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

Emi Yati[1,3*] and Shoshiro Minobe[1, 2]

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

[1] *Department of Natural History Sciences, Graduate School of Science, Hokkaido University, Japan.*

[2] *Department of Earth and Planetary Sciences, Faculty of Science, Hokkaido University, Japan.*

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

Abstract

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

Keywords: SST, seasonal forecast, multi-model ensemble, predictability, prediction skill, C3S.

1. Introduction

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

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

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

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

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

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

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)

2. Data and Methods

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

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

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

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

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

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

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

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

3. Results

3.1. SST predictability in the NP

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

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

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

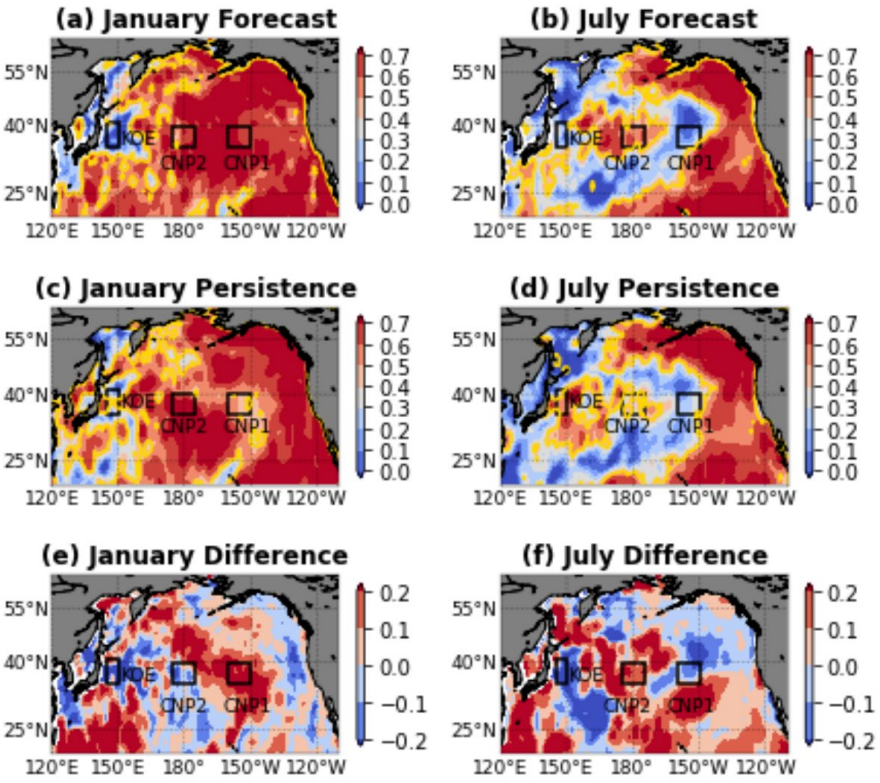

Fig. 1. (a)-(b) Prediction skill of NP based on the point-wise correlation between MMEM and observation in January and July. (c)-(d) Persistence of NP based on the point-wise correlation between observed SSTA in November (May) and January (July). (e)-(f) Difference between prediction and persistence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

