Merits of a 108-Member Ensemble System in ENSO and IOD Predictions

TAKESHI DOI, SWADHIN K. BEHERA, AND TOSHIO YAMAGATA

Application Laboratory, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

(Manuscript received 1 April 2018, in final form 17 September 2018)

ABSTRACT

This paper explores merits of 100-ensemble simulations from a single dynamical seasonal prediction system by evaluating differences in skill scores between ensembles predictions with few (~ 10) and many (~ 100) ensemble members. A 100-ensemble retrospective seasonal forecast experiment for 1983–2015 is beyond current operational capability. Prediction of extremely strong ENSO and the Indian Ocean dipole (IOD) events is significantly improved in the larger ensemble. It indicates that the ensemble size of 10 members, used in some operational systems, is not adequate for the occurrence of 15% tails of extreme climate events, because only about 1 or 2 members (approximately 15% of 12) will agree with the observations. We also showed an ensemble size of about 50 members may be adequate for the extreme El Niño and positive IOD predictions at least in the present prediction system. Even if running a large-ensemble prediction system is quite costly, improved prediction of disastrous extreme events is useful for minimizing risks of possible human and economic losses.

1. Introduction

Simulation of parallel worlds (or ensembles) in numerical seasonal prediction indicates the range of possible future seasonal climate states with the embedded uncertainties owing to the inherent chaotic nature of the system. Some previous studies have suggested that increasing the ensemble size is beneficial for improving seasonal prediction skills, particularly for the extratropical phenomena, where the signal-to-noise ratio is relatively low (Scaife et al. 2014; Chen et al. 2013; Kumar 2009; Déqué 1997; Kumar et al. 2001). However, the ensemble size is usually constrained by the cost of computational resources. Generally, an ensemble size of 10–50 members is adopted for the retrospective forecasts (or reforecast) with operational dynamical seasonal prediction systems in the world, although more ensemble members are used in the forecast systems (Kirtman et al. 2014; Wang et al. 2009; Takaya et al. 2017a,b; Saha et al. 2014; Molteni et al. 2011; Maclachlan et al. 2015; Sanna et al. 2015; Hudson et al. 2013; Vecchi

et al. 2014; Luo et al. 2005; Vernieres et al. 2012; Tompkins et al. 2017). Actually, a 10-member ensemble is the minimum size requested by the Climate-System Historical Forecast Project (CHFP) of the World Climate Research Programme (WCRP; <https://www.wcrp-climate.org/wgsip-chfp>).

It is generally expected that the prediction performance based on a 10-member ensemble system will be less efficient as compared to that of a 100-member ensemble system, when predicting the occurrence of rare and extreme climate events. This is because a fewer number of members in the 10-member system means fewer chances of capturing extreme ends in the observed distributions. Barnston and Mason (2011) showed skillful prediction of the extreme 15% tails of a seasonal climate event by a 100-ensemble multi-AGCM forced by the predicted sea surface temperature (SST). However, since it is not a coupled GCM experiment, it cannot predict a coupled phenomenon like El Niño. On the other hand, using a free long run with a coupled climate model, Lopez and Kirtman (2014) suggest that predictability of extreme El Niño events remains a challenge as the number of ensemble members required to capture these events is on the order of 100 members.

In this paper, we have practically evaluated merits of 100 ensemble simulations by a single dynamical seasonal prediction based on a coupled GCM: Application

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-18-0193.s1>.

Corresponding author: Takeshi Doi, takeshi.doi@jamstec.go.jp

Laboratory, Japan Agency for Marine-Earth Science and Technology (APL/JAMSTEC)-SINTEX-F2 ensemble dynamical seasonal prediction system (Doi et al. 2016, 2017). Since ENSO and the Indian Ocean dipole (IOD) are among the leading sources of seasonal climate predictability in the world, this study tried to access the improvements in their predictability using the large-ensemble system. Those tropical climate modes are strongly coupled with ocean variations, which contain much stronger signals relative to atmosphere variations on seasonal to interannual time scales. Therefore, it still remains an open question whether increasing the ensemble size by a coupled GCM is beneficial for improving the prediction skill in the tropics, although its merits are already shown for the extratropical phenomena, where the atmosphere mainly drives the ocean and the signal-to-noise ratio is low (Scaife et al. 2014; Chen et al. 2013; Kumar 2009; Déqué 1997; Kumar et al. 2001). For example, if we were to find no merits of increasing the ensemble size by a coupled GCM for improving the ENSO and the IOD prediction skill, the enhancement of ensemble size using a forecast system based on a stand-alone AGCM forced by the predicted SST (i.e., so-called two-tier approach), instead of a coupled GCM, could be useful for improvement of the midlatitude climate prediction as shown in Barnston and Mason (2011).

We challenge a 108-member ensemble reforecast system for the 1983–2015 period and analyze its relative benefit as compared to the original 12-ensemble outputs in terms of prediction skill scores. As far as we know, the 108-member ensemble retrospective seasonal forecast by a single system has not yet been conducted, though similar numbers are reported based on multimodel ensemble approach (Kang and Yoo 2006; Kirtman et al. 2014; Tompkins et al. 2017). Multimodel ensembles can sometimes outperform any single-model prediction and provide more robust prediction. While the multimodel approach is very useful, it is usually difficult to address each model's biases/uncertainties and conduct additional hypothesis-driven experiments, considering that those models are usually managed by different operational centers. We hope that the present study may provide a useful guideline to improve operational ensemble prediction systems based on a single-model approach.

2. Methods

a. Dynamical seasonal prediction systems

The dynamical seasonal prediction system used here is based on an ocean–atmosphere–land–sea ice coupled climate model named the SINTEX-F2 coupled model (Masson et al. 2012; Sasaki et al. 2013). The atmospheric

component (ECHAM5) has a horizontal resolution of T106 with 31 vertical levels (Roeckner et al. 2003). The oceanic component [Océan Parallélisé, version 9 (OPA9)] has a horizontal resolution of a $0.5^\circ \times 0.5^\circ$ tripolar grid (known as the ORCA05 configuration), with 31 vertical levels (Madec 2008). The dynamical sea ice model of the Louvain-la-Neuve Sea Ice Model, version 2 (LIM2; Fichefet and Morales Maqueda 1997), is embedded. The atmospheric and oceanic fluxes, including SST, sea ice fraction, freshwater, surface heat, surface current, and momentum fluxes, are exchanged every 2 h with no flux correction by means of the Ocean Atmosphere Sea Ice Soil, version 3 (OASIS3), coupler (Valcke et al. 2004). Several versions of the SINTEX-F2 coupled model are already used for various climate studies: for example, seasonal to interannual variations in both the tropics and extratropics, the Indian summer monsoon, and tropical cyclones (Prodhomme et al. 2014; Joseph et al. 2012; Terray et al. 2016, 2017; Prodhomme et al. 2015, 2016; Terray et al. 2011; Masson et al. 2012; Morioka et al. 2014; Crétat et al. 2016; Sasaki et al. 2014, 2015; Morioka et al. 2013; Sasaki et al. 2013; Ratnam et al. 2017; Crétat et al. 2017; Morioka et al. 2015, 2017).

