increase the number of observations and assimilate them into the model prediction system. The initial accuracy of the temperature in the whole Pacific, including the tropical Pacific, subtropical Pacific, and extratropical Pacific, may all be important for distinguishing El Niño types. Under these circumstances, it is urgent to address several major issues to improve the skill of the two types of El Niño predictions. First, to what extent does the initial accuracy of the extratropical Pacific matter to the two types of El Niño predictions, and is its importance comparable to that of the tropical Pacific? Second, whether and how does the sensitive area change with lead times and initial condition of predictions? Third, how do we determine an optimal observational array position that is little dependent on the model, the lead time, and the initial condition of prediction, to ensure that the practical oceanic buoy observation is long-time standing and robust?

In this study, we utilized the PF method in KD13 to seek the optimal observational locations through the whole Pacific for ENSO type predictions. The paper is organized as follows: In Sect. 2, datasets and the PF method are described. In Sect. 3, the assimilation experiments carried out in this study are depicted clearly. In Sects. 4 and 5, the core of the paper, we quantify the relative importance of the observations in the different Pacific areas for two types of El Niño predictions. In Sect. 6, we design the array of the optimal observational array for ENSO predictions. In Sect. 7, the results of the verification experiments are shown. Finally, in Sect. 8, we present our summary and discussion.

2 Datasets and PF assimilation methodology

Coupled Model Intercomparison Project Phase 5 (CMIP5) provides abundant global coupled model data resources. In this study, we use outputs from CMIP5 preindustrial control (piControl) experiments, in which the models are driven for at least 500 years after spin-up running under

a constant external forcing (greenhouse gas, solar radiation, aerosol, land use, etc.) at the level of the year 1850. Thus, the integration results in piControl experiments only include signals of internal variability. Massive analysis of the ENSO simulation ability of the CMIP5 models has been conducted (Bellenger et al. 2014; Ham and Kug 2012; Ren et al. 2016). A consensus has been reached that only some CMIP5 models can capture the main features of both flavors of El Niño events, especially the CP-El Niño. Referring to Kim and Yu (2012), six models that can reasonably simulate two types of El Niño events were chosen in our work. Specific model configurations and affiliations are listed in Table 1. SST data are obtained from the output datasets of the six coupled models. It is noted that models have different integration times and different spatial resolutions. To simplify the calculations, we choose the first 500 years of the integration in each model. The variable is interpolated onto the same grids with a resolution of $2.5^\circ \times 2.5^\circ$ by using the bilinear interpolation method. All anomalies are computed by removing their monthly climatological mean.

To assimilate observations into the prediction ensemble, we use the PF method in KD13. The PF method is a sequential Monte Carlo method using particles (samples) to estimate the probability density functions (PDFs). The core of this assimilation method is to change the weight of each particle by assimilating observation data. Specifically, the mathematical expression, based on Kramer et al. (2012) and KD13, is presented as follows: The starting point is an ensemble of size N of model states \mathbf{X}_k^i , referred to as particles, that represent the prior PDF $p_N(\mathbf{X}_k)$, as.

$$p_N(\mathbf{X}_k) \approx \sum_{i=1}^N w_k^i \delta(\mathbf{X}_k - \mathbf{X}_k^i) \tag{1}$$

Herein, $\delta(\bullet)$ is the Dirac delta function over real numbers, whose value is zero anywhere except at zero and whose integral over the entire real line is equal to one. The PDF of state vector \mathbf{X}_k is estimated by “particles”, i.e., ensemble members, $\mathbf{X}_k^i (i = 1, 2, \dots, N)$, multiplied by weights w_k^i of these particles. At the beginning ($k = 0$), the

Table 1 List of models from the IPCC AR5 analyzed in this study

Model	Institute/country	Resolution (lon × lat, vertical)	
		Atmosphere	Ocean
CCSM4	NCAR/USA	288 × 192, L26	320 × 384, L60
CESM1-BGC	NSF-DOE-NCAR/USA	288 × 192, L26	320 × 384, L60
CMCC-CMS	CMCC/Italy	192 × 96, L95	182 × 149, L31
CNRM-CM5	CNRM-CERFACS/France	256 × 128, L31	362 × 292, L42
GFDL-CM3	NOAA GFDL/USA	144 × 90, L24	360 × 200, L50
GISS-E2-R	NASA GISS/USA	144 × 90, L24	288 × 180, L32

(All the datasets are available online at <https://esgf-node.llnl.gov/search/cmip5/>, and expansions of acronyms are available online at <http://www.ametsoc.org/PubsAcronymList>)

w_0^i of each particle is identical, which is equal to $1/N$. An observation \mathbf{Y}_k then becomes available at time $t = t_k$, which can be assimilated to obtain the posterior PDF $p_N(\mathbf{X}_k|\mathbf{Y}_k)$ by Bayes' Theorem:

$$p_N(\mathbf{X}_k|\mathbf{Y}_k) = \frac{p(\mathbf{Y}_k|\mathbf{X}_k)p_N(\mathbf{X}_k|\mathbf{Y}_{k-1})}{p(\mathbf{Y}_k)} \quad (2)$$

By using Eq. (1) and (2), we can update the weight at t_k , which is

$$w_k^i = \frac{p(\mathbf{Y}_k|\mathbf{X}_k^i)}{p(\mathbf{Y}_k)} w_{k-1}^i \quad (3)$$

In Eq. (3), $p(\mathbf{Y}_k|\mathbf{X}_k^i)$ is the PDF of the observations \mathbf{Y}_k given the model state \mathbf{X}_k^i , and $p(\mathbf{Y}_k)$ is the PDF of the observation. Noted that $p(\mathbf{Y}_k)$ can be regarded as a normalization factor, which ensures that the sum of the weights is equal to one. $p(\mathbf{Y}_k|\mathbf{X}_k^i)$ is directly related to the (known) probability distribution of the observation error. If the observations are measured univariately with a Gaussian distribution for the measurement error, with variance Σ , then

$$p(\mathbf{Y}_k|\mathbf{X}_k^i) \sim \exp\left[-\frac{1}{2}(\mathbf{Y}_k - H(\mathbf{X}_k^i))^T \Sigma^{-1}(\mathbf{Y}_k - H(\mathbf{X}_k^i))\right] \quad (4)$$

Here, H is the observation operator, which can be calculated by simply selecting the model equivalents from the full state vector. The weight w_k^i can be calculated from Eqs. (3) and (4). Also, if several observations at different grids are assimilated simultaneously, the weight w_k^i is updated as follows:

$$w_k^j \sim \exp\left[-\frac{1}{2} \sum_{j=1}^m (\mathbf{Y}_k - H(\mathbf{X}_k^j))^T \Sigma^{-1}(\mathbf{Y}_k - H(\mathbf{X}_k^j))\right] \quad (5)$$

for $j = 1, 2, \dots, m$

The abovementioned method of weight updating is known as the sequential importance sampling (SIS) method, which is a useful PF algorithm for designing experiments in our work. However, the major problem of SIS is that after assimilating the observations at $t = t_k$, the weight is concentrated on only a small number of particles, which is referred to as the degeneracy of the particles. A strongly degenerated ensemble, where only a few ensemble members have weights, cannot yield a reasonable prediction ensemble for the predicted variable, such as the Niño3 and Niño4 SSTAs. A basic solution to avoid degeneracy is to perform resampling. Namely, the particles with high weight will be duplicated, and particles with low weight will be discarded. In addition, setting the proper magnitude of the error covariance is also important to

avoid degeneracy. If the observation error is set too small, only particles that are close to the observation remain, which will cause large degeneracy. However, it will be unrealistic if we set observation errors that are too large. In this study, after performing tuning experiments, we set the observational error to $0.3 \delta_T$, where δ_T is the standard deviation of SST.

The increase in prediction utility by assimilation observations can be evaluated by the predictive power (PP; Schneider and Griffies 1999) and root mean square error (RMSE). Herein, PP is defined by

$$PP = 1 - \exp(IE_{X_{new}} - IE_{X_{init}}) \quad (6)$$

where $IE_{X_{init}}$ and $IE_{X_{new}}$ are the information entropy of the prediction before assimilating observations and after assimilating observations, respectively. The information entropy can be estimated by using the PDF of the ensemble in the following manner:

$$IE_X = - \sum_a p_a(\mathbf{X}) \ln p_a(\mathbf{X}) \quad (7)$$

here, $p(\mathbf{X})$ is the PDF of the prediction ensemble, which is obtained by using the PF method following Eq. (1). To calculate the entropy, $p(\mathbf{X})$ could be cut into a bins, which are represented by $p_a(\mathbf{X})$. The entropy can be calculated if we choose a proper a . In this study, we set $a = \sqrt{N} + 1$. Thus, by calculating the entropy of the initial ensemble $IE_{X_{init}}$ and the new ensemble $IE_{X_{new}}$ after assimilation, PP can be calculated from Eq. (6). The information entropy measures the uncertainty level of the ensemble. Therefore, PP presents a decrease in uncertainty due to the assimilation of observations. The larger the PP is, the greater the decrease in the uncertainty of the ensemble prediction. In addition, the RMSE, a commonly used measure, is also calculated to assess the assimilation performance, which is defined by

$$RMSE = \sqrt{\frac{1}{z} \sum_{d=1}^z (\bar{X}_d - Y_d)^2} \quad \text{for } d = 1, 2, \dots, z \quad (8)$$

here, d denotes the grid index and z is the total number of grid points of the entire computational domain.

As previously illustrated, the core of the PF method is to change the weight of each ensemble member according to the observation information. Thus, this assimilation method can be applied not only to model forward integrations but also to offline model ensemble prediction datasets. In this paper, all assimilation experiments are conducted by using offline model datasets from CMIP5. In this way, several models can be involved comprehensively to obtain a model-independent result. The details of the assimilation experiments for detecting the sensitive area for ENSO prediction are introduced in the following section.

3 Experiment design

To identify the most sensitive area in terms of the improvement in the ENSO intensity prediction skill in its mature phase, we opt to use Niño indices of the boreal winter as the major prediction targets. The definitions of two types of El Niño events, defined by Kug et al. (2010), are employed here. Namely, we use Niño3 and Niño4 SSTA [i.e., the SST anomaly averaged over the Niño3 area and the Niño4 area] to represent the EP- and CP-El Niño events and their intensities. An El Niño event occurs if at least one of the two SSTAs exceeds 0.5 °C in the boreal winter (November, December, and January in the next year). Then, if the Niño3 (Niño4) SSTA is greater than the Niño4 (Niño3) SSTA, it is considered as an EP- (CP-) El Niño event. In this way, 13 typical EP- El Niño events and 13 typical CP- El Niño events are chosen from each of the six models. The spatial and temporal characteristics of these El Niño events are shown in Fig. 1. The spatial patterns of El Niño are similar in the six models. The Niño indices of all the typical El Niño events first increase gradually until the mature phase and then decrease in the next year.

