Clim Dyn (2011) 36:649–661
DOI 10.1007/s00382-010-0754-x

Prediction of the Madden–Julian oscillation with the POAMA dynamical prediction system

Harun A. Rashid · Harry H. Hendon · Matthew C. Wheeler · Oscar Alves

Received: 23 June 2009 / Accepted: 27 January 2010 / Published online: 19 February 2010

Abstract Predictions of the Madden–Julian oscillation (MJO) are assessed using a 10-member ensemble of hindcasts from POAMA, the Australian Bureau of Meteorology coupled ocean–atmosphere seasonal prediction system. The ensemble of hindcasts was initialised from observed atmosphere and ocean initial conditions on the first of each month during 1980–2006. The MJO is diagnosed using the Wheeler-Hendon Real-time Multivariate MJO (RMM) index, which involves projection of daily data onto the leading pair of eigenmodes from an analysis of zonal winds at 200 and 850 hPa and outgoing longwave radiation (OLR) averaged about the equator. Forecasts of the two component (RMM1 and RMM2) index are quantitatively compared with observed behaviour derived from NCEP reanalyses and satellite OLR using the bivariate correlation skill, root-mean-square error (RMSE), and measures of the MJO amplitude and phase error. Comparison is also made with a simple vector autoregressive (VAR) prediction model of RMM as a benchmark. Using the full hindcast set, we find that the MJO can be predicted with the POAMA ensemble out to about 21 days as measured by the bivariate correlation exceeding 0.5 and the bivariate RMSE remaining below ~ 1.4 (which is the value for a climatological forecast). The VAR model, by comparison, drops to a correlation of 0.5 by about 12 days. The prediction limit from POAMA increases by less than 2 days for times when the MJO has large initial amplitude, and has little sensitivity to the initial phase of the MJO. The VAR model, on the other hand, shows a somewhat larger increase in skill for times of strong MJO variability and has greater sensitivity to initial phase, with lower skill for times when MJO convection is developing in the Indian Ocean. The sensitivity to season is, however, greater for POAMA, with maximum skill occurring in the December–January–February season and minimum skill in June–July–August. Examination of the MJO amplitudes shows that individual POAMA members have slightly above observed amplitude after a spin-up of about 10 days, whereas examination of the MJO phase error reveals that the model has a consistent tendency to propagate the MJO slightly slower than observed. Finally, an estimate of potential predictability of the MJO in POAMA hindcasts suggests that actual MJO prediction skill may be further improved through continued development of the dynamical prediction system.

Keywords Madden–Julian Oscillation · Tropical intraseasonal variability · Predictability · Ensemble hindcasts · RMM index

1 Introduction

The Madden–Julian Oscillation (MJO) exerts an important influence on tropical weather, including driving variations of the Australian and Asian summer monsoons (e.g., Hendon and Liebmann 1990; Goswami 2005; Wheeler and McBride 2005) and modulating tropical cyclone development (e.g., Liebmann et al. 1994; Hall et al. 2001; Leroy and Wheeler 2008). The MJO also influences extratopical weather through modulation of large-scale sources of

H. A. Rashid · H. H. Hendon · M. C. Wheeler · O. Alves
Centre for Australian Weather and Climate Research
(A partnership between CSIRO and the Bureau of Meteorology),
Melbourne, Australia

H. A. Rashid (✉)
Centre for Australian Weather and Climate Research, CSIRO
Marine and Atmospheric Research, 107-121 Station Street,
Aspendale, VIC 3195, Australia
e-mail: harun.rashid@csiro.au

tropical diabatic heating which drive tropical-extratopical teleconnections that alter storm tracks and modulate synoptic weather events (e.g., Weickmann et al. 1985; Ferranti et al. 1990; Higgins and Mo 1997; Barlow et al. 2005; Wheeler et al. 2009). Extended range weather prediction for many regions of the globe would thus appear to benefit from the ability to make extended range prediction of the MJO. However, despite the relatively long timescale of the MJO, the upper limit of prediction skill for the MJO using global numerical weather prediction systems has been reported to be less than 2 weeks (Hendon et al. 2000; Vitart et al. 2007; Lin et al. 2008; Agudelo et al. 2009; Seo 2009). Skilful dynamical prediction of the MJO is limited, in part, by the inability to simulate the MJO with the models that are used for extended range weather prediction (e.g., Slingo et al. 1996; Hendon et al. 2000, Lin et al. 2006; Agudelo et al. 2009). Statistical forecasts of the MJO, on the other hand, have been modestly more successful, with the forecast skill extending beyond 2 weeks (e.g., Lo and Hendon 2000; Jones et al. 2004).

The limited success of dynamical predictions of the MJO compared to statistical predictions has been attributed to model deficiencies and imperfect initial conditions, rather than having already encountered some intrinsic limit of dynamical predictability of the MJO (Vitart et al. 2007; Agudelo et al. 2009). This is supported by the fact that potential predictability of the MJO using dynamical forecast models has been estimated to be up to 4 weeks (Waliser et al. 2003; Reichler and Roads 2005). The implication is that prediction skill of the MJO using dynamical models can be significantly enhanced with continued improvements of the models and their associated data assimilation systems.

As forecast systems improve, it is important to document the improvement in their ability to predict the MJO. However, different investigators have used different methods to assess predictive skill of the MJO. For example, in some studies intraseasonal anomalies have been isolated by applying band pass filters to the forecast data (e.g., Waliser et al. 1999; Jones et al. 2000; Wheeler and Weickmann 2001; Waliser et al. 2003); whereas, no intraseasonal filter was applied in others (e.g., Lo and Hendon 2000; Jiang et al. 2008; Lin et al. 2008). There are also differences in the variables used to identify the MJO, including the use of outgoing longwave radiation (OLR; e.g., Waliser et al. 1999), rainfall (Lin et al. 2008), streamfunction (Lo and Hendon 2000), zonal winds (Jones et al. 2004; Jiang et al. 2008; Seo 2009), and/or velocity potential (Vitart et al. 2007). Some recent MJO prediction studies (e.g., Vitart et al. 2007; Lin et al. 2008; Seo 2009) have employed variations of the Wheeler-Hendon Real-time Multivariate MJO index (RMM; Wheeler and Hendon 2004; hereafter WH04) for identifying the MJO. In its original application, the RMM is a two component index derived from the leading pair of empirical orthogonal functions (EOFs) of the combined fields of 15°S–15°N averaged OLR, 850-hPa zonal wind, and 200-hPa zonal wind. Importantly, it requires no time filtering. However, procedural differences in its applications to forecast models hinder meaningful comparisons and the ability to judge forecast improvements.

We describe here a systematic procedure for evaluating forecasts of the MJO using the RMM and apply it to hindcasts from the Australian Bureau of Meteorology's coupled ocean–atmosphere forecast model (called POAMA, for Predictive Ocean–Atmosphere Model for Australia). Besides a demonstration of the technique for evaluating forecast skill of the MJO that can be applied to other forecast systems, assessment of the MJO in POAMA forecasts is of interest because the POAMA model has been reported to simulate a reasonable MJO (Zhang et al. 2006). The procedure to identify and score forecasts of the MJO is similar to that developed by Lin et al. (2008), except here we use a form of the RMM indices that closely resembles WH04. The procedure is consistent with that recommended by the U.S. Climate Variability and Predictability (CLIVAR) MJO Working Group (Gottschalck et al. 2009; see also http://www.usclivar.org/mjo.php) for assessment of realtime predictions of the MJO from operational forecasting centers. Here we extend their technique to evaluate forecast skill from a large hindcast set. We also include comparison with forecasts of RMM from a simple autoregressive prediction model (Maharaj and Wheeler 2005), which is a useful baseline against which dynamical predictions of the MJO can be assessed.

The POAMA prediction system and the hindcast experiments are briefly reviewed in Sect. 2. The procedures for evaluating forecasts of the MJO using the RMM index are presented in Sect. 3. Section 4 describes the results, and conclusions are provided in Sect. 5.

