How Predictable is the Indian Ocean Dipole?

LI SHI, HARRY H. HENDON, AND OSCAR ALVES

Centre for Australian Weather and Climate Research,* Melbourne, Australia

JING-JIA LUO†

Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

MAGDALENA BALMASEDA

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

DAVID ANDERSON

Centre for Australian Weather and Climate Research,* Melbourne, Australia

(Manuscript received 15 December 2011, in final form 4 May 2012)

ABSTRACT

In light of the growing recognition of the role of surface temperature variations in the Indian Ocean for driving global climate variability, the predictive skill of the sea surface temperature (SST) anomalies associated with the Indian Ocean dipole (IOD) is assessed using ensemble seasonal forecasts from a selection of contemporary coupled climate models that are routinely used to make seasonal climate predictions. The authors assess predictions from successive versions of the Australian Bureau of Meteorology Predictive Ocean–Atmosphere Model for Australia (POAMA 15b and 24), successive versions of the NCEP Climate Forecast System (CFSv1 and CFSv2), the ECMWF seasonal forecast System 3 (ECSys3), and the Frontier Research Centre for Global Change system (SINTEX-F) using seasonal hindcasts initialized each month from January 1982 to December 2006.

The lead time for skillful prediction of SST in the western Indian Ocean is found to be about 5–6 months while in the eastern Indian Ocean it is only 3–4 months when all start months are considered. For the IOD events, which have maximum amplitude in the September–November (SON) season, skillful prediction is also limited to a lead time of about one season, although skillful prediction of large IOD events can be longer than this, perhaps up to about two seasons. However, the tendency for the models to overpredict the occurrence of large events limits the confidence of the predictions of these large events. Some common model errors, including a poor representation of the relationship between El Niño and the IOD, are identified indicating that the upper limit of predictive skill of the IOD has not been achieved.

1. Introduction

The primary source of seasonal climate predictability is sea surface temperature (SST) variations in the tropical Pacific Ocean associated with El Niño–Southern Oscillation (ENSO), for which multiseason predictive capability with dynamical coupled models is well established (e.g., Cane et al. 1986; Latif et al. 1998). Despite the success of predicting ENSO–SST variations and some of the associated global climate impacts using coupled
climate models (e.g., Stockdale et al. 1998, 2011; Kirtman et al. 2001; Alves et al. 2003; Palmer et al. 2004; Luo et al. 2005; Saha et al. 2006; Jin et al. 2008; Wang et al. 2009), there is a growing appreciation of the role of coupled atmosphere–ocean variability in the other tropical ocean basins for driving predictable climate variability. We focus here on SST variations in the tropical Indian Ocean, which are a primary source of seasonal climate variability throughout the adjoining landmasses of eastern Africa (e.g., Black et al. 2003; Clark et al. 2003), Asia (e.g., Saji and Yamagata 2003), the Maritime Continent (e.g., Hendon 2003), and Australia (e.g., Nicholls 1989; Ansell et al. 2000; Cai et al. 2011).

In particular, we focus on the Indian Ocean dipole (IOD; e.g., Saji et al. 1999; Webster et al. 1999), which is a zonal mode of SST variability in the tropical Indian Ocean. The IOD is tightly tied to the seasonal cycle, with events tending to develop in boreal summer, peak in fall, and decay rapidly in winter (Saji et al. 1999) as well as to ENSO, with peak positive phases (cold eastern Indian Ocean) tending to occur during El Niño (e.g., Shinoda et al. 2004b). The typical structure of the SST anomalies associated with the IOD is shown in Fig. 1a, which is based on a composite of six recent strong positive events (1982, 1991, 1994, 1997, 2003, and 2006) during the boreal fall [September–November (SON)]. Positive IOD events are

Fig. 1. Composite seasonal mean (September–November) SST anomaly for positive IOD events from (a) observed SST, and forecasts from POAMA24 at lead times of (b) 1 month and (c) 3 months. Contour and shading interval is 0.2°C. Boxes indicate area means for WIO, EIO, and Niño-3.4 SST indices.
associated with negative SST anomalies in the eastern tropical Indian Ocean anchored off the Java–Sumatra coast, which act to reduce rainfall there, and positive SST anomalies in the western Indian Ocean, which act to increase rainfall there (e.g., Saji et al. 1999). The association of IOD events with El Niño is evident in this compositive by the appearance of strong positive SST anomalies in the equatorial eastern Pacific.

The variations of tropical rainfall in the Indian Ocean during an IOD event not only affect the immediate local climate in the tropical Indian Ocean (e.g., Black et al. 2003; Saji and Yamagata 2003), but also act to excite Rossby wave trains (e.g., Saji and Yamagata 2003; Cai et al. 2011), giving rise to extratropical climate variability (e.g., Ansell et al. 2000; Ashok et al. 2003; Black et al. 2003; Clark et al. 2003; Ummenhofer et al. 2011). The role of SST variations due to the IOD for driving rainfall variations across southern Australia during ENSO is especially important (e.g., Cai et al. 2011). Hence, predictability of rainfall across southern Australia during ENSO is limited by the ability to predict the IOD. Furthermore, although IOD events tend to co-occur with El Niño, this is not always the case and so it is important to be able to predict variations of the IOD and, more generally, SST throughout the tropical Indian Ocean that are independent of ENSO in order to progress the capability to make seasonal climate forecasts. Forecast skill of the IOD has previously been assessed using a range of coupled climate models and verification approaches (e.g., Wajsowicz 2005, 2007; Luo et al. 2005, 2007; Song et al. 2008; Wang et al. 2009; Zhao and Hendon 2009; Sooraj et al. 2012). In general, forecast skill of SST in the Indian Ocean is found to be much less than in the tropical Pacific, with skillful prediction of SST in the east and west portions of the tropical Indian Ocean basin reported to lead times up to 5–6 and 6–9 months, respectively. The lead time for skillful prediction of the IOD is reported to be only about 3–4 months because of a strong boreal spring “predictability barrier.” On the other hand, some strong individual IOD events (i.e., 2003 and 2006) have been reported to be predictable at long lead times (e.g., Luo et al. 2007, 2008), which may stem from their association with El Niño (e.g., Song et al. 2008; Zhao et al. 2010).

Predictability of the IOD, and especially large IOD events, may be model dependent. As an example, the composite of predictions of the IOD events in Fig. 1a using the current version of the Australian Bureau of Meteorology (BoM) Predictive Ocean–Atmosphere Model for Australia (POAMA; details are supplied in section 2) are displayed in Fig. 1b for a lead time of 1 month (i.e., the SON forecasts were initialized on 1 August) and in Fig. 1c for a lead time of 3 months. Although the amplitude of the predicted SST anomaly at a lead time of 1 month in the Indian Ocean is consistent with observation, the amplitude particularly along the Java–Sumatra coast drops off rapidly at a lead time of 3 months. Furthermore, some systematic model errors are evident, such as the westward shift of the El Niño SST anomaly in the Pacific and a westward extension of the negative SST anomalies into the western Indian Ocean (see also Zhao and Hendon 2009).