All assimilation experiments are conducted in the framework of the identical twin experiment by utilizing the piControl model outputs from CMIP5. As illustrated in KD13, the synthetic truth is selected from one model realization, and the observations are produced by adding a normal-distributed observation random noise. We use the same approach to fabricate the observation data. Specifically, we divide the 500-year integration into 500 one-year segments and choose a single one-year period (e.g., from January to December) of a typical EP- or CP-El Niño year as a truth run, from which the “observation” is made by adding a normally distributed observation error to the truth. The other 499 1-year integrations can all be regarded as its “predictions” up to a lead time of 11 months. These predictions altogether compose a prediction ensemble for this specific EP- or CP-El Niño event. These ensemble members are assumed to be independent, although there is a correlation between one year and the next year. However, if we leave the odd years out, the ensemble member will be only half of the entire ensemble, which hinders sample diversity. We have to choose larger sample numbers instead of better sample independence. In addition, choosing segments less than one year is also inadvisable because ENSO dynamics are very seasonally dependent. Therefore, it is reasonable to obtain the observation and the prediction ensemble in this way.

The PF method is used to conduct offline assimilation experiments via Eq. (4). The basic principle here is to assimilate only one observation data, such as SSTA, in one single grid among the Pacific in one experiment. After assimilation, the improvement in the prediction utility is calculated

to evaluate the importance of this observation. The next assimilation is conducted by using another grid observation. In this way, the assimilation process is repeated until all observations in the Pacific are evaluated. The most sensitive area targeting ENSO prediction improvement can be located by comparing the improvement in the prediction skill among all assimilation experiments. In addition, assimilation experiments are conducted repeatedly by using different assimilation times from January to December because we also want to address the question of whether and how the sensitive areas change with different prediction lead times.

4 Impact of observations for monitoring CP-El Niño events

To obtain a less model-dependent result, the assimilation experiments are conducted by using 78 (6 models*13 events) CP-El Niño events as observations, and then the assessment is performed by making a composite of all results. Two important metrics, PP and RMSE decrease, are employed to determine the optimal observations. Our main purpose is to improve the El Niño prediction in its mature phase. Thus, the evaluation target is the improvement in the prediction skill only in December. Specifically, the weights of 499 ensemble members are updated by assimilating SST in January (or other months), and then the weights are multiplied by the ensemble of the December Niño4 index to generate the prediction of CP-El Niño in December, with a lead time of 11 months (or other leads).

The spatial pattern of the averaged PP over 78 CP-El Niño cases is shown in Fig. 2, obtained by assimilating observations at different times from January to November. For example, Fig. 2a indicates the PP value of December Niño4 SST index prediction by assimilating January SST, whereas Fig. 2k is the PP value of December Niño4 SST index prediction by assimilating November SST. It should be noted that the PP value at one grid in Fig. 2 is the resultant PP value of the Niño4 SSTA contributed by the assimilation of this grid’s observation. Thus, the location with a high PP indicates that its observation has a high impact on the prediction of the December Niño4 SSTA. Therefore, the regions with high PP can be determined as optimal observation locations. It is obvious in Fig. 2 that the signal changes along with the assimilation time. Centers of large PP values are mainly located in three areas, including the equatorial Pacific, the North Pacific, and the South Pacific. Signals continually gather around the Niño3 and Niño4 areas. However, the largest PP is located first in the Niño3 area and then moves to the Niño4 area after July. The value of PP in the equatorial Pacific decreases from January to April and then increases after May. It seems natural and intuitive for the PP value to increase as the lead time decreases in

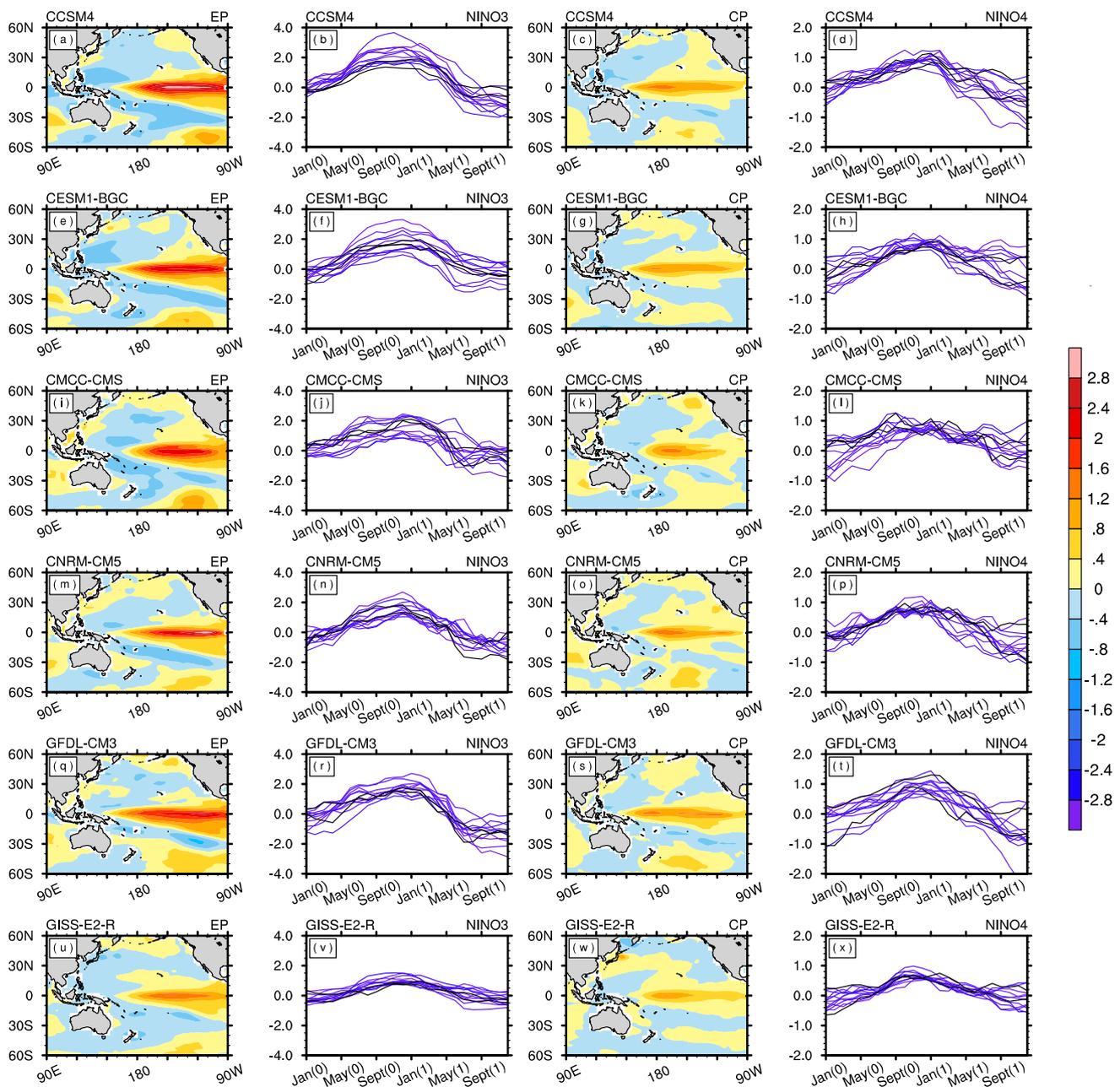


Fig. 1 The spatial and temporal composite pattern of the chosen 13 typical EP- and CP-El Niño events in six CMIP5 models. The truth field of SSTA (units: °C) in the boreal winter of 13 EP- and CP-El Niño events are given in the first and the third column. The evolution of SSTA (units: °C) in Niño3 and Niño4 areas of EP- and CP-El Niño

events are shown in the second and the fourth column. On the X-axis, month (0) represents the month of El Niño attaining peak year, and month (1) represents the month of decaying the El Niño year. Different rows correspond to different models

the equatorial Pacific from July to November, as shown in Fig. 2, in that the observation can offer more information and stronger predictable signals as it is closer to the prediction. However, the opposite situation could also occur when some key teleconnection processes contribute to predictable signals, for example, a delayed impact of the western Pacific Ocean on the ENSO, as described by the delayed-oscillator

mechanism. This may explain why the PP value decreases from January to April, as shown in Fig. 2.

In terms of the PP in the North Pacific area, one large center is located at approximately 40° N in the northwest Pacific near the Kuroshio Extension. The value there increases from January, peaks near August, and then starts to decrease. Another large center is observed over the northeast

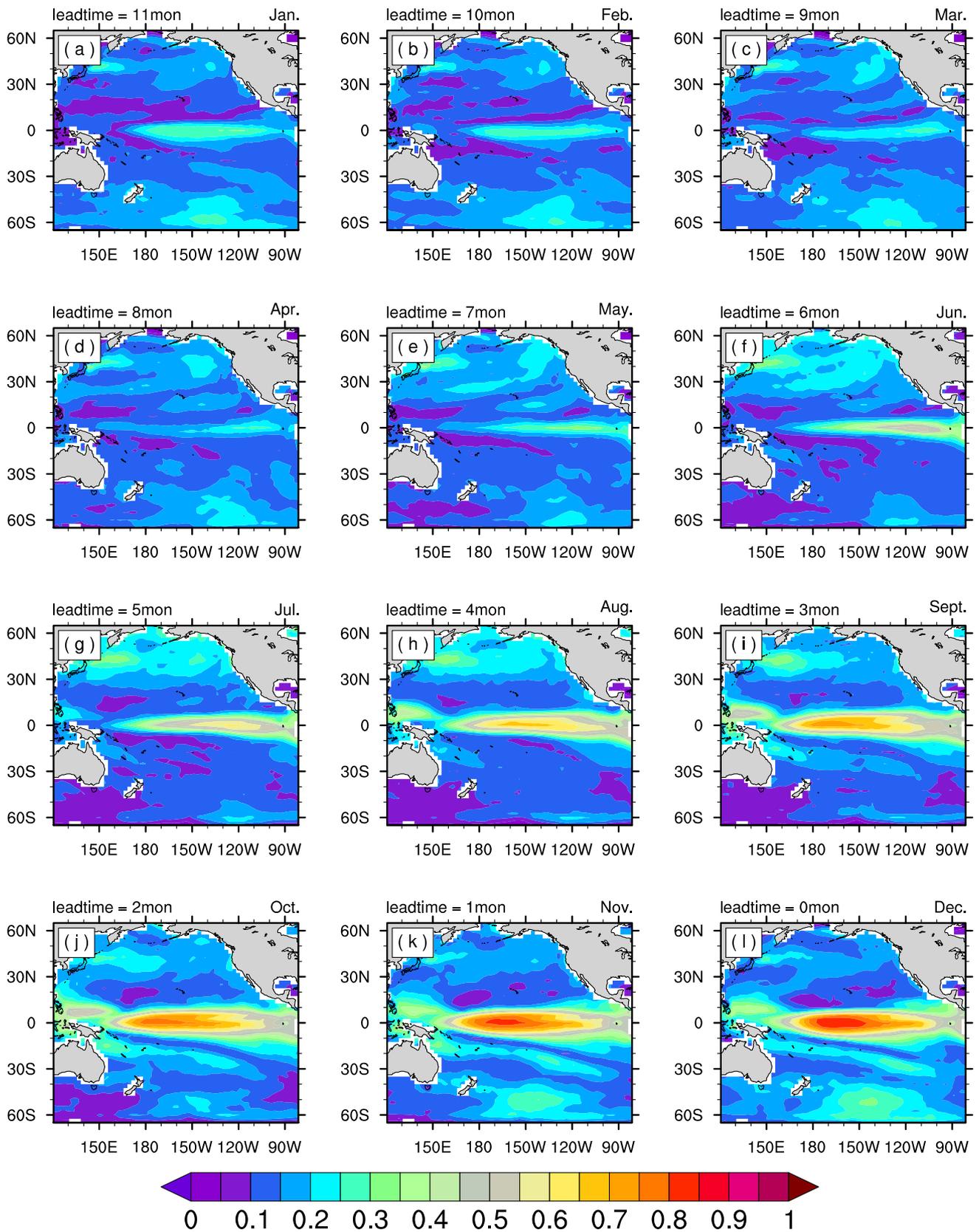


Fig. 2 The spatial pattern of the predictive power, PP(NINO4), averaged over 78 CP-El Niño cases, obtained by assimilating observation for a given location as a function of assimilation time from a January to l December

Pacific, which has a similar spatial pattern to the NPMM near Baja California in Mexico. This pattern is clearly seen from March to June. There are two large centers of PP in the South Pacific. One is located over the extratropical Pacific and is centered at approximately 130°W – 140°W , 50°S – 60°S . The other is located at approximately 30°S and has a similar spatial pattern to SPMM. The former fades from January to September and rises in October. The latter emerges in June and matures in November and December. The PP spatial pattern here shows agreement with the error pattern of SST, which interferes with CP-El Niño predictions, as shown in Fig. 7 by Hou et al. (2019). The importance of the initial accuracy in the extratropical Pacific to CP-El Niño predictions is also emphasized by Hou et al. (2019).