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

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

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

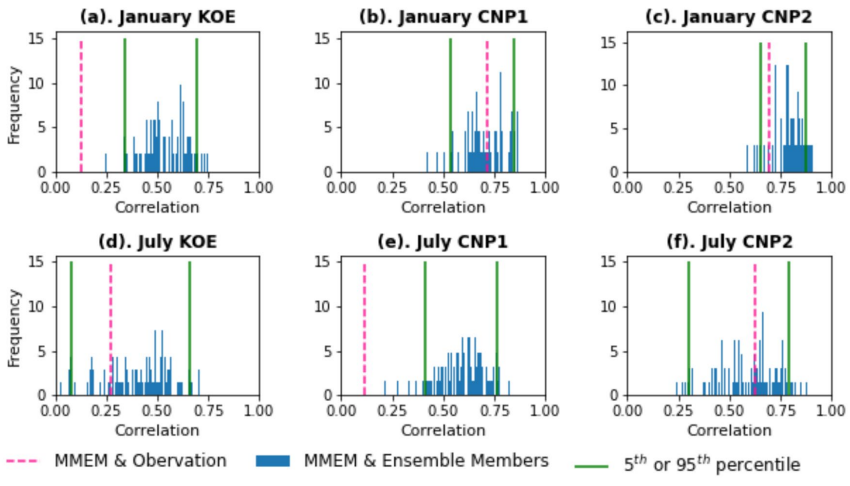

Fig. 2. Histograms of correlation distribution between the area-averaged MMEM and ensemble members (blue bars), the correlation between the MMEM and observation (vertical pink dashed line), and 5th or 95th percentile of correlation distribution between the MMEM and ensemble members (green lines) in January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

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

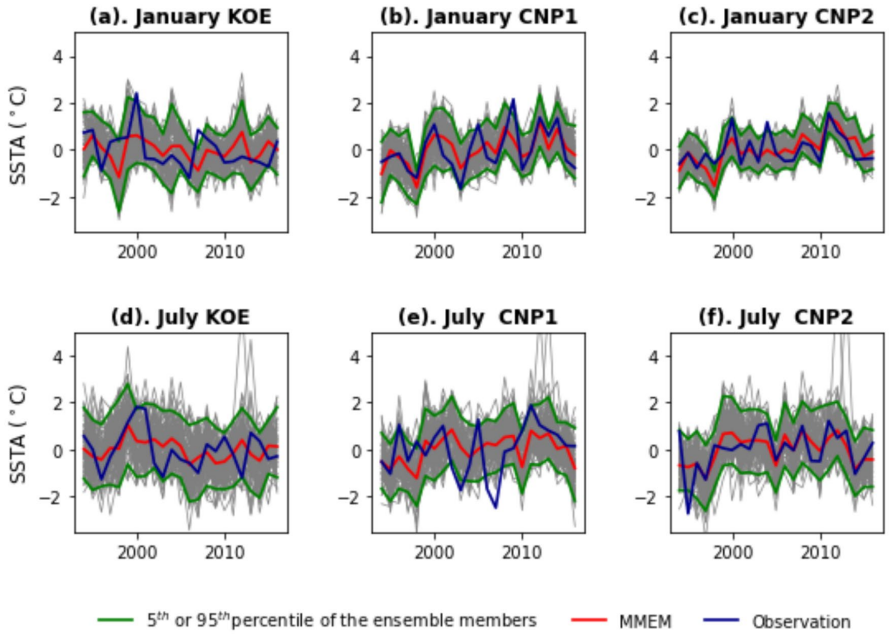

Fig. 3. Area-averaged SST anomaly (SSTA) time series of the MMEM (red), 5^{th} and 95^{th} percentile ensemble members (green), ensemble members (gray) and observation (blue) for January and July in the KOE, the CNP1, and the CNP2.

