Forecast skill of the Madden-Julian Oscillation in two Canadian atmospheric models

(*Revised version*)

Hai Lin, Gilbert Brunet

MRD/ASTD, Environment Canada, Dorval, Québec, Canada

Jacques Derome

Department of Atmospheric and Oceanic Sciences

McGill University, Montréal, Québec, Canada

Corresponding author: Dr. Hai Lin, MRD/ASTD, Environment Canada,

2121 route Trans-canadienne,

Dorval, Québec H9P 1J3, Canada. Email: hai.lin@ec.gc.ca

Abstract

The output of two global atmospheric models participating in the second phase of the Canadian Historical Forecasting Project (HFP2) is utilized to assess the forecast skill of the Madden-Julian Oscillation (MJO). The two models are the third generation of general circulation model (GCM3) of the Canadian Centre for Climate Modelling and Analysis (CCCma) and the Global Environmental Multiscale (GEM) model of Recherche en Prévision Numérique (RPN). Space-time spectral analysis of the daily precipitation in near-equilibrium integrations reveals that GEM has a better representation of the convectively coupled equatorial waves including the MJO, Kelvin, equatorial Rossby (ER) and mixed Rossby-gravity (MRG) waves. An objective of this study is to examine how the MJO forecast skill is influenced by the model's ability in representing the convectively coupled equatorial waves.

The observed MJO signal is measured by a bivariate index which is obtained by projecting the combined fields of 15°S-15°N meridionally averaged precipitation rate and zonal winds at 850 and 200 hPa onto the two leading empirical orthogonal function (EOF) structures as derived using the same meridionally averaged variables following a similar approach as Wheeler and Hendon (2004). The forecast MJO index, on the other hand, is calculated by projecting the forecast variables onto the same two EOFs.

With the HFP2 hindcast output spanning 35 years, for the first time the MJO forecast skill by dynamical models is assessed over such a long time period with a significant and robust result. The result shows that the GEM model produces a significantly better forecast skill for the MJO in the first two weeks. The difference is larger in Northern Hemisphere winter than in summer, when the correlation skill score drops below 0.50 at a lead time of 10 days for GEM whereas at 6 days for GCM3. At lead times longer than about 15 days, GCM3 performs slightly better. There are some

features that are common for the two models. The forecast skill is better in winter than in summer. Forecasts initialized with a large amplitude of the MJO are found to be more skillful than those with a weak MJO signal in the initial condition. The forecast skill is dependent on the phase of MJO at the initial condition. Forecasts initialized with an MJO that has an active convection in tropical Africa and Indian ocean sector have a better forecast skill than those initialized with a different phase of the MJO.

1   Introduction

The current predictability limit for the weather is of the order of one week. Beyond that, due to growth of initial errors and imperfection of the model, forecast error becomes so large that a prediction contains little useful information. Accordingly, great effort has been made in reducing initial errors, improving physical processes and developing more accurate and efficient numerical algorithms in order to improve weather forecasts on a time-scale shorter than a week (e.g., Van den Dool 1994). Another important development is in the area of seasonal forecasts. On this time scale, the predictability comes from influences external to the atmosphere (e.g., Shukla et al. 2000). In particular, the world oceans, with their much higher thermal and mechanical inertia, and correspondingly longer time scales, interact with the atmosphere and can induce atmospheric responses that are predictable on a seasonal time scale. There has recently been a growing interest in the extended-range (7 to 30 days) forecasts (e.g., Waliser et al. 2006), which reflects the need to fill the historical gap between the short range weather forecast and seasonal predictions. The strong amplitude and low-frequency nature of the Madden-Julian Oscillation (MJO) provide an important signal for the extended-range forecast in a global domain (e.g., Madden and Julian 1971; Lau and Phillips 1986; Matthews et al. 2004). Thus a skillful forecast of the MJO is both highly desirable and potentially achievable..

Using a "perfect model" approach, Waliser et al. (2003) estimated the limit of potential predictability of the MJO by the National Aeronautics and Space Administration (NASA) Goddard Laboratory for the Atmospheres (GLA) general circulation model (GCM). It was demonstrated that in the region of strong MJO (i.e., the Eastern Hemisphere), useful MJO predictability can reach 25-30 days for the 200 hPa velocity potential and about 10-15 days

for precipitation. In another potential predictability study of Reichler and Roads (2005), the "perfect model" approach was applied to the National Centers for Environmental Prediction (NCEP) seasonal forecasting model. It was shown that when model, initial, and boundary conditions are all perfect, the useful forecast skill of the intraseasonal variability is about 4 weeks. Forecasts of the MJO with statistical methods are also quite encouraging (e.g., Waliser et al. 1999; Lo and Hendon 2000; Wheeler and Weickmann 2001; Mo 2001; Jones et al. 2004). The useful predictive skill of the MJO from these empirical models can usually reach about 15-20 days lead time.