The PP pattern shown in Fig. 2 is a composite of all the cases and models. Composites for the cases of each model were also calculated (not shown), and all bear great resemblances to the composite of all models. That is, there are five apparent large centers of PP related to the December CP-El Niño predictions, including the tropical Pacific (TP; 120°E – 80°W , 23.5°S – 23.5°N), subtropical North Pacific (SNP; 23.5°N – 40°N), subtropical South Pacific (SSP; 23.5°S – 40°S), extratropical North Pacific (ENP; 40°N – 66.5°N) and extratropical South Pacific (ESP; 40°S – 66.5°S).

To quantify the importance of target observations in these five areas, we calculated the area averages of PP, as shown in Fig. 3. Here, the signal evolution in different areas is well illustrated by PP (NINO4) in Fig. 3. It is shown that the observations in the TP are essential after August compared with those in other areas (Fig. 3f). The PP average value over the TP starts to increase dramatically after late spring or early summer and peaks in October (Fig. 3a) in almost all cases, which may be related to the spring persistence barrier for ENSO prediction. The spring persistence barrier is a phenomenon in which the persistence of the ENSO SSTA drops significantly in late spring, which can lead to the ENSO predictability barrier in spring. Herein, the PP value can be considered as a precursor signal of ENSO events. Thus, it is reasonable that the signal in the tropical area is indeed much smaller before boreal spring than that after summer in Fig. 3a. However, the signal outside the tropical area is slightly larger in the first half-year. In terms of the PP averaged over the SSP and ESP, the PP value peaks in the boreal winter and is larger than that of TP before May. For the PP averaged over the SNP and ENP, the PP value peaks in the boreal summer and is larger than that of TP before July. As suggested early, PP is closely related to the change in the uncertainty of the ensemble prediction. Thus, Figs. 2 and 3 imply that when predicting CP-El Niño events before boreal spring, the initial conditions of SST outside the tropical area

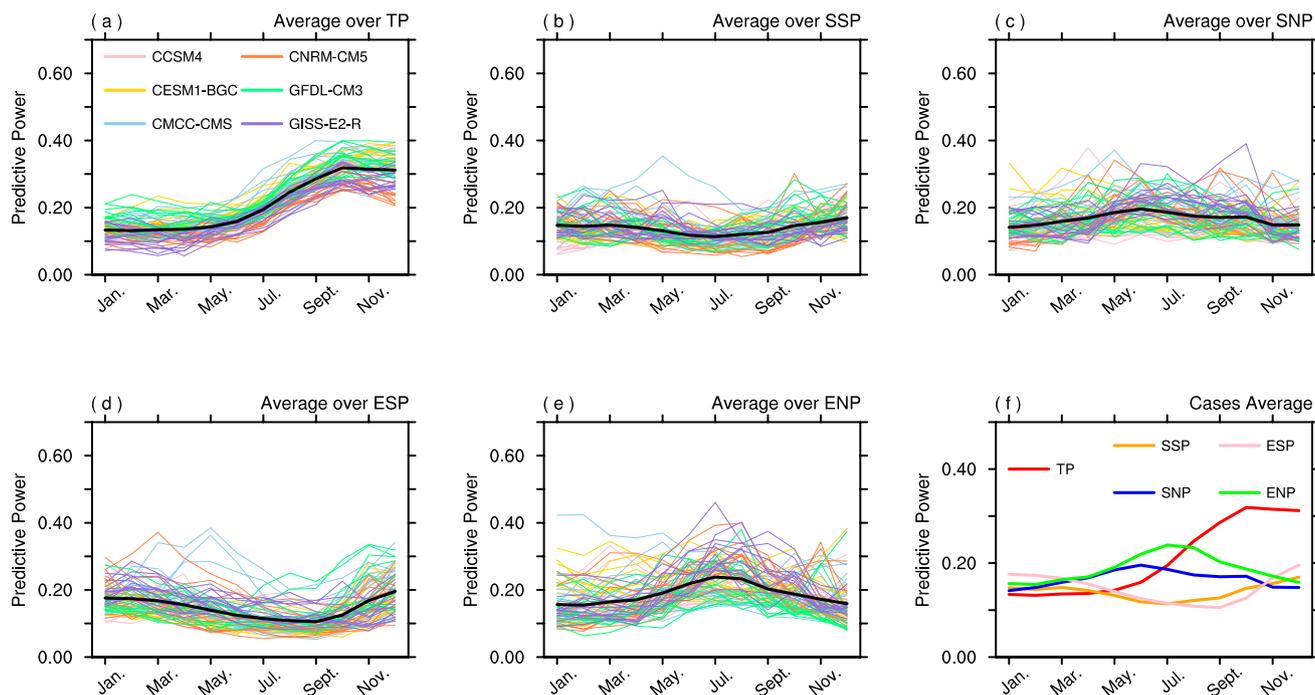


Fig. 3 Area average of PP(NINO4) over **a** tropical Pacific, **b** subtropical south Pacific, **c** subtropical North Pacific, **d** extratropical South Pacific, and **e** extratropical North Pacific as a function of assimilation time (X-axis). Different colors represent different model cases. All

colorful lines represent the results of 78 CP-El Niño prediction cases. The black lines represent the average of all cases. And all the black lines in **a**–**e** are replotted in **f** with different colors

in the Pacific are important and will help reduce the uncertainty of the prediction.

In the previous discussions, the PP measures the resultant decrease in prediction uncertainty from the decrease in initial errors by assimilation of observations. It is a metric of potential predictability in theory and could produce spurious results in an ill-designed overconfidence ensemble system, for example, a small ensemble spread and a far-from-truth ensemble mean. Thus, an actual prediction skill measure should be applied to evaluate the impact of potential observations on prediction. Here, we use RMSE for this purpose. Figure 4 shows the change in the RMSE of the Niño4 SSTA prediction due to assimilation, where the red areas represent the improvement in prediction skill after assimilating observations in these locations. Unlike the PP value, the large center of the RMSE decreases resulting from the SST observations is always in the tropical Pacific. The largest value is first located in the central equatorial Pacific, then moves eastwards in boreal spring, and finally moves back to the central equatorial Pacific. The spring persistence barrier still interferes with the prediction as shown in Fig. 4d since the background value is lowest in April. Observations in the South Pacific become more useful in the latter half of the year but still cannot be comparable to those in the equatorial Pacific. Furthermore, to check the result of all cases besides the composite result, we add up the number of the cases that give the same result (RMSE decreasing/increasing) as the composite one. The purple (or black) dots on the panels in Fig. 4 indicate that there are more than two-thirds (or three-fourths) of all cases in which the RMSE decreases after assimilating observations at that location, just the same as the composite result. It is noteworthy that there are some purple dots in the North and South Pacific regions from January to April, but the decrease in RMSE is not as large as that in the equatorial Pacific. Hence, assimilating observations in these North or South Pacific regions in late winter and early spring does improve the deterministic forecast skill of the CP-El Niño, but the improvement is quite limited.

5 Impact of observations for monitoring EP-El Niño events

Similar assimilation experiments are conducted by using 78 EP-El Niño synthetic observations. The spatial pattern of PP (NINO3), averaged over these 78 cases, is shown in Fig. 5. The spatial pattern of PP (NINO3) is similar to that of PP (NINO4) in the previous section. In the tropical Pacific, the optimal location for predicting the Niño3 index in December is more restricted to the eastern Pacific. During the beginning of the year, from January to March, the signal in the tropical Pacific is not as strong as that in the extratropical Pacific. The prediction of Niño3 suffers from a severe spring

predictability barrier, and south extratropical signals in January and north extratropical signals in JFMA can provide more predictability than the tropical Pacific. In this case, adding optimal observations in the extratropical Pacific and assimilating them into the model may attenuate the predictability barrier of EP-El Niño prediction.

The area average of PP (NINO3) is also analyzed, as illustrated in Fig. 6. It shows that the observations in the tropical Pacific after August are significant because the PP average of all cases increases from June to September (Fig. 6f). However, in the first half-year, the signals outside the tropical area are important for the prediction of the Niño3 index. By comparing Figs. 5 and 6 with Figs. 2 and 3, we can analyze the difference between the predictions of the two types of El Niño events. It seems that the seasonal predictability barrier is more severe in EP-El Niño predictions since the background value of PP in Fig. 5a–d is slightly less than that in Fig. 2a–d. Additionally, there are some spatial differences between PP (NINO4) and PP(NINO3) in the North and South Pacific. The PP (NINO3) in the northeast Pacific is quite small during the whole year compared with PP (NINO4) in the northeast Pacific (Fig. 5). The SPMM-like spatial pattern is stronger and persists longer in Fig. 5h–l than in Fig. 2h–l, which is in agreement with the notion that the SPMM is more related to the development of EP-El Niño events (Min et al. 2017) through the wind-evaporation-SST feedback (Xie and Philander 1994). Overall, this finding implies that extratropical SST initial conditions can affect the prediction of both types of El Niño events, but the extent of the effect can be different depending on the lead time of the prediction.

Similar to the assessment of the CP-El Niño assimilation experiments, the deterministic prediction skill is also evaluated by calculating the RMSE. Figure 7 shows that the signal in the tropical Pacific is the most significant at all times. There are some large centers outside 30 degrees latitude, but they are not as large as those in the tropical area. It is also noteworthy that the optimal observations for the most effectively improved Niño3 index predictions are found at approximately 170° W, 10° N in May (Fig. 7c). Similar patterns in the tropical Pacific, as shown in the CP-El Niño experiments, can be identified while comparing Fig. 7c–e with Fig. 4c–e, both of which bear some resemblances to NPMM. Thus, the observations in the NPMM-like region during spring are important for predictions of both EP- and CP-El Niño events.

