

Toward an Optimal Observational Array for Improving Two Flavors of El Niño Predictions in the Whole Pacific

Meiyi Hou

Hohai University

Youmin Tang (✉ ytang@unbc.ca)

Hohai University

Wansuo Duan

Institute of Atmospheric Physics Chinese Academy of Sciences

Zheqi Shen

Hohai University

Research Article

Keywords: ENSO, initial error, particle filter, target observation

Posted Date: January 14th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1234234/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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

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

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

7 2) College of Oceanography, Hohai University, Nanjing, China

8 3) Environmental Science and Engineering, University of Northern British Columbia, Prince
9 George, British Columbia, Canada

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

13 5) University of Chinese Academy of Sciences, Beijing, China

14
15 *Corresponding authors:

16 Prof. Youmin Tang, email: ytang@unbc.ca.

17
18 **Abstract**

19 This paper investigates the optimal observational array for improving the prediction of the El
20 Niño-Southern Oscillation (ENSO) by exploring sensitive areas for target observations of two
21 types of El Niño events in the whole Pacific. A target observation method based on the particle
22 filter and pre-industrial control runs from six coupled model outputs in Coupled Model
23 Intercomparison Project Phase 5 (CMIP5) experiments are used to quantify the relative
24 importance of the initial accuracy of sea surface temperature (SST) in different Pacific areas. The
25 initial accuracy of the tropical Pacific, subtropical Pacific, and extratropical Pacific can all exert
26 influences on both types of El Niño predictions. The relative importance of different areas changes
27 along with different lead times of predictions. Tropical Pacific observations are crucial in
28 decreasing the root mean square error of predictions of all lead times. Subtropical and
29 extratropical observations play an important role in decreasing the prediction uncertainty,
30 especially when the prediction is made before and throughout boreal spring. To consider different

31 El Niño types and different start months for predictions, a quantitative frequency method based on
32 frequency distribution is applied to determine the optimal observations of ENSO predictions. The
33 final optimal observational array contains 31 grid points, including 21 grid points in the equatorial
34 Pacific and 10 grid points in the north Pacific, suggesting the importance of the initial SST
35 conditions for ENSO predictions not only in the tropical Pacific but also in the area outside the
36 tropics. Furthermore, the predictions made by assimilating SST in sensitive areas have better
37 prediction skills in the verification experiment, which can indicate the validity of the optimal
38 observational array designed in this study.

39

40 Keywords: ENSO; initial error; particle filter; target observation

41

42 **1. Introduction**

43 The El Niño-Southern Oscillation (ENSO) is the dominant mode of interannual climate
44 variability on Earth, alternating between warm (El Niño) and cold (La Niña) conditions, which are
45 centered in the central and eastern equatorial Pacific (Bjerknes 1969; Philander 1983; Webster and
46 Yang 1992). It has long been a focus of exploration because of its profound influences on the
47 tropical climate and even the global climate (Andrews et al. 2004; Hoell et al. 2018; Wu et al.
48 2018; Zhang et al. 2016). In theory, the self-sustaining nature of the ENSO is conducive to its
49 potential predictability up to two years in advance (Liu 2021; Tang et al. 2018). However, most
50 current real-time ENSO predictions, made by the existing dynamical and statistical models, can
51 only provide useful skills two or three seasons in advance (Barnston et al. 2012; Jin et al. 2008;
52 Liu 2021). More interestingly, even though the models are constantly improving, the El Niño
53 prediction skill during 2002-2011 is the lowest for the 1981-2010 period (Barnston et al. 2012). In
54 addition, most models substantially overestimated the amplitude of warming for the 2014/15 event
55 when the prediction was initialized around June 2014 (McPhaden 2015). It is notable that although
56 much progress has been made in understanding the ENSO mechanism and improving the physical
57 parameterizations in models in recent decades, El Niño prediction is still an immense challenge
58 (Tang et al. 2018).

59 ENSO diversity is certainly a crucial factor that hampers the accuracy of the ENSO
60 prediction. Since the 1990s, a new flavor of El Niño (denoted as CP-El Niño) has occurred

61 increasingly frequently (Yu and Kao 2007). Different from the canonical El Niño (denoted as
62 EP-El Niño), the largest warming SST anomalies (SSTA) center of CP-El Niño during peak time
63 is located in the central tropical Pacific, instead of in the eastern tropical Pacific (Kao and Yu 2009;
64 Kug et al. 2009). The spatial differences of the two types of El Niño events increase the difficulty
65 of the ENSO prediction regarding both structure and intensity. Moreover, the difference between
66 these two types of El Niño events is also reflected in other aspects, including the formation
67 mechanism, evolutionary process, and climate influence. Considering these differences, it is
68 necessary to discern which type of El Niño will occur while making El Niño predictions. However,
69 it was shown that the useful predictions of El Niño types can be made only approximately 3-4
70 months in advance (Hendon et al. 2009; Jeong et al. 2012). Moreover, Ren et al. (2018)
71 emphasized that El Niño types cannot be distinguished in their mature phase in some models even
72 with a 1-month lead time.

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

84 Usually, two approaches are utilized to determine the optimal observation locations. One
85 approach seeks the largest growth of initial errors, which are most assumed to affect the prediction.
86 This approach includes methods such as singular vector (SV; Palmer et al. 1998; Tang et al. 2006),
87 breeding vector (BV; Toth and Kalnay 1997), adjoint sensitivity (Bergot 1999), conditional
88 nonlinear optimal perturbation (CNOP; Mu et al. 2003; Duan et al. 2018b) and other uncertain
89 analysis approaches of prediction (Hou et al. 2019; Zhang et al. 2015). By focusing on the error
90 growth, these methods help locate the most sensitive area where the initial errors grow the most

91 dramatically and unavoidably interfere with the prediction. Most research on detecting ENSO
92 optimal observation arrays uses this kind of method. In particular, Duan et al. (2018b) designed an
93 optimal observation array for ENSO prediction in the tropical Pacific by using the CNOP method.
94 The other kind of approach is based on assimilation methods, including the ensemble transform
95 (Bishop and Toth 1999), ensemble transform Kalman filter (ETKF; Bishop et al. 2001), and
96 ensemble Kalman filter (EnKF; Liu and Kalnay 2008, Wu et al. 2020). However, the EnKF and its
97 variants assume that both the model errors and observation errors are Gaussian. The Kalman filter,
98 which is the fundamental basis of EnKF, is only held for linear state-space systems. As such, a
99 new assimilation method referred to as the particle filter (PF; Gordon et al. 1993; Van Leeuwen
100 2009; Shen et al. 2017) has recently attracted broad attention and is appropriate for any
101 non-Gaussian and nonlinear system.