The initialization scheme is developed by blending the continuous SST-nudging scheme with monthly three-dimensional variational ocean data assimilation (3DVAR) correction (Doi et al. 2017). Modeled SSTs are strongly nudged toward the historical observations in the coupled run continuously from January 1982 after a 32-yr spinup run of strongly nudging to the mean SST climatology. In the SST-nudging run, the specified SSTs force the surface atmospheric variability through the atmospheric component that in turn forces subsurface ocean and sea ice variability through the oceanic component including the dynamical ice model. Then initial conditions for the atmospheric model and the ice model are also provided through the SST-nudging scheme. In addition, 3DVAR correction is conducted every first day of each month from January 1982 to the present using the Met Office Hadley Centre EN4 profile data of subsurface ocean temperature and salinity observations (Good et al. 2013). The details of the 3DVAR scheme used here are discussed in Storto et al. (2011, 2014) and Doi et al. (2017).

This dynamical seasonal prediction system has originally 12 ensemble members, which are used for the quasi-real-time operational seasonal predictions by the APL/JAMSTEC. The 12 ensemble members are generated by three steps: 1) two kinds of observational SST datasets: the weekly OISSTv2 data with a 1.0° latitude $\times$ 1.0° longitude global grid (Reynolds et al. 2002) and the high-resolution daily NOAA OISST analysis with 0.25°

latitude $\times 0.25^\circ$ longitude global grid (Reynolds et al. 2007); 2) three large negative feedback values (-2400 , -1200 , and $-800 \text{ W m}^{-2} \text{ K}^{-1}$) for the SST-nudging scheme in the initialization phase (Luo et al. 2005); and 3) two different modeling ways for the ocean vertical mixing induced by small vertical-scale structures (SVSs; SVS mixing) within and above the equatorial thermocline (Sasaki et al. 2012). Therefore, the 12-ensemble prediction system takes into account to some extent uncertainties of both initial conditions and model physics. Here, we focused on the retrospective forecast with 9-month lead time initialized on 1 June in 1983–2015. More details and the overview of this seasonal prediction system are shown by Doi et al. (2017). This prediction system was also used in multiyear prediction (Morioka et al. 2018b,a; Ogata et al. 2018).

Based on those 12 ensemble members, we increased the ensemble size using the lagged average forecasting (LAF) method (Dalcher et al. 1988). Although it remains unclear if the LAF method is better than the direct perturbation method (Murphy 1990), the ensemble member generated by the LAF method has the advantage of being based on the governing dynamics (Dalcher et al. 1988). We have conducted the 1–8-day integration with the 12 ensemble members by the SST nudging after the initialization date of 1 June during the 1983–2015 period and started the prediction runs from the eight initialized dates (2–9 June during the 1983–2015 period) additionally. This results in 108 members of the ensemble predictions used in this study. The 108-member ensemble includes the original 12-ensemble members from 1 June initial conditions. Early June initialization is good at predicting ENSO and the IOD, both of which are crucial as potential sources of seasonal predictability, because it is after the so-called spring prediction barrier of ENSO events (Webster and Hoyos 2010; Latif et al. 1998; Ren et al. 2016) and the May preconditioning of IOD events (Horii et al. 2008). We have calculated prediction anomalies by linearly removing the model mean climate drifts at each lead time a posteriori using the retrospective forecast outputs during the period from 1983 to 2015. Also, we have calculated the ensemble means simply by averaging results from the original 12-member ensemble and the newly increased 108-member ensemble systems.

b. Prediction skill metrics

As prediction skill scores of the phase and amplitude of the ensemble mean, anomaly correlation coefficient (ACC) and root-mean-square errors normalized by the standard deviation of the observation [normalized root-mean-square error (nRMSE)] are used. When the ACC is higher than the persistence (lag autocorrelation) or

the nRMSE is less than one, the prediction is expected to be skillful. The jackknife resampling is adopted without replacement for 33 events during the 33 years of data during 1983–2015. The statistical significances of the differences in the skills between the 12-member ensemble mean and the 108-member ensemble mean are tested based on a t test above the 95% confidence level.

A small sample size always poses a serious challenge to study the extreme and rare events. It is difficult to verify the statistical significance in our analyses using some of the traditional metrics, as they tend to become trivial in noninformative limits when the base rate tends to become zero. In this study, we have used symmetric extremal dependence index (SEDI) score to capture the extreme and rare events. SEDI is introduced in Ferro and Stephenson (2011) as

$$\text{SEDI} = \frac{\log F - \log H - \log(1 - F) + \log(1 - H)}{\log F + \log H + \log(1 - F) + \log(1 - H)},$$

where H is the hit rate and F is the false-alarm rate, which are calculated by use of different probability thresholds to trigger advisory actions. SEDI ranges between -1 and 1 , with 0 corresponding to no skill. SEDI approaches 1 when the forecast approaches perfection: when $H \rightarrow 1$ and $F \rightarrow 0$. This skill score is independent of the base rate (climatological probability) and can provide a nondegenerative limit, even for very rare events. Also, SEDI is more difficult to hedge and is asymptotically equitable. Therefore, we feel the choice of statistics of SEDI is appropriate for this study. The standard error derived by the delta method is used to assess the 95% confidence intervals of the SEDI scores (cf. Ferro and Stephenson 2011). The score is a more refined measure of determining where the tails of the probability density function are and is better predicted in comparison with the relative operating characteristics scores (ROCS; Mason and Graham 2002). The latter provides a summary of the ROC curve of hit rate and false-alarm rate and is a measure of potential performance of probabilistic prediction. Moreover, ROCS may underestimate the prediction skill in the extratropics (Eade et al. 2014). From deterministic anomaly prediction to probabilistic prediction, we simply counted ensemble members that fall above or below thresholds of the 15% tails of the historical records.

We also have estimated the potential economic value as a function of cost–loss ratio to help the decision-making process (Richardson 2003). The value of a forecast system is defined as the reduction in mean expense relative to the reduction that would be obtained by using perfect forecasts. Potential economic values are normalized so that a perfect deterministic oracle has a

FIG. 1. Time series of Niño-3.4 averaged in November–December ($^{\circ}\text{C}$) from observation (gray bar), the prediction from the 12-member ensemble (light blue: each ensemble member; blue $\times$: ensemble mean) and the 108-member ensemble (orange: each ensemble member; red $\times$: ensemble mean).

value of one and climatological probability has a value of zero. When the value is larger than zero, the decision-maker will gain some economic benefit by using the forecast information in addition to using the climatological rate information.

To evaluate the prediction results, we have used the NOAA OISSTv2 (Reynolds et al. 2002) for SST, the NCEP–NCAR reanalysis data (Kalnay et al. 1996) for 2-m air temperature and the CPC Merged Analysis of Precipitation (CMAP) dataset (Xie and Arkin 1996) for precipitation. The monthly climatologies are calculated by averaging monthly data over the period from 1983 to 2015, and then anomalies are defined as deviations from them.

3. Results

The ENSO and the IOD are the two dominant modes of interannual climate variations on the globe and can work as potential sources of seasonal predictability (Philander 1989; Saji and Yamagata 2003; Luo et al. 2007; Wang et al. 2003). Here, we examine differences in skill for predicting these two climate modes using the two different ensembles.

TABLE 1. Prediction skill scores of the ensemble-mean Niño-3.4 in November–December initialized in early June (see section 2 in the text).