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

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

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

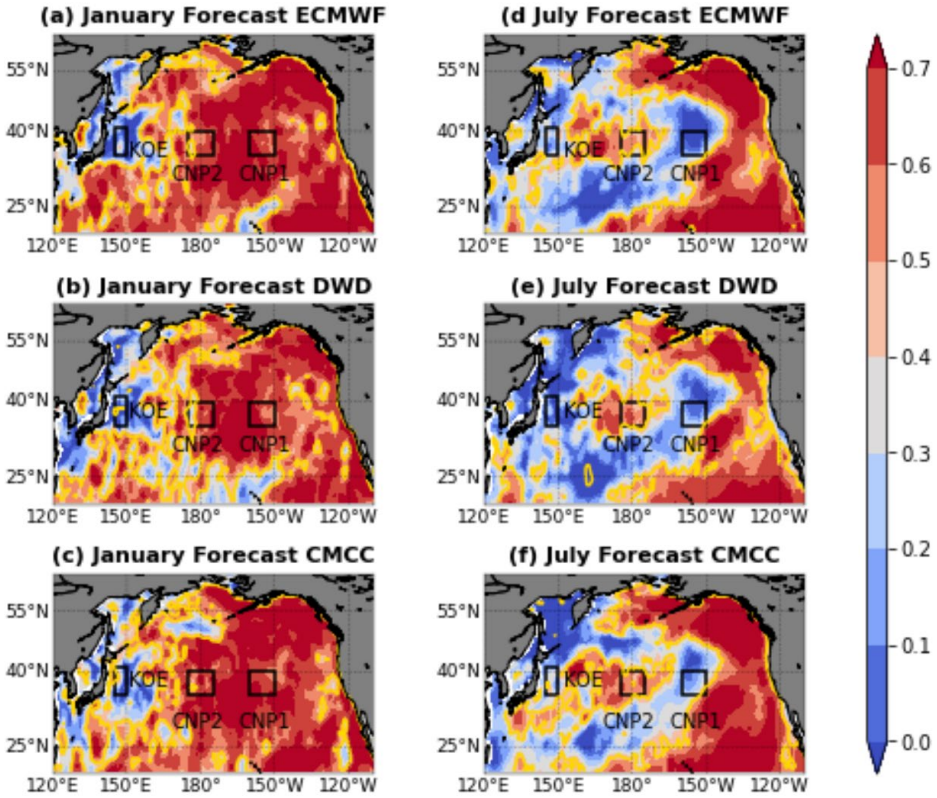

Fig. 4. Point-wise correlation between the RMEM and observed SSTA for January (a-c) and July (d-f) forecasts. Black rectangles indicate regions of interest, i.e., the KOE, the CNP1, and the CNP2. Colors indicate point-wise correlation values, with yellow contours showing areas where correlations were significant at the 95% confidence level.

We analyzed the area-average prediction skill of MMEM and RMEM with respect to the ensemble spread, both in January and July (Figure 5). Like the high SST predictability in MMEM, the high SST predictability in RMEM was also related to the small spread of RMEM that captures the observation, as seen in January and July CNP2. In contrast, low predictability due to the ensemble members not capturing the observation within their spread also occurs across the RMEM, as in January KOE and July CNP1. The widespread multi-model ensemble members in July KOE was related to the large spread of the ECMWF and CMCC ensemble members. The prediction skill of the CMCC model was higher than the multi-model prediction skill, as seen in January KOE and in July CNP1. It may be due to a