102 By using an offline numerically efficient method, Kramer and Dijkstra (2013; hereafter
103 referred to KD13) applied the PF to explore the predictability barrier for two types of El Niño
104 events in the tropical Pacific domain (also see Duan et al. 2018a). Under the framework of
105 observation system simulation experiment (OSSE), they proposed an offline approach without
106 model forward integration to update the weights of particles (ensemble members). By this means,
107 they discovered that the initial accuracy of the SST in the tropical Pacific near the Niño3 and
108 Niño4 areas is very significant for ENSO predictions.

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

120 As previously discussed, the most effective way to improve the ENSO prediction is to

121 increase the number of observations and assimilate them into the model prediction system. The
122 initial accuracy of the temperature in the whole Pacific, including the tropical Pacific, subtropical
123 Pacific, and extratropical Pacific, may all be important for distinguishing El Niño types. Under
124 this circumstance, it is urgent to address several major issues to improve the two types of El Niño
125 predictions. First, to what extent does the initial accuracy of the extratropical Pacific matter to the
126 two types of El Niño predictions, and is its importance comparable to that of the tropical Pacific?
127 Second, in considering the observations in the whole Pacific, how would a stable
128 model-independent optimal observational array be designed to improve the prediction of two
129 types of El Niño events.

130 In this study, we utilized the PF method in KD13 to seek the optimal observational locations
131 through the whole Pacific for ENSO type predictions. The paper is organized as follows: In
132 Section 2, datasets and the PF method are illustrated. In Section 3, the assimilation experiments
133 are depicted clearly. In Sections 4 and 5, the core of the paper, we quantify the relative importance
134 of the observations in the different Pacific areas for two types of El Niño predictions. In Section 6,
135 we design the array of the optimal observational array for ENSO predictions. In Section 7,
136 verification experiments are conducted. Finally, in Section 8, we present our summary and
137 discussion.

138

139 **2. Datasets and PF assimilation methodology**

140 Coupled Model Intercomparison Project Phase 5 (CMIP5) provides abundant global coupled
141 model data resources. In this study, we use outputs from CMIP5 preindustrial control (piControl)
142 experiments, in which the models are driven for more than 500 years under a constant external
143 forcing (greenhouse gas, solar radiation, aerosol, land use, etc.) at the level of 1850. Thus, the
144 integration results in piControl experiments only include signals of internal variability. Massive
145 analysis of the ENSO simulation ability of the CMIP5 models has been conducted (Bellenger et al.
146 2014; Ham and Kug 2012; Ren et al. 2016). A consensus has been reached that only some CMIP5
147 models can capture the main features of both flavors of El Niño events, especially the CP-El Niño.
148 Referring to Kim and Yu (2012), six models that can reasonably simulate two types of El Niño
149 events were chosen in our work. Specific model configurations and affiliations are listed in Table
150 1. SST data are obtained from the output datasets of the six coupled models. It is noted that

151 models have integration times and different resolutions. To simplify the calculations, we only
 152 choose the first 500 years of the integration in each model, and the variable is interpolated onto
 153 the same grids with a resolution of $2.5^\circ \times 2.5^\circ$ by using the bilinear interpolation method. All
 154 anomalies are computed by removing their monthly climatology mean.

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

$$161 \quad p_N(\mathbf{X}_k) \approx \sum_{i=1}^N w_k^i \delta(\mathbf{X}_k - \mathbf{X}_k^i). \quad (1)$$

162 Herein, $\delta(\cdot)$ is the Dirac delta function over real numbers, whose value is zero anywhere except
 163 at zero and whose integral over the entire real line is equal to one. The PDF of state vector \mathbf{X}_k is
 164 estimated by ‘‘particles’’, i.e., ensemble members, \mathbf{X}_k^i ($i = 1, 2, \dots, N$), multiplied by weights w_k^i
 165 of these particles. At the beginning ($k = 0$), the w_0^i of each particle is identical, which is equal to
 166 $1/N$. An observation y_k then becomes available at time $t = t_k$, which can be assimilated to
 167 obtain the posterior PDF $p_N(\mathbf{X}_k|\mathbf{Y}_k)$ by Bayes Theorem:

$$168 \quad p_N(\mathbf{X}_k|\mathbf{Y}_k) = \frac{p(\mathbf{Y}_k|\mathbf{X}_k^i)p_N(\mathbf{X}_k|\mathbf{Y}_{k-1})}{p(\mathbf{Y}_k)}. \quad (2)$$

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

$$170 \quad w_k^i = \frac{p(\mathbf{Y}_k|\mathbf{X}_k^i)}{p(\mathbf{Y}_k)} w_{k-1}^i. \quad (3)$$

171 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 $p(\mathbf{Y}_k)$
 172 is the PDF of the observation. Noted that $p(\mathbf{Y}_k)$ can be regarded as a normalization factor, which
 173 ensures that the sum of the weights is equal to one. $p(\mathbf{Y}_k|\mathbf{X}_k^i)$ is directly related to the probability
 174 distribution of the observation error. If the observation error is a multivariate normal distribution
 175 and its covariance matrix is Σ , then

$$176 \quad p(\mathbf{Y}_k|\mathbf{X}_k^i) \sim \exp\left[-\frac{1}{2}\left(\mathbf{Y}_k - H(\mathbf{X}_k^i)\right)^T \Sigma^{-1}\left(\mathbf{Y}_k - H(\mathbf{X}_k^i)\right)\right]. \quad (4)$$

177 Here, H is the observation operator, which can be calculated by simply selecting the model
 178 equivalents from the full state vector. Considering eq. (3) and (4), the weight w_k^i is given by

179
$$w_k^i \sim \exp\left[-\frac{1}{2}\left(\mathbf{Y}_k - H(\mathbf{X}_k^i)\right)^T \Sigma^{-1} \left(\mathbf{Y}_k - H(\mathbf{X}_k^i)\right)\right]. \quad (5)$$

180 The weight is updated when assimilating one observation data at $t = t_k$ following the eq. (5). If
 181 several observations at different grids are assimilated simultaneously, the weight w_k^i is updated
 182 as follows:

183
$$w_k^i \sim \exp\left[-\frac{1}{2}\sum_{j=1}^m \left(\mathbf{Y}_k - H(\mathbf{X}_k^i)\right)^T \Sigma^{-1} \left(\mathbf{Y}_k - H(\mathbf{X}_k^i)\right)\right] \quad \text{for } j = 1, 2, \dots, m. \quad (6)$$