2 POAMA model predictions

We use a 10-member ensemble of hindcasts from POAMA version 1.5b. This version of the model has been run operationally at the Australian Bureau of Meteorology for seasonal prediction since January 2008. Details of the model components and the initialization and generation of the retrospective hindcasts are described in Alves et al. (2003), Wang et al. (2008), and Zhao and Hendon (2009). Briefly, POAMA1.5b consists of the Bureau of Meteorology unified atmospheric model (BAM) version 3.0d coupled to the Australia Community Ocean Model version 2 (ACOM2), which was developed from the Geophysical Fluid Dynamics Laboratory Modular Ocean Model version

2. BAM3.0d is a spectral transform model, which in POAMA1.5b has triangular truncation 47 and 17 vertical levels. The ocean model grid spacing is 2 degrees zonally and 0.5° meridionally within 8° of the equator, increasing gradually to 1.5° near the poles. There are 25 levels in the vertical, with 12 in the top 185 m. The maximum depth is 5,000 m. The level thicknesses range from 15 m near the surface to almost 1,000 m near the bottom. Technical details and some aspects of the performance of the ACOM2 model, including the depiction of the Indonesian throughflow, are provided in Schiller et al. (1997) and Schiller et al. (2002). The ocean and atmosphere models are coupled using the Ocean Atmosphere Sea Ice Soil (OASIS) coupling software (Valcke et al. 2000) with no flux correction applied. Coupling occurs every 3 h.

Ocean initial conditions for the hindcasts are produced every three days by an ocean data assimilation scheme (Smith et al. 1991). Atmospheric and land initial conditions are provided daily by the atmosphere–land initialization (ALI) scheme (Hudson and Alves 2007), which involves nudging the output fields from an offline version of BAM3 toward the atmospheric reanalyses from ERA-40 (Uppala et al. 2005). This nudging process results in atmospheric fields that are very similar to the ERA-40 reanalysis but which are more in balance with the BAM 3.0d atmospheric model than would be if the ERA-40 fields were used directly as initial conditions. ALI also provides initial land surface conditions that are in balance with the atmospheric surface fields.

A 10 member ensemble of hindcasts for start times January 1980 through December 2006 was generated from the observed ocean and atmosphere initial conditions from the 1st of each month. The ensemble was generated by perturbing the atmospheric initial conditions by successively picking the atmospheric analysis from a 6 h earlier period (i.e., the tenth member was initialized 2.5 days earlier than the first member). There was no perturbation applied to the ocean initial conditions. All hindcasts were run for 9 months, although only the first 30 days are used in this study. Anomalies are formed relative to the hindcast model climatology for 1980–2006, which is a function of start month and lead time. In this fashion, the mean bias is removed.

3 Procedures for the assessment of MJO predictions

3.1 Computing the RMM indices

We identify the MJO using the RMM indices, following closely the procedure described in WH04. The RMM indices together with their associated spatial structures capture the key broadscale features of the MJO and are thus a suitable starting point for assessment of forecast skill. Assessment of forecast skill of the RMM indices also provides insight into predictive skill of tropical intraseasonal variability in general, because of the prominence of the MJO in the Tropics. The full development and justification for the index is provided in WH04. Briefly, the RMM indices are obtained from a combined EOF analysis of daily anomalies of latitudinally-averaged (15°S–15°N) zonal wind at 200 hPa, zonal wind at 850 hPa, and OLR, where OLR is used as a surrogate for organized moist convection. Anomalies are defined with respect to the climatological seasonal cycle and an estimate of interannual variability is removed. To ensure that the three variables have similar contributions to the multivariate index, the individual anomaly fields are normalized by the square root of their respective domain averaged variances before the EOF analysis. The resulting leading pair of EOFs have comparable variances (i.e., the eigenvalues are approximately equal) and their principal component time series (which are referred to as RMM1 and RMM2) exhibit a peak temporal correlation of $\sim$0.6 at $\sim$10 days lag. Together, this pair of EOFs depict the seasonally-independent eastward propagation of the MJO in the zonal wind and convection along the equator, while seasonally-dependent regional variability of the MJO can be recovered by regression onto RMM1 and RMM2. WH04 normalized RMM1 and RMM2 by the square root of their associated eigenvalues from the EOF analysis, which thus means that their standard deviations, when computed over all seasons and years, are individually equal to one. Compared to previous methods of diagnosing the MJO, the main advantage of the RMM indices are that they efficiently extract the salient features of the MJO including its intraseasonal timescale without the need for conventional time filtering.

For the present study, we use the spatial structures (eigenvectors) and variances (eigenvalues) as computed by WH04, available from http://www.bom.gov.au/bmrc/clfor/cfstaff/matw/maproom/RMM/. These were computed by WH04 using NCEP reanalysis data for the winds, and satellite observed OLR (Liebmann and Smith 1996). The spatial structures (eigenvectors) are displayed in Fig. 1. For both verification and model hindcast RMM values, we project daily data onto these observationally-derived spatial structures. For our verification of RMM, we use NCEP reanalysis winds and satellite OLR, like in WH04. However, because of differences in climatologies and the way interannual variability is removed, these verification RMM values are not exactly the same as those of WH04. In particular, the climatological seasonal cycle in this work is based on the 1979–2006 period, and we only perform one of the two steps of WH04 for the removal of interannual variability. The step we perform is the subtraction of the

previous 120-day mean from each input field, and we neglect the more complicated step of removing interannual variability that is specifically related to the El Nino-Southern Oscillation. Consistent with Lin et al. (2008), we found the results to be insensitive to this latter step, so the recommendation has been to not include it (Gottschalck et al. 2009). The individual latitudinally-averaged anomaly fields are then divided by the normalizations provided by WH04 prior to projecting onto the eigenvectors. The resulting verification RMM1 and RMM2, after dividing by the square-root of the eigenvalues obtained from WH04, still closely match those defined in WH04 (correlation coefficients exceeding 0.94 and standard deviations close to 1).

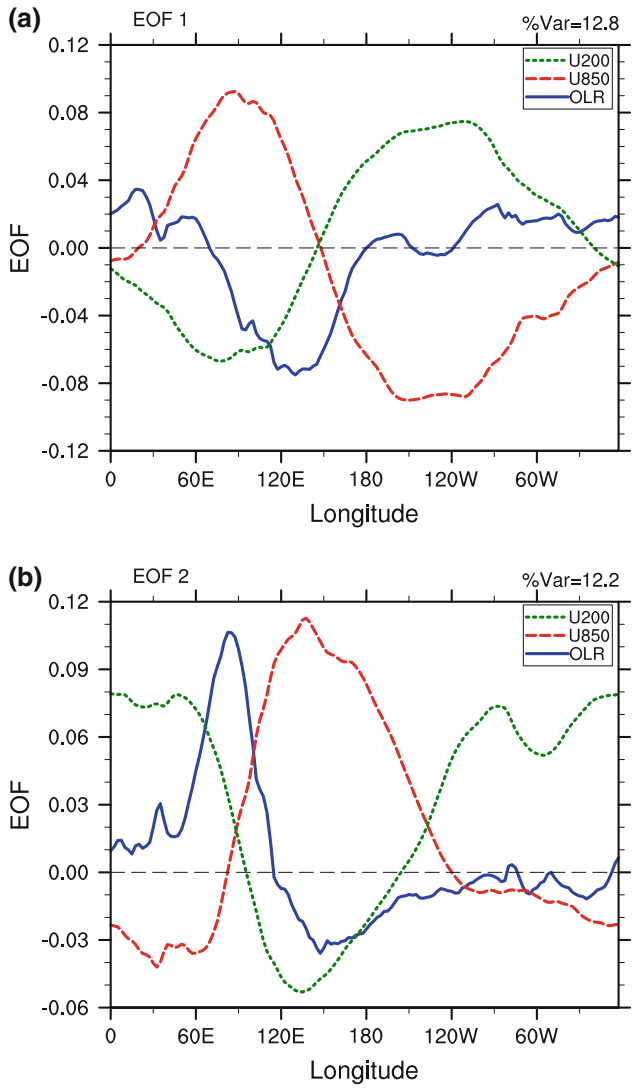


Fig. 1 Combined EOF1 (**a**) and EOF2 (**b**) of the latitudinally-averaged (15°S–15°N) anomalies of OLR (*blue curves*) and zonal winds at 200- (*green curves*) and 850-hPa (*red curves*) levels, adapted from WH04. The variances explained by EOFs 1 and 2 are 12.8 and 12.2%, respectively

The RMM indices for the dynamical hindcasts are computed similarly by projecting the hindcast anomalies of zonal winds and OLR onto the observed eigenvectors obtained from WH04. As described in Sect. 2, forecast anomalies are defined by removal of the hindcast model climatology for the hindcast period 1980–2006. Interannual variability is removed by subtracting the previous 120-day mean, but now this mean is created for a forecast at lead time n days as the mean of the previous n days of the forecast plus the previous $120 - n$ days of analyses up to the start of the forecast (Lin et al. 2008). The analyses are obtained from the forecast model initial conditions, which in the case of POAMA is the ALI analyses. Prior to projecting the model anomalies onto the observed EOFs, the individual latitudinally-averaged fields are divided by observed normalizations provided by WH04. After the projection, the resulting RMM1 and RMM2 are divided by the square root of the respective observed eigenvalues as obtained from WH04. This ensures that any systematic biases in the amplitude of the simulated MJO variability are revealed. The method described above is consistent with the approach that is being recommended by the U.S. Climate Variability and Predictability (CLIVAR) MJO Working Group (Gottschalck et al. 2009) for assessment of realtime forecasts of the MJO.