	For skillful prediction, the value should be	12-member ensemble	108-member ensemble
ACC	>0.55 (persistent)	0.88	0.87
nRMSE	<1	0.51	0.51

a. ENSO

First, we focus on the prediction of ENSO, which is the most important source of seasonal predictability in the global climate (Shukla et al. 2000; Hoerling and Kumar 2002; Kosaka et al. 2013). Niño-3.4, the most popular index of ENSO, is defined as the SST anomalies averaged over the domain of 5°S – 5°N , 170° – 120°W . The predictions are initiated in June, which is just after the spring prediction barrier (Webster and Hoyos 2010; Latif et al. 1998). Figure 1 shows the time series of the Niño-3.4 averaged in November–December, (i.e., when most ENSO events mature). At a glance, we see that the 12-member ensemble mean and the 108-member ensemble mean are almost the same. As shown in Table 1, the prediction skill score of the ensemble mean does not show any statistically significant improvement by increasing the ensemble members from 12 to 108. However, the prediction of extreme events is improved. Figure 2 shows the prediction plumes of the top three strongest events: 1988/89 La Niña, 1997/98 El Niño, and 2015/16 El Niño. For the 1988/89 La Niña, the largest-amplitude member of the original 12-member ensemble showed -2.1°C in November–December 1988 and still underestimated the actually observed amplitude of -2.4°C . On the other hand, the 108-member ensemble captured the observed amplitude as the tails of the probability density function (PDF). About 5 members of the 108-member ensemble show the similar amplitude to the observation, and a few members predicted the stronger relative to the observation. For the 1997/98 El Niño and the 2015/16 El Niño, both the 12-member ensemble and the 108-member ensemble captured the observed amplitude as the tails of the PDFs. We can also see that probability of the Niño-3.4 averaged in November–December 1997 with amplitude above 2.5°C is 8% in the 12-member ensemble, while it is 14% in the 108-member ensemble. Hence, the probability for the extreme event is slightly enhanced for the 1997/98 El Niño.

As the prediction skill score of extreme and rare events, the SEDI skill scores for the extreme 15% tails of Niño-3.4 averaged in November–December are shown in Figs. 3a and 3b. The decision (early warning or advisory action) for rare events based on the 108-member ensemble is more skillful relative to the 12-member

FIG. 2. (a) Monthly Niño-3.4 in 1988/89 ($^{\circ}\text{C}$) from the observational data of NOAA OISSTv2 (black) and the prediction from the 1 Jun 1988 initialization with the 12-member ensemble (thin light blue: each ensemble member; thick blue: ensemble mean) and the 108-member ensemble (thin orange: each ensemble member; thick red: ensemble mean). (b) As in (a), but for 1997/98. (c) As in (a), but for 2015/16.

ensemble when the trigger threshold of probability for the extreme event is beyond 35%, which is about 2.5 times the climatological probability. The key years to explain the improvement are the 2002 El Niño and 2010 La Niña. The probability prediction of 15% tails of a positive extreme event for the 2002 El Niño is 33% by the 12-member ensemble size, while its probability is enhanced to 56% by the 108-member ensemble (Fig. 1a in the online supplemental material). Also, the probability prediction of 15% tails of the negative extreme event for the 2010 La Niña is 33% by the 12-member ensemble size, while its probability is enhanced to 52% by the 108-member ensemble (supplemental Fig. 1b).

b. IOD

The IOD is an air–sea coupled climate phenomenon in the tropical Indian Ocean, which is characterized by a dipole structure along the equator in both oceanic and atmospheric anomalies (Vinayachandran et al. 1999; Saji et al. 1999). The IOD is now known to have serious impacts not only on Indian Ocean rim countries such as Australia, India, and East African countries but also on the Mediterranean Sea rim countries and East Asia including Japan (Behera et al. 1999; Ashok et al. 2001, 2003; Saji and Yamagata 2003; Guan and Yamagata 2003; Behera and Yamagata 2003; Lu et al. 2017). Therefore, prediction of the IOD is crucial for societal applications in agriculture, fisheries, marine ecosystems, human health, natural disasters, etc. (Yuan and Yamagata 2015; Akihiko et al. 2014; Hashizume et al. 2009). Very recently, Doi et al. (2017) have shown that the six ensemble of the SINTEX-F2–3DVAR system, which we used here, is skillful for the IOD prediction.

An index of the Indian Ocean dipole mode (DMI) is first introduced by (Saji et al. 1999) as the SST anomaly difference between the western pole off East Africa (10°S – 10°N , 50° – 70°E) and the eastern pole off Sumatra (10°S – 0° , 90° – 110°E). Here, we focus on the September–October average because most IOD events peak during that period. As shown in the time series (Fig. 4), the IOD prediction is noisier relative to the ENSO prediction. Also, we can see that the 108-member ensemble mean is slightly different from the 12-member ensemble mean for some events: for example, 1983 and 1985. Although comparison of the ACC of the ensemble mean did not show any significant improvement with increasing ensemble members from 12 to 108, the nRMSE slightly reduced by about 6% (Table 2). Hence, the IOD prediction is slightly improved by the 108-member ensemble mean relative to the 12-member ensemble.

Next, we focus on the prediction plumes of the top three strongest events from the 1983–2015: 1994, 1997, and 2006 positive events (Fig. 5). The 1994 event is the

FIG. 3. (a) The SEDI skill scores for the extreme positive 15% tails of Niño-3.4 averaged in November–December by use of different probability thresholds to trigger advisory actions for rare binary events (red: 108-member ensemble; blue: 12-member ensemble). The error bar shows the standard errors of the SEDI by applying the delta method. (b) As in (a), but for the negative event. (c), (d) As in (a) and (b), but for the DMI averaged in September–October.

strongest event in the analysis period as a pure positive IOD that is not associated with a conventional El Niño (Saji et al. 1999). Since a few members of the 12-member ensemble as well as the 108-member ensemble agree with the observation for the maximum peak of October 1994, both ensemble systems successfully captured the observed tails of the PDFs. On the other hand, both failed to capture the extremely strong event in October–November 1997. There is no ensemble member from the 12-member ensemble or the 108-member ensemble beyond 2.0°C . Probabilistic prediction of the DMI beyond 1.0°C in the September–October average of 1997 is 17% in the 12-member ensemble, while that is 28% in the 108-member ensemble. Therefore, we can say that the

probability of the event beyond 1.0°C is enhanced in the 108-member ensemble. For the 2006 positive event, the 108-member ensemble successfully captured the observation as the tail of the PDF, while the 12-member ensemble failed. One member of the 108-member ensemble almost perfectly agrees with the observation. It indicates that we can recognize a chance of the occurrence of the extremely strong event with increasing ensemble members from 12 to 108.

As shown in Fig. 3c, the SEDI skill scores for the extreme 15% positive tail of the DMI averaged in September–October have significantly improved by the 108-member ensemble relative to the 12-member ensemble, when the threshold of probability is 25% and

FIG. 4. As in Fig. 1, but for the DMI averaged in September–October.

35%. The difference between the 12- and the 108-member ensembles is larger at the threshold of 35%. For this improvement, the strongest positive event in 1997 is one of the important factors. The probability prediction of 15% tails of a positive extreme event for 1997 is 25% by the 12-member ensemble, while its probability is enhanced to 37% by the 108-member ensemble (Fig. 5b). Also, the false alarms for 1986 and 1995 are improved by the 108-member ensemble. Those years were actually normal years for the IOD; however, the 12-member ensemble system predicted 42% chance of 15% tails to be positive events for those years, while that chance has reduced in the case of the 108-member ensemble size (supplemental Figs. 1c,d). Unfortunately, the SINTEX-F system is not skillful for prediction of the extreme negative IOD event (Fig. 3d). This may be partly due to the amplitude asymmetry of SST anomalies between positive and negative IOD events (Hong et al. 2008a,b).

c. Global temperature and precipitation

In the previous subsection, we showed that prediction for extremely strong ENSO and positive IOD events based on the 108-member ensemble system is more skillful relative to the 12-member ensemble system for a threshold of 35% probability in SEDI skill scores with the extreme 15% tails. Here, we explore how the prediction

TABLE 2. As in Table 1, but for DMI in September–October issued in early June.