184
$$= \prod_{j=1}^m \exp\left[-\frac{1}{2}\left(\mathbf{Y}_k - H(\mathbf{X}_k^i)\right)^T \Sigma^{-1} \left(\mathbf{Y}_k - H(\mathbf{X}_k^i)\right)\right]. \quad (7)$$

185 The abovementioned method of weight updating is known as the sequential importance
 186 sample (SIS) method, which is a useful PF algorithm for designing experiments in our work.
 187 However, the major problem of SIS is that after assimilating the observations at $t = t_k$, the
 188 weight is concentrated on only a small number of particles, which is referred to as the degeneracy
 189 of the particles. A strongly degenerated ensemble, where only a few ensemble members have
 190 weights, cannot yield a reasonable prediction ensemble for the predicted variable, such as the
 191 Niño3 and Niño4 SSTAs. A basic solution to avoid degeneracy is to perform resampling. Namely,
 192 the particles with high weight will be duplicated, and particles with low weight will be discarded.
 193 In addition, setting the proper magnitude of the error covariance is also important to avoid
 194 degeneracy. If the observation error is set too small, only particles that are close to the observation
 195 remain, which will cause large degeneracy. However, it will be unrealistic if we set observation
 196 errors that are too large. In this study, after performing tuning experiments, we set the observation
 197 error to $0.3\delta_T$, where δ_T is the variance of SST.

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

201
$$PP = 1 - \exp(IE_{X_{new}} - IE_{X_{init}}), \quad (8)$$

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

205
$$IE_{p(\mathbf{X})} = \sum_a p_a(\mathbf{X}) \ln p_a(\mathbf{X}) \Delta \mathbf{X}. \quad (9)$$

206 Here, $p(\mathbf{X})$ is the PDF of the prediction ensemble. The information entropy measures the
 207 uncertainty level of the ensemble (Kleeman et al. 2003). Therefore, PP presents a decrease in

208 uncertainty due to the assimilation of observations. The larger the PP is, the greater the decrease in
209 the uncertainty of the ensemble prediction. In addition, the RMSE, a commonly used measure, is
210 also calculated to assess the assimilation performance.

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

218

219 **3. Experiment design**

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

232 all assimilation experiments are conducted in the framework of the OSSE by utilizing the
233 piControl model outputs from CMIP5. As illustrated in the data analysis approach for exploring
234 predictability dynamics proposed by Hou et al. (2019), under the circumstances that the external
235 forcing is fixed during the entire integration time, if we choose two partial time snippets with the
236 same lengths of integration and regard one time snippet as an “observation” and another one as a
237 “prediction”, then, the “prediction” error can be regarded as being caused by the initial difference

238 (error) between the two snippets. This method of fabricating “observation” and “prediction
239 ensemble” can perfectly eliminate the effect of the model error on prediction so that we can focus
240 only on the initial error problems.

241 Specifically, we divide the 500-year integration into 500 one-year segments and choose a
242 single one-year period (e.g., from January to December) of a typical EP- or CP-El Niño year as a
243 truth run, from which the “observation” is made by adding a normally distributed observation
244 error to the truth. The other 499 one-year integrations can all be regarded as its “predictions” up to
245 a lead time of 11 months. These predictions altogether compose a prediction ensemble for this
246 specific EP- or CP-El Niño event. These ensemble members are assumed to be independent,
247 although in fact, there is a correlation between one year and the next year. However, if we leave
248 the odd years out, the ensemble member will be only half of the entire ensemble, which hinders
249 sample diversity. We have to choose larger sample numbers instead of better sample independence.
250 In addition, choosing segments less than one year is also inadvisable because ENSO dynamics are
251 very seasonally dependent. Therefore, it is reasonable to obtain the observation and the prediction
252 ensemble in this way.

253 The PF method is used to conduct offline assimilation experiments via eq. (5). The basic
254 principle here is to assimilate only one observation data, such as SSTA, in one single grid among
255 the Pacific in one experiment. After assimilation, the improvement in the prediction utility is
256 calculated to evaluate the importance of this observation. The next assimilation is conducted by
257 another grid observation. In this way, the assimilation process is repeated until all observations in
258 the Pacific are evaluated. By comparing the improvement in the prediction skill among all
259 assimilation experiments, the most sensitive area targeting ENSO prediction improvement can be
260 located. In addition, assimilation experiments are conducted repeatedly by using different
261 assimilation times from January to December because we also want to address the question of
262 whether and how the sensitive areas change with different prediction lead times.

263

264 **4. Impact of observations for monitoring CP-El Niño events**

265 To obtain a less model-dependent result, the assimilation experiments are conducted by using
266 78 (6 models*13 events) CP-El Niño events as observations, and then the assessment is performed
267 by making a composite of all results. Two important metrics, PP and RMSE decrease, are

268 employed to determine the optimal observations. Our main purpose is to improve the El Niño
269 prediction in its mature phase. Thus, the evaluation target is the improvement of the prediction
270 skill only in December. Specifically, the weights of 499 ensemble members are updated by
271 assimilating SST in January (or other months), and then the weights are multiplied by the
272 ensemble of the December Niño4 index to generate the prediction of CP-El Niño in December,
273 with a lead time of 11 months (or other leads).

274 The spatial pattern of the averaged PP over 78 CP-El Niño cases is shown in Fig. 2, obtained
275 by assimilating observations at different times from January to November. For example, Fig. 2a
276 indicates the PP value of December Niño4 SST index prediction by assimilating January SST,
277 whereas Fig. 2k is the PP value of December Niño4 SST index prediction by assimilating
278 November SST. It should be noted that the PP value at one grid in Fig. 2 is the resultant PP value
279 of the Niño4 SSTA contributed by the assimilation of this grid's observation. Thus, the grid with a
280 high PP indicates that its observation has a high impact on the prediction of the December Niño4
281 SSTA. Therefore, the regions with high PP can be determined as optimal observation locations. It
282 is obvious in Fig. 2 that the signal changes along with the assimilation time. Centers of large PP
283 values are basically located in three areas, including the equatorial Pacific, the north Pacific, and
284 the south Pacific. Signals continually gather around the Niño3 and Niño4 areas. However, the
285 largest PP is located first in the Niño3 area and then moves to the Niño4 area after July. The value
286 of PP in the equatorial Pacific decreases from January to April and then increases after May. It
287 seems natural and intuitive for the PP value to increase as the lead time decreases in the equatorial
288 Pacific from July to November, as shown in Fig. 2, in that the observation can offer more
289 information and stronger predictable signals as it is closer to the prediction. However, the opposite
290 situation could also occur when some key teleconnection processes contribute to predictable
291 signals, for example, a delayed impact of the western Pacific Ocean on the ENSO, as described by
292 the delayed-oscillator mechanism. This may explain why the PP value decreases from January to
293 April, as shown in Fig. 2.