Because of the importance of climate variations linked to the IOD, we are motivated to know what our current capability to predict the IOD is and how the predictive capability varies between models. To this end we use a common methodology to assess the current predictive capability of the IOD from a selection of coupled forecast models that are routinely used to make global climate predictions. As part of this assessment, we aim to highlight both common successes and failures, which might then be useful for future development of these forecast systems.

To survey the current capability and to track some of the recent progress for forecast system development, we will evaluate the predictive capability of the IOD in two successive versions of the coupled forecast systems run at the Australian Bureau of Meteorology (POAMA1.5; Alves et al. 2003, and POAMA24; Wang et al. 2011) and at the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP) Climate Forecast System versions 1 and 2 (CFSv1; Saha et al. 2006; and CFSv2; S. Saha et al. 2011, unpublished manuscript). We will also assess the European Centre for Medium-Range Weather Forecasts (ECMWF) Seasonal Forecast System 3 (ECSys3; Stockdale et al. 2011) and the Frontier Research Centre for Global Change system (SINTEX-F) coupled model that is run experimentally by the Japan Agency for Marine Earth Science and Technology (JAMSTEC; e.g., Luo et al. 2005).

Although a thorough investigation of predictive skill of the three-dimensional state of the atmosphere–ocean climate system associated with the IOD is warranted, our approach here, which can be considered as a preliminary investigation, is to assess predictive skill of these systems using just some standard indices of SST. The seasonal forecast systems and hindcast datasets are briefly described in section 2. In section 3, we focus on assessing the general ability and common deficiencies of predicting SST anomalies. We will assess the simulation and prediction of the teleconnection of ENSO to the IOD, as the inability to faithfully represent this teleconnection could limit forecast skill. Some benchmark statistical and statistical-dynamical forecast models are also developed. Conclusions are provided in section 4.
2. Seasonal forecast systems, hindcasts, and verification datasets

Forecast skill of the SST in the eastern and western tropical Indian Ocean and the IOD event is assessed using hindcast predictions of standard indices of the SST anomalies in the western Indian Ocean (WIO; 10°S–10°N, 50°–70°E) and the eastern Indian Ocean (EIO; 10°S–0°, 90°–110°E; see Fig. 1). The Dipole Mode Index (DMI), which is used to represent IOD events, is defined as DMI = WIO – EIO. We also use the Niño-3.4 index (SST anomaly averaged over 5°S–5°N, 170°–120°W) to monitor El Niño variations in the Pacific.

Ensembles of hindcasts for a common period of 1982–2006 are available from two successive versions of the POAMA (P15b and P24), two successive versions of the NCEP CFS (CFSv1 and CFSv2; hindcast data were downloaded from http://cfs.ncep.noaa.gov), the ECSys3, and the SINTEX-F model. Table 1 provides a brief summary of each of these forecast systems, including the model resolutions and initialization procedures. The ECsys3 system became operational at ECMWF in March 2007 but was replaced by Sys4 in November 2011. The P15b became operational at BoM in June 2007 and was replaced by P24 at the end of 2011. CFSv1 became operational at NCEP in August 2004. The CFSv2 became operational in March 2011, although the CFSv1 will continue to run through at least June 2012. The SINTEX-F model has been run routinely (although experimentally) at JAMSTEC since January 2005.

Three important aspects of the choice of models that are used in this study need to be emphasized. First, we use hindcast data from modeling centers that routinely issue climate forecasts, so as to provide a representative assessment of forecast skill that can be achieved with current forecast systems. Second, we include hindcasts from two successive versions of the POAMA and CFS models, in order to track recent progress of these systems (more detail about the major changes in these systems is provided below). And, third, we include hindcasts from SINTEX-F, which does not make use of any atmosphere and subsurface ocean data assimilation but instead initializes the coupled model using a SST-nudging scheme, so as to provide a benchmark for what can be achieved in the absence of atmosphere and subsurface ocean data assimilation.

With regard to the two upgraded systems, all components of the forecast system were upgraded for CFSv2 over CFSv1, including the atmosphere and ocean component models and atmosphere and ocean assimilation systems (Table 1). Two key upgrades for the POAMA system from P15b to P24 are an improved ocean data assimilation system (Yin et al. 2011) and an increased ensemble size (30 vs 10) formed using 10 members each from three versions of the forecast model (P24a, P24b, and P24c; see Table 1 for more details).

The hindcasts from the P15b, P24, CFSv1, CFSv2, and ECsys3 systems are all initialized with observed ocean and atmosphere states but using different assimilation/ensemble perturbation procedures (Table 1). In contrast, SINTEX-F hindcasts are not initialized with any analyzed atmospheric or subsurface ocean data; rather they are initialized by nudging the coupled forecast model to observed SST in the year leading up to the initialization time. This technique does not provide any initial information about the true intraseasonally varying initial state in the atmosphere or ocean, while relying on the model simulation of the response to prescribed SST in order to spin up the slowly varying coupled component of the initial state. This method of initialization might reduce initial model shock, which perhaps overcomes a lack of knowledge of the true initial state and thus provides good forecasts. However, in a study to assess the impact of assimilating ocean and atmospheric data for forecasts of El Niño with the ECsys3 system, Balmaseda and Anderson (2009) demonstrate a distinct forecast advantage for full atmosphere–ocean data assimilation in comparison to the SST nudging technique.

Apart from the CFSv1 and CFSv2, the hindcasts from the other systems are all initialized on the first of each month. Ensemble members are generated for the ECsys3 by perturbing both the ocean and atmosphere initial conditions. P15b uses a single analyzed ocean state but perturbs the atmospheric state by taking successively 6-h earlier states (thus ensemble member 10 uses an atmospheric state that is 2.25 days earlier than ensemble member 1). For P24, perturbed ocean states are provided by the ensemble ocean data assimilation system (Yin et al. 2011), but only a single atmospheric initial state on the first of the month is used. This technique of not perturbing the atmospheric state results in reduced forecast spread, especially atmospheric spread, in the first month of the forecast, but after that internally generated variability should saturate the coupled model spread. The SINTEX-F model effectively perturbs both the atmosphere and ocean states for each of its nine ensemble members by using different model versions and initialization procedures (e.g., Luo et al. 2007, 2008). On the other hand, the CFSv1 generates 15 ensemble members by initializing one member each day for the 5 days around the 11th, 21st and 1st. The CFSv2 generates 24 ensemble members by initializing 4 members every fifth day, thus resulting in a lagged ensemble (with the lag of up to 30 days; e.g., S. Saha et al. 2011, unpublished manuscript).

