



Title	Sea surface temperature predictability in the North Pacific from multi-model seasonal forecast
Author(s)	Yati, Emi; Minobe, Shoshiro
Citation	Journal of oceanography, 77(6), 897-906 https://doi.org/10.1007/s10872-021-00618-1
Issue Date	2021-12
Doc URL	http://hdl.handle.net/2115/87348
Rights	This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature 's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/s10872-021-00618-1
Type	article (author version)
File Information	J. Oceanogr. 77-6_897-906.pdf



[Instructions for use](#)

1 Sea Surface Temperature Predictability in the North 2 Pacific from Multi-Model Seasonal Forecast

3
4 Emi Yati^{1,3*} and Shoshiro Minobe^{1,2}

5 *Email: emiyati@gmail.com or emiyati@lapan.go.id

6
7 ¹ *Department of Natural History Sciences, Graduate School of Science, Hokkaido*
8 *University, Japan.*

9 ² *Department of Earth and Planetary Sciences, Faculty of Science, Hokkaido University,*
10 *Japan.*

11 ³ *Remote Sensing Application Center, Indonesian National Institute of Aeronautics and*
12 *Space, Indonesia.*

13 14 Abstract

15 Sea surface temperature (SST) prediction based on the multi-model seasonal forecast with numerous
16 ensemble members have more useful skills to estimate the possibility of climate events than individual models.
17 Hence, we assessed SST predictability in the North Pacific (NP) from multi-model seasonal forecasts. We used
18 23 years of hindcast data from three seasonal forecasting systems in the Copernicus Climate Change Service to
19 estimate the prediction skill based on temporal correlation. We evaluated the predictability of the SST from the
20 ensemble members' width spread, and co-variability between the ensemble mean and observation. Our analysis
21 revealed that areas with low prediction skills were related to either the large spread of ensemble members or the
22 ensemble members not capturing the observation within their spread. The large spread of ensemble members
23 reflected the high forecast uncertainty, as exemplified in the Kuroshio-Oyashio Extension region in July. The
24 ensemble members not capturing the observation indicates the model bias; thus, there is room for improvements
25 in model prediction. On the other hand, the high prediction skills of the multi-model were related to the small
26 spread of ensemble members that captures the observation, as in the central NP in January. Such high
27 predictability is linked to El Niño Southern Oscillation (ENSO) via teleconnection.

28
29 *Keywords: SST, seasonal forecast, multi-model ensemble, predictability, prediction skill,*
30 *C3S.*

31 **1. Introduction**

32 Prediction skill of sea surface temperature (SST) on a seasonal time scale (one to 12
33 months) in the North Pacific (NP) shares a similar spatial pattern across different studies
34 (e.g., Becker et al. 2014; Doi et al. 2016; Johnson et al., 2019). In winter and summer,
35 prediction skills are generally high in the eastern NP. At the same time, they are usually low
36 in the region between the Kuroshio Extension and the subpolar front to the east of Japan
37 (Wen et al., 2012). Ocean dynamics and atmospheric teleconnection can cause this regional
38 discrepancy in prediction skills. SST and upper ocean heat content under anomalous
39 conditions also takes a relatively long time to decay (i.e., days to years), significantly
40 impacting the atmosphere above (Alexander 1992). This ocean-atmosphere interaction raises
41 climate variability, such as the El Niño Southern Oscillation (ENSO). The ENSO induces
42 SST variability in the NP via teleconnection (Alexander et al. 2002).

43 The coupled ocean-atmosphere models provide a set of forecasts (ensemble members)
44 generated by small perturbations in the initial condition that reflect the uncertainties (Lorenz,
45 1982; Rodwell and Doublas-Reyes, 2006). The average of the ensemble members is the
46 ensemble mean. Accordingly, an ensemble mean represents the predictable component, while
47 the spread of ensemble members represents the unpredictable component (or the uncertainty).
48 The large (small) ensemble spreads are generally associated with high (low) forecast
49 uncertainty (Kirtman et al., 2014; Miller and Wang, 2019). Thus, the predictability of coupled
50 ocean-atmosphere models can be analyzed based on the spread of ensemble members and co-
51 variability between the ensemble mean and observation.

52 Large ensemble members that can represent forecast uncertainty better have recently
53 been applied for seasonal forecasting systems. Doi et al. (2019) produced 108 ensemble
54 members only from one seasonal forecasting model. Nevertheless, multi-model ensembles
55 are more frequently used to obtain large ensemble members. The North American Multi-

56 Model Ensemble (NMME) is one of the multi-model ensemble projects for seasonal
57 forecasting (Kirtman et al., 2014). Becker et al. (2014) reported predictability of global SST
58 by using 109 ensemble members provided nine modeling centers in the NMME.

59 Since October 2019, the Copernicus Climate Change Service (C3S) in Europe
60 provides the output of operational multi-model seasonal forecasts. In May 2020, when we
61 downloaded the data for this study, there were six modeling centers in C3S, including five
62 centers in Europe and one in the United States (Min et al., 2020). The C3S forecasting system
63 is the successor of the earlier European Multi-model Seasonal to Interannual Prediction
64 (EUROSIP) project conducted by the European Center for Medium-Range Weather Forecasts
65 (ECMWF).

66 In this paper, we investigated the SST predictability of multi-model seasonal forecast
67 in the NP using the output of three forecasting systems in C3S (Table 1). The larger ensemble
68 size and a better vertical and horizontal resolution of a seasonal forecast system from C3S are
69 expected to improve the SST prediction skill. Some ensemble predictions using climate
70 models can have a signal-to-noise paradox that the ensemble mean correlates with the
71 observation more than their ensemble members (Scaife and Smith, 2018). Therefore, we
72 analyzed the ensemble members' spread and evaluated the co-variability between the
73 ensemble mean and observation. The rest of the manuscript is designed as follows. Section 2
74 defines the data and methods used in this study. Section 3 assesses the relationship of SST
75 predictability to basin-scale climate variability. A summary of our significant results and
76 discussions is provided in Section 4.

77

78

79

80

81 **Table 1.** Descriptions of three forecasting systems available in the C3S data store.