3.2 Example behavior of the RMM indices

Here we show an example of the usefulness of using the RMM indices to depict and assess forecasts of the MJO for late 1985 to early 1986 when the MJO was particularly active. Figure 2a shows a time-longitude plot of 850-hPa zonal wind anomalies for 17 December 1985 through 10 February 1986. Strong MJO activity is apparent with an episode of westerly anomalies appearing over Africa around 1 January 1986 that then propagates into the central Pacific by the end of January. The ensemble mean hindcast from POAMA that was initialized on 1 January 1986 is shown in Fig. 2b. While the overall intensity and evolution of the event is well predicted, the amplitude of the simulated event appears to be a bit too strong initially and it starts to weaken somewhat earlier than observed (which, as will be shown below, mostly stems from ensemble averaging of spreading forecasts).

The same forecasts may be analysed in terms of their projections onto the RMM indices as displayed in the two-dimensional phase–space (Fig. 3). We show the ensemble mean forecast that was initialized on 1 January 1986 and that runs through 10 February 1986, along with its verification. Systematic counter-clockwise progression of the observed RMM indices, with amplitude greater than 1, is indicative of well defined eastward propagation of the MJO during this period. The forecast captures well the initial

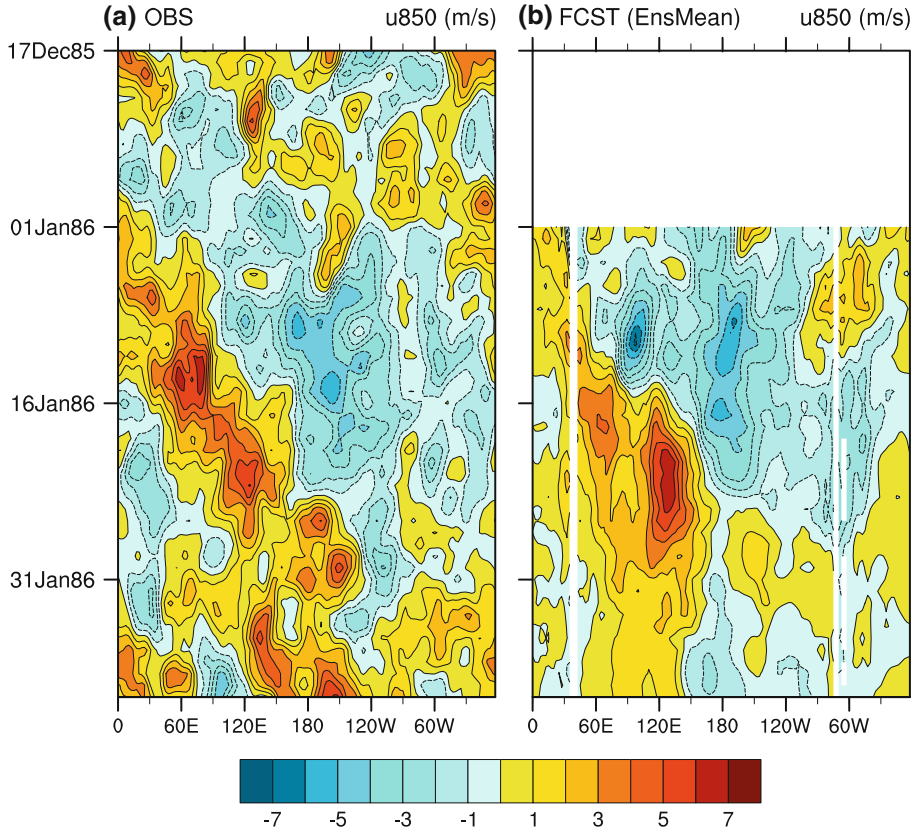


Fig. 2 Hovmoller diagrams of the latitudinally-averaged (15°S–15°N) zonal wind anomalies at 850-hPa from (**a**) NCEP/NCAR reanalysis from 17 December 1985 through 10 February 1986 and (**b**) the ensemble-mean POAMA hindcast initialised on 1 January 1986. The observed anomalies were smoothed with a 5-point local filter in space and time to facilitate comparison with the ensemble-mean

amplification, the systematic eastward propagation, the maintenance of the strength over the following few weeks, and the speed of eastward propagation (as indicated by similar rates of progression through the RMM phase–space). The loss of amplitude of the ensemble mean forecast, inferred in Fig. 2, is depicted by the spiral of the forecast toward the origin in Fig. 3. This example shows the usefulness of the RMM series as an MJO index and of the phase plot as a tool for displaying the salient features of the MJO. Quantification of skill using the RMM series is now developed.

3.3 Forecast skill, amplitude, phase, and statistical comparison

The RMM indices can be treated as a bivariate index whose individual components have near equal amplitude with unit standard deviation (at least for the observations). Following Lin et al. (2008), we can then score all of the forecasts at lead time τ with correlation (COR) and root-mean-square error (RMSE) using forms suitable for a bivariate index:

$$\mathrm{COR}(\tau) = \frac{\sum_{t=1}^{N}[a_1(t)b_1(t,\tau)+a_2(t)b_2(t,\tau)]}{\sqrt{\sum_{t=1}^{N}\left[a_1^2(t)+a_2^2(t)\right]}\sqrt{\sum_{t=1}^{N}\left[b_1^2(t,\tau)+b_2^2(t,\tau)\right]}} \quad (1)$$

$$\mathrm{RMSE}(\tau) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left([a_1(t)-b_1(t,\tau)]^2+[a_2(t)-b_2(t,\tau)]^2\right)} \quad (2)$$

Here, $a_1(t)$ and $a_2(t)$ are the verification RMM1 and RMM2 at time t, and $b_1(t, \tau)$ and $b_2(t, \tau)$ are the respective forecasts for time t for a lead time of τ days. N is the number of forecasts.

The bivariate amplitude (RMMA) for the verification is

$$\mathrm{RMMA}_{\mathrm{obs}}(t) = \sqrt{a_1(t)^2 + a_2(t)^2} \quad (3)$$

and similarly for the forecasts is

$$\mathrm{RMMA}_{\mathrm{for}}(t,\tau) = \sqrt{b_1(t,\tau)^2 + b_2(t,\tau)^2} \quad (4)$$

Average MJO amplitudes for observations and forecasts are thus defined as the time average (i.e., over t) of these two quantities respectively, and an average amplitude error can then be defined as

$$\mathrm{ERR}_{\mathrm{amp}}(\tau) = \frac{1}{N}\sum_{t=1}^{N}\left(\mathrm{RMMA}_{for}(t,\tau) - \mathrm{RMMA}_{\mathrm{obs}}(t)\right) \quad (5)$$

Note that we can infer $\mathrm{ERR}_{\mathrm{amp}}$ from the difference of the time-average of $\mathrm{RMMA}_{\mathrm{for}}$ and the time-average of