To make the assessment among all seasonal forecast systems as uniform as possible, we use only the 8 ensemble
### Table 1. Details of the seasonal prediction systems and hindcasts.

<table>
<thead>
<tr>
<th>Atmosphere model</th>
<th>Ocean model</th>
<th>Atmosphere initial scheme</th>
<th>Ocean initial scheme</th>
<th>Ensemble members</th>
<th>Forecast months</th>
<th>Climatology</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECMWF Sys3</td>
<td>Hamburg Ocean Primitive Equation (HOPE)</td>
<td>40-yr ECMWF Re-Analysis (ERA-40) direct (Uppala et al. 2005) and singular vectors perturbation</td>
<td>OI technique of sea temperature, salinity profiles, and after Jan 1993 also altimetry sea level</td>
<td>11</td>
<td>7</td>
<td>1982–2006</td>
</tr>
<tr>
<td>(EC Sys3)</td>
<td>T159L62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stockdale et al. (2011)</td>
<td>Primitive Equation (HOPE) 1.4° x 0.5°–1.4°; L29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINTEX-F</td>
<td>T106L19</td>
<td>2.0° x 0.5°–2.0°; L31</td>
<td>Forced by nudged SST</td>
<td>No ocean assimilation except SST nudged to observations with a time scale of 1 day</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Luo et al. (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCEP CSv1</td>
<td>T62L64</td>
<td>Modular Ocean Model, version 3 (MOM3) (1.0° x 1/3°–1.0°; L40)</td>
<td>NCEP Reanalysis-2 direct</td>
<td>SST relaxed to obs with a time scale of 5 days Synthetic salinity created by T-S relationship</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>(CSv1)</td>
<td>T126L64</td>
<td>MOM4</td>
<td>NCEP Reanalysis-2 direct</td>
<td>Same as above except the ocean model of GODAS is upgraded from MOM3 to MOM4 SST strongly relaxed to obs every 6 h [see Saha et al. (2010) for more details]</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Saha et al. (2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCEP CSv2</td>
<td>T62L64</td>
<td>MOM2</td>
<td>Atmospheric and Land Initialization (ALI) (ERA-40 nudged; Hudson et al. 2010)</td>
<td>OI technique (sea temperature profiles only; Smith et al. 1991) SST relaxed to obs with a time scale of 3 days</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>(CSv2)</td>
<td>T126L64</td>
<td>MOM2</td>
<td>ALI (ERA-40 nudged)</td>
<td>POAMA Ensemble Ocean Data Assimilation System (PEODAS; ensemble Kalman filter, sea temperature, salinity profiles)</td>
<td>30</td>
<td>9</td>
</tr>
<tr>
<td>S. Saha et al. (2011, unpublished manuscript)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POAMA1.5b</td>
<td>T47L17</td>
<td>MOM2</td>
<td>ALI (ERA-40 nudged)</td>
<td>POAMA Ensemble Ocean Data Assimilation System (PEODAS; ensemble Kalman filter, sea temperature, salinity profiles)</td>
<td>30</td>
<td>9</td>
</tr>
<tr>
<td>(P15b)</td>
<td>BAM3.0d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alves et al. (2003)</td>
<td>T47L17</td>
<td>(2.0° x 0.5°; L25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POAMA2.4</td>
<td>T47L17</td>
<td>MOM2</td>
<td>ALI (ERA-40 nudged)</td>
<td>POAMA Ensemble Ocean Data Assimilation System (PEODAS; ensemble Kalman filter, sea temperature, salinity profiles)</td>
<td>30</td>
<td>9</td>
</tr>
<tr>
<td>multimodel ensemble (P24)</td>
<td>BAM3.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al. (2011)</td>
<td>T47L17</td>
<td>(2.0° x 0.5°; L25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POAMA2.4</td>
<td>T47L17</td>
<td>(2.0° x 0.5°; L25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>multimodel ensemble (P24)</td>
<td>BAM3.0d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al. (2011)</td>
<td>(P24a, b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POAMA2.4</td>
<td>(P24a, b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>multimodel ensemble (P24)</td>
<td>BAM3.0d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al. (2011)</td>
<td>(P24c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unauthenticated | Downloaded 07/11/22 04:35 AM UTC
members of the complete 15/24 ensemble members for
the CFSv1/CFSv2 that were initialized no more than 8
days prior to the first of each month. We note that
we should use as many ensemble members as possible.
However, the forecast skill of the measures used in this
study for the CFSv2 using the complete 24 ensemble
members is only slightly better than that using the most
recent 8 ensemble members, even though some of the
forecasts are up to 30 days lagged (not shown). The
difference between the smaller but less lagged ensemble and
the complete ensemble for CFSv1 is negligible (not shown).

In this study, we assess forecast skill using predicted
anomalies that are formed with respect to the corre-
sponding hindcast climatology of each model from
January 1982 to December 2006. We will also assess
mean state drift using the total predicted fields. We de-
fine a lead time of 0 month for the first full month of
forecast initialized on the first of the month (or in the
case of CFS, initialized up to 8 days earlier). For our
assessment of forecast skill, we will make use of 3-month
running mean (hereafter referred to as 3mM), where the
3-month mean is formed for each start time [note this is
in contrast to Luo et al. (2007) who applied running means
across different start times at a constant lead time]. The
filter is applied in order to remove intraseasonal noise
and to be consistent with the typical seasonal forecast
products issued by national meteorological agencies.
We define the 0-month lead forecast after the 3-month
running mean is applied as the mean of the first 3 months
of the forecast. For example, the 0-month lead SON fore-
cast is the 3-month mean of the September, October, and
November forecasts that were initialized on 1 September.
The remaining lead times are treated analogously.

Prediction of the WIO, EIO, and DMI indices are
verified against the same indices formed using the monthly
NOAA Optimum Interpolation (OI) SST V2 analyses
(Reynolds et al. 2002). The climatology of OI SST is
calculated for the period from January 1982 to Decem-
ber 2006. Hereafter, the OI SST will be simply referred
to as the “observed” data.

3. Skill assessment

a. Prediction skill in the WIO and EIO regions

We begin by examining forecast skill with the corre-
lation (COR; Figs. 2a,b) and root-mean-square-error
(RMSE; Figs. 2c,d) for anomalies of the ensemble mean
predictions of WIO and EIO indices using all start months.
The forecast RMSE is normalized by the observed stan-
dard deviation ($\sigma_{\text{obs}}$) of the indices.