6 Sensitive area for target observations

The previous sections evaluated the relative importance of the tropical, subtropical, and extratropical Pacific for two types of El Niño predictions at different lead times. We then

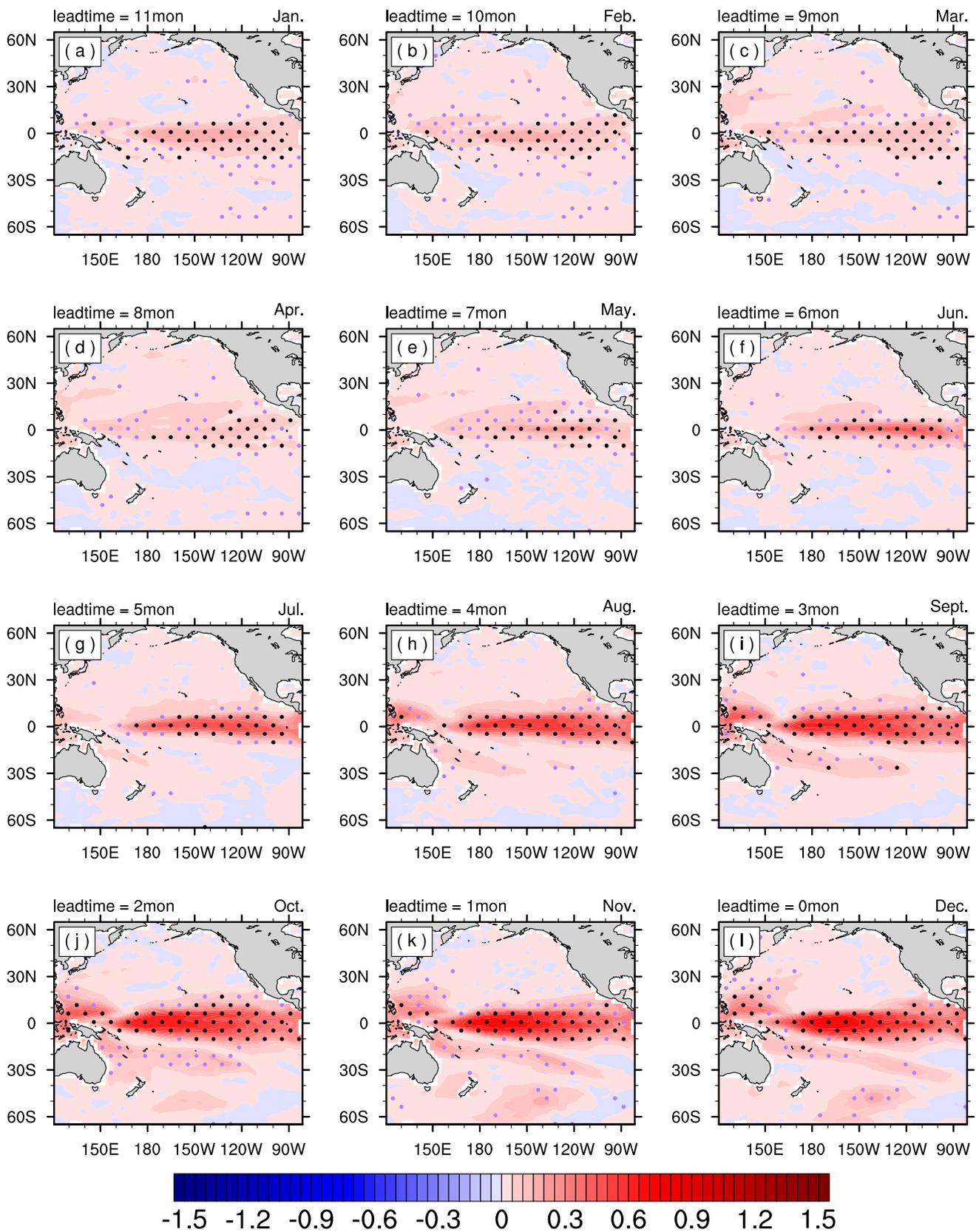


Fig. 4 As in Fig. 2, but for the results of the decrease in the RMSE of the prediction of Niño4 SSTA in December (units: °C). The area with purple dots means that the RMSE of the prediction decreases

(increases) in more than 2/3 of prediction cases while the composite results of all cases also decrease (increase) after assimilating the observation here. Similarly, the black dots mean greater than 3/4

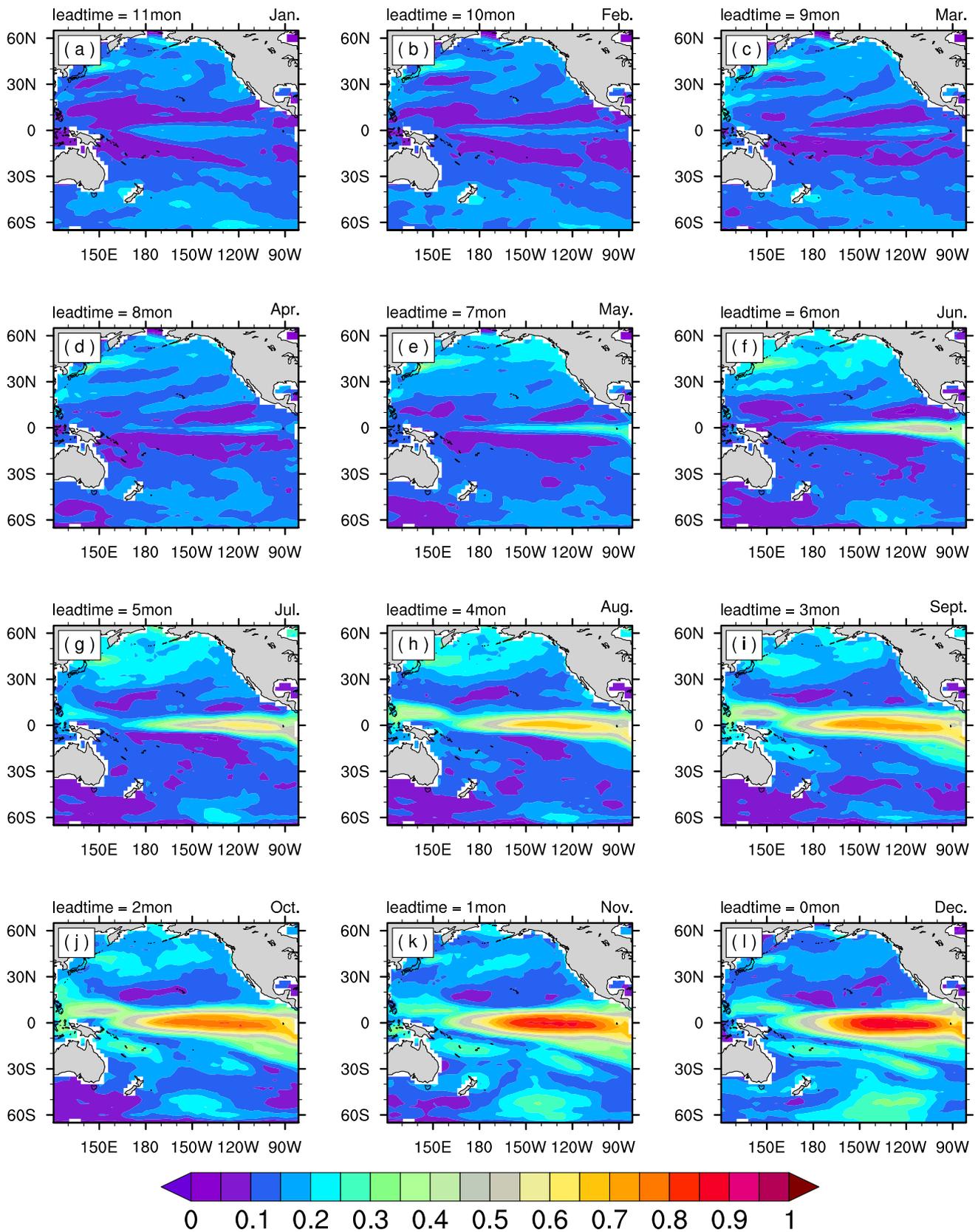


Fig. 5 As in Fig. 2, but for the results of PP (NINO3) targeting at EP-El Niño predictions

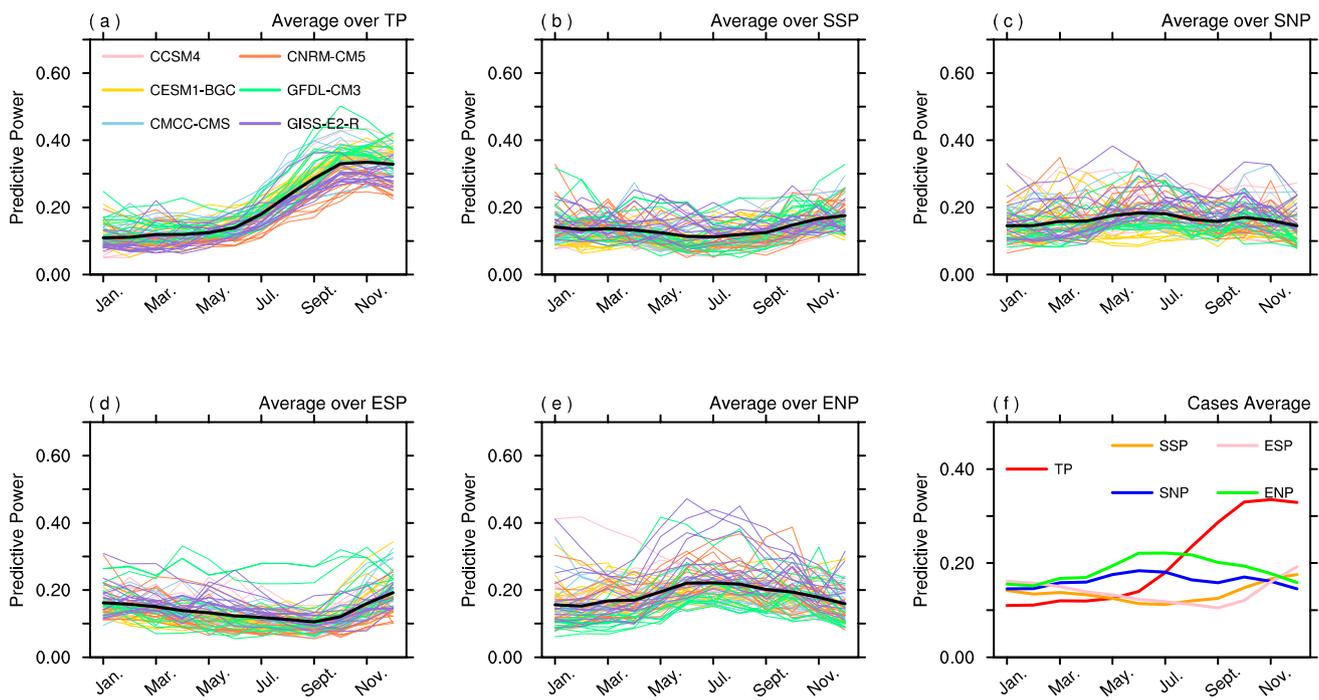


Fig. 6 As in Fig. 3, but for the results of PP (NINO3) targeting at EP-El Niño predictions

attempted to locate the optimal observations of SST for ENSO prediction considering these. As shown in the previous sections, the signals change with different lead times, which means that the optimal observations for ENSO prediction should be considered as a function of the start time of the prediction. In addition, it is unknown what type of El Niño events will occur when issuing a prediction; thus, the sensitive area should cover both types of El Niño events. Hence, we propose to locate optimal observations case by case, as in the general operations adopted in the target observation for ENSO in Duan et al. (2018b).