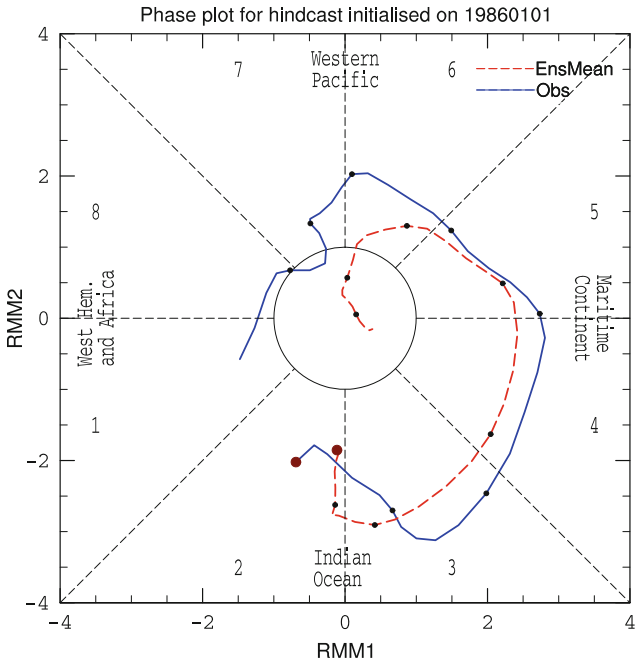


Fig. 3 Phase space plots of RMM1 and RMM2 computed from NCEP/NCAR reanalysis and satellite OLR (*blue*) and the ensemble-mean POAMA hindcast initialized on 1 January 1986 (*red*), for the period 1 January to 10 February 1986 The *black dots* are every 5 days. Each octant of the phase diagram is numbered (from 1 to 8) according to the phase definitions of WH04. Also labelled are the approximate locations of the enhanced convective signal of the MJO for that location of the phase space, e.g., the "Indian Ocean" for phases 2 and 3. The RMM1 and RMM2 values were smoothed with a 1–2–1 filter in time prior to plotting

$\mathrm{RMMA_{obs}}$, that is, a difference of the average bivariate amplitudes.

The average phase error ($\mathrm{ERR_{phs}}$) as a function of forecast lead time is:

$$\mathrm{ERR_{phs}}(\tau) = \frac{1}{N}\sum_{t=1}^{N} \tan^{-1}\left(\frac{a_1 b_2 - a_2 b_1}{a_1 b_1 + a_2 b_2}\right) \quad (6)$$

Here we have made use of the properties of the cross product (numerator) and dot products (denominator) for finding the angle between the vectors of the verification ($a_1(t)$, $a_2(t)$) and the forecasts ($b_1(t, \tau)$, $b_2(t, \tau)$) for time t and lead time of τ days. $\mathrm{ERR_{phs}}$ is therefore equivalent to the average phase angle difference between the forecasts and observations with a positive angle indicating the forecast leads the observations. For practical purposes, the phase error is computed only when the RMM amplitude is large so that it is not affected by possible large phase differences that may occur near the origin of the RMM phase space. We also note that the computation of (6) is not linear (i.e., $\mathrm{ERR_{phs}}$ is not equal to the difference between an average phase of the forecasts and observations), unlike the case for $\mathrm{ERR_{amp}}$ above.

As discussed in Lin et al. (2008), the upper limit of COR for a perfect forecast is 1, while the lower limit is -1. Similarly, for a perfect forecast the bivariate RMSE $= 0$. For a climatological forecast (i.e., zero MJO anomalies, thus $b_1 = b_2 = 0$), the bivariate RMSE $= \sqrt{2}$ because the standard deviation of each of the observed RMM indices is 1. Similarly, for forecasts with observed amplitude but completely random phase (e.g., a persistence forecast at very long lead time), the bivariate RMSE asymptotes to 2. Forecasts are deemed to be skilful for RMSE $< \sqrt{2}$ (i.e., they have lower RSME than a climatological forecast). Examination of the lead-time dependence of COR (discussed below in Fig. 4) indicates COR drops to about 0.5 at the same lead time when RMSE $\sim \sqrt{2}$. Following the discussion of Hollingsworth et al. (1980), we thus also use COR $= 0.5$ as an indication of the limit of a skilful forecast.

We also compare the dynamical predictions of the MJO from POAMA with the simple autoregressive prediction scheme for RMM1 and RMM2 developed by Maharaj and Wheeler (2005). They refer to their scheme as vector autoregressive (VAR), and it takes the form

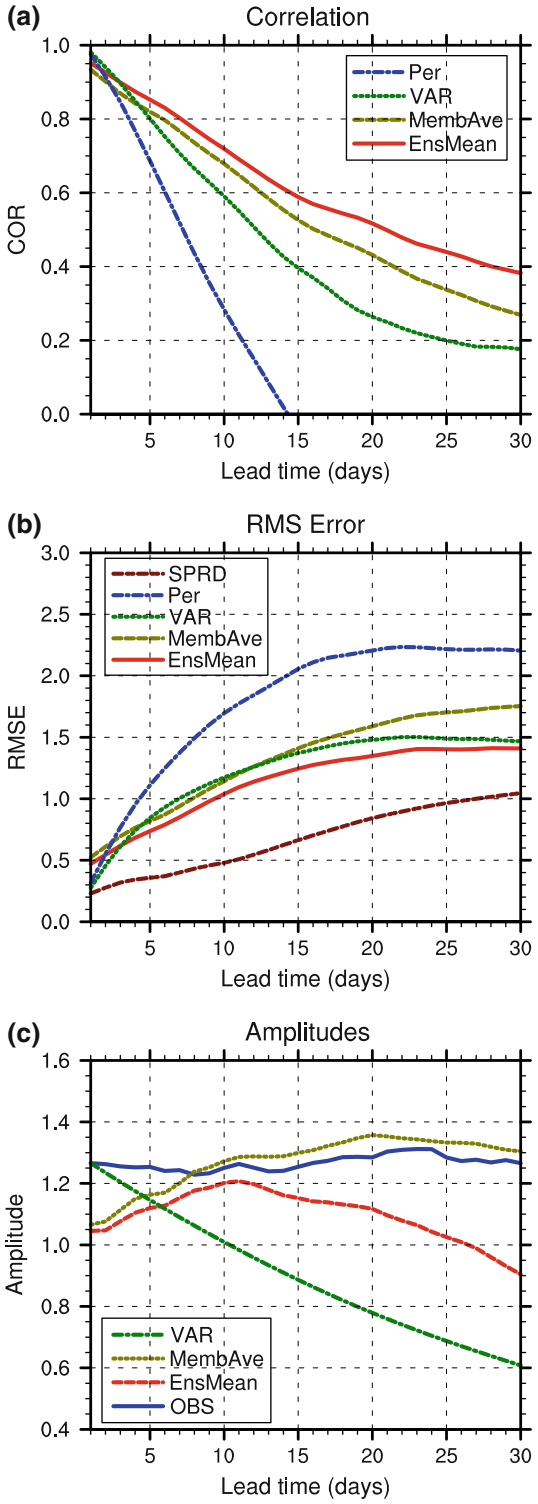
$$b_i(\tau) = \alpha_i b_1(\tau - 1) + \beta_i b_2(\tau - 1) \quad (7)$$

where $i = 1, 2$ and the coefficients α and β, which are developed using linear regression for two 6-month seasons (Nov–Apr and May–Oct), are provided in Maharaj and Wheeler (2005). Note that the initial forecast at lead time $\tau = 1$ is calculated, for each forecast case, from the observed values of $a_1(t)$ and $a_2(t)$. Forecasts at longer lead times are then obtained by successive application of (7). We refer to (7) as the VAR statistical model.

4 Results

We compute the bivariate COR and RMSE from the dynamical predictions using both the ensemble mean and the individual ensemble members. For the latter we display the average COR and RMSE for all members, which is not to be confused with the COR and RMSE of the ensemble mean. The average COR and RMSE from the individual members allows for an indication of the improvement of the ensemble mean over the individual members. It also provides insight into the changes in skill that occur as the amplitude of the ensemble mean drops to zero due to ensemble spread at longer lead times. For further comparison beyond the VAR statistical model, we also score persistence forecasts.