Modeling Center	Model name	Ensemble size (members)	Atmospheric Horizontal resolution /vertical levels	Ocean Horizontal resolution /vertical levels	References
ECMWF	SEAS5	25	T319 (~36km)/ L91	0.25° /L75	Johnson et al. (2019)
DWD	GCFS 2.0	30	T127 (~100km)/ L95	0.4° /L40	Fröhlich et al. (2021)
CMCC	SPS3	40	1°/L46	0.25° /L50	Sanna et al. (2017)

82

83 **2. Data and Methods**

84 We analyzed seasonal forecast data from three forecasting systems available in the
 85 C3S data store (<https://cds.climate.copernicus.eu/>), namely, ECMWF SEAS5, GCFS 2.0 of
 86 DWD (Deutscher Wetterdienst), and SPS3 of CMCC (Centro-Euro-Mediterraneo Sui
 87 Cambiamenti Climatici). These forecasting systems were initialized on the 1st day of the
 88 starting month, unlike other models in C3S. The forecasting system specification and the
 89 ensemble size of each model are shown in Table 1. SEAS5 improved physics, horizontal and
 90 vertical resolution, and sea-ice reanalysis with up-to-date processing than SEAS4 (Johnson et
 91 al., 2019). The GCFS 2.0 has a larger ensemble size and higher horizontal and vertical
 92 resolution of the model parameters than GCFS 1.0 (Fröhlich et al., 2021). The SPS3 has a
 93 larger ensemble size than the previous system (Sanna et al., 2017). In the following part of
 94 the paper, the names of the modeling centers (ECMWF, DWD, CMCC) are used to
 95 distinguish the models for simplicity.

96 We used 23 years of monthly averaged hindcast data in January and July of 1994-
 97 2016, with a lead time of 3 months on a global 1° × 1° grid. For January (July) forecasts with
 98 the 3-month lead time, the initialization date was November (May) 1st. The analysis presented
 99 here focuses on SST in the NP.

100 We generated a multi-model ensemble (MME) by combining the ensemble members
101 of hindcast data (i.e., reforecast data) produced by the three forecasting systems. Thus, the
102 total ensemble size of the MME was 95 ensemble members. The multi-model ensemble mean
103 (MMEM) was calculated by averaging together those ensemble members. Furthermore, we
104 defined the respective model ensemble means (RMEM) as the average of the ensemble
105 members of each forecasting system.

106 NOAA's Optimum Interpolation Sea Surface Temperature (OISST) version 2 dataset,
107 described by Reynolds et al. (2007), was used to verify the SST prediction. OISST version 2
108 dataset, produced by NOAA/OAR/ESRL Physical Science Laboratory, uses satellite data
109 (i.e., Advanced Very High-Resolution Radiometer (AVHRR) satellite) and in situ records
110 (i.e., from ships and buoys). It is grided at $1^\circ \times 1^\circ$ resolution and is available online at
111 <https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html#detail>.

112 We examined temporal correlation to evaluate SST prediction skills. Temporal
113 correlations were calculated between ensemble mean (e.g., MMEM or RMEM) and observed
114 SST anomalies (SSTA). The temporal correlation is the anomaly correlation coefficient and
115 is widely used to estimate the prediction skill (e.g., Becker et al., 2014; Hervieux et al., 2019;
116 Jacox et al., 2019). We calculated the temporal correlation coefficient between two time
117 series, either at each grid point (point-wise correlation) or for the area-averaged regions of
118 interest.

119 There are two steps to analyze the relation between ensemble members and
120 observation. First, we examined the histogram of temporal correlation distribution between
121 MMEM with ensemble members for averaged SSTA time series over the specific regions.
122 Second, the histogram of those correlation distributions is compared with the corresponding
123 temporal correlation between MMEM and observation. If the MMEM and observation
124 correlation is located within the range of correlations distribution of the MMEM and the

125 respective ensemble members, the forecast system reasonably captures the observation as an
126 ensemble member. If this is not the case (i.e., the observation is an outlier), the forecast
127 system fails to capture the observation. Here, an outlier refers to a correlation less (higher)
128 than the 5th (95th) percentiles of correlations distribution between the MMEM and the
129 ensemble members. The difference of the 5th and 95th percentile between MMEM and
130 ensemble members is defined as the spread of ensemble members. Additionally, we analyze
131 the correlation map between SSTA area-averaged time series and SSTA grid points in the
132 North and the tropical Pacific Ocean for MMEM and observation, in January and July, to
133 characterize the relation between SST predictability in NP and the Niño 3.4 region (5°S-5°N,
134 170°W-120°W).

135 Statistical significance was evaluated using a Monte-Carlo simulation. We perform
136 the following steps in the Monte-Carlo simulation to assess the significance of the
137 relationship between MMEM and observation. First, we generated 100 surrogate time series
138 for the MMEM with the lag-1 correlation of the MMEM time series. Here, a lag-1 correlation
139 refers to the correlation between values that are one month apart. The surrogate correlation
140 coefficients were calculated for each grid point between the observation and the surrogate
141 MMEM time series. The percentile of the absolute value of the observed correlation among
142 the surrogate correlation absolute values was used to estimate the confidence level.

143

144

145

146

147

148 **3. Results**

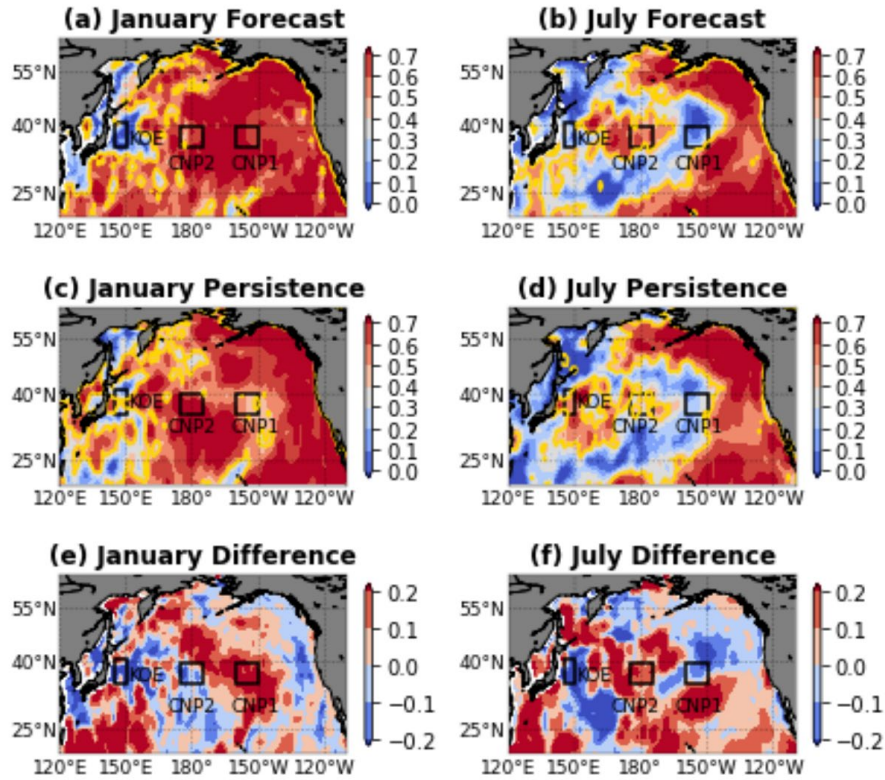
149 **3.1. SST predictability in the NP**

150

151 Figure 1(a) and (b) shows the prediction skills estimated by the point-wise correlation
152 between the MMEM and observation in January and July. The prediction skills were high
153 (>0.5) in the eastern NP, both January and July. In contrast, the prediction skills in January
154 and July were low in the east of Japan. In January, the prediction skills were also high in the
155 central NP, but the high prediction skills existed only in a small area of the central NP
156 (around 180°) in July. These results were essentially the same as the detrended data.