We see in Figs. 2a–d that except for CFSv1 all systems
have higher skill that drops off more slowly for WIO
than for EIO. For instance, all forecast models, except
for CFSv1, have a correlation of at least 0.6 to a lead
time of 5–6 months for WIO, whereas the correlation
drops below 0.6 for EIO after only 2–3 months of lead
time. Similarly, the normalized RMSE of all forecast
models, except for CFSv1, remains less than 1 (the usual
limit of predictive skill, whereby the magnitude of the
error is smaller than that of the observed variability) to
6-month lead for the WIO. For the EIO, by contrast, the
normalized RMSE hits 1 beginning at a 1-month lead for
CFSv1, at a 3-month lead for SINTEX-F, and at about
a 5-month lead for P15b. This systematically higher
skill in WIO compared to EIO is a consequence of the
stronger influence of the more predictable ENSO on the
WIO: ENSO remotely forces the SST variations in WIO
through the tropical atmospheric bridge (e.g., Klein
et al. 1999; Shinoda et al. 2004a; Saji et al. 2006; Luo et al.
2010; Zhao et al. 2010). In contrast, SST variations in the
EIO are strongly influenced by less predictable coupled
processes that are intrinsic to the eastern Indian Ocean
(e.g., Wajsowicz 2005, 2007; Luo et al. 2007; Zhao and
Hendon 2009; Sooraj et al. 2012).

There are some indications in Figs. 2a–d that the
systems that used data assimilation have higher forecast
skill at short lead time. This is especially evident in the
EIO where ECSys3, and to a lesser degree the two
POAMA models, immediately exhibit higher skill than
the SINTEX-F at the shortest lead time. This greater
skill is maintained for a 2–4-month lead time. On the
other hand, the considerably lower skill for CFSv1
compared to all other systems in both the EIO and WIO
might be caused by the longer relaxation time to the
observed SST used in the Global Ocean Data Assimi-
lation System (GODAS) analysis for CFSv1 (5 days vs 3
days or less for the other system; Table 1; S. Saha 2011,
personal communication). Further investigation reveals
depiction of a spurious trend in the ocean initial condi-
tions (as expressed by the WIO and EIO SST indices)
used in CFSv1 that is then carried into the forecasts. If
the forecast data from CFSv1 are detrended, forecast
skill comes more into line with the other system (not
shown).

Despite a clear improvement in prediction of ENSO
with the P24 system compared to P15b (not shown, but
see Lim et al. 2010), there is little indication of any sig-
nificant improvement for predictions in the Indian Ocean
with the new POAMA system (Figs. 2a–d). Although
there has been a clear improvement in the ocean analyses
used for initializing P24 (Yin et al. 2011), model error

---

1 Definitions of all forecast metrics such as COR, RMSE, SPD, BIASS, HR, and FAR, etc., are provided in the appendix.
FIG. 2. Hindcast correlation for ensemble mean prediction of (a) WIO SST anomaly and (b) EIO SST anomaly using all start months 1982–2006. Forecasts and verification were smoothed with a 3-month running mean prior to computing correlation. Verification is based on Reynolds OI v2 SST analyses. Abscissa is lead time in months. (c),(d) As in (a),(b), but for RMSE normalized by the observed standard deviation of the index. The heavy dashed curve is for a persistence forecast. The magnitude of the observed standard deviation (OBS STD) is indicated in top right. Ordinate is standardized units. (e),(f) Standard deviation of the individual ensemble members normalized by the observed standard deviation.
probably prohibits any realization of this reduced error in improved forecast skill.

Another important aspect of these forecasts is the amplitude of the predicted SST indices because the associated tropical rainfall anomalies that are the source of teleconnections to the extratropics (e.g., Hoskins and Karoly 1981) are proportional to the magnitude of the SST anomaly. The normalized standard deviation of the individual ensemble members $\sigma_{\text{Mem}}/\sigma_{\text{obs}}$ for the WIO and EIO indices is shown in Figs. 2e,f. Note that $\sigma_{\text{Mem}}/\sigma_{\text{obs}} > 1$ (<1) in Figs. 2e,f indicates simulated variability that is too strong (weak) relative to observed. For WIO, $\sigma_{\text{Mem}}/\sigma_{\text{obs}}$ from both the ECSys3 and SINTEX-F is around 1.0 at all lead times, whereas that from the other seasonal forecast systems is greater than 1.0 (i.e., these models simulate too strong variability in the WIO). In the EIO region, all models except the POAMA models simulate variability that is too strong. The CFSv1 is the outlier model in both regions, which suggests that it suffers not only from a problem with initialization (discussed above), but also from errors in representations of the dynamics of the coupled climate in the Indian Ocean. Clearly, most of these problems are largely overcome with CFSv2. Finally, there is little systematic correspondence between forecast skill as measured by correlation (Figs. 2a,b) and the simulated level of variability (Figs. 2e,f), suggesting that deficiency in the simulated level of SST variability is not the sole limitation of forecast skill.

Another key aspect of evaluation of forecast performance is the relative spread (SPD) of the forecasts compared to the error, which indicates how reliable the forecast system is. For a perfectly reliable ensemble forecasting system, the spread should track error so that every forecast member can be considered to be representative of reality (e.g., Doblas-Reyes et al. 2009; Johnson and Bowler 2009). Typically, forecast models are under-dispersive so that the ratio of spread to error is less than one. The SPD-to-RMSE ratio for both the WIO and EIO indices at the 0-month lead time is shown in the Table 2. The SPD-to-RMSE ratio of all seasonal forecast systems is smaller than one. While the best systems have a ratio of about 0.8 in the WIO (CFSv2) and EIO (SINTEX-F), the typical ratio is between 0.5 and 0.7. Hence, all models have scope for forecast improvement by increasing spread and reducing error in the Indian Ocean. However, a better SPD-to-RMSE ratio does not guarantee higher prediction skill. For instance, this ratio for all of the seasonal forecast systems becomes closer to one in both the WIO and EIO region at longer forecast lead time (5–6 months; not shown), whereas, the forecast skill continues to decline, especially in the EIO region. Therefore, improved SPD-to-RMSE ratio needs to be achieved in conjunction with reduced forecast error. The SPD-to-RMSE ratio for the two versions of POAMA are generally the lowest of all of the models considered here and are similar for both the older and updated versions of POAMA, suggesting that neither of the ensemble initialization strategies used for POAMA (perturbing the atmosphere using lagged atmospheric states or perturbing the ocean, but using a single atmospheric state) is optimal and/or that model error is dominating.