	For skillful prediction, the value should be	12-member ensemble	106-member ensemble
ACC	>0.53 (persistent)	0.54	0.58
nRMSE	<1	0.94	0.88

skill of global temperature and precipitation is improved. The SEDI skill scores of the extreme 15% in the positive tail part of 2-m air temperature anomalies over land and SST anomalies over ocean have significantly improved over some parts of Indochina, East Africa, north-central tropical Indian Ocean, off the west coast of Australia, eastern Australia, central southern Pacific Ocean, and central tropical Atlantic (Fig. 6). The improvement of 2-m air temperature anomalies over land is also seen more clearly when the decision threshold sets 25% (supplemental Fig. 2). We looked into a part of Indochina, which is strongly influenced by ENSO and IOD events. The improvement of the SEDI skill score there is mainly due to the 1997 and 2002 warm events. Supplemental Fig. 3a shows the histogram of 2-m air temperature anomalies averaged over the domain (0° – 20° N, 90° – 110° E) during November–December 1997 by the 12- and the 108-member ensemble predictions initiated in early June. The domain average value is $+0.55^{\circ}\text{C}$ in the NCEP–NCAR reanalysis data. The probability prediction of the value to be above $+0.4^{\circ}\text{C}$ is 33% in the 12-member ensemble, while its probability is enhanced to 54% in the 108-member ensemble. For 2002, when the domain average value in the reanalysis data is $+0.53^{\circ}\text{C}$, the probability of the event to be on the higher side of $+0.4^{\circ}\text{C}$ is also enhanced by the 108-member ensemble prediction (supplemental Fig. 3b).

Although the improvement in precipitation is limited relative to that of temperature, we find some improvements in the extreme dry events over western tropical Pacific, the Maritime Continent, and northern Brazil (Fig. 7). When the decision threshold sets 25%, the improvement is also shown in eastern China and Mexico (supplemental Fig. 4). We looked more at northern Brazil, which is strongly influenced by ENSO events. Supplemental Fig. 3c shows the histogram of precipitation anomalies averaged over the domain 5°S – 5°N , 65° – 50°W during November–December 2010 for the 12- and the 108-member ensemble predictions initiated in early June. The CMAP observational data show that the domain average was $+1.7\text{ mm day}^{-1}$. The probability prediction of the value to be above $+1.5\text{ mm day}^{-1}$ is 50% in the 12-member ensemble, while its probability is enhanced to 65% in the 108-member ensemble.

FIG. 5. As in Fig. 2, but for the DMI for (a) 1994, (b) 1997, and (c) 2006 positive events.

The improvement of regional extreme climate may be partly due to the improvement of prediction of extreme ENSO and IOD events as shown in the previous section. Particularly, the improvement around the west coast of Australia may be related to the regional air–sea coupled climate phenomenon called the Ningaloo Niño (Feng et al. 2013; Doi et al. 2013, 2015).

As mentioned in the introduction, the major focus of the present study is tropical climate predictions. However, a broader assessment of the model's skills was also done for the whole globe using the horizontal distribution of the ACC and nRMSE skill scores of the ensemble mean. Supplemental Figs. 5 and 6 show the prediction skill scores of global temperature and rainfall averaged in July–August and initiated in early June. Some positive improvements are seen on 2-m air temperature in the extratropics, in particular, in the North Atlantic, relative to the tropics in the 108-member ensemble. This looks similar to the North Atlantic horseshoe pattern associated with the North Atlantic Oscillation (NAO). This may support an improvement in the skill score of ensemble-mean prediction of the NAO with higher ensemble size as shown in Scaife et al. (2014). We have also checked the prediction skill of the NAO index. Though there was a marginal increase in the skill score by the 108-member ensemble, it was not skillful even after increasing the ensemble size (figure not shown). We may need a well-resolved stratosphere (high top) and/or high-resolution model, and fine-initialization schemes (particularly an atmospheric initialization) like the Met Office Global Seasonal Forecast System, version 5 (GloSea5), system (Scaife et al. 2014) to be able to produce skillful NAO prediction. For rainfall prediction, the nRMSE is significantly reduced over many parts of the globe by increasing the ensemble size. However, those positive impacts are not enough to provide any real skillful predictions there.

4. Discussion

We have tested a large-ensemble (of about 100 members) retrospective seasonal forecast for 1983–2015. It has turned out that about 10-member ensemble could not perform as well as about 100-member ensemble when predicting the occurrence of 15% tails extreme climate events of ENSO and the IOD. Also, the predictions of extreme temperature and precipitation events are improved over some regions. Those improvements may be because large-ensemble members capture tails of the PDFs and effectively retain the probability of the extremely strong but rare events.

The presented results may depend on the model and the manner of generating ensemble members but will

SEDI skill score (decision threshold: 35%)
for the extreme 15% tails of positive events in ND_2mTemp. & SST anom.

FIG. 6. (a) A horizontal map of SEDI skill scores for the extreme positive 15% tails of 2-m air temperature anomalies over land and sea surface temperature anomalies over ocean averaged in November–December during 1983–2015 between the NCEP–NCAR reanalysis data and the 12-member ensemble prediction from the 1 Jun initialization. The probability threshold is 35%. (b) As in (a), but for the 108-member ensemble. (c) The difference, defined by (b) minus (a). Only values that are significant beyond the standard errors are shown. Region where the SEDI of the 108-member ensemble is less than 0.3 is also masked out.

have implications for the optimal design of seasonal forecasting systems. While the computational cost of operationally running a dynamical prediction system of 100-ensemble members is very high, there might be benefits if the prediction system is useful for skillful prediction of extreme but rare events that have big societal impacts. In this regard, we have evaluated the potential economic value as a function of user cost-loss

ratio, based on a deterministic binary forecast of the extreme 15% tails of the Niño-3.4 index averaged during November–December (Fig. 8a).

A similar analysis is done for DMI averaged during September–October for a 35% probability threshold to trigger an advisory action (Fig. 8b). It is noted that some users with cost-loss ratios less than 0.8 will find benefits for extreme El Niño predictions. The values are higher

SEDI skill score (decision threshold: 35%)
for the extreme 15% tails of negative events in ND_precip. anom.

FIG. 7. As in Fig. 6, but for precipitation.

for the 108-member ensemble system relative to the 12-member ensemble system for the considered range of cost-loss ratio. Particularly, users with cost-loss ratios around 0.1 will find much benefit from the forecasts of the rare events based on the 108-member ensemble system: for example, users who incur large losses when the rare and extreme El Niño events occur and no protection has been taken in anticipation. For extreme La Niña cases, we can also find the similar improvement by the larger ensemble size as seen in extreme El Niño cases, although the values are reduced relative to the positive event.

We also found that the 108-member ensemble prediction of extreme positive IOD events is useful for users with cost-loss ratios range of 0.2–0.3. On the other hand, the 12-member ensemble prediction system has no clear benefit, as it is almost the same as the value from a forecast based on the simple climatology. However, prediction of the negative IOD is not beneficial from both the 12- and the 108-member ensemble systems. Those features of the potential economic values are consistent with the skill verification of the SEDI. Nevertheless, the appropriate action that will minimize the expected losses, based on the cost-benefit analyses, will

FIG. 8. (a) Potential economic value as a function of user cost–loss ratio, of deterministic binary forecast of the extreme 15% tails of Niño-3.4 averaged in November–December by use of 35% probability thresholds to trigger advisory actions (red: positive; blue: negative; thick: 108-member ensemble; thin: 12-member ensemble). (b) As in (a), but for the DMI averaged in September–October.

strongly depend on which societal applications we choose to apply the predictions: for example, agriculture, infectious diseases, and water resources (e.g., Yuan and Yamagata 2015; Ikeda et al. 2017; Oettli et al. 2018). Therefore, further analyses and transdisciplinary research with experts from those areas of societal applications are necessary.