157 Further, we evaluated the persistence skill as the correlation between observed SSTA
158 in November (May) and January (July). In January and July, the high persistence exists in the
159 eastern NP (Fig.1c and d). In July, the high persistence skill also appears in the Kuroshio-
160 Oyashio Extension region (KOE: 35° - 41° N, 145° - 150° E). In contrast, the low persistence skill
161 is found around 150° - 160° W of the central NP in July.

162 Figure 1(e) and (f) show the difference between prediction and persistence skills. The
163 prediction skill was higher than persistence skill in the central NP and the southwestern NP in
164 January and July. In contrast, the prediction skill was lower than the persistence skill in the
165 KOE region. To understand how ensemble members and observation are related with respect
166 to prediction skill, we focus our attention on three areas of interest (i.e., the KOE, the CNP1:
167 35° - 40° N, 150° - 160° W and the CNP2: 35° - 40° N, 175° E- 175° W).



168

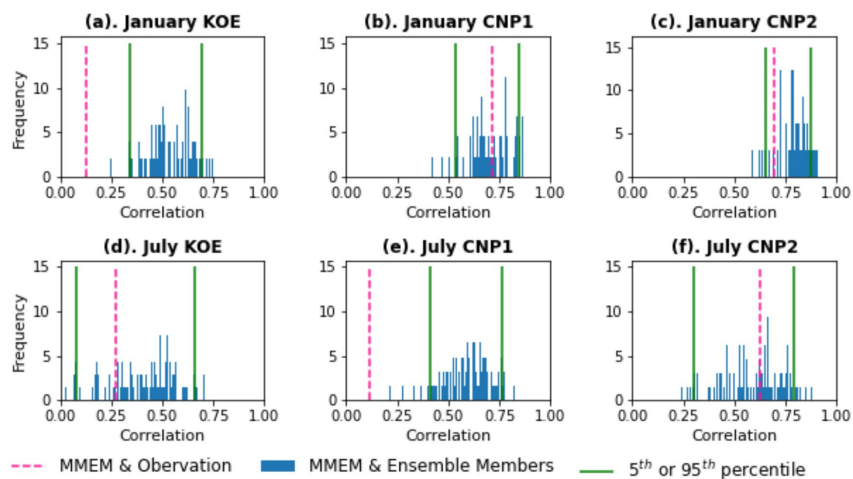
169 Fig. 1. (a)-(b) Prediction skill of NP based on the point-wise correlation between MMEM and observation in
 170 January and July. (c)-(d) Persistence of NP based on the point-wise correlation between observed SSTA in
 171 November (May) and January (July). (e)-(f) Difference between prediction and persistence skill. Black
 172 rectangles indicate regions of interest, i.e., the KOE, the CNP1, and the CNP2. Colors show point-wise
 173 correlation values, with yellow contours showing areas where correlations were significant at the 95%
 174 confidence level.

175

176 Figure 2 (a)-(c) shows histograms of temporal correlation for area-averaged SSTA
 177 between MMEM and ensemble members, along with the correlation between MMEM and
 178 observation in the three regions of interest in January. For the January KOE, the MMEM and
 179 observation correlation was less than the 5th percentiles of the correlations between the
 180 MMEM and ensemble members, indicating that the observed variability was an outlier.
 181 Further, we defined the ensemble spread as a distance of the 5th to 95th percentiles. In
 182 January, the ensemble spread of the CNP1 and CNP2 (0.31 and 0.23; Fig. 2b and 2c) was
 183 smaller than the KOE (0.36; Fig.2a). Correlations of MMEM and observation in the January
 184 CNP1 and CNP2 were high and located between the correlation distribution of the MMEM

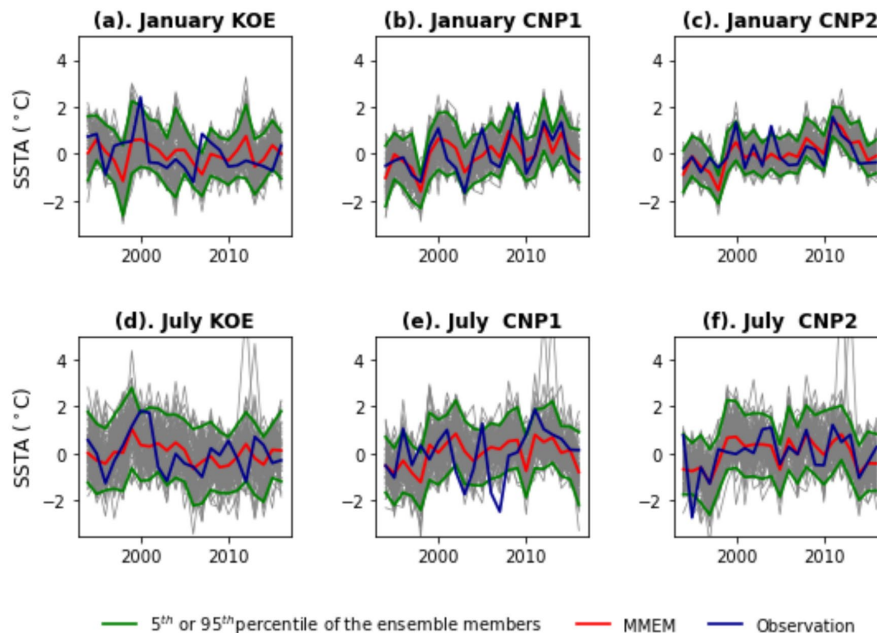
185 and the ensemble members. These results indicate that the observation can be considered as
 186 an ensemble member for the CNP1 and CNP2.