The actual forecast skill of the MJO by dynamical models, however, is considerably lower than that of the empirical models. Using the early National Meteorological Center (NMC) medium range forecasts, Chen and Alpert (1990) showed that the model forecast skill of the MJO reaches about 10 days when the MJO amplitude is large. However, when the MJO amplitude is weak, the forecast skill is poor. Jones et al. (2000) evaluated the MJO prediction skill for the NCEP's Dynamical Extended Range Forecasts (DERF), and found that useful predictive skill extends out to only 5-7 days. A similar result was reported in Hendon et al. (2000) based on the output of the same model. The low forecast skill of the MJO by dynamical models is most likely caused by problems associated with convective parameterizations in the models. Also the poor representation or neglect of air-sea interactions on this time scale may be another reason (e.g., Flatau et al. 1997). Slingo et al. (1996) compared the intraseasonal oscillation in 15 atmospheric general circulation models for the AMIP integrations, and found that the models do not simulate well the MJO. Significant model problems in simulating the tropical intraseasonal variability was also reported in a recent comparison of climate simulations by 14 coupled models participating in the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) (Lin et al. 2006). Furthermore,

due to scarce direct observations in the tropics, the dynamical MJO forecast is affected by large initial errors in the tropical analyses.

Up to now, there have been only a small number of MJO forecast studies using dynamical models and most of the models applied in studies of the MJO are similar to each other. Forecast studies of the MJO using different models are necessary to further explore the capability of the dynamical approach. In this study, we examine the forecast skill of the MJO by two Canadian models. From the near-climate-state integrations of the two models, it is observed that they have quite different behaviors in simulating the tropical waves including the MJO. One objective of this study is to see how this climate state behavior influences the MJO forecast skill in the extended range of 7-30 days.

Wheeler and Hendon (2004; hereafter WH04) developed an MJO index based on the first two combined EOFs of the meridionally averaged equatorial outgoing longwave radiation (OLR) and zonal winds at 850 and 200 hPa. An advantage of this index is that most of the low-frequency signal and structure of the MJO can be isolated from the observational and forecast data without the use of temporal filtering as used in many previous MJO-related studies, and is thus feasible for real-time MJO monitoring and prediction. We will apply this approach to analyze the predictive skill in the two dynamical models.

In section 2 the model and datasets are briefly described. The model climatology, low-frequency variability, and their comparisons with the observations are shown in section 3. The isolation of the MJO using a combined EOF analysis is discussed in section 4. In section 5, measures of forecast skill are introduced. Section 6 presents the results of MJO forecast skill produced by the two models. Section 7 gives a summary and discussion.

2 Models and data description

The forecast data are the output of two atmospheric models participating in the seasonal forecast experiments conducted under the second phase of the Canadian Historical Forecasting Project (HFP2) (Derome et al. 2001; Lin et al. 2007). The two models are the third generation of general circulation model (GCM3) of the Canadian Centre for Climate Modelling and Analysis (CCCma) and the Global Environmental Multiscale (GEM) model of Recherche en Prévision Numérique (RPN).

GCM3 shares many basic features with its precedent version (GCM2) that was used in previous studies for climate simulations (e.g., Boer et al., 1984; McFarlane et al., 1992). It is a global spectral model, with a triangular 63 (T63) horizontal resolution and 32 levels in the vertical. In GCM3 the cumulus parameterization of Zhang and McFarlane (1995) is implemented. GEM is an operational model at the Canadian Meteorological Centre (Côté et al., 1998a, 1998b). In HFP2, the GEM model was run at a horizontal resolution of $2° \times 2°$ and 50 vertical levels. It uses the Kain-Fritsch (KF) convective parameterization (Kain and Fritsch 1990) with only the threshold parameters being adjusted for the specific resolution. Neither global model has been tuned specifically for the tropics.

An ensemble of 10 parallel integrations of four-month duration was conducted for the 1969-2003 period using each model starting from the beginning of each month. The initial atmospheric conditions were at 12-hour intervals preceding the start of the forecasts, taken from the NCEP/NCAR reanalysis (Kalnay et al. 1996). The first member started from 12 hours before the start of the forecast, whereas the 10th member started from 120 hours (5 days) before. This is an acceptable practice for a seasonal forecast where signals from influences external to the atmosphere (e.g., SST anomalies) are more important than the

contribution from initial conditions for a seasonal mean condition after about 20 days. In the current study, since we are looking at the MJO forecast skill which is mainly in the first two weeks where the role of the initial condition is crucial, a 5-day lag in initial condition from one ensemble member to another is too big and the ensemble mean cannot be used as a forecast. Therefore, in this study, only one ensemble member is used that was initialized 12 hours before the start of the forecast season, i.e., for the integration of DJFM (December, January, February and March), the initial condition is at 12Z November 30. Therefore, we have a total of 420 (12 seasons × 35 years) forecasts from each model. Since the starting dates are one month apart, the MJO sample members are less dependent to each other than those that come from a set of more frequent forecasts (for example, every day). Therefore in the statistical significance tests which will be applied in the following sections, we will have a larger number of degree of freedom (dof) than a sample of the same size which comes from a set of more frequent forecasts. Global sea surface temperatures (SST) were predicted using the persistence of the anomaly of the preceding month, i.e., the SST anomaly from the previous month was added to the time-evolving climate of the forecast period. The variables analyzed are the daily data at 00Z for precipitation rate (PR) and zonal winds at 850 (u850) and 200 hPa (u200).

Since we do not have a long control run for each model, the last three months of each forecast is used to analyze the climatological behavior of each model, when the memory from the initial condition is largely forgotten. In the case of the wavenumber-spectrum analysis, the first 20 days of data are not used. Therefore a total of 1260 (12 seasons × 3 months × 35 years) months of data are available to represent a near-climate-state run for each model. Forecast data of the first 30 days are utilized to examine the forecast skill of the MJO.