294 In terms of the PP in the north Pacific area, one large center is located at approximately 40°
295 N in the northwest Pacific near the Kuroshio Extension. The value there increases from January,
296 peaks near August, and then starts to decrease. Another large center is observed over the northeast
297 Pacific, which has a similar spatial pattern to the NPMM near Baja California in Mexico. This

298 pattern is clearly seen from March to June. There are two large centers of PP in the south Pacific.
299 One is located over the extratropical Pacific and is centered at approximately 130° W-140° W,
300 50° S-60° S. The other is located at approximately 30° S and has a similar spatial pattern to
301 SPMM. The former fades from January to September and rises in October. The latter emerges in
302 June and matures in November and December. The PP spatial pattern here shows agreement with
303 the error pattern of SST, which interferes with CP-El Niño predictions, as shown in Fig. 7 of Hou
304 et al. (2019). The importance of the initial accuracy in the extratropical Pacific to CP-El Niño
305 predictions is also emphasized by Hou et al. (2019).

306 The PP pattern shown in Fig. 2 is a composite of all the cases and models. Composites for the
307 cases of each model are also checked (not shown), which all bear great resemblances to the
308 composite of all models. That is, there are five apparent large centers of PP related to the
309 December CP-El Niño predictions, including the tropical Pacific (TP; 120° E-80° W, 23.5°
310 S-23.5° N), subtropical north Pacific (SNP; 23.5° N-40° N), subtropical south Pacific (SSP;
311 23.5° S-40° S), extratropical north Pacific (ENP; 40° N-66.5° N) and extratropical south
312 Pacific (ESP; 40° S-66.5° S).

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

328 predicting CP-El Niño events before boreal spring, the initial conditions of SST outside the
329 tropical area in the Pacific are important and will help reduce the uncertainty of the prediction.

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

353

354 **5. Impact of observations for monitoring EP-El Niño events**

355 Similar assimilation experiments are conducted by using 78 EP-El Niño synthetic
356 observations. The spatial pattern of PP(NINO3), averaged over these 78 cases, is shown in Fig. 5.
357 The spatial pattern of PP(NINO3) is similar to that of PP(NINO4) in the previous section. In the

358 tropical Pacific, the optimal location for predicting the Niño3 index in December is more
359 restricted to the eastern tropical Pacific. During the beginning of the year, from January to March,
360 the signal in the tropical Pacific is not as strong as that in the extratropical Pacific. The prediction
361 of Niño3 suffers from a severe spring predictability barrier, and south extratropical signals in
362 January and north extratropical signals in JFMA can provide more predictability than the tropical
363 Pacific. In this case, adding optimal observations in the extratropical Pacific and assimilating them
364 into the model may attenuate the predictability barrier of EP-El Niño prediction.

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

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

388 during spring are important for predictions of both EP- and CP-El Niño events.

389

390 **6. Sensitive area for target observations**

391 In the previous sections, we evaluated the relative importance of the tropical, subtropical, and
392 extratropical Pacific for two types of El Niño predictions at different lead times. We then
393 attempted to locate the optimal observations of SST for ENSO prediction considering the above
394 results. As shown in the previous sections, the signals change with different lead times, which
395 means that the optimal observations for ENSO prediction should be considered as a function of
396 the start time of the prediction. In addition, it is unknown what type of El Niño events will occur
397 when issuing a prediction; thus, the sensitive area should cover both types of El Niño events.
398 Hence, we propose to locate optimal observations case by case, as in the general operations
399 adopted in the target observation for ENSO in Duan et al. (2018b). The main idea is to sort PP
400 values. The details of the operation are illustrated as follows: First, we visit all the 156 prediction
401 cases (78 CP-El Niño events and 78 EP-El Niño events) as previously mentioned and only select
402 the spatial grid points with decreases in the RMSE. These selected spatial grid points are further
403 sorted in descending order according to their PP value, and the top 15 grids, which are called PP
404 max points hereafter, are identified. As a result, we can obtain 12*156 (12 months and 156 cases)
405 series of 15 PP max points. Second, we split these series into 4 groups, each containing 3*156
406 series (3 months and 156 cases), in order to use different months (January, April, July, and October)
407 to start the prediction. Third, from each group of 3*156 samples, we compute the frequency of PP
408 max that occurred for each grid point across the Pacific domain. To express this procedure more
409 clearly, we use a formula to show the calculation of frequency, which is denoted by F , as follows:

$$410 \quad F_{i,j}^t = \frac{c_{i,j}^t}{N} \times 100\%, \quad (10)$$

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

416 As shown in Fig. 8, the optimal observations in JFM are all located in the north Pacific, with

417 8 grid points in the extratropical Pacific near the Kuroshio Extension region and 2 grid points
418 along the west coast of North America. In AMJ, the sensitive area contains 3 grid points in the
419 northwest Pacific and 7 grid points on the equator in the eastern Pacific. In JAS and OND, all
420 optimal observations are located on the equator. To obtain a stable target observation strategy, we
421 propose combining the sensitive areas in different seasons to get an array with 31 grid points, as
422 shown in Fig. 9, which includes 21 grid points in the equatorial Pacific and 10 grid points in the
423 north Pacific.

424

425 **7. Verification experiments**

426 To verify that the optimal observational array can efficiently improve ENSO predictions, we
427 still use the PF and the framework of OSSE to be consistent with the previous experiments.
428 However, the challenge here is that the ensemble will degenerate dramatically if the observations
429 on the optimal observational array are assimilated simultaneously. To mitigate the degeneration, a
430 large ensemble is created by combining the ensembles of the 6 models, containing 3000 (6 models
431 * 500 years) one-year prediction ensemble members. For this circumstance, we use the real
432 observation dataset instead of the fabricated observations because the model errors will be
433 involved anyway. Thus, 21 El Niño events from 1950-2020 are chosen as observations by using
434 the monthly mean oceanic dataset from the Extended Reconstructed Sea Surface Temperature
435 (ERSST) version 5 data. The assimilation experiments are conducted using the PF method, as
436 explained in Eq. (6) and (7), in the sensitive area for three months to calculate the weights of 3000
437 members and use the same weights in the following months to give predictions. In addition, the
438 observation error is set to $0.6\delta_T$ to diminish degeneracy of particles.