187 Figure 2 (d)-(f) shows histograms of temporal correlations between the MMEM and
 188 ensemble members, along with the MMEM and observation for area-averaged over the
 189 regions of interest in July. The high temporal correlation between the MMEM and
 190 observations of the CNP2 was within the range of correlations distribution between the
 191 MMEM and ensemble members (Fig. 2f). Although low prediction skills exist in both the
 192 KOE and the CNP1 (Fig. 2d and 2e), the relations between ensemble members and
 193 observations differed in these two regions. In the July CNP1, the correlation between the
 194 MMEM and observation was an outlier of correlations distribution between the MMEM and
 195 ensemble members (Fig. 2e), as in the KOE in January (Fig. 2a). In the July KOE, the
 196 correlation between the MMEM and observation was in the middle of a wide correlation
 197 distribution between the MMEM and ensemble members (Fig. 2d). The ensemble spread for
 198 the KOE in July was wider (i.e., 0.58) than in January. Such a large ensemble spread was
 199 related to the low prediction skill in the July KOE.



200
 201 Fig. 2. Histograms of correlation distribution between the area-averaged MMEM and ensemble members (blue
 202 bars), the correlation between the MMEM and observation (vertical pink dashed line), and 5th or 95th percentile
 203 of correlation distribution between the MMEM and ensemble members (green lines) in January (a-c) and July
 204 (d-f) for the KOE, the CNP1, and the CNP2. The distance of the 5th and 95th percentile describes the ensemble
 205 spread.

206 Figure 3 shows the area-averaged time series for the MMEM, the ensemble members,
 207 and the observation of each area of interest in January and July. Consistent with the
 208 histogram analysis of correlation distribution of respective regions for the January and July
 209 forecast (Fig.2), the time series in January and July showed different features in different
 210 areas. For the January KOE (Fig. 3a), the distance of the 5th – 95th percentile of ensemble
 211 members showed a smaller spread than the July KOE (Fig. 3d). In January and July, the
 212 temporal variability of the KOE SSTA time series for the MMEM and observation was little
 213 correlated ($r = 0.13$ in January, $r = 0.27$ in July). Unlike the CNP1 in January (Fig. 3b), the
 214 CNP1 SSTA time series for MMEM in July (Fig. 3e) did not share a similar variation with
 215 observation ($r = 0.71$ in January, $r = 0.12$ in July). In January and July, the MMEM of the
 216 CNP2 SSTA time series shares a common variation with observation ($r = 0.69$ in January, $r =$
 217 0.62 in July). The distance of the 5th – 95th percentile of ensemble members of the CNP1 and
 218 CNP2 in January showed a smaller spread (Fig. 3b and 3c) than in July (Fig. 3e and 3f),
 219 indicating July forecast has higher uncertainty than in January.



220
 221 Fig. 3. Area-averaged SST anomaly (SSTA) time series of the MMEM (red), 5th and 95th percentile ensemble
 222 members (green), ensemble members (gray) and observation (blue) for January and July in the KOE, the CNP1,
 223 and the CNP2.

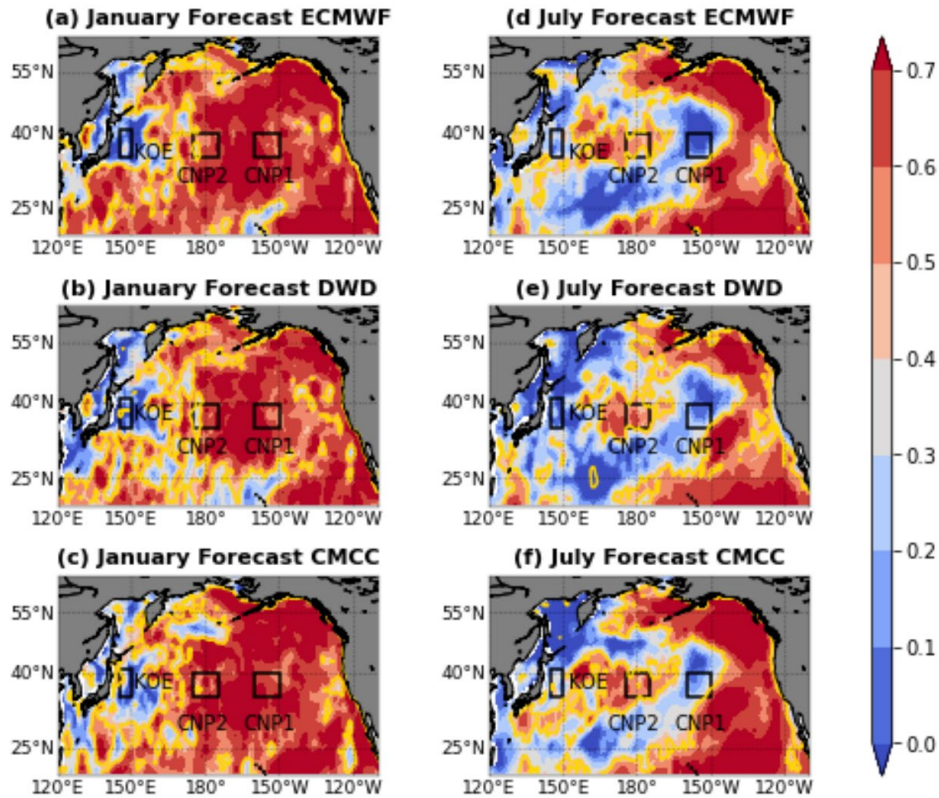
224 Consequently, low predictability is associated with two different types of relations
225 between ensemble members and observations. One is the successful capture of observation
226 features by ensemble members but with a large ensemble spread. The other is the spread of
227 ensemble members unsuccessfully capture the observation. The former case explains the low
228 SST predictability in the multi-model ensemble for the July KOE, while the latter applies to
229 the July CNP1 and the January KOE.

230 Next, we analyze the point-wise correlation between the RMEM and observation. It is
231 valuable to know whether the features found in the MMEM are commonly found in each
232 RMEM or not. It is generally expected that the prediction skill of the MMEM is higher than
233 that of the RMEM because the MMEM's have a large number of ensembles (e.g., Kirtman et
234 al., 2014; Becker et al., 2014).

235 Figure 4 shows the point-wise correlation between each RMEM and observation for
236 January and July forecasts. Generally, the patterns of point-wise correlations for RMEM were
237 similar to those of MMEM (Fig. 1a and 1b). In January, the RMEMs for ECMWF, DWD, and
238 CMCC showed low prediction skills (<0.1) in the KOE and high prediction skills in the
239 CNP1 and the CNP2. In July, the low prediction skill (<0.3) of RMEM was found widely
240 distributed in the KOE and the CNP1, and high prediction skill in the CNP2. This result
241 indicates common mechanisms that operate across models robustly determine the regionality
242 and seasonality of prediction.