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

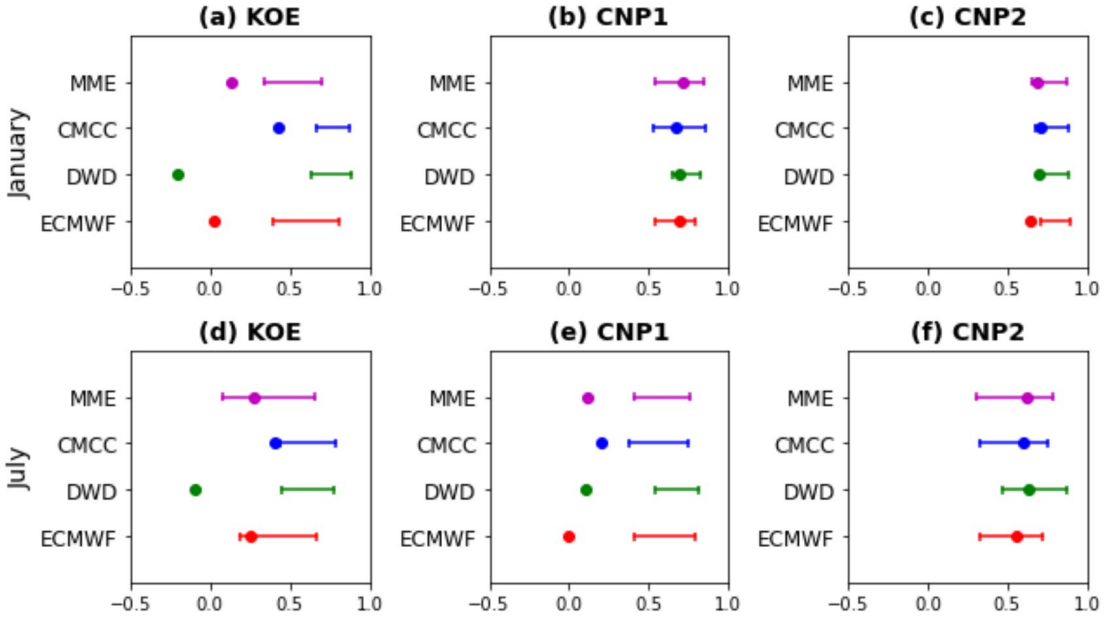

Fig. 5. Prediction skill for MMEM and RMEM (dots) and corresponding ensemble spread (horizontal lines). The prediction skills are the correlations between observation and ensemble mean (either MMEM and RMEM) for area-averaged SSTA time series. The ensemble spread is the width between 5th and 95th percentiles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

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

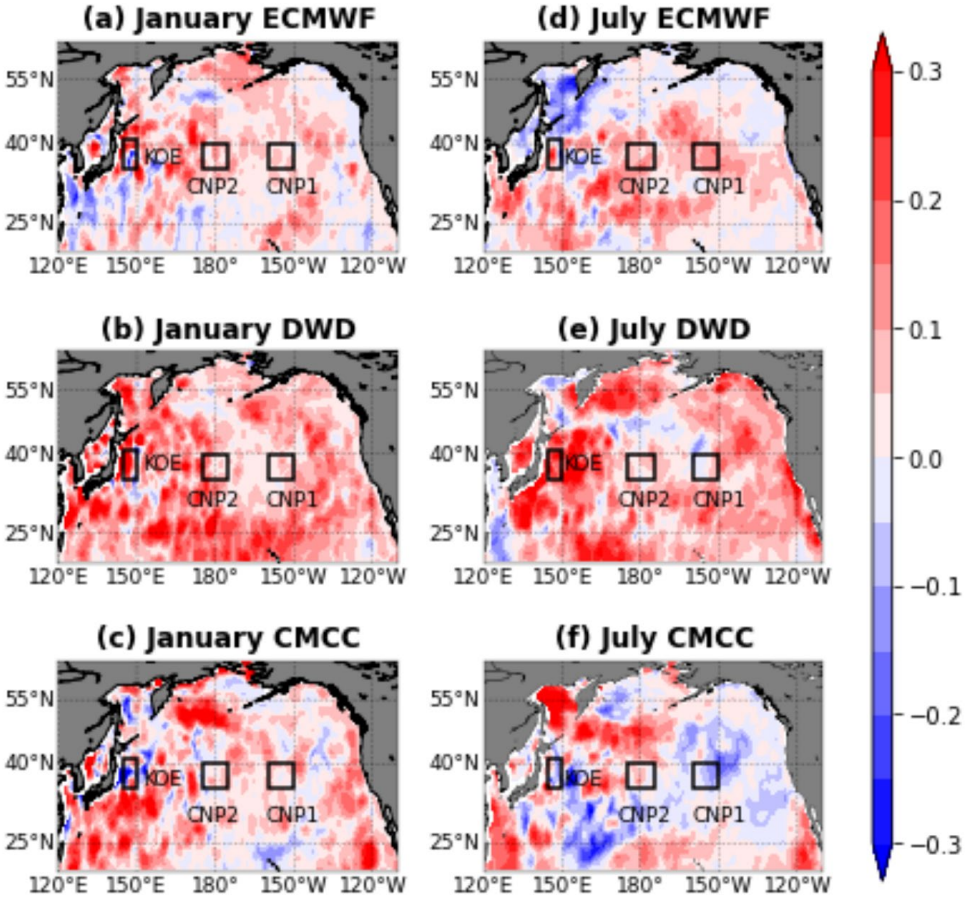

Fig. 6. The difference between MMEM prediction skill and RMEM prediction skill. The prediction skills are the point wise correlations between ensemble mean (MMEM or RMEM) and observation at each grid point. Black rectangles indicate regions of interest.