\textbf{b. Mean state bias}

One possible explanation for the range of forecast skill seen in Fig. 2 is the mean state bias of each system. All of the models simulate a too cold EIO (not shown) and all but SINTEX-F simulate a cold WIO in varying degrees (not shown). Interestingly, the model with the greatest cold bias (both versions of POAMA; see e.g., Zhao and Hendon 2009) has comparable skill to the models with the smallest bias (SINTEX-F and ECSys3). And, although the models exhibit a range of cold biases in the WIO and EIO, the models exhibit a more common bias in the mean gradient in SST across the Indian Ocean that, except for CFSv2, is generally too strong relative to that observed. This bias in SST gradient would go along with too weak surface westerlies along the equator, which is a common coupled model problem in the Indian Ocean (e.g., Fischer et al. 2005; Saji et al. 2006; Luo et al. 2007; Zhao and Hendon 2009).

Although there is no obvious relationship between the mean state bias in the Indian Ocean and prediction skill for the tropical Indian Ocean (not shown), it does not necessarily mean that the bias has negligible impact on IOD prediction. The slope of the thermocline along the equatorial Indian Ocean (as reflected for instance by the bias in west to east SST gradient) can affect the upwelling off Sumatra (e.g., too steep a slope will result in stronger than observed impacts of upwelling on surface temperature) that then influences the structure and amplitude of IOD events, which in turn might lead to the spurious formation of IOD events at long lead time (e.g., Wajsowicz 2005; Luo et al. 2007, 2010; Zhao and Hendon 2009). Moreover, the mean state of tropical SST, in particular the equatorial SST gradient, is strongly tied to the mean state of rainfall and circulation of the Australasian monsoon (e.g., Krishnamurthy and

Table 2. The SPD-to-RMSE ratio for WIO and EIO indices at a 0-month lead time.

<table>
<thead>
<tr>
<th></th>
<th>ECSys3</th>
<th>SINTEX-F</th>
<th>CFSv1</th>
<th>CFSv2</th>
<th>P15b</th>
<th>P24</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIO</td>
<td>0.68</td>
<td>0.65</td>
<td>0.54</td>
<td>0.84</td>
<td>0.56</td>
<td>0.55</td>
</tr>
<tr>
<td>EIO</td>
<td>0.69</td>
<td>0.79</td>
<td>0.57</td>
<td>0.72</td>
<td>0.50</td>
<td>0.51</td>
</tr>
</tbody>
</table>
Kirtman 2003; Annamalai et al. 2005), which is a crucial factor for the development of IOD events. Therefore, we conclude that simulation of the mean state in the Indian Ocean is a major challenge for coupled climate models.

c. IOD prediction skill in SON

Although there is no indication of anticorrelated behavior (i.e., dipole behavior) between the WIO and EIO considering all months of the year (e.g., Baquero-Bernal et al. 2002; Dommenget and Latif 2002), if only the SON season is considered, the COR between WIO and EIO is moderately negative, COR (WIO, EIO) = −0.44; or COR (WIO, EIO) = −0.51 if detrended data are used. This moderate anticorrelation suggests that dipole behavior is not dominant even in SON, but a dipole does emerge as the leading empirical orthogonal function of SST in the Indian Ocean basin if only the SON season is considered (e.g., Zhao and Hendon 2009). Since the impact of the IOD on tropical convection and associated teleconnections is greater than the sum of the two poles (e.g., Cai et al. 2011), focusing on the DMI during the SON season when the IOD is best defined and most active is warranted and useful.

An overall impression of the occurrence of IOD events and the ability to forecast them is provided in Fig. 3, which shows the observed time series of the DMI index along with the plume of the ensemble mean forecasts out to a 6-month lead time from each forecast system for start times of 1 May (Fig. 3a), 1 July (Fig. 3b), and 1 September (Fig. 3c). IOD events clearly tend to peak in the boreal autumn season (SON), with large amplitude events having occurred in 1982, 1991, 1994, 1997, 2003, and 2006. There is little evidence of systematic forecast skill for the DMI for the 1 May starts, and in particular there is a common problem of false alarms. On the other hand, there appears to be good skill in predicting the peak phase and demise of the events for forecasts initialized on 1 September.

A summary of the seasonal variation of forecast skill for the DMI is provided for a 2-month lead time (Fig. 4a) and 4-month lead time (Fig. 4b). The highest skill occurs in the seasons when the IOD has peak amplitude (August–November), with most models exhibiting skill greater than 0.6 at a 2-month lead time (e.g., initialized on 1 July for verification in the SON season), except for the SINTEX, which is close to persistence. All models fall below 0.6 at a 4-month lead time (e.g., initialized on 1 May for the SON season). The steep rise in forecast skill leading into the SON season and the drop off in forecast skill going from SON into austral summer [December–February (DJF)] reflects the strong phase locking to the seasonal cycle of the development and decay of IOD events (e.g., Saji et al. 1999; Wajsowicz 2007; Luo et al. 2007; Zhao and Hendon 2009).
Concentrating just on the SON season when the IOD tends to peak, Fig. 5a shows that forecast skill to predict the DMI with a correlation of 0.6 ranges from as short as 2 months (SINTEX-F), which is only comparable to a persistence forecast, to as long as 4 months (ECSys3 and CFSv2). The normalized standard deviation of the predicted DMI (based on individual ensemble members) is shown in Fig. 5b. Comparing Figs. 5a,b suggests that there is some relationship between forecast skill and the amplitude of the predicted DMI variability (note that each model’s simulated DMI variability behaves similarly to the simulated amplitude of the EIO, which was shown in Fig. 2f): the models that simulate overly strong EIO and DMI amplitudes (e.g., SINTEX-F and CFSv1) have the lowest DMI forecast skill at short lead times. However, the ECSy3 model is an exception, as it has the highest forecast skill, but also overestimates the amplitude of the DMI, though not by as much as SINTEX-F and CFSv1.