243

244



245

246

Fig. 4. Point-wise correlation between the RMEM and observed SSTA for January (a-c) and July (d-f) forecasts.

247

Black rectangles indicate regions of interest, i.e., the KOE, the CNP1, and the CNP2. Colors indicate point-wise

248

correlation values, with yellow contours showing areas where correlations were significant at the 95%

249

confidence level.

250

251

We analyzed the area-average prediction skill of MMEM and RMEM with respect to

252

the ensemble spread, both in January and July (Figure 5). Like the high SST predictability in

253

MMEM, the high SST predictability in RMEM was also related to the small spread of

254

RMEM that captures the observation, as seen in January and July CNP2. In contrast, low

255

predictability due to the ensemble members not capturing the observation within their spread

256

also occurs across the RMEM, as in January KOE and July CNP1. The widespread multi-

257

model ensemble members in July KOE was related to the large spread of the ECMWF and

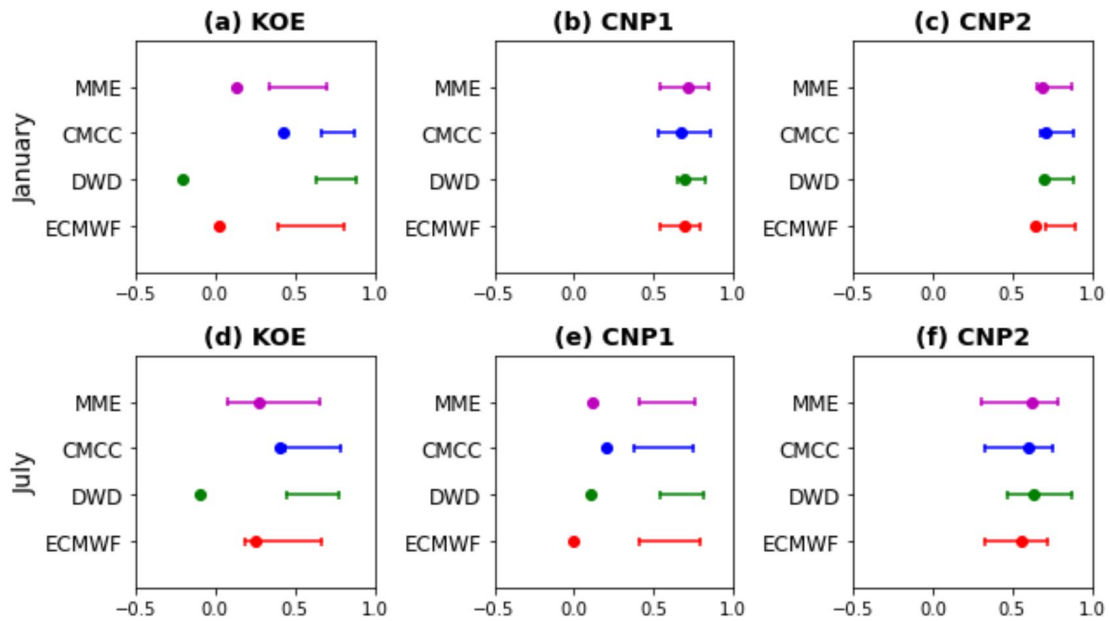
258

CMCC ensemble members. The prediction skill of the CMCC model was higher than the

259

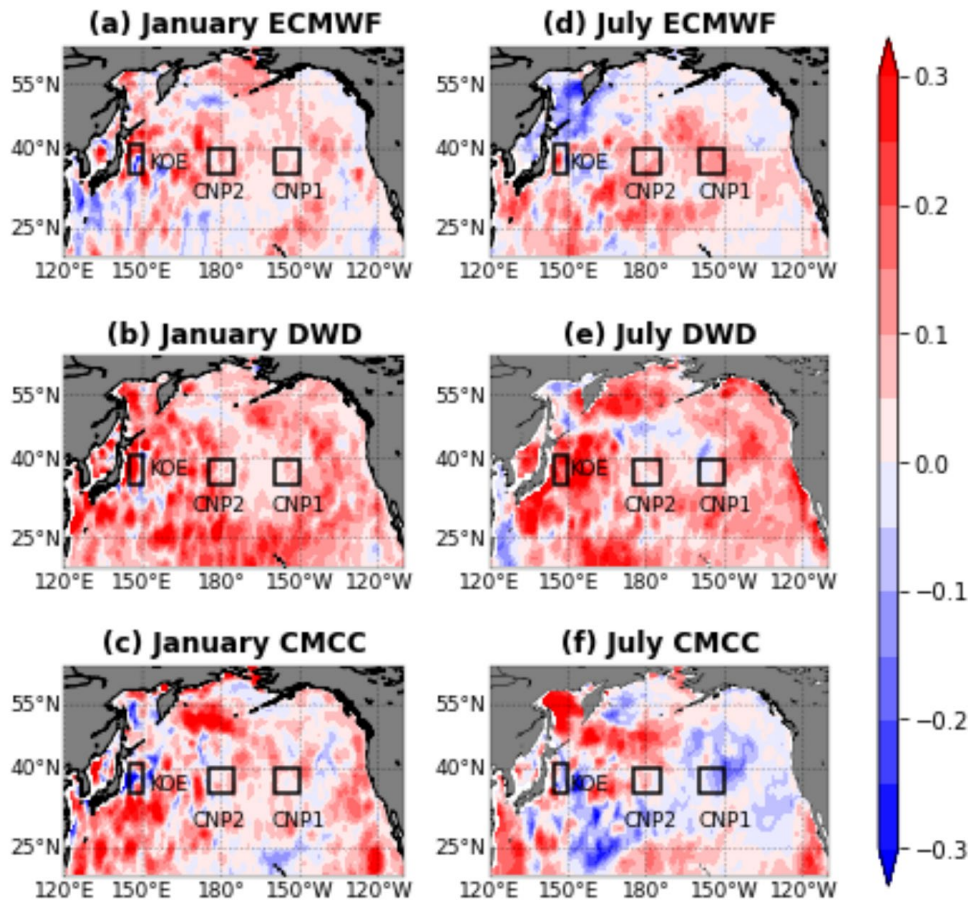
multi-model prediction skill, as seen in January KOE and in July CNP1. It may be due to a

260 better vertical and horizontal resolution of the CMCC than the DWD model and a larger
 261 ensemble size of the CMCC than the ECMWF and DWD models.