The 4-times daily precipitation rate (PR), zonal winds on 850 and 200 hPa of NCEP/NCAR

reanalysis are used to calculate the climatological features for comparison with the models, to derive the combined EOFs which will serve as projection bases to get the MJO index, and as the verification. Although the precipitation rate in the NCEP/NCAR reanalysis data is model output instead of observations, it is presumably consistent with other variables in the model used to generate the reanalysis. It has been shown that on an interannual time scale the tropical PR agrees reasonably well with the observations (Lin et al. 2005). The NASA Global Precipitation Climatology Project (GPCP) 1-degree daily precipitation estimates (Huffman et al. 2001) are used as the observed precipitation rate when comparing the models' low-frequency variability in PR with the observations. The GPCP data of eight years (1997-2004) are available.

3   Simulated tropical low-frequency variability

3.1   Simulated low-frequency variability

The MJO is the dominant mode of intraseasonal variability in the tropics. The simulated tropical low-frequency variability in GEM and GCM3 are presented and compared with the observations in this subsection. As described in section 2, the data of the last three months of each forecast is used to represent the climatological behavior in the model, since we do not have a long control run for each model. Analysis is done for the winter half year and summer half year separately, where the winter half year refers to November to the next April, whereas the summer half year spans from May to October.

To extract the tropical low-frequency variability, a Fourier filter that passes variabilities with 20-90 day periods is applied to the observational and model data. The variances of the filtered u200 for the models and the NCEP reanalysis in boreal winter and summer are

6

presented in Fig. 1. As expected, the variance of the upper zonal wind variability is large in middle latitudes and relatively small in the tropics. The minimum variance occurs over the Indian Ocean and western Pacific. In boreal winter, a large low-frequency zonal wind variance can be found over the equatorial eastern Pacific, which may be associated with an extratropical influence and a variability of the zonal outflow of MJO-related convection in the Maritime continent and western Pacific regions. GEM simulated the upper zonal wind variability very well, whereas the large variability over the eastern equatorial Pacific is missing in the GCM3 simulations.

The variances of the filtered u850 are shown in Fig. 2. In boreal winter, a large low-frequency variance of low level zonal wind is found in the tropical Indian Ocean, the Maritime continent and the western Pacific sector. In boreal summer, the maximum variance is shifted to about 10°N, reflecting the low-frequency variability of the summer Indian monsoon circulation. Both GEM and GCM3 seem to be able to represent reasonably well the low level zonal wind low-frequency variability. From Fig. 2e, it can be seen that GEM has an above normal bias of low-frequency variance in the western Pacific during the boreal summer.

Figure 3 presents the variances of the filtered precipitation rate in the two model simulations and in GPCP observations. In general, the variance of low-frequency variability in PR is consistent with that in the low level zonal wind (Fig. 2). Strong variability in the MJO frequency range is reproduced in the tropical Indian Ocean, Maritime continent and western Pacific regions by the two models. The variance of PR in the models tends to have a positive bias compared with the observations, although the fidelity of the observed PR is also difficult to assess. Also the GPCP observational data are over a shorter period than the model data.

In summary, both GEM and GCM3 have a tropical climatology that agrees in general

with the observations. GEM is slightly better in simulating the time mean zonal wind in magnitude. Zonal winds related to the summer Indian monsoon are too strong in GCM3. Both models reproduce strong low-frequency variability of low level wind and precipitation rate in the tropical Indian Ocean, Maritime continent and western Pacific regions, that may be related to the MJO.

3.2 Wavenumber-frequency spectral analysis

It has been observed that a significant part of tropical convective activity is organized in waves that are consistent with the theory of equatorially trapped waves in the linear shallow water framework of Matsuno (1966) and Lindzen (1967). Through the analysis of wavenumber-frequency spectrum using the OLR data, Wheeler and Kiladis (1999) (WK hereafter) were able to identify the convectively coupled tropical waves from the background noise, that include the Kelvin, $n =1$ equatorial Rossby (ER), mixed Rossby-gravity (MRG), $n =0$ eastward inertia-gravity (EIG), and $n =1$ westward inertia-gravity (WIG) and $n =2$ WIG waves, where n is the meridional mode number. It was found that the wavenumber-frequency spectrum distributions of the waves match well the dispersion curves of the equatorially trapped wave modes with shallow equivalent depths of the order of 25 m.

In order to compare the behavior of tropical wave activity in GCM3 and GEM, the technique of WK is applied to the precipitation rate record for the two models. The data used for each 4-month forecast is the 96-day segment after the first 20 days of integration when the models fluctuate about a near-equilibrium state. As the linear equatorial waves are either symmetric or antisymmetric about the equator, the data is first decomposed into these two parts. A gridded field PR which is a function of latitude, ϕ, can be expressed as $PR(\phi) = PRA(\phi) + PRS(\phi)$, where $PRA(\phi) = [PR(\phi) - PR(-\phi)]/2$ is the antisymmetric

component, and $PRS(\phi) = [PR(\phi) + PR(-\phi)]/2$ is the symmetric component. We first decompose the precipitation rate into its antisymmetric and symmetric components for each latitude from 15°S to 15°N. Then the wavenumber-frequency spectrum is calculated for each latitude. Finally the power is averaged over all available time segments and is summed for the latitudes between 15°S and 15°N.

The background spectrum is estimated by averaging the power of the symmetric and antisymmetric spectra and smoothing many times with a 1-2-1 filter in frequency and wavenumber as in WK. The raw spectra are then divided by the background spectrum to obtain the estimation of the signal standing above the background noise. For each model 67 segments from a random selection are used. Based on the crude estimation of the degrees of freedom by WK with the same amount of nonoverlapping segments, the signal can be considered statistically significant at a 95% level if it stands at 1.1 times the background level. Calculations using different 67 segments were conducted and it was found that the wavenumber-frequency spectrum is not sensitive to the choice of segments.