Although the dynamical models provide forecasts that are superior to a persistence forecast at all lead times
(Fig. 5a), a persistence forecast may not be a very stringent baseline because of the strong association of the IOD with ENSO, which itself is highly persistent. Dommenget and Jansen (2009) suggest a more stringent test of forecast systems is to compare to a statistical forecast that takes into account the strong relationship of the IOD with ENSO. A baseline statistical forecast for the DMI anomaly in SON using persistence and the observed lagged relationship of the DMI with the Niño-3.4 SST index is formulated as

\[
\text{DMI}_\text{SON}(\text{iy}, \text{lt}) = B(\text{lt})\text{DMI}_\text{obs}(\text{im} - \text{lt}, \text{iy}) + C(\text{lt})\text{NINO34}_\text{obs}(\text{im} - \text{lt}, \text{iy}).
\]

Here we specify \(\text{im} = 9\) for forecasts that verify in SON. Here \(B(\text{lt})\) and \(C(\text{lt})\) are developed separately for each lead month using multiple linear regression via least squares method (i.e., the cross correlation of \(\text{DMI}_\text{obs}\) and \(\text{NINO34}_\text{obs}\) is accounted for) and are cross validated via the leave-one-out approach. The cross-validated forecast skill of this purely statistical algorithm in (1) is displayed as the dashed dot curve in Fig. 5a. It is clear that this statistical scheme is better than a persistence forecast especially at a longer lead time when the association of ENSO with the IOD is accounted for. Note, however, that most of forecast models are still able to outperform the statistical scheme at short lead times.

Another interesting aspect shown in Fig. 5a is that the CFSv1, which displays the lowest forecast skill for the individual WIO and EIO indices (Figs. 2a,b), has comparable skill to the other models for the DMI. This feature may result because the spurious trend in both the WIO and EIO indices has been effectively removed in the calculation of their difference (i.e., \(\text{DMI} = \text{WIO} - \text{EIO}\)). In addition to the prediction of the amplitude of the DMI, the degree of covariation of the WIO and EIO should be assessed in the forecasts. In Fig. 6a we show observed and simulated correlation of the WIO and EIO indices in SON. Again, we compute the simulated correlation using the individual ensemble members and display the average of the correlation over all members. Compared to the observation \((r = -0.44)\), all the dynamical models but ECSys3 and CFSv1 simulate a weaker than observed anticorrelation. The strong negative correlation in CFSv1 is consistent with the stronger than observed DMI amplitude. Similarly, the two versions of POAMA (P15b and P24) exhibit a progressive weakening of the anticorrelation between WIO and EIO consistent with the underestimated amplitude of the predicted DMI of the P15b and P24 (Fig. 5b).

As another step in assessing the capability of the dynamical models, we examine the observed and simulated relationship of the DMI with Niño-3.4 in SON (Fig. 6b). For the observed record considered here (1982–2006), the correlation is strongly positive \((r = 0.71)\). The simulated correlations from the forecast models indicate that all of the models simulate a too weak simultaneous relationship between the IOD and ENSO; however, a wide range of correlation is simulated. Since a number of observational and modeling studies have indicated that the IOD events, in particular extreme IOD events, during the SON season are tightly related to the ENSO (e.g., Saji and Yamagata 2003;
Shinoda et al. 2004b; Luo et al. 2008, 2010; Zhao and Hendon 2009), the weak representation of the ENSO–IOD relationship exhibited by some of the models (e.g., SINTEX-F, ECSys3, and CFSv1) suggests again that there is ample scope for improved prediction of the DMI if model error in the representation of the ENSO–IOD relationship is reduced. However, we note that the ENSO–IOD relationship cannot explain all the occurrence of the IOD events (e.g., Shinoda et al. 2004b; Yamagata et al. 2004; Meyers et al. 2007). Thus, it is interesting that the ECSys3, which provides the best prediction skill of the IOD also exhibits the weakest ENSO–IOD relationship, thereby suggesting scope for improved prediction with the ECSys3 if this relationship is improved.

The ENSO–IOD relationship can also be exploited to make a statistical-dynamical prediction of the DMI, because the teleconnection between ENSO and IOD can be considered a fast process (e.g., Dommenget and Jansen 2009). That is, we can develop a statistical forecast scheme similar to (1) but instead of using the lagged observed relationship between Niño-3.4 and the DMI, we use the simultaneous relationship of the DMI with Niño-3.4 in SON but then use the predicted value of the Niño-3.4 index to make the prediction of the DMI:

\[
DMI'_{\text{SON}}(iy, lt) = B(lt)DMI'_{\text{obs}}(im - lt, iy) + C'_{lt}(0)\text{NINO34}_{\text{SON}}(iy, lt).
\]

(2)

Again, we set \(im = 9\) for forecasts for SON. Here, \(C_{lt}(0)\) is developed using the simultaneous relationship of observed NINO34\(_{\text{SON}}\) and DMI\(_{\text{SON}}\), (indicated by the 0 in parentheses), but \(C_{lt}\) is a function of lead time because it is computed in a multiple linear regression that includes the persisted value of the observed DMI as another predictor. Forecasts with (2) are then made by plugging in the predicted NINO34\(_{\text{SON}}\) rather than the observed NINO34\(_{\text{obs}}\). The regression model is developed and scored with a leave-one-out cross validation although the model’s hindcast climatology used to form the predicted anomalies of Niño-3.4 is not cross validated.

Forecast skill for (2) is displayed in Fig. 7a. At short lead times, this scheme does not differ much from the pure statistical scheme (1), but at longer lead times some of the systems using this statistical-dynamical approach have more skill than the purely statistical scheme. This results from exploitation of the good prediction of Niño-3.4 at longer lead times and the simultaneous relationship of Niño-3.4 with the DMI.

The improvement of the forecasts directly from the models over the statistical-dynamical predictions can be quantified using a skill score. We assess this using a skill score formed using correlation as the forecast metric for forecasts that verify in SON:

\[
SScor(lt) = \frac{\text{COR}_{\text{model}}(lt) - \text{COR}_{\text{stat-dyn}}(lt)}{1 - \text{COR}_{\text{stat-dyn}}(lt)} \times 100\%,
\]

(3)

where \(\text{COR}_{\text{model}}\) is the correlation scores shown in Fig. 5a and \(\text{COR}_{\text{stat-dyn}}\) are the scores shown in Fig. 7a. The SScor is the percentage improvement of the direct
prediction of the DMI from the models over that based on the statistical-dynamical scheme (2). Here SScor > 0 means that the direct prediction from the dynamical system has higher correlation than the statistical-dynamical prediction and vice versa for SScor < 0. The SScor is displayed in Fig. 7b. To a lead time of about 3 months, all of the forecast systems, except SINTEX-F and CFSv1, directly provide more skillful forecasts of the DMI than does the statistical-dynamical scheme that exploits the relationship with ENSO and the prediction of Niño-3.4 from the models. However, after a 3-month lead time, the statistical-dynamical scheme provides a better forecast than does any of the direct predictions from the dynamical forecast systems. This implies there is ample scope to improve the CFSv1 and SINTEX-F predictions at short lead times and that the upper limit of predictability at longer lead times has probably not yet been achieved by any of the forecast systems because none of the systems are fully exploiting the teleconnection of ENSO to the IOD.

d. Prediction of IOD events

So far, we have assessed skill of predicting the IOD regardless of its magnitude. However since large-amplitude IOD events are most climatically important and thus most important to predict, we assess, following Stanski et al. (1989), the ability to predict three categories of IOD events: positive IOD events (PIOD) that have DMI amplitude that exceed ½ standard deviation, negative IOD events (NIO) that have DMI amplitude less than −½ standard deviation, and neutral IOD events (NEUT) that fall in between. For normally distributed data, the ½ standard deviation threshold is roughly the tercile threshold (the actual tercile threshold for normally distributed data is 0.43 standard deviations). A contingency table (Table 3) is made for the occurrence of observed and predicted IOD events using the DMI (note that we use the observed ½ standard deviation threshold to categorize the model forecasts). We note that the observed frequency of occurrence of the PIOD, NIO, and NEUT events in SON using our ½ standard deviation threshold are respectively 9/25 (=J/T), 7/25 (=L/T) and 10/25 (=K/T), which is close to a ratio of 1:1:1 (as expected for equal chances for each tercile).