262
 263 Fig. 5. Prediction skill for MMEM and RMEM (dots) and corresponding ensemble spread (horizontal lines).
 264 The prediction skills are the correlations between observation and ensemble mean (either MMEM and RMEM)
 265 for area-averaged SSTA time series. The ensemble spread is the width between 5th and 95th percentiles of
 266 correlations between ensembles members and ensemble mean. The top and bottom rows are for January and
 267 July forecasts, and left, middle and right columns are for KOE, CNP1 and CNP2.

268
 269 Figure 6 shows the difference between the point-wise correlation of the MMEM with
 270 observation and the point-wise correlation of the RMEM and observation in January and July.
 271 Our results show that the difference is generally positive, indicating that the point-wise
 272 correlation between the MMEM and observation is usually higher than the RMEM and
 273 observation, as expected. The magnitudes of difference are similar between the ECMWF and
 274 CMCC model but more extensive for the DWD model relative to the other two. Relatively
 275 coarse model resolution of the DWD forecast system than to the other two forecast systems
 276 (Table 1) might be related to the large difference between the MMEM, and the DWD
 277 ensemble mean.



278

279 Fig. 6. The difference between MMEM prediction skill and RMEM prediction skill. The prediction skills are the
 280 point wise correlations between ensemble mean (MMEM or RMEM) and observation at each grid point. Black
 281 rectangles indicate regions of interest.

282

283 3.2. SST Predictability related to ENSO variability

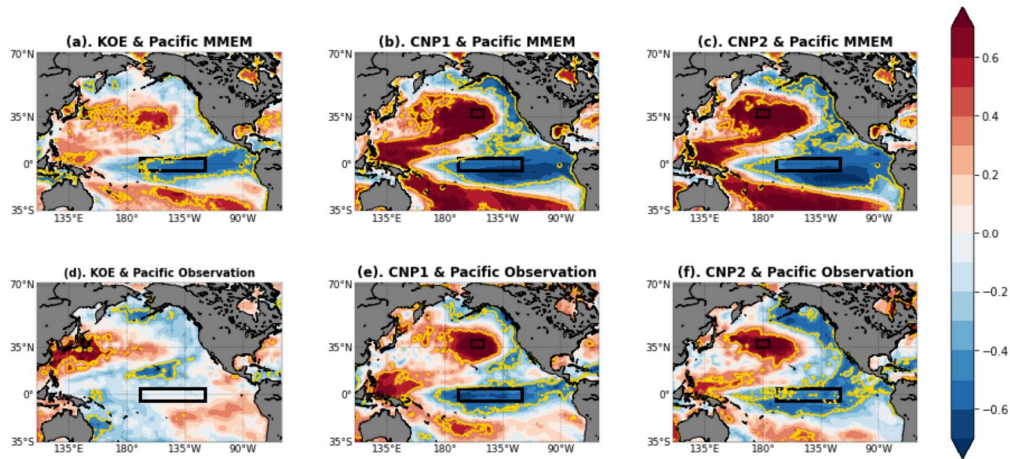
284

285 Figures 7 and 8 show the MMEM and observation correlation maps for January and
 286 July forecasts. These correlation maps are generated based on correlations between averaged
 287 regions of interest (the KOE, the CNP1, and the CNP2) and grid points from the North
 288 Pacific and the tropical Pacific Oceans.

289 The correlation map between the KOE SSTA time series and SSTA grid point for the
 290 MMEM in January (Fig. 7a) exhibited significant positive correlations in the KOE region to
 291 140°W and a significant negative correlation with the Niño 3.4 region ($r=-0.41$; Table 2). The
 292 correlation maps between the KOE SSTA time series and the SSTA grid point for the

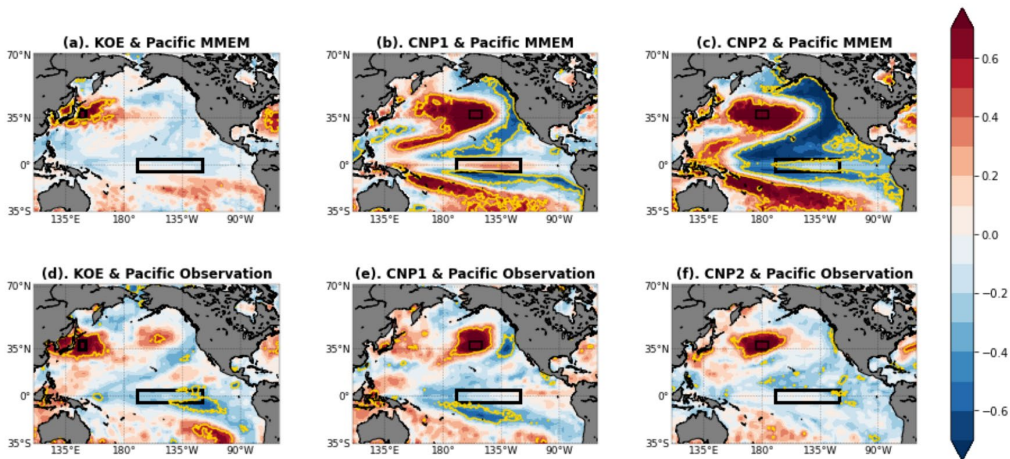
293 observation in January (Fig. 7d) showed a significant positive correlation in the KOE and a
294 low correlation with ENSO ($r = -0.06$). In July, the correlation map between the KOE SSTA
295 time series and SSTA grid points for the MMEM (Fig. 8a) shows a robust significant
296 correlation in the vicinity of the KOE and a very low correlation in the Niño 3.4 region.
297 Indeed, the correlation between the KOE SSTA time series for the MMEM and the Niño 3.4
298 index was 0.09, and the correlation between the KOE SSTA time series for the observation
299 and the Niño 3.4 index was -0.30 (Table 2). The low correlations between KOE and Niño 3.4
300 time series for both the MMEM and the observation in the two seasons would be related to
301 low predictability in the KOE region. These results suggest that the influence of
302 teleconnection associated with ENSO is too weak in the KOE region to yield high
303 predictability.

304 The correlation maps between the SSTA time series for the MMEM and observation
305 in January CNP1 show significant positive correlations in the vicinity of the CNP1 region to
306 the western tropical Pacific and a significant negative correlation in the eastern tropical
307 Pacific (Fig.7b and 7e). The SSTA time series of the CNP1 MMEM and observation in
308 January were significantly correlated with Niño 3.4 (Table 2). The SSTA time series for the
309 MMEM and the observation in January CNP2 also show a similar pattern with CNP1 (Fig.7c
310 and 7f). These results indicate that the high SST predictability of the CNP1 and CNP2 in
311 January is linked to ENSO. ENSO in the tropical Pacific influences the January SST
312 variability of CNP1 and CNP2 through teleconnection (Alexander et al., 2002; Yeh et al.,
313 2018).