439 The ensemble prediction of the Niño3 index for the 21 El Niño events, which are obtained by
440 assimilating optimal observation data from April to June, is shown in Fig. 10. It is shown that the
441 spread of the Niño3 ensemble decreases significantly when simultaneously assimilating several
442 optimal observation data from April to June. However, the spread gradually increases with a
443 longer lead time. Although the spread of the prediction is large in December, the ensemble
444 members for most cases are distributed on both sides of the observation, and the ensemble mean is
445 closer to the truth in most cases (Fig. 10). Similar predictions are also conducted for the Niño4
446 SSTA index for different seasons (JFM, JAS, and OND), showing similar results (not shown). In

447 summary, though interfering with model errors, most of the ENSO predictions improve after
448 assimilating target observation data in all seasons, especially when the predictions are made after
449 June, and the warm phase in December is correctly predicted for all events.

450 A further examination is conducted using a random experimental strategy. We randomly
451 choose 31 grid points in the whole Pacific as a random array 100 times and repeat the PF
452 assimilation procedure and ensemble prediction. Fig. 11 shows the prediction errors from the
453 random experiment against the same errors from the assimilation of the above optimal
454 observations. The prediction errors from the assimilation of optimal observations are smaller than
455 those from random cases for both the Niño3 index and Niño4 index. Moreover, the result from the
456 optimal observational array is superior to that from 98% (95%) of the randomly selected arrays in
457 terms of the Niño3 (Niño4) indices through significant examinations. Overall, these results
458 indicate that the improvement in prediction skills by optimal observation is effective and
459 significant.

460

461 **8. Conclusion and discussion**

462 In this study, we quantify the relative importance of the SST observations in different areas of
463 the Pacific for two types of El Niño predictions and explore the sensitive areas for target
464 observations for CP- and EP-El Niño events by using the PF. Two measurements, PP and RMSE,
465 are used to describe the relative importance of observations in different areas. The initial
466 uncertainty of the SST in the tropical Pacific, subtropical Pacific, and extratropical Pacific can
467 exert influences on both CP-El Niño predictions and EP-El Niño predictions. The relative
468 importance of different areas changes along with different lead times of predictions. The tropical
469 Pacific is the most sensitive area during the latter half-year. During spring, the extratropical
470 signals denoted by PP cannot be disregarded and can even surpass the tropical signals, especially
471 in the north Pacific. Subtropical and extratropical observations do play important roles in
472 decreasing the prediction uncertainty, although their impact on the decrease in the RMSE of
473 predictions is limited. Then, a quantitative method based on frequency distribution is used to
474 determine the optimal observations of El Niño predictions with the consideration of different
475 measurements, different El Niño types, and different start months of predictions. Four optimal
476 observational arrays are designed in regard to four start months of predictions. The optimal

477 observations move from the extratropical Pacific Ocean to the tropical Pacific Ocean with the start
478 month of the prediction set before and after summer. Moreover, a stable optimal observation array
479 for ENSO prediction is designed by combining these four optimal observation arrays. The final
480 optimal observation array contains 21 grid points in the equatorial Pacific and 10 grid points in the
481 north Pacific. It is shown that the optimal observation array can achieve the largest improvement
482 in the El Niño prediction skills compared with other randomly chosen arrays in the verification
483 examinations.

484 The sensitive area for target observations identified in the present study emphasizes the
485 importance of the SST in the north Pacific, the central equatorial Pacific, and the eastern
486 equatorial Pacific. By using the Zebiak-Cane model, Duan et al. (2018b) designed an array of
487 target observations to improve two types of ENSO predictions. The spatial pattern of the F indices
488 in JAS, as shown in Fig. 8c, presents a pattern similar to that in Fig. 6a in Duan et al. (2018b),
489 emphasizing the importance of the SSTA in the central equatorial Pacific and in the eastern
490 tropical Pacific. However, due to the ZC model limitation, they can only consider the tropical
491 Pacific, whereas we take considerations of the whole Pacific and the prediction of different lead
492 times.

493 The PF-based target observation method can detect the signal of El Niño events in different
494 lead times. Hence, this research also helps to understand the ENSO mechanisms in its developing
495 phase. The NPMM-like signal peaks in spring, while the SPMM-like signal peaks in late fall in
496 both types of El Niño events, and their signal strengths are not as strong as that of the tropical
497 signal. This finding indicates that the tropical Pacific Ocean is always crucial to the formation of
498 both types of El Niño events, while the subtropical and extratropical Pacific can help adjust the
499 zonal maximum SST center and the amplitude of the El Niño events during their developing
500 stages. This is in agreement with Fan et al. (2020), who indicated that the NPMM can be a trigger,
501 instead of a generator for ENSO.

502 The PF assimilation method applied in this paper has advantages, including easy operation,
503 offline implementation, and less model dependence. The drawback of this method is the
504 degeneracy of particles, which prevents us from using a sequential assimilation-based approach
505 because only a few members will remain if too many observations are assimilated. Thus, some
506 techniques including adding localization can be considered and employed to ameliorate the target

507 observation method in the future. If the degeneracy problem can be solved, a sequential
508 assimilation-based approach and multiple variables including sea ocean temperature and sea
509 surface wind can all be considered when detecting target observations in future studies.

510

511 **Data availability**

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

514

515 **Acknowledgements**

516 This study is supported by the National Natural Science Foundation of China (Grant Nos.
517 42130409, 42106004 and 41606012).

518

519 **References**

- 520 Andrews ED, Antweiler RC, Neiman PJ, Ralph FM (2004) Influence of ENSO on Flood Frequency
521 Along the California Coast. *J Clim* 17:337-348
- 522 Barnston AG, Tippett MK, L'Heureux ML, Li S, Dewitt DG (2012) Skill of Real-Time Seasonal ENSO
523 Model Predictions during 2002-11: Is Our Capability Increasing? *Bull Am Meteorol Soc*
524 93:631-651
- 525 Bellenger H, Guilyardi E, Leloup J, Lengaigne M, Vialard J (2014) ENSO representation in climate
526 models: from CMIP3 to CMIP5. *Clim Dynam* 42:1999-2018 doi:10.1007/s00382-013-1783-z
- 527 Bergot T (1999) Adaptive observations during FASTEX: A systematic survey of upstream flights. *Q J*
528 *R Meteorol Soc* 125:3271-3298
- 529 Bishop C, Toth Z (1999) Ensemble Transformation and Adaptive Observations. *J Atmos Sci*
530 56:1748-1765 doi:10.1175/1520-0469(1999)056<1748:ETAAO>2.0.CO;2
- 531 Bishop CH, Etherton BJ, Majumdar SJ (2001) Adaptive Sampling with the Ensemble Transform
532 Kalman Filter. Part I: Theoretical Aspects. *Mon Weather Rev* 129:420-436
- 533 Bjerknes J (1969) Atmospheric Teleconnections from the Equatorial Pacific. *Mon Weather Rev*
534 97:163-172
- 535 Chen DK, Cane MA, Kaplan A, Zebiak SE, Huang D (2004) Predictability of El Niño over the past 148
536 years. *Nature* 428:733-736 doi:10.1038/nature02439