Models that systematically under- or overpredict the amplitude of the DMI might be expected to under- or overpredict occurrence of IOD events. For instance, we might expect the CFSv1 to “cry wolf” often because it simulates overly strong DMI variability relative to the observed variability. To account for each model’s level of simulated IOD variability, another approach to computing hit rate (HR) and the false-alarm rate (FAR) is to first calibrate the forecast models by using the simulated rather than observed standard deviation of the DMI index as the threshold for an event. We have done this (results not shown) and the results are qualitatively similar to using the observed standard deviation as the threshold, but with differences as expected (e.g., the FAR for CFSv1 is less extreme if the model thresholds are used).

The HR for prediction of PIOD and NIO and the FAR (using thresholds based on ½ observed standard deviation) are displayed in Fig. 8 as a function of forecast lead time for forecasts that verify in SON. We also include HR and FAR using a reference climatological forecast. Here, we use the climatological probability derived from observed behavior (e.g., Stanski et al. 1989). For example, the climatological probability of occurrence of PIOD events is J/T and of NIO events is L/T.

It can be seen from the Figs. 8a,b that the HR for PIOD and NIO events in the SON season is much larger than the climatological rate of occurrence, with hit rates exceeding 50% extending to at least 4 months lead times for all systems and exceeding 70% for the best systems. That is, IOD events that occur in SON can be correctly predicted to a lead time well beyond one season. This is a much higher level of forecast skill than was suggested by Fig. 5, in which all forecasts of the IOD regardless of magnitude were considered. This higher level of skill for predicting the occurrence of IOD events is in line with the previous findings of Luo et al. (2007)

| Table 3. Contingency table of forecasting positive, neutral, and negative IOD events. |
|----------------|----------------|----------------|----------------|----------------|
|                | Observed positive IOD | Observed neutral | Observed negative IOD | Total          |
| Forecasted positive IOD | a               | d              | g               | M (a + d + g)  |
| Forecasted neutral       | b               | e              | h               | N (b + e + h)  |
| Forecasted negative IOD  | c               | f              | i               | O (c + f + i)  |
| Total                     | J (a + b + c)   | K (d + e + f) | L (g + h + i)   | T (a + b + c + d + e + f + g + h + i) |

2 See Table 3 for the definitions of the letters a–i and capital letters J–T.
who showed that some large IOD events are predictable three or more seasons in advance.

However, HR is not a complete score, even for extreme events. The FAR, which represents the fraction of observed NEUT events that were incorrectly predicted as either the PIOD or NIOD events, is greater than 50% for all systems after a 1-month lead time (Fig. 8c). It also exceeds the expected climatological occurrence of false alarm (gray dashed line in Fig. 8c) after a 2-month lead time for all systems except CFSv2 (but whose FAR is already close to the climatological rate of false alarm at a 0-month lead time). Here, the climatological expected FAR is estimated to be $\frac{3}{25}$ because the observed frequency of occurrence of the PIOD ($f = 8$), NIOD ($L = 7$), and NEUT ($K = 10$) events in SON is fairly close to a ratio of 1:1:1. Hence, although the forecast models usually correctly predict to a lead time of about 4 months the occurrence of IOD events when an event actually occurs (i.e., the HR is above 50%), they also often wrongly predict an event when none occurs; so that there is reduced confidence of an event occurring when one is forecast.

4. Conclusions

We have assessed the status of predicting SST anomalies in the tropical Indian Ocean and in particular the anomalies associated with the IOD using a representative selection of contemporary coupled climate models that are routinely used to make global seasonal climate forecasts. We have identified both common deficiencies and capabilities among these systems that are limiting and contributing to the ability to predict the IOD, thus providing some insights and guidance for future development of these forecast systems.

Prediction of the peak phase of the IOD in the boreal autumn SON season, which is typically when the IOD is best defined and most climatically important, is found to be limited to a lead time of about 3 months, which is much shorter than the lead time for prediction of ENSO. This limit of about a 3-month lead time implies that there is a strong barrier to predicting the development of the IOD prior to about May. This barrier stems in part from the strong seasonality of the development and demise of the IOD, which results from a strong seasonality in the atmosphere–ocean coupling in the eastern Indian Ocean that is responsible for the IOD (e.g., Li et al. 2003). Although the occurrence of large-amplitude IOD events can be predicted with these forecast systems to a much longer lead time (~6 months), consistent with the prediction experiments by Luo et al. (2007), all of the forecast models suffer from “crying wolf” so that confidence in prediction of large IOD events is diminished.

The skill for predicting the IOD appears to be largely limited by the skill of predicting SST in the eastern Indian Ocean, as evidenced by much reduced skill for predicting the EIO compared to the WIO SST index (e.g., Sooraj et al. 2012). Skill in the WIO is largely determined by predicting the forced response from ENSO (e.g., Zhao et al. 2010), while skill in predicting the EIO appears limited by the ability to simulate the coupled atmosphere–ocean interactions at the root of the IOD (e.g., Li et al. 2003).

In the small sample of forecast systems considered here, the overall best predictions of the IOD and SST...
more generally in the Indian Ocean were obtained from the systems that use data assimilation, indicating that future improvement of data assimilation systems is an important pathway for improved prediction in the Indian Ocean. Similarly, the best predictions of the IOD were obtained with the models that use the highest resolution. However, by some skill measures, the model with by far the lowest resolution (POAMA, versions P15b and P24) produced forecasts that were on a par with the higher-resolution models. Clearly, higher resolution is not a universal panacea for improved prediction but there appears to be a clear benefit of higher resolution. Recently, Sooraj et al. (2012) indicated that the prediction skill of SST anomaly in the tropical Indian Ocean may also be limited by the model’s capability in representing the monsoon circulation and that higher resolution is required to properly represent the upwelling along the coast of Java–Sumatra.