Figure 4a and b show, respectively, the ratio of the raw symmetric and antisymmetric wavenumber-frequency spectra to the background spectrum for GEM. The theoretical dispersion curves for the shallow water wave modes are superimposed on the spectra, corresponding to equivalent depths of 12, 25 and 50 m. In the symmetric spectra, signals of Kelvin and equatorial Rossby (ER) waves can be clearly identified. In the antisymmetric spectra, there are mixed Rossby-gravity (MRG) and eastward inertio-gravity (EIG) waves. In the wavenumber range of 1 to 6, statistically significant large values of spectra at a constant frequency with an eastward propagation and a period near 40 days represent the MJO signal, especially in the symmetric spectra. The spectra distribution in the GEM simulations agrees reasonably well with that of the observations (see Fig. 3 of WK), indicating that

9

GEM, with its current physics parameterization schemes, is able to represent a variety of convectively coupled equatorial waves. As pointed out by previous studies (e.g., Majda and Biello 2005; Moncrieff 2004), interactions between different scales may provide an important energy source to the low-frequency variability. The spectrum peaks in GEM correspond well with shallow water modes that have an equivalent depth between 25 and 50 m, whereas in the observations the equivalent depth is around 25 m. This means that the phase speeds for tropical waves in GEM are a little too fast compared to that in the observations, e.g., for Kelvin waves the phase speed is about 19 m s^{-1} for GEM compared to about 16 m s^{-1} for the observations.

The ratio of the raw wavenumber-frequency spectrum to the background spectrum for GCM3 is illustrated in Fig.5. Almost no significant wave signals that are standing out from the background spectrum can be found in GCM3. The lack of wave signal in Fig.5 does not mean that there is no low-frequency variability in GCM3. In fact, as shown in the last subsection, GCM3 is similar to GEM in simulating the 20-90 day variability variance of PR and u850. Instead, Fig.5 indicates that this low-frequency variability is not discernible from the model's background noise.

In Lin et al. (2006), a similar analysis was conducted to compare the tropical intraseasonal variability in 14 coupled climate models participating in IPCC AR4. About half of the models appear to have signals of convectively coupled equatorial waves, but the waves in most of the models are too weak and the phase speeds are generally too fast. It is still an open question why some models are doing better than others. The coupled version of GCM3 (CGCM), though with a lower resolution than the AGCM in the present study, is among the 14 models compared. Almost the same result were obtained for CGCM as in GCM3 that it has very weak wave signals standing out from the background spectrum, indicating that the coupling

process with the ocean does not help to improve the situation. Through the analysis of a climate integration of the CCCMA GCM2, the precedent version of GCM3 which employs the same convective parameterization, Sheng (1995) also found that the simulated MJO signal is too weak compared with the observations.

4   Isolation of the MJO signal

A combined EOF analysis is performed based on the technique of WH04, except that instead of using OLR to represent the tropical convection, we use the precipitation rate (PR), because GCM3 and GEM did not output OLR in HFP2. Starting from the unfiltered observed daily averaged data of the NCEP/NCAR reanalysis for PR and zonal winds at 850 and 200 hPa from 1979 to 2003, the seasonal cycle which is the time mean and first three harmonics of the daily climatology is first removed for each grid point. Then the 120-day mean of the last 120 days for each day is removed. By removing the previous 120-day average most of the interannual variability, including those related to the ENSO and variability with even longer time scales, is removed. Next, a meridional band average near the equator (15°S to 15°N) is performed for the PR, u850 and u200, while retaining the longitudinal variation for the three fields. After that, each variable is normalized by its own zonal average of temporal standard deviation, and the three fields are then combined. As pointed out in WH04, this normalization is necessary to ensure that each field contributes equally to the variance of the combined vector.

Besides the use of PR instead of OLR, the other difference between our analysis and that of WH04 is that we do not remove the variability part that is linearly associated with an ENSO index before removing the previous 120 day average as WH04 did. The ENSO

11

index was a time series of the first rotated EOF of SST over the Indian and Pacific sectors calculated operationally at the Australian Bureau of Meteorology. As will be discussed later, this step is sensitive to the definition of the ENSO index. On the other hand, the subtraction of the previous 120 average is effective to remove the long-term variabilities that include the ENSO signal.

Shown in Fig. 6 are the longitudinal distribution of the leading two EOFs of the combined fields of PR, u850 and u200. EOF1 and EOF2 explain 12.2% and 11.1% of the total combined variance, respectively. They are well separated from the rest of the EOFs according to the criterion of North et al. (1982) (EOF3 explains only 5.3% of the variance). Both EOF1 and EOF2 are characterized by wavenumber one in the zonal wind. The zonal winds at 200 hPa are out of phase with that at 850 hPa, indicating a baroclinic structure. For EOF1, the zonal winds at both levels change sign near 150°E and the Greenwich longitude. Near 150°E, upper zonal divergence (easterly wind anomalies to the west over the Indian Ocean and westerly wind anomalies to the east across the Pacific) and lower zonal convergence (westerlies to the west and easterlies to the east) occur, which favours an enhanced upward motion and precipitation. The opposite situation is observed near the Greenwich longitude, where there are upper zonal convergence and lower zonal divergence, and thus a downward air flow anomaly is implied. The structure of EOF2 is almost in quadrature to that of EOF1, with the zonal divergent/convergent centers moved eastward. The precipitation structure is a little noisy. In EOF1, enhanced precipitation can be seen near the Maritime continent longitudes, consistent with the upper zonal divergence and lower level zonal convergence near 150°E. In the case of EOF2, reduced precipitation is observed in the eastern Indian Ocean area which is colocated with upper zonal convergence and lower zonal divergence. The power spectra of the principal components of EOF1 and EOF2 (PC1 and PC2) have a peak in the

30-80 day band, with about 63% of total variance in this band. The lag correlation between PC1 and PC2 peaks (0.74) when PC1 leads PC2 by 10 days, and reaches the maximum negative value (-0.77) when PC1 lags PC2 by 10 days, indicating that EOF1 and EOF2 represent an eastward-propagating signal with a period of about 40 days.