537 Chen DK, Zebiak SE, Busalacchi AJ, Cane MA (1995) An Improved Procedure for El Niño
538 Forecasting: Implications for Predictability. *Science* 269:1699-1702

539 Ding RQ, Li JP, Tseng YH, Sun C, Guo YP (2015) The Victoria mode in the North Pacific linking
540 extratropical sea level pressure variations to ENSO. *J Geophys Res: Atmos* 120:27-45
541 doi:10.1002/2014JD022221

542 Ding RQ, Li JP, Tseng YH, Sun C, Xie F (2017) Joint impact of North and South Pacific extratropical
543 atmospheric variability on the onset of ENSO events. *J Geophys Res: Atmos* 122:279-298

544 Duan WS, Feng F, Hou MY (2018a) An application of particle filter assimilation approach in target
545 observations for ENSO predictions (in Chinese). *Atmospheric Science* 42:677-695

546 Duan WS, Hu JY (2016) The initial errors that induce a significant “spring predictability barrier” for El
547 Niño events and their implications for target observation: results from an earth system model.
548 *Clim Dynam* 46:1-17 doi:10.1007/s00382-015-2789-5

549 Duan WS, Li XQ, Tian B (2018b) Towards optimal observational array for dealing with challenges of
550 El Niño-Southern Oscillation predictions due to diversities of El Niño. *Clim Dynam* 51:3351–
551 3368 doi:10.1007/s00382-018-4082-x

552 Fan HJ, Huang BH, Yang S, Dong WJ (2020) Influence of Pacific Meridional Mode on ENSO
553 evolution and predictability: Asymmetric modulation and ocean preconditioning. *J Clim*
554 34:1881-1901 doi:10.1175/JCLI-D-20-0109.1

555 Gordon NJ, Salmond DJ, Smith AFM (1993) Novel Approach to Nonlinear/Non-Gaussian Bayesian
556 State Estimation. *Radar and Signal Processing, IEE Proceedings F* 140:107-113
557 doi:10.1049/ip-f-2.1993.0015

558 Ham YG, Kug JS (2012) How well do current climate models simulate two types of El Niño? *Clim*
559 *Dynam* 39:383-398 doi:10.1007/s00382-011-1157-3

560 Hendon HH, Eunpa L, Wang G, Oscar A, Debra H (2009) Prospects for predicting two flavors of El
561 Niño. *Geophys Res Lett* 36:L19713 doi:10.1029/2009GL040100

562 Hoell A, Barlow M, Xu T, Zhang T (2018) Cold Season Southwest Asia Precipitation Sensitivity to El
563 Niño–Southern Oscillation Events. *J Clim* 31:4463-4482 doi:10.1175/jcli-d-17-0456.1

564 Hou MY, Duan WS, Zhi XF (2019) Season-dependent predictability barrier for two types of El Niño
565 revealed by an approach to data analysis for predictability. *Clim Dynam* 53:5561-5581
566 doi:10.1007/s00382-019-04888-w

567 Jeong HI et al. (2012) Assessment of the APCC coupled MME suite in predicting the distinctive
568 climate impacts of two flavors of ENSO during boreal winter. *Clim Dynam* 39:475-493
569 doi:10.1007/s00382-012-1359-3

570 Jin EK et al. (2008) Current status of ENSO prediction skill in coupled ocean-atmosphere models.
571 *Clim Dynam* 31:647-664 doi:10.1007/s00382-008-0397-3

572 Kao HY, Yu JY (2009) Contrasting Eastern-Pacific and Central-Pacific Types of ENSO. *J Clim*
573 22:615-632 doi:10.1175/2008JCLI2309.1

574 Kim ST, Yu JY (2012) The two types of ENSO in CMIP5 models. *Geophys Res Lett* 39:221-228
575 doi:10.1029/2012GL052006

576 Kleeman R, Tang Y, Moore AM (2003) The Calculation of Climatically Relevant Singular Vectors in
577 the Presence of Weather Noise as Applied to the ENSO Problem. *J Atmos Sci* 60:2856-2868
578 doi:10.1175/1520-0469(2003)060<2856:TCOCRS>2.0.CO;2

579 Kramer W, Van Leeuwen P, Pierini S, Dijkstra HA (2012) Measuring the impact of observations on the
580 predictability of the Kuroshio Extension in a shallow-water model. *J Phys Oceanogr* 42:3-17
581 doi:10.1175/JPO-D-11-014.1

582 Kug JS, Choi J, An SI, Jin FF, Wittenberg AT (2010) Warm pool and cold tongue El Niño events as
583 simulated by the GFDL CM2.1 coupled GCM. *J Clim* 23:1226-1239
584 doi:10.1175/2009JCLI3293.1

585 Kug JS, Jin FF, An SI (2009) Two types of El Niño events: cold tongue El Niño and warm pool El Niño.
586 *J Clim* 22:1499-1515 doi:10.1175/2008JCLI2624.1

587 Liu J, Kalnay E (2008) Estimating observation impact without adjoint model in an ensemble Kalman
588 filter. *Q J R Meteorol Soc* 134:1327-1335 doi:10.1002/qj.280

589 Liu T (2021) ENSO Predictability over the Past 137 Years Based on a CESM Ensemble Prediction
590 System. *J Clim* 35:763-777 doi:10.1175/JCLI-D-21-0450.1

591 Lorenz E (1975) Climate predictability. In: *The physical bases of climate and climate modelling*. vol
592 16. WMO GARP, Geneva, pp 132-136

593 McPhaden MJ (2015) Playing hide and seek with El Niño. *Nat Clim Change* 5:791-795
594 doi:10.1038/nclimate2775

595 Min QY, Su JZ, Zhang RH (2017) Impact of the South and North Pacific Meridional Modes on the El
596 Niño–Southern Oscillation: Observational Analysis and Comparison. *J Clim* 30:1705-1720

597 doi:10.1175/jcli-d-16-0063.1

598 Moore AM, Kleeman R (1996) The dynamics of error growth and predictability in a coupled model of
599 ENSO. *Q J R Meteorol Soc* 122:1405-1446