In the case of a smaller ensemble, it might be possible that a probability prediction generated by fitting a distribution to the ensemble is more successful than a simple count. We calculated the SEDI based on a probability forecast generated by fitting a distribution to the ensemble by applying polynomial approximation of degree 3 by least squares (supplemental Fig. 8). It is found that the SEDI variations by the 12-member ensemble are smoothed by the fitting, and the improvements in the SEDI by increasing the ensemble size for extreme La Niña and extreme positive IOD events are clearer relative to the simple counting. However, the main conclusion about the merits of the 100-member ensemble does not change. More complicated methods, for example kernel density estimation, may be more useful in this regard and would be less expensive for the ensemble generation. This needs further exploration.

The predictability dependence on the ensemble size is further studied by making different prediction groups in the range of 12–108 members. We have randomly chosen ensemble members from the 108-member ensemble and created 1000 subsets each for 12, 24, 48, and 84 ensemble members. Figure 9 shows the ACC and the nRMSE skill scores of the ensemble-mean prediction as a function of ensemble size for the Niño-3.4 and the DMI. We cannot find any significant differences in the values averaged among 12-, 24-, 48-, and 84-member ensemble subsets. Nevertheless, increasing the ensemble size has reduced

the ranges between the highest and the lowest skill scores of the ENSO prediction. It is noted that a 48-member ensemble (i.e., based on the uncertainties among the 1000 subsets of randomly chosen ensemble members) may be adequate to reduce the highest–lowest range in our prediction system. Those features are also found to be similar for the DMI prediction.

Figure 10 shows the SEDI skill scores as a function of ensemble size for the extreme positive 15% tails of the Niño-3.4 averaged in November–December and the DMI averaged in September–October. We cannot find any significant differences in the values averaged among 12-, 24-, 48-, and 84-member ensemble subsets. However, the ranges between the highest and the lowest skill scores of the extreme ENSO prediction are reduced by increasing the ensemble size, as similar to Fig. 9. For the positive ENSO event, the 48-member ensemble may be enough to reduce the SEDI range (i.e., uncertainty among the 1000 subsets of randomly chosen ensemble members). On the other hand, the range of the 48-member ensemble for the negative event is still reduced by increasing the ensemble size to 84 members. For the DMI prediction, the range between the highest and the lowest skill scores is almost the same among different ensemble sizes. Figure 11 shows the histogram of the SEDI skill scores for the 15% tail positive DMI based on the 1000 subsets for randomly chosen 12-, 24-, 48-, and 84-member ensembles. There are three peaks for the high concentration of scores for a randomly chosen 12-member ensemble. The score of the original 12-member ensemble falls in one of them: the 0.1–0.2 range. The probability density of the SEDI in the range of 0.1–0.2 is reduced by increasing ensemble size from 12 to 48 members. Although the range between the highest

ACC and nRMSE with ensemble size

FIG. 9. (a) The ACC skill scores as a function of ensemble size for Niño-3.4 averaged in November–December predicted from early June. The value shows the 1000-subset average for randomly chosen 12-, 24-, 48-, and 84-member ensembles. For comparison, the red line shows the value of the 108-member ensemble, the blue line shows that of the 12-member ensemble, and the gray line shows the values of the persistent prediction. The error bar shows the maximum value and the minimum value among the 1000 subsets. (b) As in (a), but for the nRMSE. (c),(d) As in (a) and (b), but for the DMI averaged in September–October.

and the lowest skill scores is almost the same among different ensemble sizes, the probability that the scores fall in the highest concentration range of 0.4–0.5 is enhanced by increasing ensemble size. The probability density of the 48-member ensemble is similar to that of the 84-member ensemble. Those results may partly suggest an ensemble size of about 50 members may be adequate for the El Niño and positive IOD predictions at least in the present prediction system.

In this study, we have not addressed the issue of differences that arise owing to different ensemble generation methods. Supplemental Fig. 7 shows comparisons of the ensemble generation methods in terms of RMSE and ensemble standard deviation (or ensemble spread)

for Niño-3.4 and DMI predictions. For the ENSO prediction, the differences among the ensemble members are smaller than the error in the forecast. It implies that these sources of error are underestimated, and the forecast is overconfident, which is a common deficiency for many climate prediction systems in the world (Tompkins et al. 2017). The differences in the RMSE and the spread among the ensemble generation methods are less than 10% of the values and are not statistically significant. However, it is interesting to know that there are some differences for two different modeling ways for the ocean vertical mixing induced by small vertical-scale structures in November–December forecasts initialized in early June. In addition, we can see some differences in RMSE

SEDI for the extrem 15% tails (Decision threshold of probability:35%)

FIG. 10. (a) The SEDI skill scores as a function of ensemble size for the extreme positive 15% tails of Niño-3.4 averaged in November–December by use of 35% probability thresholds to trigger advisory actions for rare binary events. The value shows the 1000-subset average for randomly chosen 12-, 24-, 48-, and 84-member ensembles. For comparison, the red line shows the value of the 108-member ensemble, and the blue line shows that of the 12-member ensemble. The error bar shows the maximum value and the minimum value among the 1000 subsets. (b) As in (a), but for the negative event. (c),(d) As in (a) and (b), but for the DMI averaged in September–October.

for two kinds of observational SST datasets used for the initialization. Therefore, the choice of the ocean vertical-mixing parameterization and the observational SST datasets for the initialization may be key factors for the ensemble generation methods relative to the simple lagged-average method. For DMI prediction, the RMSE and the ensemble spread are comparable, and the prediction is not overconfident. The differences in the RMSE and the spread among the methods are still only about 10% of the values and are not statistically significant, although the differences among the methods are larger relative to the ENSO prediction. The differences in the RMSE are larger by the lagged-average method

and choice of the nudging strength and the SST datasets for the initialization relative to choice of the ocean vertical-mixing parameterization. At this stage, we do not know which combinations of ensemble generation methods are the best for the ENSO and the IOD predictions. We need further model experiments to evaluate different combinations.

The results in this paper do not exclude further scopes to improve the predictability of extreme events through other methods of initializations of the ocean and atmosphere. Further extensive modeling and careful analyses are certainly necessary to reach a definite conclusion on the ensemble size of a prediction system.

FIG. 11. (a) Histogram of the SEDI skill scores for the extreme positive 15% tails of DMI averaged in September–October by use of 35% probability thresholds to trigger advisory actions for rare binary events based on the 1000 subsets for randomly chosen 12-, 24-, 48-, and 84-member ensembles. For comparison, the red shade shows the range that covers the value of the 108-member ensemble, and the blue line shows that of the 12-member ensemble.