314

315 Fig. 7. (a-c) Correlation maps between area-averaged SSTA of a specific region (KOE, the CNP1, and the
 316 CNP2) and the SSTA grid point for MMEM in January. (d-f) Correlation maps with SSTA grid point for the
 317 observation, otherwise following (a-c). The black rectangles mark the areas of interest and the Niño 3.4 region.
 318 The colors indicate correlation values, and the yellow contours are significant correlations at the 95%
 319 confidence level.



320

321 Fig. 8. Same as Fig 7, but for July.

322

323 **Table 2.** Correlations between SSTA time series of area interest and Nino 3.4 index for the MMEM in January
 324 and July, and correlations between SSTA time series of area interest and Nino 3.4 index for the observation in
 325 January and July. Obs indicate the observation. Bold fonts show significant correlations at a 95% confidence
 326 level.

Forecast month	KOE		CNP1		CNP2	
	MMEM	Obs	MMEM	Obs	MMEM	Obs
January	-0.41	-0.06	-0.62	-0.57	-0.59	-0.48
July	0.09	-0.30	0.11	-0.17	-0.23	-0.10

327 For the MMEM in July, correlation maps between the SSTA time series of CNP1 and
328 the SSTA grid point for the MMEM in July exhibit a significant positive correlation in the
329 vicinity of the CNP1 region and a significant negative correlation in the eastern NP (Fig. 8b).
330 However, the correlations between the SSTA time series of CNP1 and the SSTA grid point in
331 the Niño 3.4 region for the MMEM in July were less than in January (Table 2). Consistently,
332 the correlation maps between the SSTA time series of CNP1 and the SSTA grid point for the
333 observation in July (Fig. 8e) show the significant correlations only in the vicinity of the
334 CNP1 and weaker correlation in the Niño 3.4 region. The correlation map between the CNP2
335 SSTA time series and the SSTA grid point for the MMEM in July shows a strong significant
336 positive correlation in the vicinity of the CNP2 area and a significant negative correlation in
337 the eastern NP (Fig. 8c). In contrast, the correlation map between the CNP2 SSTA time series
338 and the SSTA grid point for the observation in July showed a significant correlation in the
339 vicinity of the CNP2 (Fig.8f). It indicates that the teleconnection related to ENSO between
340 the tropical Pacific and the CNP2 observed SSTA did not exist in July. Indeed, Table 2 shows
341 that the CNP1 and CNP2 SSTA time series for the MMEM in July has a low correlation with
342 Niño 3.4 index. The low correlation with Niño 3.4 index also exists in the CNP1 and CNP2
343 SSTA time series for the observations in July (Table 2).

344

345 **4. Summary and Discussions**

346 We analyzed the SST predictability in MMEM over the NP using seasonal forecast
347 data from ECMWF, DWD, and CMCC for the winter (January) and summer (July) with a
348 lead time of three months and focus on three regions of interest, namely, KOE, CNP1, and
349 CNP2. High SST predictability was linked to a small ensemble spread capturing the
350 observation, as in the January CNP1 and the January and July CNP2. In contrast, the low
351 predictability of the KOE in January and the CNP1 in July were related to the ensemble

352 members not capturing the observation within their spread, indicating bias variability.
353 Besides that, the low SST prediction skill of the July KOE was related to a large ensemble
354 spread showing the high uncertainty.

355 The prediction skill differences (MMEM prediction skill - RMEM prediction skill)
356 were generally positive in January and July. Similar patterns between the prediction skills of
357 RMEM and MMEM indicate that the regionality and seasonality of predictability are robustly
358 determined by mechanisms that commonly occur across models. The positive value of the
359 difference is mainly due to more ensemble members in MMEM than RMEM (Becker et al.,
360 2014; Kirtman et al., 2014). According to Scaife and Smith (2018), the ensemble prediction
361 skill of the climate model grows with the size of ensemble members, although the resolution
362 of seasonal forecast and another physical parameter may also determine the prediction skill.

363 The high SST predictability of the CNP1 and CNP2 of MMEM in January and the
364 CNP2 of MMEM in July, but low SST predictability in July CNP1 were consistent with
365 previous studies (Becker et al., 2014; Doi et al., 2016; Johnson et al., 2019). Our results show
366 that high SST predictability of January CNP1 and CNP2 of MMEM were related to the small
367 ensemble spread that capturing the observation. Moreover, the high prediction skill in the
368 January CNP1 and CNP2 is linked to ENSO through teleconnection. Indeed, the SSTA time
369 series of the CNP1 and CNP2 in January significantly correlates with Niño 3.4 index. ENSO
370 in the tropical Pacific also contributed to the predictability when sea level pressure anomalies
371 were considered (Doi et al., 2020). In contrast, the low predictability in the CNP1 July was
372 related to bias variability across the model, and fixing bias will improve the results.

373 The low prediction skills of the KOE both in January and July were also consistent
374 with the previous studies (Becker et al., 2014; Doi et al., 2016; Johnson et al., 2019).
375 However, the previous studies did not examine the relation between ensemble members and
376 observation. Our analysis exhibit that low SST predictability in the seasonal forecast was

377 either a wide ensemble spread (e.g., the July KOE) or the model not capturing the observation
378 (e.g., the January KOE). This low predictability of the KOE in January and July occurred
379 across the model. It is important to note that the SST variations in the KOE are challenging to
380 be predicted, owing to the chaotic oceanic variability caused by strong currents (Kelly et al.,
381 2010; Wen et al., 2012). Such chaotic variability also generates a wide ensemble spread on
382 the interannual to decadal time scale (Nonaka et al., 2020). The wide ensemble spread
383 indicates high uncertainty in the July KOE. Our result also shows that the teleconnection
384 associated with ENSO is too weak in the KOE region to yield high predictability (e.g., the
385 January KOE).

386

387 **5. Author Contributions**

388 EY and SM designed this study, collected the data, analyzed the results, wrote, and discussed the
389 manuscript.

390

391 **6. Acknowledgements**

392 EY was supported by the doctoral program scholarship from the Research and Innovation Science and
393 Technology Project, Ministry of Research and Technology/ National Research and Innovation Agency of the
394 Republic of Indonesia. SM was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI
395 (JP18H04129, 19H05704). We thank C3S, ECMWF, DWD, and CMCC center for provided the seasonal
396 forecast data and NOAA for the OISST Version 2 dataset. We also thank the reviewers and editors for their
397 constructive comments.