The COR and RMSE using all available start dates for 1980–2006 are presented in Fig. 4a, b. COR for the ensemble mean prediction from POAMA drops to 0.5 at about 21 days (Fig. 4a). The VAR statistical model, by



◀ **Fig. 4** The bivariate (**a**) correlation and (**b**) RMSE for RMM 1 and 2, as a function of forecast lead time for all hindcasts initialized 1980–2006. Displayed in (**a**) and (**b**) are scores for the ensemble mean (*red curve*), the VAR statistical forecast (*green curve*), the persistence forecast (*blue curve*), and the average value of COR or RMSE of the 10 individual ensemble members (*light green curve*). Also plotted in (**b**) is the bivariate standard deviation of the ensemble members about the ensemble mean (*brown curve*), which is also known as the ensemble spread. In (**c**) the average bivariate amplitudes are shown as a function of lead time for the ensemble mean (*red*), individual ensemble members (*light green*), VAR statistical forecasts (*green*), and observed (*blue*). All curves are smoothed with a 3-point running mean before plotting

comparison, reaches this level at 12 days, and VAR is consistently less skilful than the dynamical predictions at all lead times beyond the first few days. Both the dynamical and statistical predictions, however, clearly beat persistence. The correlation skill of the ensemble mean prediction is higher than the average correlation skill of the individual members, which shows the usefulness of the ensemble prediction over single-member predictions.

For the first 2–3 days, RMSE is smallest for persistence and VAR, but thereafter the ensemble mean forecast from POAMA has the smallest RMSE, consistent with the behavior of correlation skill discussed above. For lead times of about 10–20 days, the ensemble prediction is about 5 days better than for VAR or persistence. We also note that the RMSE for a persistence forecast grows to about $\sqrt{2}$ (i.e., the RMSE expected for a climatological forecast) by about 7 days, at which time COR for persistence is at about 0.5. Similarly, the average RMSE for the individual POAMA members reaches $\sqrt{2}$ at about 15 days, at which point the average COR for the individual members is also about 0.5, thus supporting our argument that a COR of about 0.5 is a useful indicator of the limit of skill.

Both the persistence and individual member forecasts have RMSE values that asymptote to ~ 2 at long lead times (longer than 30 days; Fig. 4b) as a result of their amplitude being realistically maintained into the future (Fig. 4c). Despite the fact that POAMA maintains a realistic amplitude of the MJO at long lead times for the individual members, the amplitude of the ensemble mean approaches zero at longer lead times (Fig. 4c) because the phase of the individual members becomes random. The amplitude of the VAR forecasts also asymptotes to zero, but this is a consequence of the amplitude decay stemming from the regression technique used to develop the statistical forecast. We also note in Fig. 4c that the average amplitude of the POAMA forecasts (both individual members and the ensemble mean) is initially smaller than observed, and then exhibits some spin-up over the first 10–12 days of the forecast. We attribute this spin-up to the fact that we initialize POAMA using ALI to a state which is very close to ERA-40, which apparently depicts a weaker MJO than is intrinsically supported by the POAMA model.

We also plot the ensemble spread about the ensemble mean in Fig. 4b, which can be compared to the RMSE of

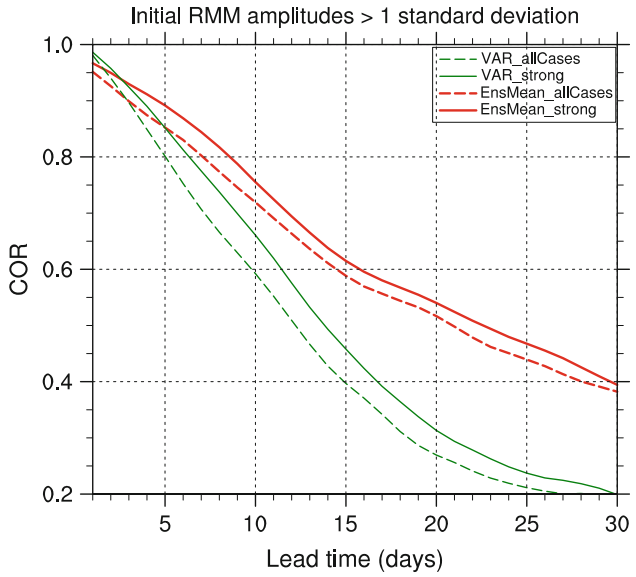


Fig. 5 As in Fig. 4a, except for those hindcasts for which the initial bivariate amplitude $RMMA_{obs} > 1$ (i.e., "strong" cases; *solid curves*). Curves are shown for the ensemble-mean (*red*) and the VAR statistical forecasts (*green*), and values using all initial conditions (*dashed curves*) are also plotted for comparison. The "strong" cases represent 62% of all forecasts

the ensemble mean. Even after only 1 day lead time, the RMSE of the ensemble mean exceeds the ensemble spread, indicating that the POAMA system is deficient in that the spread does not encompass the error. This is a common problem among dynamical ensemble prediction systems but is often remedied by inflating the ensemble spread (e.g., Bechtold et al. 2009), which we have not done here.

The assessment of skill discussed above considered all hindcasts, irrespective of whether or not there was a well defined MJO event at the initial time. If we select only those hindcasts that are associated with an already developed MJO event at the initial time, the predictive skill increases somewhat. Shown in Fig. 5 are the bivariate correlations for a subset of the dynamical and statistical hindcasts for which the initial $RMMA_{obs} > 1$. Similar results are inferred from RMSE (not shown). For both dynamical and statistical forecasts, the predictive skill is higher when the MJO is strong at the initial time, as has been reported previously for both statistical and dynamical predictions (Lo and Hendon 2000; Wheeler and Weickmann 2001; Lin et al. 2008). Compared to the assessment using all available start times (Fig. 4a), the lead time at which the COR drops to 0.5 for the ensemble mean prediction increases modestly by about 1–2 days lead time, while that for the VAR statistical model increases by at least 2 days lead time. Nonetheless, the skill of the dynamical predictions still far exceeds that of the VAR statistical model after 2–3 day lead time.

Next we examine the dependence of skill on season (Fig. 6). A significant seasonal variation in prediction skill is detected in both POAMA and the statistical VAR model, with generally the highest COR occurring during December–January–February (DJF) and lowest COR during June–July–August (JJA), consistent with other studies (e.g., Wheeler and Weickmann 2001; Lin et al. 2008; Seo 2009). Interestingly, the seasonal variation of prediction skill is greater in POAMA than the statistical model. Using a COR of 0.5 for comparison, predictive skill from POAMA extends to 22–23 days during DJF (and MAM) and 15 days in JJA, whereas for the statistical model it is 15 days in DJF and 10 days in JJA. Although we have no definite explanation for the cause of this seasonal

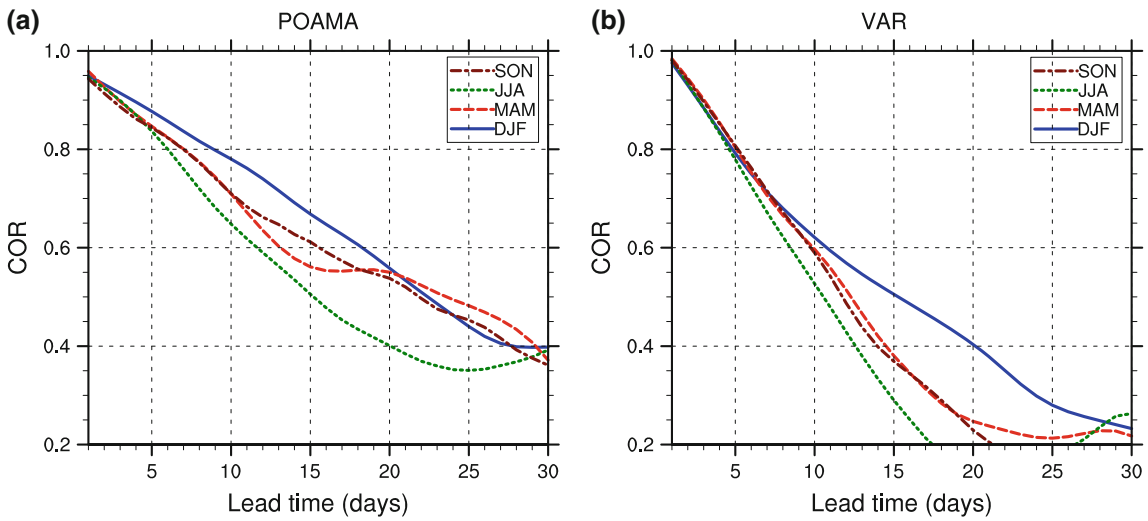


Fig. 6 As in Fig. 4a, except for (**a**) the ensemble-mean and (**b**) VAR statistical forecast for forecasts initialized in DJF (*blue*), MAM (*red*), JJA (*green*), and SON (*brown*)