Despite the difference in approach, our resulting structures of EOF1 and EOF2 for u200 and u850 are almost identical to those of WH04 (their Fig. 1). The temporal correlation between the daily principal component time series of the first EOF (PC1) with that of WH04 in the period of 1979-2003 (obtained from the Australian Bureau of Meteorology website) is 0.96, and 0.97 for PC2. The two leading EOFs are consistent with those obtained in previous studies using bandpass filtered single variable data (e.g., Lau and Chan 1986), that represent the MJO at different phases, and their corresponding PCs vary mostly on the intraseasonal time scale of the MJO. Since no temporal filtering is used here in obtaining the combined EOFs, they are suitable for real-time application. As in WH04, our pair of PC time series are also called the Real-time Multivariate MJO series 1 (RMM1) and 2 (RMM2). It should be noted that although the MJO dominates the RMM indices, some of the variance in actual MJO events may not be included directly in the indices (e.g., its seasonality and northward propagation in the Asian monsoon region during summer, etc.).

In the verification for the MJO forecasts, the observational data are the once daily at 00Z of the NCEP/NCAR reanalysis, that match the output time of the models. The same steps as described above are applied to both the observed and the forecast data to calculate the observed and forecast RMM1 and RMM2. In forecasts, when we remove the interannual variability part, we do not have 120 days of data. So we use the observations before the start of integration to replace those days missing, i.e., when removing the 120-day mean for the forecast at day n, the average of the $120 - n + 1$ days of observational data preceding the

forecast and the forecast data from day 1 to day $n-1$ is used. Both the observed and forecast RMM1 and RMM2 are normalized by the standard deviations of the observed RMM1 and RMM2.

5   Measures of skill

Temporal correlations between the forecasts and verification fields at each grid point are used to evaluate the general model forecast quality. Before doing so, the annual cycle and interannual variability are removed with the approach as described in the last section, to focus on the intraseasonal time scale. The analysis is confined to the global tropics (30°S-30°N), as the MJO occurs mainly in the tropical region.

To measure the forecast quality of the bivariate MJO index, three metrics are used: the correlation skill COR, the root mean squared error ($RMSE$), and the mean square skill score ($MSSS$).

1. Correlation skill

The correlation skill COR is defined as

$$COR(\tau) = \frac{\sum_{i=1}^{N} [a_{1i}(t)b_{1i}(t) + a_{2i}(t)b_{2i}(t)]}{\sqrt{\sum_{i=1}^{N} \left[a_{1i}^2(t) + a_{2i}^2(t)\right]}\sqrt{\sum_{i=1}^{N} \left[b_{1i}^2(t) + b_{2i}^2(t)\right]}}, \tag{1}$$

where $a_{1i}(t)$ and $a_{2i}(t)$ are the observed RMM1 and RMM2 at day t, and $b_{1i}(t)$ and $b_{2i}(t)$ are their respective forecasts, for the ith forecast with a τ day lead. N is the number of forecasts.

$COR(\tau)$ measures the skill in forecasting the phase of the MJO, which is insensitive to amplitude errors. $COR(\tau)$ is equivalent to a spatial pattern correlation between the observations and the forecasts when they are expressed by the two leading combined EOFs.

14

2. Root mean squared error

The root mean square error $RMSE$ can be written as

$$RMSE(\tau) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\{[a_{1i}(t) - b_{1i}(t)]^2 + [a_{2i}(t) - b_{2i}(t)]^2\}}. \tag{2}$$

It takes into account errors in both phase and amplitude. $RMSE$ is equivalent to the spatially averaged root mean squared difference between the observations and the forecasts when they are expressed by the two leading combined EOFs.

As RMM1 and RMM2 are normalized to their observed standard deviations, the saturated value of $RMSE$ when τ is so big that the forecast and observation are no longer correlated is 2, assuming the forecast RMMs have the same standard deviations as the observations. Similarly, for a climatological forecast ($RMM1 = 0$ and $RMM2 = 0$), $RMSE$ is $\sqrt{2}$.

3. Mean square skill score

The mean square skill score ($MSSS$) [$Murphy$, 1988] is defined as:

$$MSSS(\tau) = \left[1 - \frac{MSE_f(\tau)}{MSE_c}\right], \tag{3}$$

where

$$MSE_f(\tau) = \frac{1}{N}\sum_{i=1}^{N}\{[a_{1i}(t) - b_{1i}(t)]^2 + [a_{2i}(t) - b_{2i}(t)]^2\}$$

is the mean square error of the model forecast, and

$$MSE_c = \frac{1}{N}\sum_{i=1}^{N}\left[a_{1i}^2(t) + a_{2i}^2(t)\right]$$

is the climatological variance. $MSSS$ provides a relative skill of the MJO forecast compared to a climatological forecast that predicts no MJO signal. A perfect forecast ($MSE_f = 0$) has an $MSSS$ of 1, a forecast with an error as big as the climatological variance ($MSE_f = MSE_c$) has a zero $MSSS$, and when a forecast is doing worse than the climatological forecast a negative $MSSS$ is obtained.