The main idea is to sort PP values. The details of the idea are illustrated as follows: First, we visit all the 156 prediction cases (78 CP-El Niño events and 78 EP-El Niño events) as previously mentioned and only select the spatial grid points with decreases in the RMSE. These selected spatial grid points are further sorted in descending order according to their PP value, and the top 15 grids, which are called PP max points hereafter, are identified. As a result, we can obtain a 12×156 (12 months and 156 cases) series of 15 PP max points. Second, we split these series into 4 groups, each containing 3×156 series (3 months and 156 cases), to use different months (January, April, July, and October) to start the prediction. Third, from each group of 3×156 samples, we compute the frequency of PP max that occurred for each grid point across the Pacific domain. To express this procedure more clearly, we use a formula to show the calculation of frequency, which is denoted by F , as follows:

$$F_{i,j}^t = \frac{c_{i,j}^t}{L} \times 100\% \quad (9)$$

where $t = 1, 2, 3$, and 4, represent different groups; $c_{i,j}^t$ is the number of the grid points (i, j) being the PP max points in the 3×156 series in the t group; and L is the “ 3×156 ” series. Finally, we choose grids with the first 10 (largest) F values as the optimal observation area for two flavors of El Niño predictions. The spatial distributions of the F value in different seasons are shown in Fig. 8, where the red dots represent the sensitive areas.

As shown in Fig. 8, the optimal observations in JFM are all located in the North Pacific, with 8 grid points in the extratropical Pacific near the Kuroshio Extension region and 2 grid points along the west coast of North America. In AMJ, the sensitive area contains 3 grid points in the northwest Pacific and 7 grid points on the equator in the eastern Pacific. In JAS and OND, all optimal observations are located on the equator. To obtain a long-time standing observational position and to consider all the seasons to start predictions, we propose combining the sensitive areas in different seasons to get an array with 31 grid points, as shown in Fig. 9, which includes 21 grid points in the equatorial Pacific and 10 grid points in the North Pacific.

The optimal observational array identified here implies the importance of the uncertainties of SST outside the tropical Pacific, especially the North Pacific, to both types of El Niño predictions. Several recent studies have shed light on

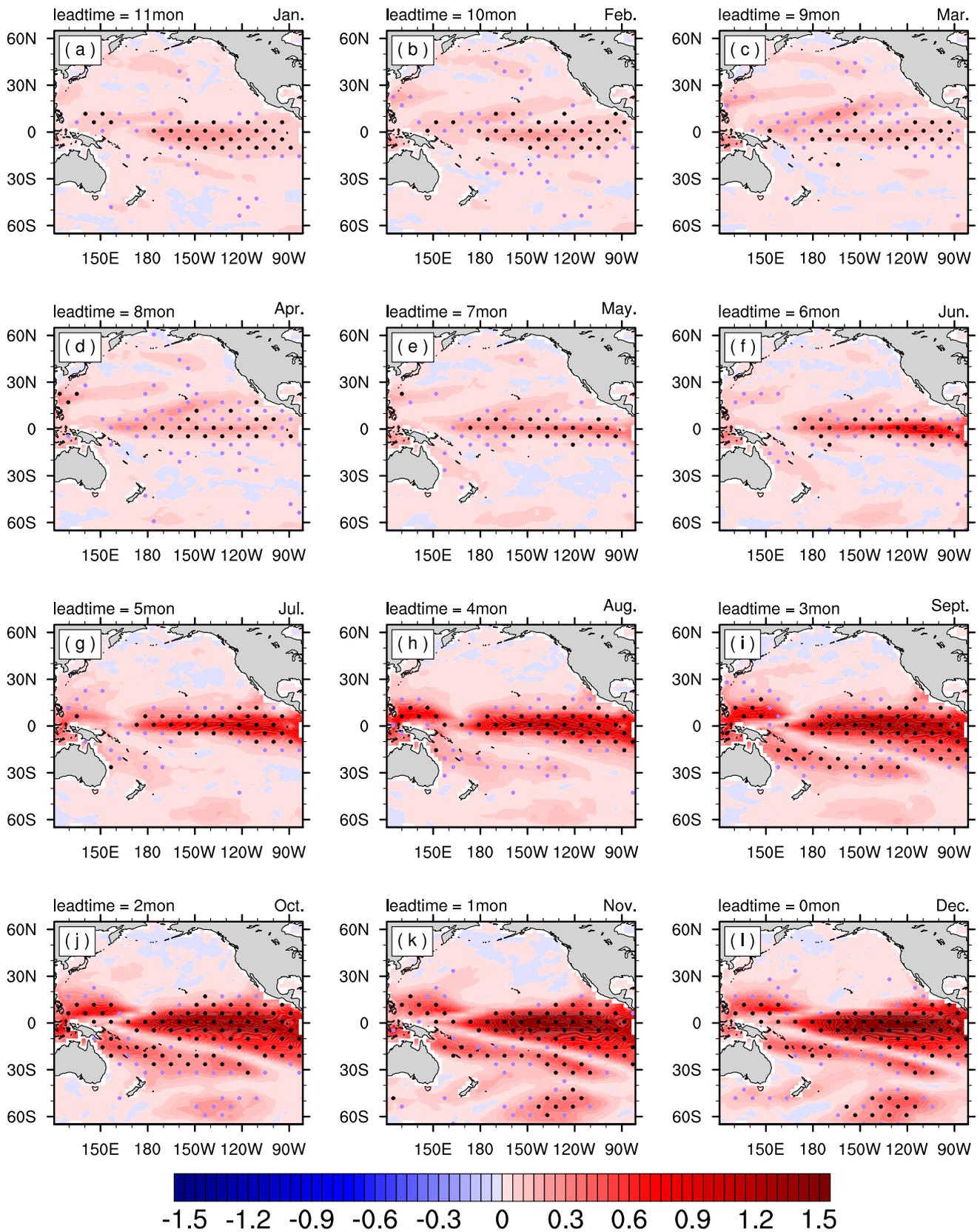


Fig. 7 As in Fig. 4, but for the results of the decrease in the RMSE of the prediction of Niño3 SSTA in December (units: °C)

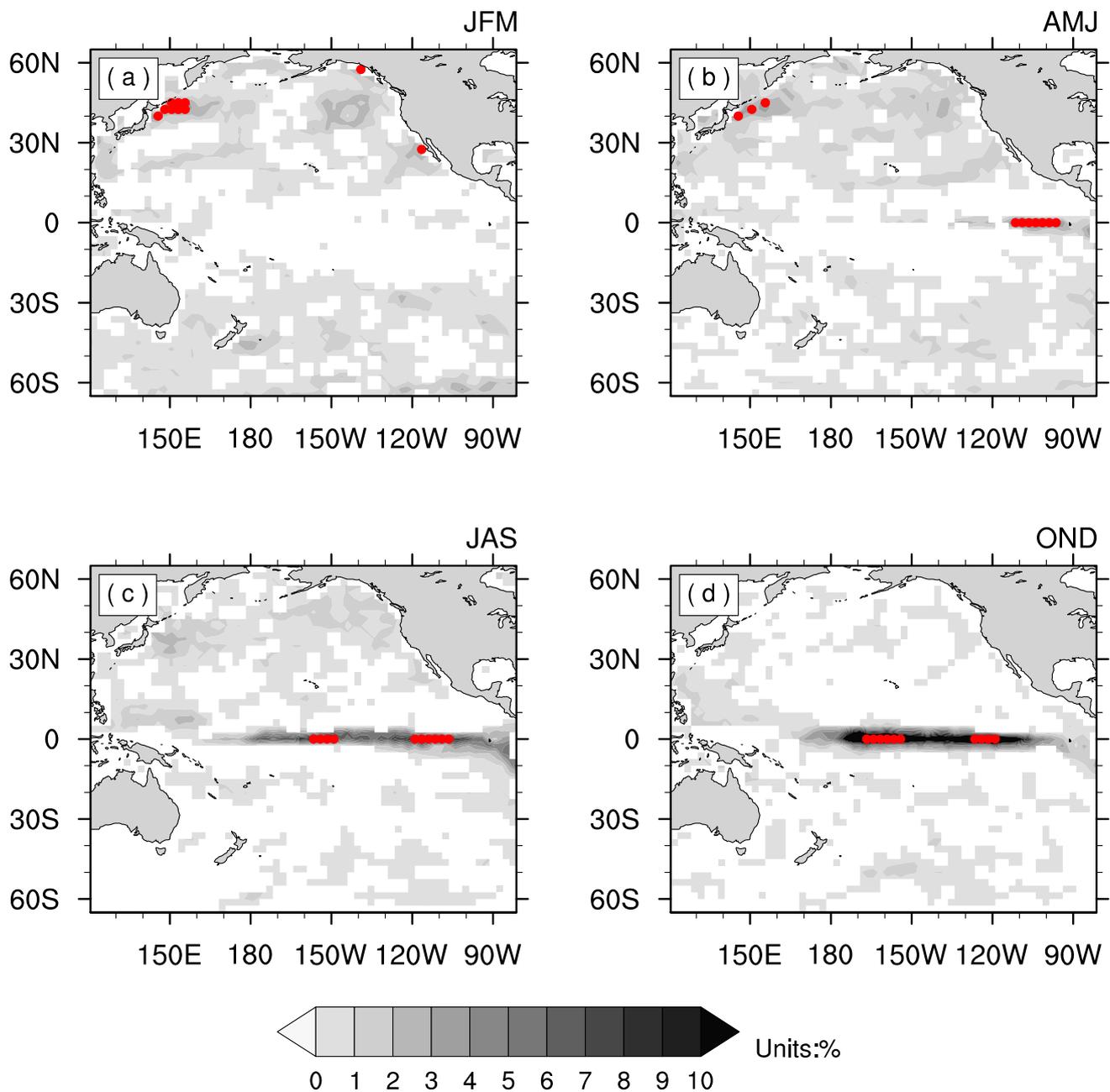


Fig. 8 The F indices (units: %) computed based on the Predictive Power and the decrease in RMSE in all assimilation experiments in **a** JFM, **b** AMJ, **c** JAS, **d** OND. The red dots represent the top 10 sensitive grid points determined by the F indices

the mechanism of the extratropical and tropical interaction, which support our findings (Alexander et al. 2010; Amaya 2019; Ding et al. 2015; Hou et al. 2019; Jin 1997; Vimont et al. 2003). Specifically, it is believed that the abnormal atmospheric status associated with the North Pacific Oscillation (NPO) forces the SST in mid-latitudes and leaves a “footprint” in the boreal winter (Vimont et al. 2003). Thus, the optimal observational array detected in JFM is mainly located in the large center of the NPO-forced SSTA area. In boreal spring, the NPO-forced SSTA variability maintains

and propagates into the tropics southwestwards through interaction between the SSTAs, surface wind anomalies, and latent heat flux anomalies, which is known as WES feedback (Xie and Philander 1994), corresponding to the increase of the F value in the eastern North Pacific near Baja California as shown in Fig. 9b. In terms of ocean circulation, off-equatorial wind stress curl anomalies help transport the water mass meridionally, which can charge or discharge the heat content in the equatorial Pacific (Anderson et al. 2013). In addition, the propagation and reflection of ocean Rossby