Although forecast skill for the IOD is much less than for ENSO, every forecast system appears to suffer from systematic error that acts to limit forecast skill of the IOD, thus suggesting that the upper limit of predictability of the IOD with these systems has yet to be achieved. For instance, many of the models simulate IOD variability that is systematically too large. This probably relates to the mean state bias in the Indian Ocean (too shallow thermocline in the equatorial eastern Indian Ocean as a result of too weak surface westerly winds; e.g., Fischer et al. 2005). Simulation of too strong IOD variability will contribute to false alarms and thus acts to limit the confidence of the good predictions of strong events.

The ENSO–IOD relationship is also universally underdone in the models, which appears to act to limit the longer-range prediction of the IOD because ENSO is predictable to lead times beyond 6 months. Although the cause of the weaker-than-observed ENSO–IOD relationship in the models cannot be diagnosed with the datasets available here, some previous work (e.g., Zhong et al. 2005; Zhao et al. 2010) suggests that it can be due to overly active atmosphere–ocean coupling in the Indian Ocean and biases in structure of the SST and circulation anomalies during ENSO. Such biases, which are acting to limit forecast skill but which we would hope should be alleviated with steady development of the coupled models, imply that the upper limit of prediction of the IOD has yet to be reached. In this study, we have emphasized the impact of ENSO on IOD prediction skill. However, the interaction goes the other way as well (i.e., the IOD impacts ENSO; see Ashok et al. 2003; Annamalai et al. 2010; Kug and Kang 2006; Luo et al. 2010; Izumo et al. 2010) and that an improper representation of this interaction may also contribute to reduced forecast skill of the IOD because of reduced skill in predicting ENSO.

Marked improvement in forecast skill can clearly be achieved through a program of forecast system development, as demonstrated by the CFS system. CFSv2 provides demonstrably better forecast skill for Indian Ocean SST than does its predecessor CFSv1. Our analysis suggests that this stems both from improved initialization and improved representation of the Indian Ocean coupled climate, although based on the analysis here we cannot provide detailed insights as to what changes matter most. On the other hand, the experience with the development of the POAMA system emphasizes the challenge: the substantial improvement in ocean initial conditions provided by the new ocean analyses in the P24 system (Yin et al. 2011) have not resulted in a marked improvement in forecast skill for Indian Ocean SST compared to the older P15b system. Model error in the newer version of POAMA is similar to that in the older version, and this error is probably the limiting factor for improved prediction of the IOD. Clearly the pathway to improved predictions of the IOD is to improve both the initialization of the coupled ocean–atmosphere state and to reduce systematic model error in the representation of the coupled ocean–atmosphere climate (both mean state and variability). Improving the coupled models used to make predictions is especially important because as more remotely sensed and in situ observational data in the Indian Ocean becomes available through implementation of programs like the Climate Variability and Predictability (CLIVAR) Indian Ocean Observing System (IndOOS; see http://www.clivar.org/organization/indian/IndOOS/obs.php), the paucity of quality data in the Indian Ocean should not be a limiting factor for prediction of the IOD.

Acknowledgments. Support for this work was provided by the Managing Climate Variability Program GRDC and the South East Australia Climate Initiative. Comments and encouragement by the WCRP Working Group for Seasonal-to-Interannual Prediction are acknowledged. Assistance with POAMA data by G. Liu, A. Charles, M. Zhao, G. Wang, E.-P. Lim, and Y. Yin is gratefully acknowledged. Comments by S. Saha and assistance from Weiqing Qu and the CFS team for accessing CFS data are acknowledged.

APPENDIX

Computational Formulas for COR, RMSE, $\sigma_{\text{Mem}}$, SPD, BIAS, HR, and FAR

$a$. COR

Forecast anomalies $X_{im}(im, iy, ie, lt)$ from model version iv and ensemble member (i.e., that verify at month
im for year iy at lead time lt) are formed relative to each model’s 25-yr hindcast climatology \(X_{iv}(im, lt)\) as follows:

\[
X'_{iv}(im, iy, ie, lt) = X_{iv}(im, iy, ie, lt) - [X]_{iv}(im, lt),
\]

(A1)

where

\[
\text{COR}(lt) = \left\{ \frac{1}{nyrs} \sum_{iy=1}^{nyrs} \sum_{im=1}^{12} \frac{[X'(im, iy, lt) \text{OBS}'(im, iy)]^2}{[\sigma_{\text{EnsM}}(lt) \sigma_{\text{obs}}]^{-1}} \right\}^{1/2}. \tag{A3}
\]

b. RMSE

The root-mean-square-error (RMSE) for anomalies of the ensemble mean is defined as

\[
\text{RMSE}(lt) = \sqrt{\frac{1}{nyrs} \sum_{iy=1}^{nyrs} \sum_{im=1}^{12} [X'(im, iy, lt) - \text{OBS}'(im, iy)]^2}. \tag{A6}
\]

c. \(\sigma_{\text{Mem}}\)

The standard deviation of the individual ensemble members, which gives an indication of the amplitude of the simulated intrinsic variability by the models, is defined as

\[
\sigma_{\text{Mem}}(lt) = \left\{ \frac{1}{nvers} \sum_{iv=1}^{nvers} \frac{1}{nens} \sum_{iy=1}^{nyrs} \sum_{im=1}^{12} \frac{[X'_{iv}(im, iy, ie, lt)]^2}{[\sigma_{\text{EnsM}}(lt) \sigma_{\text{obs}}]^{-1}} \right\}^{1/2}. \tag{A7}
\]

d. SPD

The spread (SPD) of the ensemble members about the ensemble mean is given by

\[
\text{SPD}(lt) = \sqrt{\frac{1}{nyrs} \sum_{iy=1}^{nyrs} \frac{1}{nvers} \sum_{iv=1}^{nvers} \frac{1}{nens} \sum_{ie=1}^{nens} \sum_{im=1}^{12} [X'_{iv}(im, iy, ie, lt) - X'(im, iy, lt)]^2}. \tag{A8}
\]

e. \(HR^{A1}\)

The hit rate (HR) for correctly forecasting the occurrence of an IOD event is defined as

\[
\text{HR} = a/(a + b + c) \times 100\% \quad \text{or} \quad i/(g + h + i) \times 100\%. \tag{A9}
\]

f. FAR

The false-alarm rate (FAR), which is a measure of “crying wolf” (i.e., incorrectly forecasting an event when in reality a neutral event occurred), is formed as

\[
\text{FAR} = (d + f)/(d + e + f) \times 100\%. \tag{A10}
\]

REFERENCES