Acknowledgments. The SINTEX-F seasonal climate prediction system was run by the Earth Simulator at JAMSTEC (see <http://www.jamstec.go.jp/es/en/index.html> for the system overview). We are grateful to Drs. Wataru Sasaki, Jing-Jia Luo, Sebastian Masson, and Andrea Storto and our European colleagues of INGV/CMCC, L'OCEAN, and MPI for their contribution to developing the prototype of the systems. We also thank Dr. Masami Nonaka and Antonio Navarra for helpful comments and suggestions. This research was supported by the Environment Research and Technology Development Fund (2–1405) of the Ministry of the Environment, Japan, the Japan Agency for Medical Research and Development (AMED) and Japan International Cooperation Agency (JICA) through the Science and Technology Research Partnership for Sustainable Development (SATREPS) project for iDEWS South Africa, and JSPS KAKENHI Grants 16H04047 and 16K17810. The GrADS software was used for creating the figures and the maps.

REFERENCES

- Akihiko, T., Y. Morioka, and S. K. Behera, 2014: Role of climate variability in the heatstroke death rates of Kanto region in Japan. *Sci. Rep.*, **4**, 5655, <https://doi.org/10.1038/srep05655>.
- Ashok, K., Z. Guan, and T. Yamagata, 2001: Impact of the Indian Ocean dipole on the relationship between the Indian monsoon rainfall and ENSO. *Geophys. Res. Lett.*, **28**, 4499–4502, <https://doi.org/10.1029/2001GL013294>.
- , —, and —, 2003: A look at the relationship between the ENSO and the Indian Ocean dipole. *J. Meteor. Soc. Japan*, **81**, 41–56, <https://doi.org/10.2151/jmsj.81.41>.
- Barnston, A. G., and S. J. Mason, 2011: Evaluation of IRI's seasonal climate forecasts for the extreme 15% tails. *Wea. Forecasting*, **26**, 545–554, <https://doi.org/10.1175/WAF-D-10-05009.1>.
- Behera, S. K., and T. Yamagata, 2003: Influence of the Indian Ocean dipole on the Southern Oscillation. *J. Meteor. Soc. Japan*, **81**, 169–177, <https://doi.org/10.2151/jmsj.81.169>.
- , R. Krishnan, and T. Yamagata, 1999: Unusual ocean-atmosphere conditions in the tropical Indian Ocean during 1994. *Geophys. Res. Lett.*, **26**, 3001–3004, <https://doi.org/10.1029/1999GL010434>.
- Chen, M., W. Wang, and A. Kumar, 2013: Lagged ensembles, forecast configuration, and seasonal predictions. *Mon. Wea. Rev.*, **141**, 3477–3497, <https://doi.org/10.1175/MWR-D-12-00184.1>.
- Crtat, J., P. Terray, S. Masson, K. P. Sooraj, and M. K. Roxy, 2016: Indian Ocean and Indian summer monsoon: Relationships without ENSO in ocean–atmosphere coupled simulations. *Climate Dyn.*, **49**, 1429–1448, <https://doi.org/10.1007/s00382-016-3387-x>.
- , —, —, and —, 2017: Intrinsic precursors and timescale of the tropical Indian Ocean Dipole: Insights from partially decoupled numerical experiment. *Climate Dyn.*, **51**, 1311–1332, <https://doi.org/10.1007/s00382-017-3956-7>.
- Dalcher, A., E. Kalnay, and R. N. Hoffman, 1988: Medium range lagged average forecasts. *Mon. Wea. Rev.*, **116**, 402–416, [https://doi.org/10.1175/1520-0493\(1988\)116<0402:MRLAF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116<0402:MRLAF>2.0.CO;2).
- Dqu, M., 1997: Ensemble size for numerical seasonal forecasts. *Tellus*, **49A**, 74–86, <https://doi.org/10.1034/j.1600-0870.1997.00005.x>.
- Doi, T., S. K. Behera, and T. Yamagata, 2013: Predictability of the Ningaloo Nio/Nia. *Sci. Rep.*, **3**, 2892, <https://doi.org/10.1038/srep02892>.
- , —, and —, 2015: An interdecadal regime shift in rainfall predictability related to the Ningaloo Nio in the late 1990s. *J. Geophys. Res. Oceans*, **120**, 1388–1396, <https://doi.org/10.1002/2014JC010562>.
- , —, and —, 2016: Improved seasonal prediction using the SINTEX-F2 coupled model. *J. Adv. Model. Earth Syst.*, **8**, 1847–1867, <https://doi.org/10.1002/2016MS000744>.
- , A. Storto, S. K. Behera, A. Navarra, and T. Yamagata, 2017: Improved prediction of the Indian Ocean dipole mode by use of subsurface ocean observations. *J. Climate*, **30**, 7953–7970, <https://doi.org/10.1175/JCLI-D-16-0915.1>.
- Eade, R., D. M. Smith, A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson, 2014: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.*, **41**, 5620–5628, <https://doi.org/10.1002/2014GL061146>.
- Feng, M., M. J. McPhaden, S. Xie, and J. Hafner, 2013: La Nia forces unprecedented Leeuwin Current warming in 2011. *Nature*, **3**, 1277, <https://doi.org/10.1038/srep01277>.
- Ferro, C. A., and D. B. Stephenson, 2011: Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events. *Wea. Forecasting*, **26**, 699–713, <https://doi.org/10.1175/WAF-D-10-05030.1>.
- Fichefet, T., and M. A. Morales Maqueda, 1997: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *J. Geophys. Res.*, **102**, 12 609–12 646, <https://doi.org/10.1029/97JC00480>.
- Good, S. A., M. J. Martin, and N. A. Rayner, 2013: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Oceans*, **118**, 6704–6716, <https://doi.org/10.1002/2013JC009067>.
- Guan, Z., and T. Yamagata, 2003: The unusual summer of 1994 in East Asia: IOD teleconnections. *Geophys. Res. Lett.*, **30**, 1544, <https://doi.org/10.1029/2002GL016831>.