6 MJO forecast skill in GEM and GCM3

As discussed in section 3, GEM simulates reasonably well the tropical wave spectral peaks, whereas in GCM3 the tropical waves cannot be easily separated from the background noise. Here we compare the forecast skill in these two models to see how the MJO forecast is influenced by the model climatological behavior.

6.1 Skill calculated at grid points

Two types of forecast skill are assessed: 1) the original daily forecasts against the total fields of observations; and 2) the 2-dimensional (2-D) MJO forecasts against the observed MJO signal. The purpose of the first type of skill is to assess the overall quality of the models in predicting tropical variabilities that includes all time scales shorter than a season, while the second is designed to concentrate on the MJO forecasts. The 2-D MJO signal is reconstructed based on a linear regression with the two leading PCs of the combined EOF. Taking u200 as an example, we first use the linear regression equation between the observed u200 and the two PCs to obtain the two regression coefficients at each grid point. Then, with these regression coefficients and RMM1 and RMM2 of the forecasts and observations, 2-D u200 fields can be constructed that contain the MJO signal. Temporal correlation between the forecast and the verification at each grid point is calculated for all the forecasts over the whole period (a total of 12 months $\times$35 years =420 forecasts).

Figure 7 illustrates the area averaged (30°S-30°N) correlation skill for PR, u850 and u200 by GEM and GCM3 as a function of forecast lead time. Shown in thin solid and dashed curves are the correlation skill of the total tropical flow by GEM and GCM3, respectively, whereas in thick solid and dashed lines are the MJO forecast skill by GEM and GCM3. As

described above, the total tropical flow here contains many scales of variability in addition to the MJO. It is clear that the forecast skill for the MJO is better than that for the tropical flow as a whole. The skill for the total flow varies significantly from one variable to another. PR has the lowest skill score, and the forecast skill for u200 is the best. By construction, the MJO forecast skill is the same for all the variables. We also note that GEM is doing somewhat better than GCM3 in forecasting both the total tropical flow and the MJO when the lead time is shorter than 15 days. After 15 days, there is indication that the correlation skill for the MJO forecast by GCM3 is higher than GEM. As will be discussed (in Fig. 10), in some areas the correlation skill in GCM3 lasts to a longer lead time).

The spatial distributions of correlation skill of the total tropical u200 field are presented in Fig. 8 for GEM and GCM3 at forecast lead times of 4.5, 7.5 and 10.5 days. To estimate the significance level for the correlation, we consider the fact that the correlation is calculated from all the 420 forecast cases that start from initial conditions one month apart. Assuming that the MJO signal has no autocorrelation at a lag of 75 days, there is one independent case for every 2.5 forecasts. Therefore, we have about 168 independent cases. According to a Student-t test, a correlation of 0.2 is required to pass the significance level of 0.01. The skill distribution in GEM is similar to that in GCM3. In general, the forecast skill for the total flow of u200 is lower in the tropics than in the middle latitudes, as is evident in the skill for lead times shorter than about a week (Fig. 8a, b, d and e). The synoptic-scale baroclinic eddies in the middle latitudes are better predicted than perturbations in the tropics. The lowest forecast skill is observed in the equatorial Indian ocean and western Pacific where the MJO is active. In the tropical region, the best forecast skill can be found over the eastern Pacific and tropical Atlantic. This is probably caused by the penetration of middle latitude baroclinic waves in these regions where climatological westerlies are located which is favourable for the

17

equatorward propagation of extratropical waves (e.g., Webster and Holton 1982; Hoskins and Yang 2000). On the other hand, in the tropical Indian ocean and western Pacific, upper easterlies constitute a barrier for tropics-extratropics interactions (e.g., Brunet and Haynes 1996) and thus almost no waves exist in these regions of extratropical origin which could provide a higher forecast skill. The spatial distribution of forecast skill of the total tropical flow for u850 is similar to that of u200, and the skill for the total field of PR is low (not shown).

Figure 9 shows the spatial distribution of forecast skill for the reconstructed u200 of the MJO. High correlation skill is found in the tropical Maritime continent longitudes, and over a large tropical area extending from the eastern Pacific eastward to the western Indian Ocean. Significant forecast skill over 0.5 is present in these regions even at a lead time of 10.5 days. Minimum skill is observed in the eastern Indian Ocean and western Pacific. Although the spatial distributions of skill in the two models are similar, GEM is doing noticeably better than GCM3. At the 4.5-day (7.5-day) lead time, GEM has a correlation skill greater than 0.6 (0.4) everywhere, while in GCM3 correlations less smaller than 0.6 (0.4) are found over the western Pacific, Africa and western Indian Ocean. At the 10.5-day lead time, bigger areas of correlation greater than 0.5 are observed in GEM than in GCM3. The skill distribution for the MJO reconstructed u850 (not shown) is similar to that of the u200. From the forecast skill for the reconstructed PR of the MJO (not shown), significant forecast skill in the tropical Indian Ocean is present in the GEM forecasts. Even at a lead time of 10.5 days, a large area of correlation skill greater than 0.5 is found in this region. GCM3, however, is less skillful. Almost no correlation larger than 0.5 can be found at a lead time of 10.5 days for GCM3.