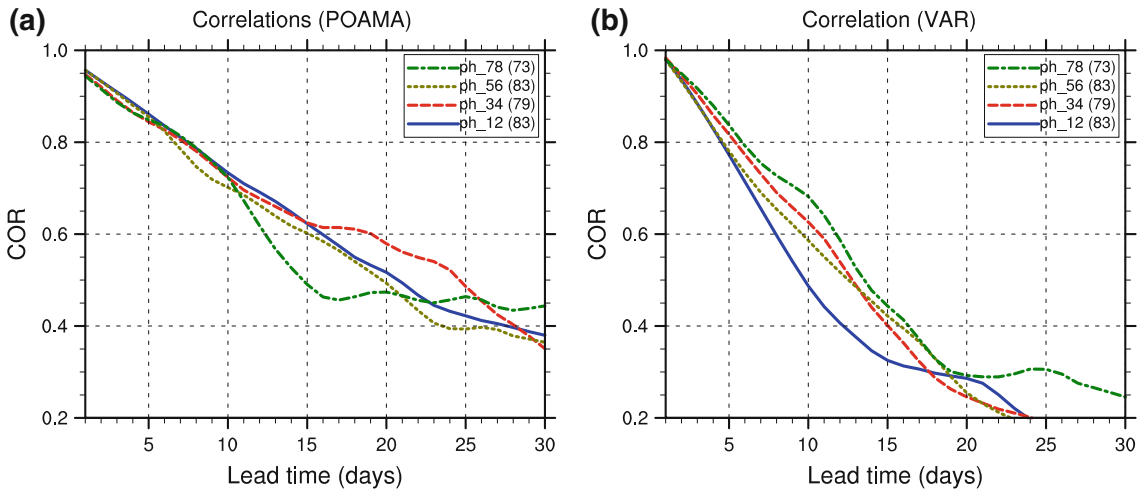


Fig. 7 As in Fig. 4a, except for (**a**) the ensemble-mean and (**b**) VAR statistical forecasts for forecasts stratified by the initial phase of the MJO corresponding to the four quadrants in Fig. 3. The *numbers* within parentheses show the number of forecasts that fall into each quadrant

dependence in skill, skill may be expected to depend on season because the MJO is usually best defined and strongest in southern summer, while during northern summer its behaviour is complicated by poleward propagation in the Indian monsoon region (e.g., Zhang 2005). Changes in model biases with season may also impact the seasonality of MJO prediction skill in POAMA.

We also examine the dependence of the predictive skill on the initial phase of the MJO as given by its location in the RMM phase–space (e.g., Fig. 3). Although the MJO may propagate around the entire globe along the equator and events may start or end in any phase (see also Matthews 2008), there are important changes in the structure of the MJO that occur between phases that may impact its prediction. In particular, the enhanced convective signature of the MJO is mostly restricted to the Indian and western Pacific Oceans (e.g., Hendon and Salby 1994), whereas the wind signal is detectable at all longitudes. Thus, as the MJO progresses through each RMM phase, there is the possibility that its different convective signal may impact upon its predictability. For example, the Indonesian archipelago has been reported to act as a barrier for propagation of the convective component of the MJO from the Indian Ocean into the western Pacific in some models (e.g., Inness et al. 2003). On the other hand, Lin et al. (2008) report distinctly better skill for forecasts initialized when MJO-convection is in the Indian Ocean as compared to when MJO-convection is in the western Pacific.

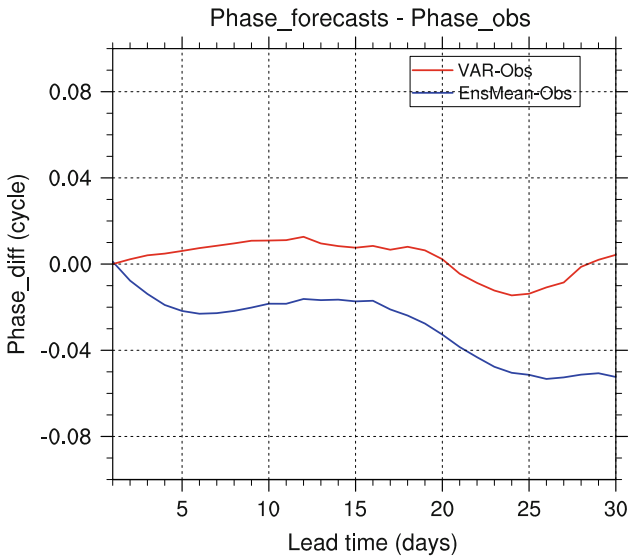


Fig. 8 Average phase errors (in cycles) as a function of forecast lead time for the ensemble mean (*blue*) and VAR statistical forecasts (*red*), using only those forecasts for which the initial bivariate amplitude $RMMA_{obs} > 1$. A 5-point running mean is applied to the curves prior to plotting

In order to investigate sensitivity to the initial phase of the MJO, we separate all the available forecasts into four groups depending on the RMM phase (corresponding to four quadrants in Fig. 3) at the forecast start-time. The CORs for these four groups of forecasts were computed separately (Fig. 7). Interestingly, the skill of the dynamical model shows little sensitivity to the initial phase of the MJO for forecasts out to about 10 days, while the VAR statistical model shows distinctly lower skill for forecasts initialized in phases 1 and 2 (i.e., when MJO-convection is developing in the Indian Ocean) after about 6 days. Hence, the POAMA forecast model appears not to suffer from the previously reported bias of developing and propagating the MJO-convection from the Indian Ocean into the west Pacific. A good example of this capability was displayed in Fig. 2.

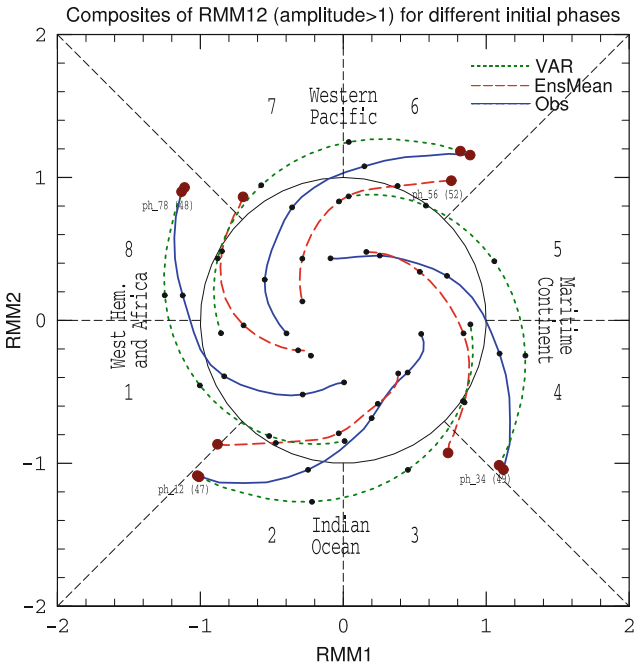


Fig. 9 Phase–space evolution of forecast and observed RMM1 and 2 composited for initial dates as in Fig. 7, but only for $RMMA_{obs} > 1$ at the initial time. The initial point is marked by a *big brown dot* for each curve. The *numbers* within parentheses near the initial points show the number of forecasts that fall into each quadrant. The evolution is shown for 21 days and the *black dots* are 5 days apart. A 5-point running mean is applied to the composite RMM values prior to plotting

Finally, we assess the average phase error (ERR_{phs}) for the POAMA ensemble mean and VAR forecasts (Fig. 8). Here, the average phase error is computed for times when the MJO is large at the initial time as in Fig. 5. As expected, the average phase error of the VAR model is near zero, consistent with its development based on observations. For the POAMA model, on the other hand, its average phase error is consistently negative, and gradually grows with time, equating to a systematic delay of about 2 days for a typical MJO with 50 day period by about 30 day lead time. This indicates that the coupled model tends to systematically propagate the MJO more slowly to the east than observed, consistent with the period of the intrinsic MJO in the POAMA model being slightly longer than observed (Marshall et al. 2009).