- Hashizume, M., T. Terao, and N. Minakawa, 2009: The Indian Ocean dipole and malaria risk in the highlands of western Kenya. *Proc. Natl. Acad. Sci. USA*, **106**, 1857–1862, <https://doi.org/10.1073/pnas.0806544106>.
- Hoerling, M. P., and A. Kumar, 2002: Atmospheric response patterns associated with tropical forcing. *J. Climate*, **15**, 2184–2203, [https://doi.org/10.1175/1520-0442\(2002\)015<2184:ARPAWT.2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2184:ARPAWT.2.0.CO;2).
- Hong, C.-C., T. Li, LinHo, and J.-S. Kug, 2008a: Asymmetry of the Indian Ocean dipole. Part I: Observational analysis. *J. Climate*, **21**, 4834–4848, <https://doi.org/10.1175/2008JCLI2222.1>.
- , —, and J.-J. Luo, 2008b: Asymmetry of the Indian Ocean dipole. Part II: Model diagnosis. *J. Climate*, **21**, 4849–4858, <https://doi.org/10.1175/2008JCLI2223.1>.
- Horii, T., H. Hase, I. Ueki, and Y. Masumoto, 2008: Oceanic precondition and evolution of the 2006 Indian Ocean dipole. *Geophys. Res. Lett.*, **35**, L03607, <https://doi.org/10.1029/2007GL032464>.
- Hudson, D., A. G. Marshall, Y. Yin, O. Alves, and H. H. Hendon, 2013: Improving intraseasonal prediction with a new ensemble generation strategy. *Mon. Wea. Rev.*, **141**, 4429–4449, <https://doi.org/10.1175/MWR-D-13-00059.1>.
- Ikeda, T., S. Behera, Y. Morioka, N. Minakawa, M. Hashizume, A. Tsuzuki, R. Maharaj, and P. Kruger, 2017: Seasonally lagged effects of climatic factors on malaria incidence in South Africa. *Sci. Rep.*, **7**, 2458, <https://doi.org/10.1038/s41598-017-02680-6>.
- Joseph, S., A. K. Sahai, B. N. Goswami, P. Terray, S. Masson, and J. J. Luo, 2012: Possible role of warm SST bias in the simulation of boreal summer monsoon in SINTEX-F2 coupled model. *Climate Dyn.*, **38**, 1561–1576, <https://doi.org/10.1007/s00382-011-1264-1>.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kang, S., and J. H. Yoo, 2006: Examination of multi-model ensemble seasonal prediction methods using a simple climate system. *Climate Dyn.*, **26**, 285–294, <https://doi.org/10.1007/s00382-005-0074-8>.
- Kirtman, B. P., and Coauthors, 2014: The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Kosaka, Y., S.-P. Xie, N.-C. Lau, and G. A. Vecchi, 2013: Origin of seasonal predictability for summer climate over the northwestern Pacific. *Proc. Natl. Acad. Sci. USA*, **110**, 7574–7579, <https://doi.org/10.1073/pnas.1215582110>.
- Kumar, A., 2009: Finite samples and uncertainty estimates for skill measures for seasonal prediction. *Mon. Wea. Rev.*, **137**, 2622–2631, <https://doi.org/10.1175/2009MWR2814.1>.
- , A. G. Barnston, and M. P. Hoerling, 2001: Seasonal predictions, probabilistic verifications, and ensemble size. *J. Climate*, **14**, 1671–1676, [https://doi.org/10.1175/1520-0442\(2001\)014<1671:SPPPVAE>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<1671:SPPPVAE>2.0.CO;2).
- Latif, M., and Coauthors, 1998: A review of the predictability and prediction of ENSO. *J. Geophys. Res.*, **103**, 14 375–14 393, <https://doi.org/10.1029/97JC03413>.
- Lopez, H., and B. P. Kirtman, 2014: WWBs, ENSO predictability, the spring barrier and extreme events. *J. Geophys. Res. Atmos.*, **119**, 10 114–10 138, <https://doi.org/10.1002/2014JD021908>.
- Lu, B., and Coauthors, 2017: An extreme negative Indian Ocean dipole event in 2016: Dynamics and predictability. *Climate Dyn.*, **51**, 89–100, <https://doi.org/10.1007/s00382-017-3908-2>.
- Luo, J. J., S. Masson, S. Behera, S. Shingu, and T. Yamagata, 2005: Seasonal climate predictability in a coupled OAGCM using a different approach for ensemble forecasts. *J. Climate*, **18**, 4474–4497, <https://doi.org/10.1175/JCLI3526.1>.
- , —, —, and T. Yamagata, 2007: Experimental forecasts of the Indian Ocean dipole using a coupled OAGCM. *J. Climate*, **20**, 2178–2190, <https://doi.org/10.1175/JCLI4132.1>.
- Maclachlan, C., and Coauthors, 2015: Global Seasonal Forecast System version 5 (GloSea5): A high-resolution seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 1072–1084, <https://doi.org/10.1002/qj.2396>.
- Madec, G., 2008: NEMO ocean engine, version 3.0. Institut Pierre-Simon Laplace Note du Pole de modélisation 27, 209 pp.
- Mason, S. J., and N. E. Graham, 2002: Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quart. J. Roy. Meteor. Soc.*, **128**, 2145–2166, <https://doi.org/10.1256/003590002320603584>.
- Masson, S., P. Terray, G. Madec, J. J. Luo, T. Yamagata, and K. Takahashi, 2012: Impact of intra-daily SST variability on ENSO characteristics in a coupled model. *Climate Dyn.*, **39**, 681–707, <https://doi.org/10.1007/s00382-011-1247-2>.
- Molteni, F., and Coauthors, 2011: The new ECMWF seasonal forecast system (system 4). ECMWF Tech. Memo. 656, 51 pp.
- Morioka, Y., J. V. Ratnam, W. Sasaki, and Y. Masumoto, 2013: Generation mechanism of the South Pacific subtropical dipole. *J. Climate*, **26**, 6033–6045, <https://doi.org/10.1175/JCLI-D-12-00648.1>.
- , S. Masson, P. Terray, C. Prodhomme, S. K. Behera, and Y. Masumoto, 2014: Role of tropical SST variability on the formation of subtropical dipoles. *J. Climate*, **27**, 4486–4507, <https://doi.org/10.1175/JCLI-D-13-00506.1>.
- , F. Engelbrecht, and S. K. Behera, 2015: Potential sources of decadal climate variability over southern Africa. *J. Climate*, **28**, 8695–8709, <https://doi.org/10.1175/JCLI-D-15-0201.1>.
- , B. Taguchi, and S. K. Behera, 2017: Eastward propagating decadal temperature variability in the South Atlantic and Indian Oceans. *J. Geophys. Res. Oceans*, **122**, 5611–5623, <https://doi.org/10.1002/2017JC012706>.
- , T. Doi, and S. K. Behera, 2018a: Decadal climate predictability in the southern Indian Ocean captured by SINTEX-F using a simple SST-nudging scheme. *Sci. Rep.*, **8**, 1029, <https://doi.org/10.1038/s41598-018-19349-3>.
- , —, A. Storto, S. Masina, and S. K. Behera, 2018b: Role of subsurface ocean in decadal climate predictability over the South Atlantic. *Sci. Rep.*, **8**, 8523, <https://doi.org/10.1038/s41598-018-26899-z>.
- Murphy, J. M., 1990: Assessment of the practical utility of extended range ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **116**, 89–125, <https://doi.org/10.1002/qj.49711649105>.
- Oettli, P., S. K. Behera, and T. Yamagata, 2018: Climate based predictability of oil palm tree yield in Malaysia. *Sci. Rep.*, **8**, 2271, <https://doi.org/10.1038/s41598-018-20298-0>.
- Ogata, T., T. Doi, Y. Morioka, and S. Behera, 2018: Mid-latitude source of the ENSO-spread in SINTEX-F ensemble predictions. *Climate Dyn.*, <https://doi.org/10.1007/s00382-018-4280-6>, in press.
- Philander, S., 1989: *El Niño, La Niña, and the Southern Oscillation*. S. G. Philander, Ed., Academic Press, 293 pp.
- Prodhomme, C., P. Terray, S. Masson, T. Izumo, T. Tozuka, and T. Yamagata, 2014: Impacts of Indian Ocean SST biases on the Indian monsoon: As simulated in a global coupled model. *Climate Dyn.*, **42**, 271–290, <https://doi.org/10.1007/s00382-013-1671-6>.
- , —, —, G. Bosch, and T. Izumo, 2015: Oceanic factors controlling the Indian summer monsoon onset in a coupled model. *Climate Dyn.*, **44**, 977–1002, <https://doi.org/10.1007/s00382-014-2200-y>.
- , L. Batté, F. Massonnet, P. Davini, O. Bellprat, V. Guemas, and F. J. Doblas-Reyes, 2016: Benefits of increasing the model