To further investigate the tropical MJO forecast skill as a function of forecast lead time, plotted in Fig. 10 is the longitude-lead time distribution of correlation skill for the MJO

18

reconstructed u850, meridionally averaged between 15°S and 15°N. The longitudinal distribution of the forecast skill is consistent with Fig.9, where relatively high correlation skill is located in the Maritime continent longitudes, and over longitudes extending from the eastern Pacific eastward to the western Indian Ocean. Interestingly, there is an indication that the forecast skill over the Indian Ocean and Maritime continent area moves eastward with the increase of forecast lead time, from the eastern Indian Ocean at the beginning of forecast to near the dateline in about 20 days. GCM3 has a better skill after about two weeks. Near the Greenwich longitude and in the western Pacific region, correlation skill greater than 0.3 remains until a 24 day lead time in GCM3, while in GEM the same skill lasts about 20 days. This result seems consistent with that of Kharin and Zwiers (2004), who analyzed the forecast skill as a function of time scale in HFP1 using GCM2 and an earlier version of the numerical weather prediction model of the Canadian Meteorological Centre (SEF). They found that SEF produces better Z500 and T700 forecasts over Northern Hemisphere than GCM2 in the first one-two weeks whereas GCM2 performs slightly better at longer time leads. It was argued that the initial conditions are important in the first one-two weeks and the atmospheric response to boundary forcing becomes more dominant for longer time leads. In the present study, however, the interannual variability has been largely removed. The contribution from the atmospheric response to boundary forcing is not clear. It is more likely that GCM3 is able to predict well the pattern of some large amplitude MJO events, which contributes to an improved correlation skill. The eastward propagation of the forecast skill is likely to be related to some strong eastward-propagating MJO cases that are consistently predictable along their paths. It might also suggest that the models do not initiate new MJO events well. Similar propagation of forecast skill is found for the MJO u200, but less clear for the precipitation (not shown).

6.2 Skill of the MJO index

We now concentrate on the forecasts of the MJO bivariate index. The skill scores presented in this subsection are calculated for annual data (12 months), winter season of 6 months from November to April, and summer season of 6 months from May to October.

Shown in Fig. 11 is the correlation skill as defined in (1). The annual skill (Fig. 11a) is identical to the area averaged correlation skill of the 2-D MJO reconstructed fields (thick curves in Fig.7). In general, the correlation skill is higher in winter than in summer, which might be expected since the MJO is usually better organized and is stronger during boreal winter (e.g., Madden 1986). It is evident that GEM has a better forecast skill than GCM3 in the first 15 days, and the difference is larger in winter than in summer. If we take a correlation of 0.5 as the minimum for useful skill, GEM produces skillful forecasts upto 10 days in winter, while in GCM3 the skill drops below 0.5 at a lead time of 6 days. After 15 days, the correlation skill for the MJO forecasts by GCM3, which was designed as a climate model, is higher than those by GEM.

The results for the $RMSE$ as defined in (2) are presented in Fig. 12. As can be seen, GEM is superior in predicting the MJO to GCM3. The error grows much more slowly in GEM than in GCM3, reaching the saturated value at a lead time of about 20 days in GEM compared to about 10 days in GCM3. Again, the forecast skill is better in winter than in summer, and the difference between GEM and GCM3 is more obvious in winter. In winter forecasts, $RMSE$ reaches the climatological forecast value at 10 and 5 days for GEM and GCM3, respectively. Similar conclusions can also be drawn from the $MSSS$ results (not shown).

6.3 Skill stratified by the MJO strength

The amplitude of the MJO is defined as $amp = \sqrt{RMM1^2 + RMM2^2}$. In this section, we look at the dependence of forecast skill on the MJO amplitude in the initial conditions. Cases of strong and weak MJO forecasts are determined based on the observed MJO amplitude in the initial conditions. For a strong MJO, the initial MJO amplitude is greater than 1, whereas for a weak MJO the amplitude is smaller than 1. In all the 420 forecasts during the whole period, we have 250 strong MJO cases and 170 weak MJOs.

Figure 13 shows the correlation skill for the MJO index of forecast cases with strong and weak MJO signals at the initial state by GEM and GCM3. It is seen that for both models up to a forecast lead time of about 15 days the forecast skill is significantly higher when the initial condition has a strong MJO signal. After 15 days, the correlation skill for weak MJO cases seems to overpass that for strong MJOs. However, the usefulness of the MJO forecast beyond 15 days is questionable with such a low value of correlation skill. The same conclusion can be reached from the predictive skill measured by $MSSS$ (not shown).

The dependence of forecast skill on the MJO amplitude is consistent with some of the previous studies. For example, using the early National Meteorological Center (NMC) medium range forecasts, Chen and Alpert (1990) showed that the model forecast skill of the MJO reached about 10 days when the MJO amplitude was large. However, when the MJO amplitude was weak, the forecast skill was poor. Similar results were obtained in other studies (e.g., von Storch and Xu 1990; Jones et al. 2000).

In addition to the forecast skill of the MJO, we also look at the predictive skill of the total flow, i.e., including all time scales besides the MJO. This is to assess how the overall performance of the dynamical models in the tropics depends on the amplitude of the MJO.

Fig. 14 presents area averaged (30°S-30°N) correlation skill for total u200 field by GEM and GCM3 with strong and weak MJO signals at the initial forecast time. Interestingly, both models perform slightly better when there is a weak MJO signal in the initial condition, though its statistical significance is unclear. A strong initial MJO signal does not help to improve the overall forecast skill on the intraseasonal time scale in the tropics. Similar results are obtained for the skill of total u850 and PR fields (not shown).