600 Mu M, Duan WS, Chen DK, Yu WD (2015) Target observations for improving initialization of
601 high-impact ocean-atmospheric environmental events forecasting. *Natl Sci Rev* 2:226-236
602 doi:10.1093/nsr/nwv021

603 Mu M, Duan WS, Wang B (2003) Conditional nonlinear optimal perturbation and its applications.
604 *Nonlinear Proc Geoph* 10:493-501 doi:10.5194/npg-10-493-2003

605 Palmer T, Gelaro R, Barkmeijer J, Buizza R (1998) Singular Vectors, Metrics, and Adaptive
606 Observations. *J Atmos Sci* 55:633-653
607 doi:10.1175/1520-0469(1998)055<0633:SVMAAO>2.0.CO;2

608 Philander SGH (1983) El Niño Southern Oscillation phenomena. *Nature* 302:295-301

609 Qi Q, Duan W, Xu H (2021) The most sensitive initial error modes modulating intensities of CP- and
610 EP- El Niño events. *Dynam Atmos Oceans*:101257 doi:10.1016/j.dynatmoce.2021.101257

611 Ren H et al. (2018) Seasonal predictability of winter ENSO types in operational dynamical model
612 predictions. *Clim Dynam* 52:3869-3890 doi:10.1007/s00382-018-4366-1

613 Ren H-L, Zuo J, Jin F-F, Stuecker MF (2016) ENSO and annual cycle interaction: the combination
614 mode representation in CMIP5 models. *Clim Dynam* 46:3753-3765
615 doi:10.1007/s00382-015-2802-z

616 SchneiderT, GriffiesSM (1999) A Conceptual Framework for Predictability Studies. *J Clim*
617 12:3133-3155

618 Tang Y, Kleeman R, Miller S (2006) ENSO Predictability of a Fully Coupled GCM Model Using
619 Singular Vector Analysis. *J Clim* 19:3361-3377 doi:10.1175/JCLI3771.1

620 Tang Y et al. (2018) Progress in ENSO prediction and predictability study. *Natl Sci Rev* 5:826–839
621 doi:10.1093/nsr/nwy105

622 Toth Z, Kalnay E (1997) Ensemble Forecasting at NCEP and the Breeding Method. *Mon Weather Rev*
623 125:3297-3319

624 Vimont DJ, Alexander MA, Newman M (2014) Optimal growth of Central and East Pacific ENSO
625 events. *Geophys Res Lett* 41:4027-4034 doi:10.1002/2014GL059997

626 Webster PJ, Yang S (1992) Monsoon and Enso: Selectively Interactive Systems. *Q J R Meteorol Soc*

627 118:877-926

628 Wu L, Zhang HJ, Chen J-M, Feng T (2018) Impact of Two Types of El Niño on Tropical Cyclones over
629 the Western North Pacific: Sensitivity to Location and Intensity of Pacific Warming. *J Clim*
630 31:1725-1742 doi:10.1175/jcli-d-17-0298.1

631 Wu Y, Shen Z, Tang Y (2020) A Flow-Dependent Targeted Observation Method for Ensemble Kalman
632 Filter Assimilation Systems. *Earth Space Sci* 7:e2020EA001149 doi:10.1029/2020EA001149

633 Xie S-P, Philander SGH (1994) A coupled ocean-atmosphere model of relevance to the ITCZ in the
634 eastern Pacific. *Tellus A* 46:340-350 doi:10.3402/tellusa.v46i4.15484

635 Yu J-Y, Kao HY, Lee T (2010) Subtropics-Related Interannual Sea Surface Temperature Variability in
636 the Central Equatorial Pacific. *J Clim* 23:2869-2884 doi:10.1175/2010JCLI3171.1

637 Yu JY, Kao HY (2007) Decadal changes of ENSO persistence barrier in SST and ocean heat content
638 indices: 1958-2001. *J Geophys Res: Atmos* 112:125-138 doi:10.1029/2006JD007654

639 Zhang J, Duan WS, Zhi XF (2015) Using CMIP5 model outputs to investigate the initial errors that
640 cause the “spring predictability barrier” for El Niño events. *Sci China: Earth Sci* 58:685–696
641 doi:10.1007/s11430-014-4994-1

642 Zhang WJ et al. (2016) Unraveling El Niño's impact on the East Asian Monsoon and Yangtze River
643 summer flooding. *Geophys Res Lett* 43:11375-11382 doi:10.1002/2016GL071190

644

Figures

Figure 1

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: $^{\circ}\text{C}$) 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: $^{\circ}\text{C}$) in Niño3 and Niño4 area 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.