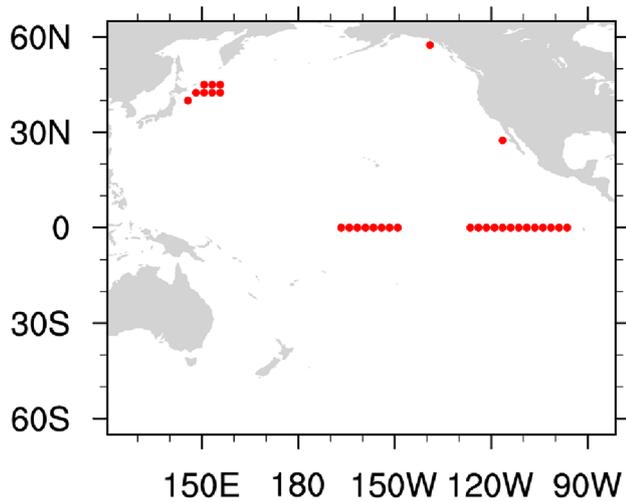


Fig. 9 Stable optimal observational array for two types of El Niño predictions

waves triggered by the thermocline perturbation in vertical and their corresponding Kelvin waves help bring up the SSTA from the western equatorial subsurface Pacific to the eastern equatorial surface Pacific (Alexander et al. 2010). Under these circumstances, the F value in the equatorial Pacific becomes larger after late spring.

7 Verification experiments

After designing the optimal observational array, we perform a verification test to verify that the optimal observational array can efficiently improve ENSO predictions. We still use the PF method to be consistent with the previous experiments. However, the challenge here is that the ensemble will degenerate dramatically if the observations on the optimal observational array are assimilated simultaneously. To mitigate the degeneration, a large ensemble is created by combining the ensembles of the 6 models, containing 3000 (6 models * 500 years) one-year prediction ensemble members. For this circumstance, we use the real observation dataset instead of the fabricated observations because the model errors will be involved anyway. Thus, 21 El Niño events from 1950 to 2020 are chosen as observations by using the monthly mean oceanic dataset from the Extended Reconstructed Sea Surface Temperature (ERSST) version 5 data. The assimilation experiments are conducted using the PF method, as explained in Eq. (5), in the sensitive area for three months to calculate the weights of 3000 members and use the same weights in the following months to give predictions. In addition, the observation error is set to $0.6 \delta_T$ to diminish the degeneracy of particles.

The ensemble prediction of the Niño3 index for the 21 El Niño events, which are obtained by assimilating optimal observation data from April to June, is shown in Fig. 10. It is shown that the spread of the Niño3 ensemble decreases significantly when simultaneously assimilating several optimal observation data from April to June. However, the spread gradually increases with a longer lead time. Although the spread of the prediction is large in December, the ensemble members for most cases are distributed on both sides of the observation, and the ensemble mean is closer to the truth in most cases (Fig. 10). Similar predictions are also conducted for the Niño4 SSTA index for different seasons (JFM, JAS, and OND), showing similar results (not shown). In summary, though interfering with model errors, most of the ENSO predictions improve after assimilating target observation data in all seasons, especially when the predictions are made after June, and the warm phase in December is correctly predicted for all events.

A further examination is conducted using a random experimental strategy. We randomly choose 31 grid points in the whole Pacific as a random array 100 times and repeat the PF assimilation procedure and ensemble prediction. Figure 11 shows the prediction errors from the random experiment against the same errors from the assimilation of the above optimal observations. The prediction errors from the assimilation of optimal observations are smaller than those from random cases for both the Niño3 index and Niño4 index. Moreover, the result from the optimal observational array is superior to that from 98% (95%) of the randomly selected arrays in terms of the Niño3 (Niño4) indices through significant examinations. In addition, the effective sample size at the last assimilation step in all cases is calculated. It is worth mentioning that the assimilation of randomly chosen observations shows a more severe degeneracy than that of optimal observations. For the former, the average effective sample size is only about 2 in contrast to 15 for the optimal observation assimilation. This may be due to different observation samples and also the involved model error. However, stochastic universal resampling (Van Leeuwen 2015) steps are taken after the assimilation step, which can alleviate degeneracy in all cases. However, a resampling strategy is taken after the assimilation step in all cases to increase the effective sample size, which prevents all cases from the most severe degeneracy. Overall, these results indicate that the improvement in prediction skills by optimal observation is effective and significant.

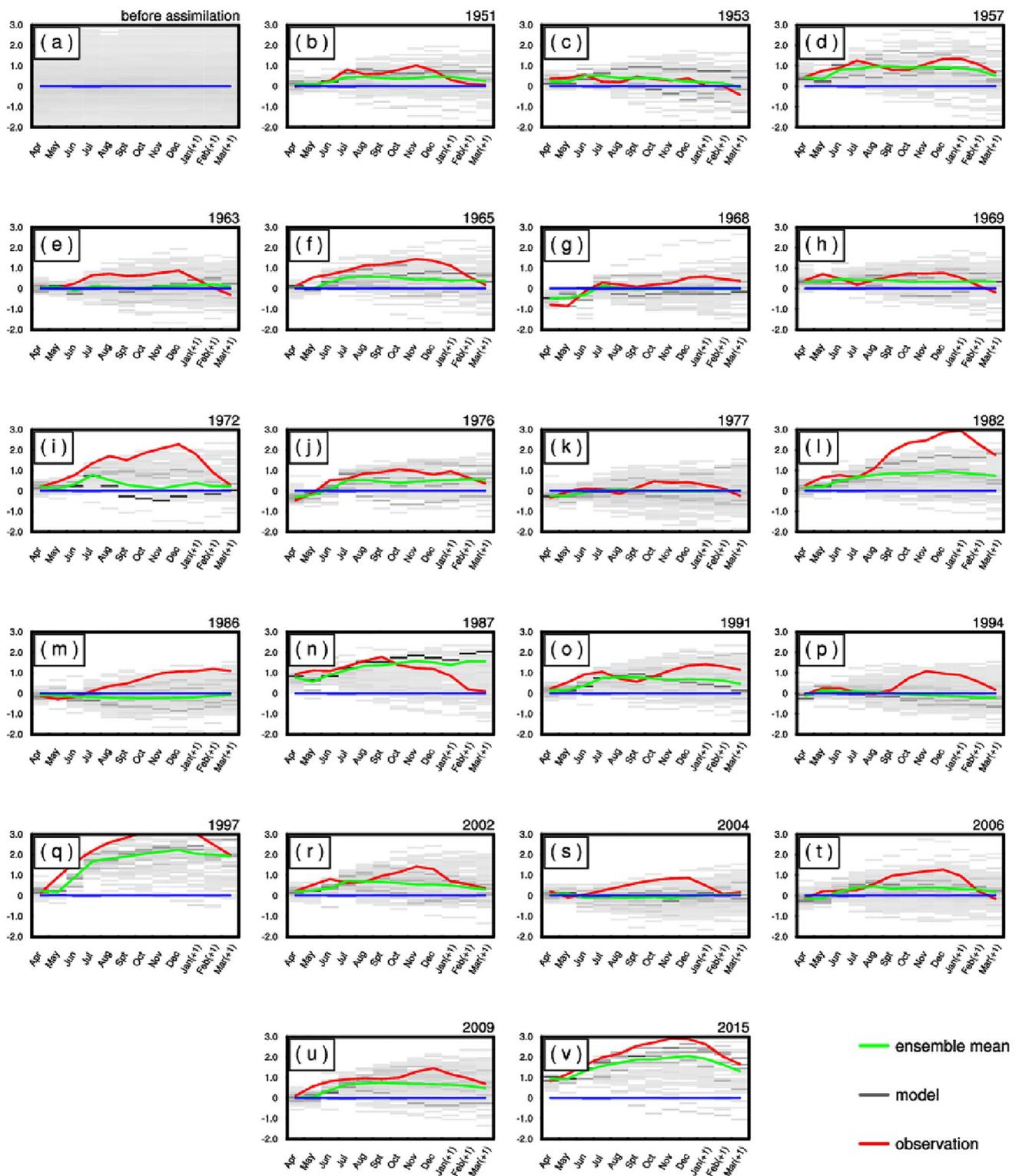


Fig. 10 Ensemble prediction of Niño3 index (units: °C) in model ensembles: **a** the origin ensemble forecast before assimilation; **b–v** new ensemble predictions after assimilating optimal observed ocean

temperature during April, May, and June. Red lines represent the observations; green lines represent the ensemble mean forecast after assimilation; the areas shaded in gray represent the probability

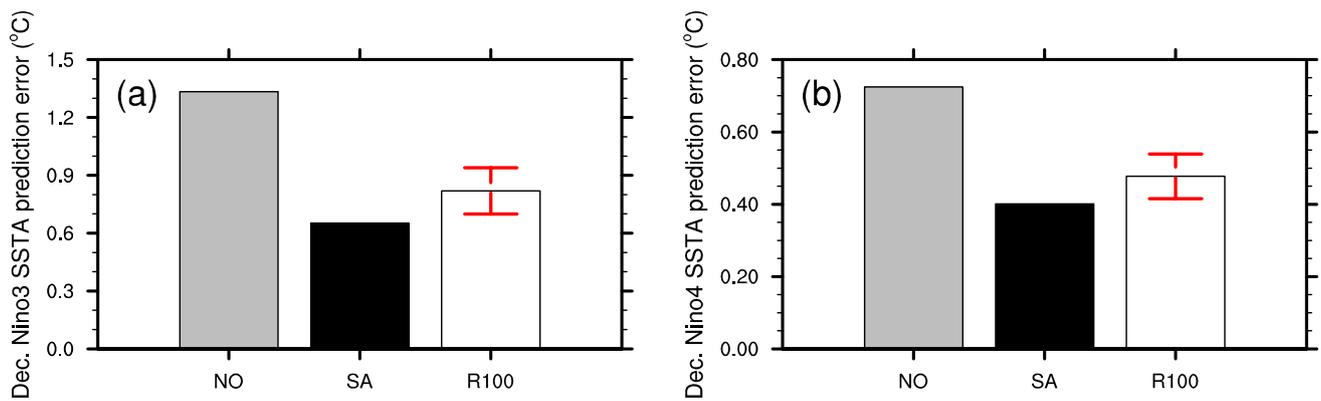


Fig. 11 The bar charts of the skills of the “hindcast” forecast when data assimilations were made in the sensitive area (SA, black bars) and the other 100 randomly selected arrays (R100, white bars). **a** Average prediction errors of Niño3 SSTA in December (units: °C); **b**

average prediction error of the Niño4 SSTA in December (units: °C) among all 84 prediction cases. Grey bars denote the prediction errors of the climatology predictions. Red lines denote the standard deviation among all 100 random cases

8 Conclusion and discussion

In this study, we quantify the relative importance of the SST observations in different areas of the Pacific for two types of El Niño predictions and explore the sensitive areas for target observations for CP- and EP-El Niño events by using the PF. Two measures, PP and RMSE, are used to describe the relative importance of observations in different areas. The initial uncertainty of the SST in the tropical Pacific, subtropical Pacific, and extratropical Pacific can exert influences on both CP-El Niño predictions and EP-El Niño predictions. The relative importance of different areas changes with the lead times of predictions. The tropical Pacific is the most sensitive area during the latter half-year. During spring, the extratropical signals cannot be disregarded and can even surpass the tropical signals, especially in the North Pacific. Subtropical and extratropical observations do play important roles in decreasing the prediction uncertainty, although their impact on the decrease in the RMSE of predictions is limited. A quantitative method based on frequency distribution is used to determine the optimal observations of El Niño predictions with the consideration of different measurements, different El Niño types, and different start months of predictions. Four optimal observational arrays are designed concerning four start months of predictions. The optimal observations move from the extratropical Pacific Ocean to the tropical Pacific Ocean with the start month of the prediction set before and after summer. Moreover, a robust and long-time standing optimal observational array for ENSO prediction is designed by combining these four optimal observational arrays. The final optimal observational array contains 21 grid points in the equatorial Pacific and 10 grid points in the North Pacific. It is shown that the optimal observational array can achieve more improvement in the El Niño prediction than almost all randomly chosen arrays.