398

399 **7. References**

400

401 Alexander, M. A. 1992. Midlatitude atmosphere–ocean interaction during El Niño. Part I: The North Pacific
402 Ocean. *Journal of Climate* 5 (9): 944–958. doi:10.1175/1520-
403 0442(1992)005<0944:MAIDEN>2.0.CO;2.

404 Alexander M. A., Bladé I., Newman M., Lanzante, J.R., Lau, N-C., Scott, J.D. 2002. The atmospheric bridge:
405 the influence of ENSO teleconnections on air-sea interaction over the global oceans. *Journal of Climate*
406 15(16): 2205–2231. doi: 10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2.

407 Becker, E. J., Van den Dool, H. M., and Zhang, Q. 2014. Predictability and forecast skill in NMME. *Journal of*
408 *Climate* 27(15): 5891–5906. doi:10.1175/JCLI-D-13-00597.1.

409 Doi, T., Behera, S.K., and Yamagata, T. 2016. Improved seasonal prediction using the SINTEX-F2 coupled
410 model. *Journal of Advances in Modeling Earth Systems* 8(4): 1847–1867. Blackwell Publishing Ltd.
411 doi:10.1002/2016MS000744.

412 Doi, T., S.K. Bahera and Toshio Yamagata. 2019. Merits of a 108-member ensemble system in ENSO and IOD
413 Prediction. *Journal of Climate*. 32: 957-972. doi: 10.1175/JCLI-D-18-0193.1.

414 Doi T, Nonaka M and Behera S. 2020. Skill Assessment of Seasonal-to-Interannual Prediction of Sea Level
415 Anomaly in the North Pacific Based on the SINTEX-F Climate Model. *Front. Mar. Sci.* 7:546587. doi:
416 10.3389/fmars.2020.546587.

417 Fröhlich, K., Dobrynin, M., Isensee, K., Gessner, C., Paxian, A., Pohlmann, H., et al. (2021). The german
418 climate forecast system: GCFs. *Journal of Advances in Modeling Earth Systems*, 13, e2020MS002101.
419 <https://doi.org/10.1029/2020MS002101>.

420 Hervieux G, et al. 2019. More reliable coastal SST forecasts from the North American Multimodel Ensemble.
421 *Climate Dynamics* 53: 7153–7168. doi:10.1007/s00382-017-3652-7.

422 Jacox, M.G., Alexander, M.A., Stock, C.A., and Hervieux, G. 2019. On the skill of seasonal sea surface
423 temperature forecasts in the California Current System and its connection to ENSO variability. *Climate*
424 *Dynamics* 53(12): 7519–7533. Springer Verlag. doi:10.1007/s00382-017-3608-y.

425 Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L., Tietsche, S.,
426 Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S.P.E., Mogensen, K., Zuo, H., and Monge-Sanz,
427 B.M. 2019. SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*
428 12(3): 1087–1117. Copernicus GmbH. doi:10.5194/gmd-12-1087-2019.

429 Kelly, K. A., Small, R. J., Samelson, R., Qiu B., Joyce, T. M., Kwon, Y. O., and Cronin, M. F. 2010. Western
430 boundary currents and frontal air–sea interaction: Gulf Stream and Kuroshio Extension. *Journal of*
431 *Climate* 23 (21): 5644–5667. doi:10.1175/2010JCLI3346.1.

432 Kirtman, B.P., et al. 2014. The North American multimodel ensemble: Phase-1 seasonal-to-interannual
433 prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American*
434 *Meteorological Society* 95(4): 585–601. American Meteorological Society. doi:10.1175/BAMS-D-12-
435 00050.1.

436 Lorenz, E. N. 1982. Atmospheric predictability experiments with a large numerical model. *Tellus* 34A: 505–
437 513. doi:10.1111/j.2153-3490.1982.tb01839.x.

438 Miller, D. E., and Wang, Z. 2019. Assessing Seasonal Predictability Source and Windows of High Predictability
439 in the Climate Forecast System, Version 2. *Journal of Climate* 32(4): 1307-1326. doi: 10.1175/JCLI-D-
440 18-0389.1.

441 Min, Y.M., Ham, S., Yoo, J.H., and Han, S.H. 2020. Recent progress and future prospects of subseasonal and
442 seasonal climate predictions. In *Bulletin of the American Meteorological Society*. American
443 Meteorological Society: E640–E644. doi:10.1175/BAMS-D-19-0300.1.

444 Nonaka M, Sasaki H, Taguchi B and Schneider N (2020) Atmospheric-Driven and Intrinsic Interannual-to-
445 Decadal Variability in the Kuroshio Extension Jet and Eddy Activities. *Front. Mar. Sci.* 7:547442. doi:
446 10.3389/fmars.2020.547442

447 Reynolds, Richard W., Thomas M. Smith, Chunying Liu, Dudley B. Chelton, Kenneth S. Casey, Michael G.
448 Schlax, 2007: Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *J. Climate*, 20,
449 5473-5496. doi: <https://doi.org/10.1175/2007JCLI1824.1>.

450 Rodwell, M. J., and Doblas-Reyes, F. J. 2006: Medium-range, monthly, and seasonal prediction for Europe and
451 the use of forecast information. *Journal of Climate* 19: 6025–6046. doi:10.1175/JCLI3944.1.

452 Sanna, A., Borelli, A., Athanasiadis, P., Materia, S., Storto, A., Navarra, A., Tibaldi, S., and Gualdi, S. 2017.
453 RP0285-CMCC-SPS3: The CMCC Seasonal Prediction System 3. Technical report Centro Euro-
454 Mediterraneo sui Cambiamenti Climatici.

455 Scaife, A.A., and Smith, D. A signal-to-noise paradox in climate science. *npj Clim Atmos Sci* 1, 28
456 (2018). <https://doi.org/10.1038/s41612-018-0038-4>.

457 Wen, C.H., Xue, Y., and Kumar A. 2012. Seasonal prediction of North Pacific SSTA and PDOI in the NCEP
458 CFS hindcast. *Journal of Climate* 25 (17): 5689-5710. Doi: 10.1175/JCLI-D-11-00556.1.

459 Yeh, S.-W., Cai, W., Min, S.-K., McPhaden, M. J., Dommenges, D., Dewitte, B., ...Kug, J.-S. 2018. ENSO
460 atmospheric teleconnections and their response to greenhouse gas forcing. *Reviews of Geophysics*, 56,
461 185–206. <https://doi.org/10.1002/2017RG00056>

462

463