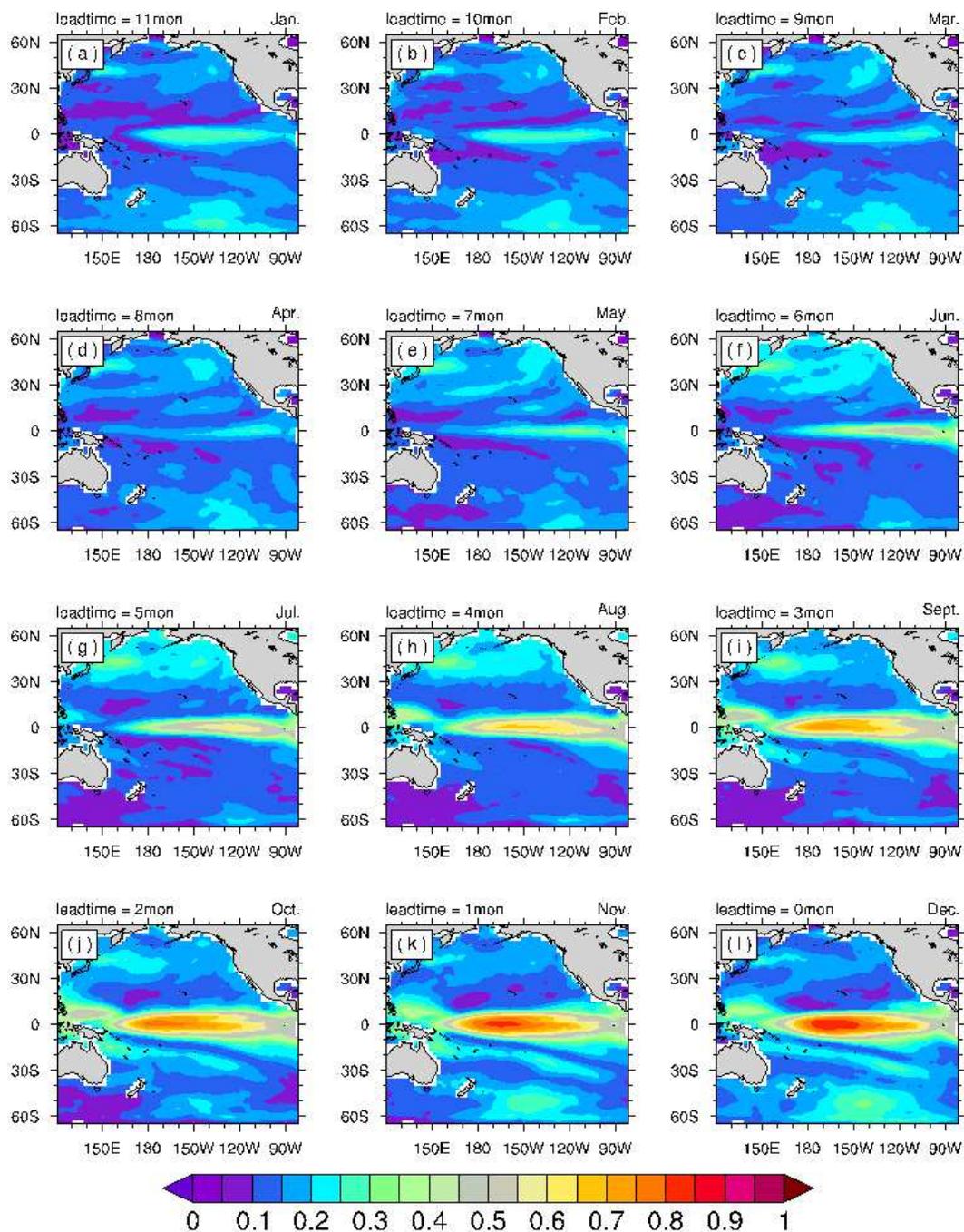


Figure 2

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.

Figure 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.

Figure 4

As in Fig. 2, but for the results of the decrease in the RMSE of the prediction of Niño4 SSTA in December (units: K^2). The area with black dots means that the RMSE of the prediction decrease(increase) in more than 2/3 prediction cases while the composite results over all cases also decrease(increase) after assimilating the observations here. Similarly, the green dots mean greater than 3/4.

Figure 5

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

Figure 6

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

Figure 7

As in Fig. 4, but for the results of the decrease in the RMSE of the prediction of Niño3 SSTA in December (units: K^2).

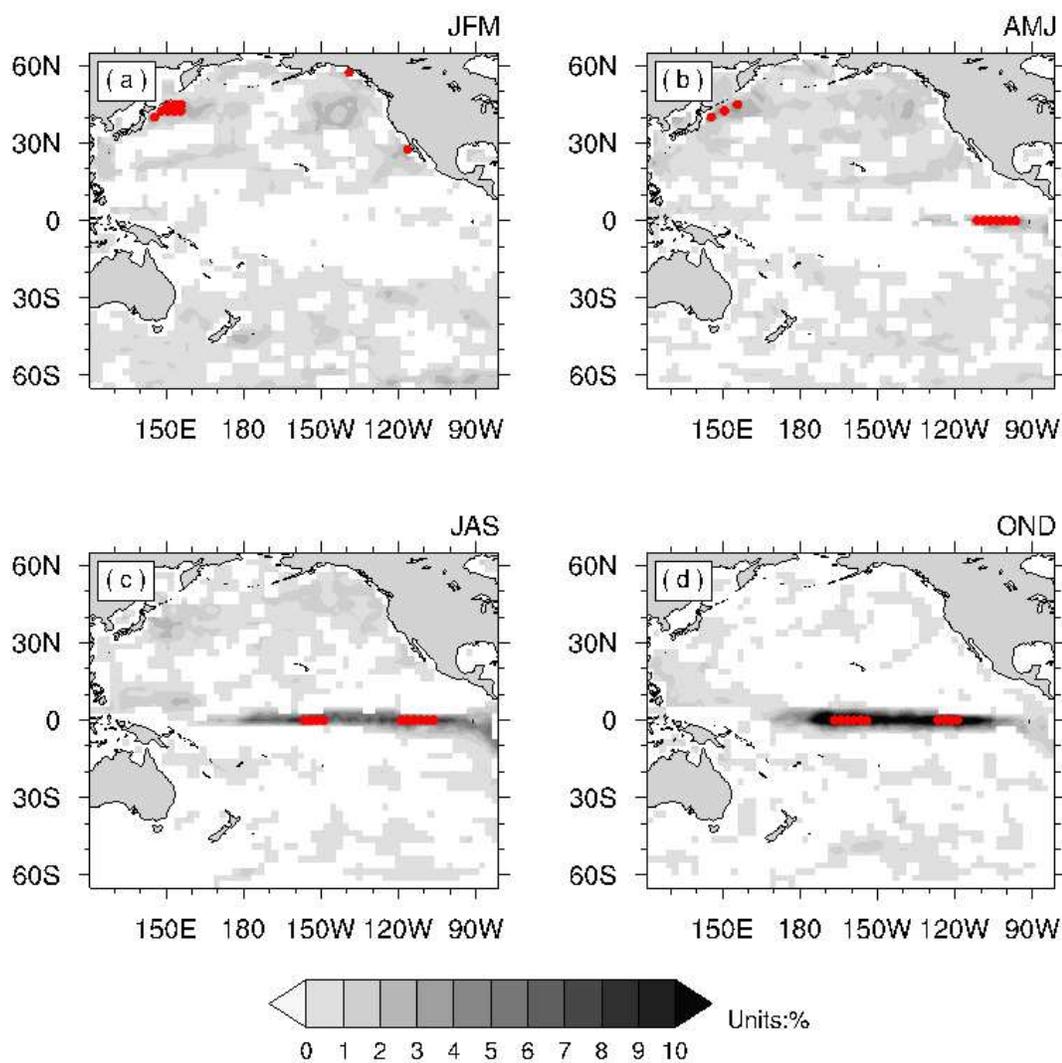


Figure 8

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.

Figure 9

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

Figure 10

Ensemble prediction of Niño3 index (units: $^{\circ}\text{C}$) 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 truth value; green lines represent the ensemble mean forecast after assimilation; and the areas shaded in gray represent the probability.

Figure 11

The bar charts of the skills of the “hindcast” forecast when data assimilations made in the sensitive area (SA, black bars) and other 100 randomly selected arrays (R100, white bars). **(a)** Average prediction errors of Niño3 SSTA in December (units: $^{\circ}\text{C}$); **(b)** average prediction error of the Niño4 SSTA in December (units: $^{\circ}\text{C}$) 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.