The predictability of the two types of El Niño events has been explored for years. By using the Zebiak-Cane model, Duan et al. (2018b) designed an array of target observations to improve two types of ENSO predictions. The spatial pattern of the *F* indices in JAS, as shown in Fig. 8c, presents a pattern similar to that in Fig. 6a of Duan et al. (2018b), emphasizing the importance of the SSTA in the central equatorial Pacific and the eastern tropical Pacific. By using the intermediate coupled model (ICM) and CNOP method, Tao et al. (2017) and Mu et al. (2019) found that the CNOP-related initial errors that affect the ENSO prediction show seasonal dependence. We had similar findings of the spatial pattern of PP changing with lead times of predictions. However, due to the limitation of the ZC model and the ICM, they could only consider the tropical Pacific. In addition, our finding of the spatial structure of predictive power in the tropical Pacific is in agreement with the spatial structures of optimal SST precursors in Fig. 8 by Mu et al. (2019). By using the PF method, KD13 found that the initial errors at Niño3 and Niño4 areas are the most sensitive for the prediction of the EP- and CP-El Niño events at the lead time of three months, respectively. Utilizing the same assimilation method, we have similar findings. However, we extend their work in three aspects. First, we focused on the whole Pacific instead of only the tropical Pacific in KD13, finding that the initial errors in the extratropical Pacific should not be ignored in ENSO predictions; Second, we use two measures, not only potential skill (predictive power) but also deterministic skill to detect the sensitive areas. Third, we consider the optimal sensitivity areas as a function of lead time, allowing us to explore the continuous evolution of optimal error growth with the lead time of prediction from one month to 12 months.

The PF-based target observation method can detect the signal of El Niño events in different lead times. Hence, this

research also helps to understand the ENSO mechanism in its developing phase. The NPMM-like signal peaks in spring, while the SPMM-like signal peaks in late fall in both types of El Niño events, and their signal strengths are not as strong as that of the tropical signal. This finding indicates that the tropical Pacific Ocean is always crucial to the formation of both types of El Niño events, while the subtropical and extratropical Pacific can help adjust the zonal maximum SST center and the amplitude of the El Niño events during their developing stages. This is in agreement with Fan et al. (2020), who indicated that the NPMM can be a modulator, rather than a generator for ENSO.

The PF assimilation method applied in this paper has advantages, including easy operation, offline implementation, and less model dependence. The drawback of this method is the degeneracy of particles, which prevents us from using a sequential assimilation-based approach because only a few members will remain if too many observations are assimilated. Thus, some techniques including adding localization can be considered and employed to ameliorate the target observation method in the future. If the degeneracy problem can be solved, a sequential assimilation-based approach and multiple variables including sea ocean temperature and sea surface wind can all be considered when detecting target observations in future studies.

Acknowledgements The authors thank two anonymous reviewers who provided constructive comments that greatly improved the overall quality of the paper. All figures in this study were generated by the NCAR Command Language (version 6.4.0, <https://doi.org/10.5065/D6WD3XH5>).

Funding This study is supported by the National Natural Science Foundation of China (Grant Nos. 42130409, 42106004, and 41606012).

Data availability The Climate Model Intercomparison Project (CMIP) datasets applied in this study are available online (<https://esgf-node.llnl.gov/search/cmip5/>).

Declarations

Conflict of interest The authors have not disclosed any competing interests.

References

- Alexander MA, Bladé I, Newman M, Lanzante JR, Lau N-C, Scott JD (2002) The atmospheric bridge: the influence of ENSO Teleconnections on air–sea interaction over the global oceans. *J Clim* 15:2205–2231. [https://doi.org/10.1175/1520-0442\(2002\)015%3c2205:Tabtio%3e2.0.Co;2](https://doi.org/10.1175/1520-0442(2002)015%3c2205:Tabtio%3e2.0.Co;2)
- Alexander M, Vimont D, Chang PP, Scott J (2010) The impact of extratropical atmospheric variability on ENSO: testing the seasonal footprinting mechanism using coupled model experiments. *J Clim* 23:2885–2901. <https://doi.org/10.1175/2010JCLI3205.1>
- Amaya D (2019) The Pacific meridional mode and ENSO: a review. *Curr Clim Change Rep* 5:1–12. <https://doi.org/10.1007/s40641-019-00142-x>
- Anderson BT, Perez RC, Karspeck A (2013) Triggering of El Niño onset through trade wind–induced charging of the equatorial Pacific. *Geophys Res Lett* 40:1212–1216. <https://doi.org/10.1002/grl.50200>
- Andrews ED, Antweiler RC, Neiman PJ, Ralph FM (2004) Influence of ENSO on flood frequency along the California Coast. *J Clim* 17:337–348. [https://doi.org/10.1175/1520-0442\(2004\)017%3c0337:IOEOFF%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017%3c0337:IOEOFF%3e2.0.CO;2)
- Ashok K, Behera SK, Rao SA, Weng H, Yamagata T (2007) El Niño Modoki and its possible teleconnection. *J Geophys Res: Oceans* 112:1–25. <https://doi.org/10.1029/2006JC003798>
- Barnston AG, Tippett MK, L’Heureux ML, Li S, Dewitt DG (2012) Skill of real-time seasonal ENSO model predictions during 2002–11: is our capability increasing? *Bull Am Meteorol Soc* 93:631–651. <https://doi.org/10.1175/BAMS-D-11-00111.2>
- Bellenger H, Guilyardi E, Leloup J, Lengaigne M, Vialard J (2014) ENSO representation in climate models: from CMIP3 to CMIP5. *Clim Dynam* 42:1999–2018. <https://doi.org/10.1007/s00382-013-1783-z>
- Bergot T (1999) Adaptive observations during FASTEX: a systematic survey of upstream flights. *Q J R Meteorol Soc* 125:3271–3298
- Bishop C, Toth Z (1999) Ensemble transformation and adaptive observations. *J Atmos Sci* 56:1748–1765. [https://doi.org/10.1175/1520-0469\(1999\)056%3c1748:ETAAO%3e2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056%3c1748:ETAAO%3e2.0.CO;2)
- Bishop CH, Etherton BJ, Majumdar SJ (2001) Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Mon Weather Rev* 129:420–436
- Bjerknes J (1969) Atmospheric teleconnections from the equatorial Pacific. *Mon Weather Rev* 97:163–172
- Capotondi A et al (2015) Understanding ENSO diversity. *Bull Am Meteorol Soc* 96:921–938. <https://doi.org/10.1175/BAMS-D-13-00117.1>
- Chang P et al (2007) Pacific meridional mode and El Niño—Southern Oscillation. *Geophys Res Lett* 34:130–144. <https://doi.org/10.1029/2007GL030302>
- Chen DK, Zebiak SE, Busalacchi AJ, Cane MA (1995) An improved procedure for El Niño forecasting: implications for predictability. *Science* 269:1699–1702
- Chen DK, Cane MA, Kaplan A, Zebiak SE, Huang D (2004) Predictability of El Niño over the past 148 years. *Nature* 428:733–736. <https://doi.org/10.1038/nature02439>
- Chen DK et al (2015) Strong influence of westerly wind bursts on El Niño diversity. *Nat Geosci* 8:339–345. <https://doi.org/10.1038/NGEO2399>
- Dijkstra HA, Petersik P, Hernández-García E, López C (2019) The application of machine learning techniques to improve El Niño prediction skill. *Front Phys*. <https://doi.org/10.3389/fphy.2019.00153>
- Ding RQ, Li JP, Tseng YH, Sun C, Guo YP (2015) The Victoria mode in the North Pacific linking extratropical sea level pressure variations to ENSO. *J Geophys Res: Atmos* 120:27–45. <https://doi.org/10.1002/2014JD022221>
- Ding RQ, Li JP, Tseng YH, Sun C, Xie F (2017) Joint impact of North and South Pacific extratropical atmospheric variability on the onset of ENSO events. *J Geophys Res: Atmos* 122:279–298
- Duan WS, Hu JY (2016) The initial errors that induce a significant “spring predictability barrier” for El Niño events and their implications for target observation: results from an earth system model. *Clim Dyn* 46:1–17. <https://doi.org/10.1007/s00382-015-2789-5>
- Duan WS, Feng F, Hou MY (2018a) An application of particle filter assimilation approach in target observations for ENSO predictions. *Atmos Sci* 42:677–695 (in Chinese)