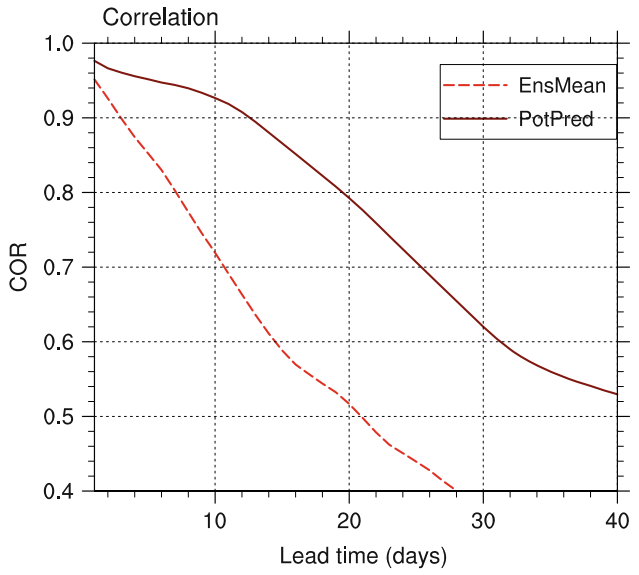


Fig. 10 Actual (*dashed curve*) and an estimate of potential skill (*solid curve*) for all hindcasts from POAMA for 1980–2006 using COR. A 5-point running mean is applied to the curves prior to plotting

Many of these important characteristics of both the observed and predicted behaviour of the MJO, including the phase error and maintenance of amplitude, may be summarized by the creation of the composite evolution of observed and predicted RMM1 and RMM2 for different starting phases, as displayed in Fig. 9. This view consistently shows that the dynamical predictions of RMM lie closer to reality than do the VAR predictions after about 5–10 days lead time. The slower phase speed of the dynamical model is also apparent in these composites. Interestingly, however, the amplitudes for both the observed and ensemble mean composites decay faster than for the statistical model. This may seem surprising at first, given that the amplitude for the statistical model decays monotonically with increasing lead time (Fig. 4c). However, the statistical forecasts of RMM are more consistent (no spread), therefore less variable, than the dynamical RMM forecasts. This ensures that the composite RMM forecast for the statistical model maintains larger amplitude than for the composite dynamical forecast, because in the latter case there are large cancellations among forecast values at longer lead times. The same explanation also applies to the rapid decay of the observed composite RMM values at longer lead times.

An approximate upper bound for MJO predictability can be estimated by the potential predictability of a perfect model (e.g., Anderson et al. 1999). In this approach, one member of the ensemble of forecasts is selected in turn as "truth" and the skill of the ensemble mean computed from the rest of the members is scored against it (instead of the observation). The effects of errors in the model and initial conditions on predictability are ignored in this approach. The difference between the potential predictability and the actual predictability can be viewed as the extent by which the predictive skill can be expected to be increased through improved initial conditions and reduced model error. Here we assess potential predictability using the average COR for each of the 10 ensemble sub samples (Fig. 10). Not surprisingly, a large difference is found between potential and actual predictability similar to that reported by Waliser et al. (2003). For example, potential COR skill remains

above 0.5 at 40 day lead time while actual COR skill drops below 0.5 by 22 days. The potential predictive skill exhibits a similar seasonal variation as the actual predictive skill, with the potential COR being highest during northern winters (not shown).

5 Discussion and conclusions

In this work, we propose and demonstrate a method to assess predictions of the MJO using hindcasts from the POAMA dynamical seasonal prediction system. We also compared the forecast skill from the dynamical model to a benchmark statistical model (a first-order vector autoregressive model) of the MJO.

The method to assess MJO forecasts using the RMM indices is summarized as:

1. Obtain observed eigenvectors and eigenvalues and normalisations for individual fields as calculated by WH04: (http://www.bom.gov.au/bmrc/clfor/cfstaff/matw/maproom/RMM/).
2. Form anomalies of the forecast fields of latitudinally-averaged zonal winds at 850 and 200 hPa and OLR by subtraction of the forecast model climatology (function of lead time and start date) and by removal of the previous 120-day mean (combination of n-days of forecast and $120 - n$ days of previous analyses for a forecast at $n + 1$ day lead time). Divide the individual fields by the observed normalizations obtained from WH04.
3. Project forecast anomalies onto observed eigenvectors from WH04 and then divide by the square root of the observed eigenvalues to obtain forecasts of RMM1 and RMM2.
4. Score the forecasts of the RMM indices using the bivariate correlation (Eq. 1) and root-mean-square-error (Eq. 2.). Examine also the average amplitude of forecasts compared to observations (Eqs. 3–5), and average phase error (Eq. 7).

Our investigation of the forecast performance of the POAMA system shows some promising results. The POAMA model is able to maintain realistic amplitude of the MJO throughout the forecast cycle, although there is an indication of some initial spin-up. For the ensemble mean, the bivariate correlation remains above 0.5 and bivariate RMSE remains less than $\sqrt{2}$ out to a lead time of about 21 days, which compares favourably to the statistical model ($\sim$12 days) and to persistence ($\sim$7 days). POAMA forecasts also show little sensitivity to the initial phase of the MJO, which indicates little bias in the representation of the MJO as it develops and propagates from the Indian Ocean into the western Pacific. The superiority of the dynamical forecasts over the statistical model is seen at all lead times beyond the first 2–3 days. Nonetheless, skill from the dynamical model at short lead time is less than from the statistical model, indicating scope for improvement of the dynamical forecasts especially relating to initialization.

Although we demonstrate that dynamical forecasts from POAMA are superior to simple statistical forecasts, predictive skill from the dynamical model is still limited to lead times less than about half the cycle of the MJO. However, our analysis of potential predictability indicates that the MJO may be predictable to lead times approaching a full cycle of the MJO. Whether skilful predictions of the MJO to lead times of 40 days can be realised through continual improvements in the model performance and initial conditions remains to be seen, but we argue that a systematic method of assessing MJO skill can facilitate forecast model improvement. For instance, our analysis of forecast skill from POAMA has highlighted some model deficiencies that need to be addressed, including eastward propagation of the MJO that is slower than observed (Figs. 8 and 9) and the initial amplitude is weaker than observed (Fig. 4c). The systematic method of assessing MJO forecast skill will also allow for comparison between different forecast models, thereby aiding model improvement. Such an activity of forecast model inter-comparison is being promoted by the CLIVAR MJO Working group (http://www.usclivar.org/mjo.php).

Finally, we have focused on assessing forecast skill of the tropical wide signature of the MJO. Ultimately, extended range climate prediction will require regional-scale prediction of tropical climate variability. As a next step, the role of the planetary-scale component of the MJO for regional climate prediction should be ascertained. This can be done, for instance, by assessing how much of the predictable component of regional tropical climate variability is accounted for by prediction of the MJO-component as determined by the RMM indices. Such an assessment with the POAMA forecasts is underway and will be reported elsewhere.

Acknowledgments Development of the forecast skill assessment for the MJO stemmed from the participation of HHH and MCW in the CLIVAR MJO Working Group. Drs. Debbie Hudson and Andrew Marshall reviewed an earlier version of the manuscript and provided useful comments. Comments from two anonymous reviewers have led to improvements in presentation of the results.