- resolution for the seasonal forecast quality in EC-Earth. *J. Climate*, **29**, 9141–9162, <https://doi.org/10.1175/JCLI-D-16-0117.1>.
- Ratnam, J. V., T. Doi, and S. K. Behera, 2017: Dynamical downscaling of SINTEX-F2v CGCM seasonal retrospective austral summer forecasts over Australia. *J. Climate*, **30**, 3219–3235, <https://doi.org/10.1175/JCLI-D-16-0585.1>.
- Ren, H.-L., F.-F. Jin, B. Tian, and A. A. Scaife, 2016: Distinct persistence barriers in two types of ENSO. *Geophys. Res. Lett.*, **43**, 10 973–10 979, <https://doi.org/10.1002/2016GL071015>.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625, [https://doi.org/10.1175/1520-0442\(2002\)015<1609:AHSAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AHSAS>2.0.CO;2).
- , T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution-blended analyses for sea surface temperature. *J. Climate*, **20**, 5473–5496, <https://doi.org/10.1175/2007JCLI1824.1>.
- Richardson, D. S., 2003: Economic value and skill. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., Wiley, 165–187.
- Roeckner, E., and Coauthors, 2003: The atmospheric general circulation model ECHAM5. Part I: Model description. Max-Planck-Institut für Meteorologie Rep. 349, 140 pp., https://www.mpimet.mpg.de/fileadmin/publikationen/Reports/max_scierep_349.pdf.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Saji, N. H., and T. Yamagata, 2003: Possible impacts of Indian Ocean dipole mode events on global climate. *Climate Res.*, **25**, 151–169, <https://doi.org/10.3354/cr025151>.
- , B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature*, **401**, 360–363, <https://doi.org/10.1038/43854>.
- Sanna, A., A. Borrelli, S. Matera, P. Athanasiadis, A. Bellucci, P. G. Fogli, E. Scoccimarro, and S. Gualdi, 2015: The new CMCC–Seasonal Prediction System. CMCC Rep. RP0253, 13 pp., <https://www.cmcc.it/publications/rp0253-the-new-cmcc-seasonal-prediction-system>.
- Sasaki, W., K. J. Richards, and J. J. Luo, 2012: Role of vertical mixing originating from small vertical scale structures above and within the equatorial thermocline in an OGCM. *Ocean Modell.*, **57–58**, 29–42, <https://doi.org/10.1016/j.ocemod.2012.09.002>.
- , —, and —, 2013: Impact of vertical mixing induced by small vertical scale structures above and within the equatorial thermocline on the tropical Pacific in a CGCM. *Climate Dyn.*, **41**, 443–453, <https://doi.org/10.1007/s00382-012-1593-8>.
- , T. Doi, K. J. Richards, and Y. Masumoto, 2014: Impact of the equatorial Atlantic sea surface temperature on the tropical Pacific in a CGCM. *Climate Dyn.*, **43**, 2539–2552, <https://doi.org/10.1007/s00382-014-2072-1>.
- , —, —, and —, 2015: The influence of ENSO on the equatorial Atlantic precipitation through the Walker circulation in a CGCM. *Climate Dyn.*, **44**, 191–202, <https://doi.org/10.1007/s00382-014-2133-5>.
- Scaife, A. A., and Coauthors, 2014: Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.*, **41**, 2514–2519, <https://doi.org/10.1002/2014GL059637>.
- Shukla, J., and Coauthors, 2000: Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.*, **81**, 2593–2606, [https://doi.org/10.1175/1520-0477\(2000\)081<2593:DSP>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2593:DSP>2.3.CO;2).
- Storto, A., S. Dobricic, S. Masina, and P. Di Pietro, 2011: Assimilating along-track altimetric observations through local hydrostatic adjustment in a global ocean variational assimilation system. *Mon. Wea. Rev.*, **139**, 738–754, <https://doi.org/10.1175/2010MWR3350.1>.
- , S. Masina, and S. Dobricic, 2014: Estimation and impact of nonuniform horizontal correlation length scales for global ocean physical analyses. *J. Atmos. Oceanic Technol.*, **31**, 2330–2349, <https://doi.org/10.1175/JTECH-D-14-00042.1>.
- Takaya, Y., and Coauthors, 2017a: Japan Meteorological Agency/Meteorological Research Institute-Coupled Prediction System version 1 (JMA/MRI-CPS1) for operational seasonal forecasting. *Climate Dyn.*, **48**, 313–333, <https://doi.org/10.1007/s00382-016-3076-9>.
- , and Coauthors, 2017b: Japan Meteorological Agency/Meteorological Research Institute-Coupled Prediction System version 2 (JMA/MRI-CPS2): Atmosphere–land–ocean–sea ice coupled prediction system for operational seasonal forecasting. *Climate Dyn.*, **50**, 751–765, <https://doi.org/10.1007/s00382-017-3638-5>.
- Terray, P., S. Masson, and K. Kakitha, 2011: Role of the frequency of coupling in the simulation of the monsoon-ENSO relationship in a global coupled model. *European Geosciences Union General Assembly*, Vienna, Austria, European Geosciences Union, 13355.
- , —, C. Prodhomme, M. K. Roxy, and K. P. Sooraj, 2016: Impacts of Indian and Atlantic Oceans on ENSO in a comprehensive modeling framework. *Climate Dyn.*, **46**, 2507–2533, <https://doi.org/10.1007/s00382-015-2715-x>.
- , K. P. Sooraj, S. Masson, R. P. M. Krishna, G. Samson, and A. G. Prajeesh, 2017: Towards a realistic simulation of boreal summer tropical rainfall climatology in state-of-the-art coupled models: Role of the background snow-free land albedo. *Climate Dyn.*, **50**, 3413–3439, <https://doi.org/10.1007/s00382-017-3812-9>.
- Tompkins, A. M., and Coauthors, 2017: The Climate-System Historical Forecast Project: Providing open access to seasonal forecast ensembles from centers around the globe. *Bull. Amer. Meteor. Soc.*, **98**, 2293–2301, <https://doi.org/10.1175/BAMS-D-16-0209.1>.
- Valcke, S., A. Caubel, R. Vogelsang, and D. Declat, 2004: OASIS3 Ocean Atmosphere Sea Ice Soil user's guide. CERFACS Tech. Rep. TR/CMGC/04/68, 70 pp.
- Vecchi, G. A., and Coauthors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27**, 7994–8016, <https://doi.org/10.1175/JCLI-D-14-00158.1>.
- Vernieres, G., M. M. Rienecker, R. Kovach, and C. L. Keppenne, 2012: The GEOS-iodAS: Description and evaluation. NASA Tech. Rep. NASA/TM-2012-104606/VOL30, 73 pp.
- Vinayachandran, P. N., N. H. Saji, and T. Yamagata, 1999: Response of the equatorial Indian Ocean to an unusual wind event during 1994. *Geophys. Res. Lett.*, **26**, 1613–1616, <https://doi.org/10.1029/1999GL900179>.
- Wang, B., R. Wu, and T. Li, 2003: Atmosphere–warm ocean interaction and its impacts on Asian–Australian monsoon variation. *J. Climate*, **16**, 1195–1211, [https://doi.org/10.1175/1520-0442\(2003\)16<1195:AOIAII>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)16<1195:AOIAII>2.0.CO;2).
- , and Coauthors, 2009: Advance and prospectus of seasonal prediction: Assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate Dyn.*, **33**, 93–117, <https://doi.org/10.1007/s00382-008-0460-0>.
- Webster, P. J., and C. D. Hoyos, 2010: Beyond the spring barrier? *Nat. Geosci.*, **3**, 152–153, <https://doi.org/10.1038/ngeo800>.
- Xie, P., and P. A. Arkin, 1996: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. Climate*, **9**, 840–858, [https://doi.org/10.1175/1520-0442\(1996\)009<0840:AOGMPU>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<0840:AOGMPU>2.0.CO;2).
- Yuan, C., and T. Yamagata, 2015: Impacts of IOD, ENSO and ENSO Modoki on the Australian winter wheat yields in recent decades. *Sci. Rep.*, **5**, 17252, <https://doi.org/10.1038/srep17252>.