3.2. SST Predictability related to ENSO variability

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

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

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

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

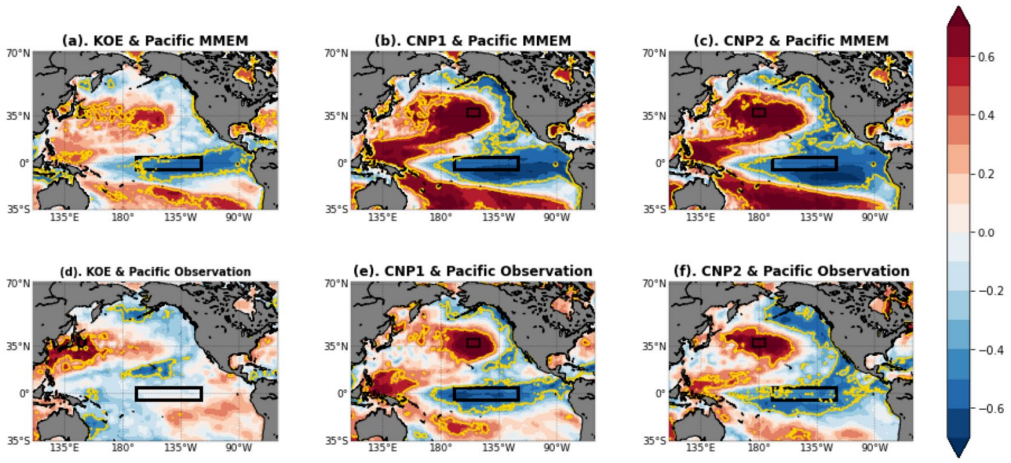

Fig. 7. (a-c) Correlation maps between area-averaged SSTA of a specific region (KOE, the CNP1, and the CNP2) and the SSTA grid point for MMEM in January. (d-f) Correlation maps with SSTA grid point for the observation, otherwise following (a-c). The black rectangles mark the areas of interest and the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

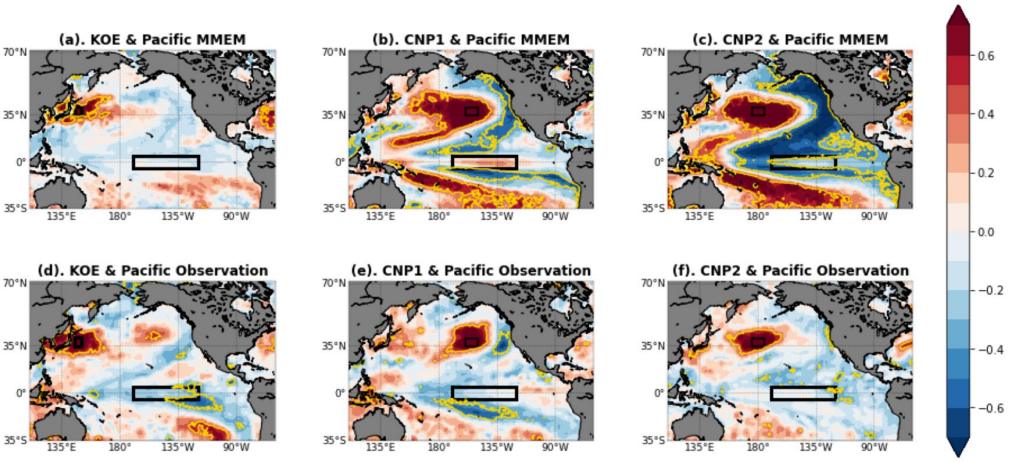

Fig. 8. Same as Fig 7, but for July.

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

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

4. Summary and Discussions

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

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

5. Author Contributions

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

6. Acknowledgements

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

7. References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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