References

Agudelo PA, Hoyos CD, Webster PJ, Curry JA (2009) Application of a serial extended forecast experiment using the ECMWF model to interpret the predictive skill of tropical intraseasonal variability. Clim Dyn 32:855–872

Alves O, Wang G, Zhong A, Smith N, Tseitkin F, Warren G, Schiller A, Godfrey S, Meyers G (2003) POAMA: Bureau of Meteorology operational coupled model seasonal forecast system. In: Proceedings of national drought forum, Brisbane, April 2003, pp 49–56. Available from DPI Publications, Department of Primary Industries, GPO Box 46, Brisbane, Qld 4001, Australia

Anderson J, van den Dool H, Barnston A, Chen W, Stern W, Ploshay J (1999) Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and prediction. Bull Am Meteorol Soc 80:1349–1361

Barlow M, Wheeler M, Lyon B, Cullen H (2005) Modulation of daily precipitation over Southwest Asia by the Madden–Julian oscillation. Mon Weather Rev 133:3579–3594

Bechtold P, Kohler M, Jung T, Doblas-Reyes F, Leutbecher M, Rodwell MJ, Vitart F, Balsamo G (2009) Advances in simulating atmospheric variability with the ECMWF model: from synoptic to decadal time-scales. Q J R Meteorol Soc 134:1337–1351

Ferranti L, Palmer TN, Molteni F, Klinker K (1990) Tropical–extratropical interaction associated with the 30–60-day oscillation and its impact on medium and extended range prediction. J Atmos Sci 47:2177–2199

Goswami BN (2005) South Asian monsoon. In: Lau WKM, Waliser DE (eds) Intraseasonal variability in the atmosphere-ocean climate system. Springer, New York, pp 19–62

Gottschalck J, Wheeler M, Weickmann K, Vitart F, Savage N, Lin H, Hendon H, Waliser D, Sperber K, Prestrelo C, Nakagawa M, Flatau M, Higgins W (2010) A framework for assessing operational model MJO forecasts: a project of the CLIVAR Madden-Julian oscillation working group. Bull Amer Meteor Soc (accepted)

Hall JD, Matthews AJ, Karoly DJ (2001) The modulation of tropical cyclone activity in the Australian region by the Madden–Julian oscillation. Mon Weather Rev 129:2970–2982

Hendon HH, Liebmann B (1990) A composite study of onset of the Australian summer monsoon. J Atmos Sci 47:2227–2240

Hendon HH, Salby ML (1994) The life cycle of the Madden Julian oscillation. J Atmos Sci 51:2225–2237

Hendon HH, Liebmann B, Newman M, Glick JD, Schemm JE (2000) Medium range forecast errors associated with active episodes of the MJO. Mon Weather Rev 128:69–86

Higgins RW, Mo KC (1997) Persistent north Pacific circulation anomalies and the tropical intraseasonal oscillation. J Clim 10:223–244

Hollingsworth A, Arpe K, Tiedtke M, Capaldo M, Savijärvi H (1980) The performance of a medium-range forecast model in winter-impact of physical parameterizations. Mon Weather Rev 108:1736–1773

Hudson D, Alves O (2007) The impact of land-atmosphere initialisation on dynamical seasonal prediction. CAWCR research report no. 133, Bur. Met., Melbourne, Australia, 4 pp

Inness PM, Slingo JM, Guilyardi E, Cole J (2003) Simulation of the Madden–Julian oscillation in a coupled general circulation model. Part II: the role of the basic state. J Clim 16:365–382

Jiang X, Waliser DE, Wheeler MC, Jones C, Lee MI, Schubert SD (2008) Assessing the skill of an all-season statistical forecast model for the Madden–Julian oscillation. Mon Weather Rev 136:1940–1956

Jones C, Waliser DE, Schemm J-KE, Lau WKM (2000) Prediction skill of the Madden and Julian oscillation in dynamical extended range forecasts. Clim Dyn 16:273–289

Jones C, Carvalho LMV, Higgins RW, Waliser DE, Schemm J-KE (2004) A statistical forecast model of tropical intraseasonal convective anomalies. J Clim 17:2078–2095

Leroy A, Wheeler MC (2008) Statistical prediction of weekly tropical cyclone activity in the Southern Hemisphere. Mon Weather Rev 136:3637–3654

Liebmann B, Smith CA (1996) Description of a complete (interpolated) OLR dataset. Bull Am Meteorol Soc 77:1275–1277

Liebmann B, Hendon HH, Glick JD (1994) The relationship between tropical cyclones of the western Pacific and Indian Oceans and the Madden–Julian oscillation. J Meteorol Soc Jpn 72:401–412

Lin J-L et al (2006) Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: convective signals. J Clim 19:2665–2690

Lin H, Brunet G, Derome J (2008) Forecast skill of the Madden–Julian oscillation in two Canadian atmospheric models. Mon Weather Rev 136:4130–4149

Lo F, Hendon HH (2000) Empirical prediction of the Madden–Julian oscillation. Mon Weather Rev 128:2528–2543

Maharaj EA, Wheeler MC (2005) Forecasting an index of the Madden-oscillation. Int J Climatol 25:1611–1618

Marshall AG, Alves O, Hendon HH (2009) A coupled GCM analysis of MJO activity at the onset of El Niño. J Atmos Sci 66:966–983

Matthews AJ (2008) Primary and successive events in the Madden-Julian oscillation. Q J R Meteorol Soc 134:439–453

Reichler T, Roads JO (2005) Long-range predictability in the tropics. Part II: 30–60-day variability. J Clim 18:634–650

Schiller A, Godfrey JS, McIntosh P, Meyers G (1997) A global ocean general circulation model climate variability studies. CSIRO Marine Research Report No. 227, 60 pp

Schiller A, Godfrey JS, McIntosh PC, Meyers G, Smith NR, Alves O, Wang G, Fiedler R (2002) A new version of the Australian community ocean model for seasonal climate prediction. CSIRO Marine Research Report No. 240

Seo K-H (2009) Statistical-dynamical prediction of the Madden-Julian oscillation using NCEP Climate Forecast System (CFS). Int J Climatol. doi:10.1002/joc.1845

Slingo JM et al (1996) Intraseasonal oscillations in 15 atmospheric general circulation models: results from an AMIP diagnostic subproject. Clim Dyn 12:325–357

Smith NR, Blomley JE, Meyers G (1991) A univariate statistical interpolation scheme for subsurface thermal analyses in the tropical oceans. Prog Oceanogr 28:219–256

Uppala SM, Kallberg PW, Simmons AJ, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA (2005) The ERA-40 re-analysis. Q J R Meteorol Soc 131(612):2961–3012

Valcke S, Terray L, Piacentini A (2000) OASIS 2.4 Ocean Atmospheric Sea Ice Soil users guide, version 2.4. CERFACS technical report, CERFACS TR/CMGC/00-10, 85 pp

Vitart F, Woolnough S, Balmaseda MA, Tompkins AM (2007) Monthly forecast of the Madden–Julian oscillation using a coupled GCM. Mon Weather Rev 135:2700–2715

Waliser DE, Lau KM, Kim JH (1999) The influence of coupled sea surface temperatures on the Madden Julian oscillation: a model perturbation experiment. J Atmos Sci 56:333–358

Waliser DE, Lau KM, Stern W, Jones C (2003) Potential predictability of the Madden-Julian oscillation. Bull Am Meteorol Soc 84:33–50

Wang G, Alves O, Hudson D, Hendon H, Liu G, Tseitkin F (2008) SST skill assessment from the new POAMA-1.5 system. BMRC Res Lett 8:2–6

Weickmann KM, Lussky GR, Kutzbach JE (1985) Intraseasonal (30–60 day) fluctuations of outgoing longwave radiation and 250 mb streamfunction during northern winter. Mon Weather Rev 113:941–961

Wheeler MC, Hendon HH (2004) An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. Mon Weather Rev 132:1917–1932

Wheeler MC, McBride JL (2005) Australian–Indonesian monsoon. In: Lau WKM, Waliser DE (eds) Intraseasonal variability in the atmosphere-ocean climate system. Springer-Verlag, New York, pp 125–173

Wheeler M, Weickmann KM (2001) Real-time monitoring and prediction of modes of coherent synoptic to intraseasonal tropical variability. Mon Weather Rev 129:2677–2694

Wheeler MC, Hendon HH, Cleland S, Meinke H, Donald A (2009) Impacts of the Madden–Julian oscillation on Australian rainfall and circulation. J Clim 22:1482–1498

Zhang C (2005) Madden-Julian Oscillation. Rev Geophys 43. doi: 10.1029/2004RG000158

Zhang C, Dong M, Gualdi S, Hendon HH, Maloney ED, Marshall A, Sperber KR, Wang W (2006) Simulations of the Madden-Julian oscillation in four pairs of coupled and uncoupled global models. Clim Dyn 27:573–592

Zhao M, Hendon HH (2009) Representation and prediction of the Indian Ocean dipole in the POAMA seasonal forecast model. Q J R Meteorol Soc 135:337–352