- Duan WS, Li XQ, Tian B (2018b) Towards optimal observational array for dealing with challenges of El Niño-Southern Oscillation predictions due to diversities of El Niño. *Clim Dynam* 51:3351–3368. <https://doi.org/10.1007/s00382-018-4082-x>
- Fan HJ, Huang BH, Yang S, Dong WJ (2020) Influence of Pacific Meridional Mode on ENSO evolution and predictability: asymmetric modulation and ocean preconditioning. *J Clim* 34:1881–1901. <https://doi.org/10.1175/JCLI-D-20-0109.1>
- Fang XH, Mu M (2018) A three-region conceptual model for central Pacific El Niño including zonal advective feedback. *J Clim* 31:4965–4979. <https://doi.org/10.1175/jcli-d-17-0633.1>
- Gao C, Wu X, Zhang R-H (2016) Testing a four-dimensional variational data assimilation method using an improved intermediate coupled model for ENSO analysis and prediction. *Adv Atmos Sci* 33:875–888. <https://doi.org/10.1007/s00376-016-5249-1>
- Gordon NJ, Salmond DJ, Smith AFM (1993) Novel Approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc F Radar Signal Process* 140:107–113. <https://doi.org/10.1049/ip-f-2.1993.0015>
- Guo W, Zhang R, Wang X (2021) Impacts of diverse El Niño events on north tropical Atlantic warming in their decaying springs. *J Geophys Res Oceans* 126:e2021JC017514. <https://doi.org/10.1029/2021JC017514>
- Ham YG, Kug JS (2012) How well do current climate models simulate two types of El Niño? *Clim Dyn* 39:383–398. <https://doi.org/10.1007/s00382-011-1157-3>
- Ham Y-G, Kim J-H, Luo J-J (2019) Deep learning for multi-year ENSO forecasts. *Nature* 573:568–572. <https://doi.org/10.1038/s41586-019-1559-7>
- Hoell A, Barlow M, Xu T, Zhang T (2018) Cold season southwest Asia precipitation sensitivity to El Niño-southern oscillation events. *J Clim* 31:4463–4482. <https://doi.org/10.1175/jcli-d-17-0456.1>
- Hou MY, Duan WS, Zhi XF (2019) Season-dependent predictability barrier for two types of El Niño revealed by an approach to data analysis for predictability. *Clim Dyn* 53:5561–5581. <https://doi.org/10.1007/s00382-019-04888-w>
- Jin FF (1997) An equatorial ocean recharge paradigm for ENSO. Part I: conceptual model. *J Atmos Sci* 54:811–829. [https://doi.org/10.1175/1520-0469\(1997\)054%3c0811: Aeorpf%3e2.0.Co;2](https://doi.org/10.1175/1520-0469(1997)054%3c0811: Aeorpf%3e2.0.Co;2)
- Jin EK et al (2008) Current status of ENSO prediction skill in coupled ocean-atmosphere models. *Clim Dyn* 31:647–664. <https://doi.org/10.1007/s00382-008-0397-3>
- Kao HY, Yu JY (2009) Contrasting Eastern-Pacific and Central-Pacific types of ENSO. *J Clim* 22:615–632. <https://doi.org/10.1175/2008JCLI2309.1>
- Kim ST, Yu JY (2012) The two types of ENSO in CMIP5 models. *Geophys Res Lett* 39:221–228. <https://doi.org/10.1029/2012GL052006>
- Kramer W, Dijkstra HA (2013) Optimal localized observations for advancing beyond the ENSO predictability barrier. *Nonlinear Proc Geophys* 20:221–230. <https://doi.org/10.5194/npg-20-221-2013>
- Kramer W, Van Leeuwen P, Pierini S, Dijkstra HA (2012) Measuring the impact of observations on the predictability of the Kuroshio extension in a shallow-water model. *J Phys Oceanogr* 42:3–17. <https://doi.org/10.1175/JPO-D-11-014.1>
- Kug JS, Jin FF, An SI (2009) Two types of El Niño events: cold tongue El Niño and warm pool El Niño. *J Clim* 22:1499–1515. <https://doi.org/10.1175/2008JCLI2624.1>
- Kug JS, Choi J, An SI, Jin FF, Wittenberg AT (2010) Warm pool and cold tongue El Niño events as simulated by the GFDL CM2.1 coupled GCM. *J Clim* 23:1226–1239. <https://doi.org/10.1175/2009JCLI3293.1>
- Lin CY, Yu JY, Hsu HH (2015) CMIP5 model simulations of the Pacific meridional mode and its connection to the two types of ENSO. *Int J Climatol* 35:2352–2358
- Liu T (2021) ENSO predictability over the past 137 years based on a CESM ensemble prediction system. *J Clim* 35:763–777. <https://doi.org/10.1175/JCLI-D-21-0450.1>
- Liu J, Kalnay E (2008) Estimating observation impact without adjoint model in an ensemble Kalman filter. *Q J R Meteorol Soc* 134:1327–1335. <https://doi.org/10.1002/qj.280>
- Lorenz E (1975) Climate predictability. The physical bases of climate and climate modelling, vol 16. WMO GARP, Geneva, pp 132–136
- Lu F, Liu Z, Liu Y, Zhang S, Jacob R (2017) Understanding the control of extratropical atmospheric variability on ENSO using a coupled data assimilation approach. *Clim Dyn* 48:1–22
- McPhaden MJ (2015) Playing hide and seek with El Niño. *Nat Clim Change* 5:791–795. <https://doi.org/10.1038/nclimate2775>
- Min QY, Su JZ, Zhang RH (2017) Impact of the south and north Pacific meridional modes on the El Niño-Southern oscillation: observational analysis and comparison. *J Clim* 30:1705–1720. <https://doi.org/10.1175/jcli-d-16-0063.1>
- Moore AM, Kleeman R (1996) The dynamics of error growth and predictability in a coupled model of ENSO. *Q J R Meteorol Soc* 122:1405–1446
- Mu M, Duan WS, Wang B (2003) Conditional nonlinear optimal perturbation and its applications. *Nonlinear Proc Geophys* 10:493–501. <https://doi.org/10.5194/npg-10-493-2003>
- Mu M, Duan WS, Chen DK, Yu WD (2015) Target observations for improving initialization of high-impact ocean-atmospheric environmental events forecasting. *Natl Sci Rev* 2:226–236. <https://doi.org/10.1093/nsr/nwv021>
- Mu B, Ren J, Yuan S, Zhang R-H, Chen L, Gao C (2019) The optimal precursors for ENSO events depicted using the gradient-definition-based method in an intermediate coupled model. *Adv Atmos Sci* 36:1381–1392. <https://doi.org/10.1007/s00376-019-9040-y>
- Palmer T, Gelaro R, Barkmeijer J, Buizza R (1998) Singular vectors, metrics, and adaptive observations. *J Atmos Sci* 55:633–653. [https://doi.org/10.1175/1520-0469\(1998\)055%3c0633:SVMAAO%3e2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055%3c0633:SVMAAO%3e2.0.CO;2)
- Philander SGH (1983) El Niño southern oscillation phenomena. *Nature* 302:295–301
- Qi Q, Duan W, Xu H (2021) The most sensitive initial error modes modulating intensities of CP- and EP- El Niño events. *Dyn Atmos Oceans*. <https://doi.org/10.1016/j.dynatmoce.2021.101257>
- Ren H, Jin FF (2013) Recharge oscillator mechanisms in two types of ENSO. *J Clim* 26:6506–6523. <https://doi.org/10.1175/JCLI-D-12-00601.1>
- Ren H, Zuo J, Jin F-F, Stuecker MF (2016) ENSO and annual cycle interaction: the combination mode representation in CMIP5 models. *Clim Dyn* 46:3753–3765. <https://doi.org/10.1007/s00382-015-2802-z>
- Ren H et al (2018) Seasonal predictability of winter ENSO types in operational dynamical model predictions. *Clim Dynam* 52:3869–3890. <https://doi.org/10.1007/s00382-018-4366-1>
- Schneider T, Griffies SM (1999) A conceptual framework for predictability studies. *J Clim* 12:3133–3155
- Shen ZQ, Tang YM, Li XJ (2017) A new formulation of vector weights in localized particle filter. *Q J R Meteorol Soc* 143:3269–3278. <https://doi.org/10.1002/qj.3180>
- Tang Y, Kleeman R, Miller S (2006) ENSO predictability of a fully coupled GCM model using singular vector analysis. *J Clim* 19:3361–3377. <https://doi.org/10.1175/JCLI3771.1>
- Tang Y et al (2018) Progress in ENSO prediction and predictability study. *Natl Sci Rev* 5:826–839. <https://doi.org/10.1093/nsr/nwy105>
- Tao LJ, Zhang RH, Gao C (2017) Initial error-induced optimal perturbations in ENSO predictions, as derived from an intermediate coupled model. *Adv Atmos Sci* 34:791–803. <https://doi.org/10.1007/s00376-017-6266-4>

- Tao LJ, Gao C, Zhang RH (2018) ENSO predictions in an intermediate coupled model influenced by removing initial condition errors in sensitive areas: a target observation perspective. *Adv Atmos Sci* 35:853–867. <https://doi.org/10.1007/s00376-017-7138-7>
- Toth Z, Kalnay E (1997) Ensemble forecasting at NCEP and the breeding method. *Mon Weather Rev* 125:3297–3319
- Van Leeuwen PJ (2009) REVIEW particle filtering in geophysical systems. *Mon Weather Rev* 137:4089–4114. <https://doi.org/10.1175/2009MWR2835.1>
- Van Leeuwen PJ (2015) Nonlinear data assimilation for high-dimensional systems. Springer, Cham, pp 1–73. https://doi.org/10.1007/978-3-319-18347-3_1
- Vimont DJ, Wallace JM, Battisti DS (2003) The seasonal footprinting mechanism in the Pacific: implications for ENSO. *J Clim* 16:2668–2675
- Vimont DJ, Alexander MA, Newman M (2014) Optimal growth of Central and East Pacific ENSO events. *Geophys Res Lett* 41:4027–4034. <https://doi.org/10.1002/2014GL059997>
- Webster PJ, Yang S (1992) Monsoon and ENSO: selectively interactive systems. *Q J R Meteorol Soc* 118:877–926
- Weng HY, Ashok K, Behera SK, Rao SA, Yamagata T (2007) Impacts of recent El Niño Modoki on dry/wet conditions in the Pacific rim during boreal summer. *Clim Dyn* 29:113–129
- Wu L, Zhang HJ, Chen J-M, Feng T (2018) Impact of two types of El Niño on tropical cyclones over the Western North Pacific: sensitivity to location and intensity of Pacific warming. *J Clim* 31:1725–1742. <https://doi.org/10.1175/jcli-d-17-0298.1>
- Wu Y, Shen Z, Tang Y (2020) A flow-dependent targeted observation method for ensemble Kalman filter assimilation systems. *Earth Space Sci* 7:e2020EA001149. <https://doi.org/10.1029/2020EA001149>
- Xie SP, Philander SGH (1994) A coupled ocean–atmosphere model of relevance to the ITCZ in the eastern Pacific. *Tellus A* 46:340–350. <https://doi.org/10.3402/tellusa.v46i4.15484>
- Yu JY, Kao HY (2007) Decadal changes of ENSO persistence barrier in SST and ocean heat content indices: 1958–2001. *J Geophys Res: Atmos* 112:125–138. <https://doi.org/10.1029/2006JD007654>
- Yu JY, Kao HY, Lee T (2010) Subtropics-related interannual sea surface temperature variability in the central equatorial Pacific. *J Clim* 23:2869–2884. <https://doi.org/10.1175/2010JCLI3171.1>
- Zhang RH, Rothstein LM, Busalacchi AJ (1998) Origin of upper-ocean warming and El Niño change on decadal scales in the tropical Pacific Ocean. *Nature* 391:879–883. <https://doi.org/10.1038/36081>
- Zhang W, Jin FF, Li JP, Ren HL (2011) Contrasting impacts of two-type El Niño over the western north Pacific during Boreal Autumn. *J Meteorol Soc Jpn* 89:563–569. <https://doi.org/10.2151/jmsj.2011-510>
- Zhang J, Duan WS, Zhi XF (2015) Using CMIP5 model outputs to investigate the initial errors that cause the “spring predictability barrier” for El Niño events. *Sci China: Earth Sci* 58:685–696. <https://doi.org/10.1007/s11430-014-4994-1>
- Zhang WJ et al (2016a) Unraveling El Niño’s impact on the East Asian Monsoon and Yangtze River summer flooding. *Geophys Res Lett* 43:11375–11382. <https://doi.org/10.1002/2016GL071190>
- Zhang WJ, Wang YL, Jin FF, Stuecker MF, Turner AG (2016b) Impact of different El Niño types on the El Niño/IOD relationship. *Geophys Res Lett* 42:8570–8576. <https://doi.org/10.1002/2015GL065703>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.