The above result is clearly in contrast to forecasting the MJO signal itself where a better skill is achieved when the model is initialized with a strong MJO. Using five winters of the NCEP dynamical extended range forecasts, Hendon et al. (2000) analyzed the systematic forecast errors associated with strong MJO episodes. They found that the forecast skill is systematically reduced during active periods of the MJO compared to quiescent times. They attributed this to the failure of the dynamical model in capturing the eastward propagation of the tropical precipitation and circulation anomalies associated with the MJO when the forecast is initialized with an active MJO. Our results partly support their findings in that the overall forecast skill for the total tropical field is slightly reduced when the forecast is initialized with a strong MJO compared to a weak MJO. However, this reduction in overall skill is not caused by a deteriorated MJO forecast. On the contrary, in both of our models, the MJO is better predicted when a strong MJO signal exists in the initial condition. It is likely that tropical variabilities on time scales different from the MJO are less predictable when the MJO is active. It also implies that strong MJO episodes may be initialized from strong interactions with variabilities of shorter time scales that are less predictable, and that the MJO amplifies those less-predictable, shorter timescale activities.

6.4 Skill stratified by the MJO phase

The MJO phase space is defined in a similar way as in WH04. An MJO state at a given time is represented as a point in the two-dimensional phase space of RMM1 and RMM2. The distance of a point from the origin can be considered as the MJO amplitude, and an eastward propagation of the MJO is reflected by an anticlockwise movement in the phase space plot. As an example, plotted in Fig. 15 are the trajectories of the observed (solid line) and forecast (dashed line) MJO in the phase space for the 45 days starting from 1 December 1997. The MJO state at 00Z for each day is marked by a filled (empty) circle for the observations (forecast). The forecast is initialized at 12Z 30 November 1997.

To demonstrate the dependence of forecast skill on the phase of the MJO, presented in Fig. 16 are the skill scores of the MJO index for forecasts initialized with an MJO that is strong (amplitude greater than 1) and has a phase in one of the two phase groups: 1) phases 1, 2 or 3; 2) phases 4 to 8. An MJO in phase group 1 has an anomalous convective activity in tropical Africa and the Indian Ocean, while that in phase group 2 has enhanced convection in other tropical regions. There are 109 forecasts in phase group 1 and 145 in phase group 2 among all the forecasts during the whole period. It can be seen that both GEM and GCM3 have a better MJO forecast skill when the initial condition has an MJO signal with a phase in phase group 1 (i.e. phase 1, 2 or 3). To test the influence of the difference in sample sizes, a random selection of 109 forecasts from the 145 forecasts in phase group 2 was used to recalculate the skill. Almost the same result as Fig. 16 was obtained. It may not be surprising to find good forecast skill when the MJO is in tropical Africa and Indian Ocean since these are the regions where the MJO signal is strong, well developed, and has a relatively slow eastward propagating speed (e.g., Zhang 2005). However, the average initial

MJO amplitude for the selected forecasts in phase group 1 is 1.63, which is not dramatically larger than that for phase group 2, which is 1.58. This indicates that the skill dependence on the MJO phase is largely determined by the convection location.

The difference of correlation skill between phase groups 1 and 2 (left panel of Fig. 16) is larger for GCM3 than for GEM. For forecasts of phases 1-3 (4-8), GCM3 has a better (worse) correlation skill than GEM. As the correlation measures the skill in forecasting the pattern and is insensitive to amplitude errors, this indicates that GCM3 predicts better the pattern when it is initialized with an MJO of phases 1-3. As for the $MSSS$ (right panel of Fig. 16) that takes into account errors in both the pattern and amplitude, GEM has a better skill for both phase groups.

7   Summary and discussions

In this study we have examined the forecast skill of the MJO produced by two atmospheric models in the Canadian seasonal hindcast experiment HFP2, i.e., GCM3 and GEM. The overall skill for the MJO forecasts in the two dynamical models is at about the same level as obtained by previous studies (e.g., Chen and Alpert 1990; Jones et al. 2000; Hendon et al. 2000), with a skillful forecast up to about a one week lead time. This level of skill seems to be lower than that obtained with purely statistical models (e.g., Waliser et al. 1999). A significant difference in skill is found between these two models. The difference is larger in winter than in summer, when the GEM model produces skillful MJO forecasts to a lead time close to 10 days, while the same level of skill in GCM3 extends only to a lead time about 6 days.

As the same setup for the hindcast experiment was used in the two models, i.e., same

SST anomaly and same initial conditions, the difference in forecast skill is likely coming from that in the models' ability to sustain the intraseasonal variability. The model that forecasts the MJO with a better skill has a significantly better representation of the tropical convectively coupled waves (including the MJO itself) in its near climatological integration. This highlights the importance of improving numerical model's representation of tropical waves.

The connection between the tropical waves and the MJO is possibly through an upscale transfer of energy (e.g., Majda and Biello 2004). The present study shows that a numerical model with a better representation of the tropical waves does simulate and forecast the MJO better, in agreement with the upscale energy transfer hypothesis. On the other hand, the possibility can not be excluded that the characteristics of GEM allow the model to have a better representation of both the equatorial waves and the MJO without an upscale energy transfer. A further detailed analysis of the multiscale interaction in the tropics is necessary to fully understand the connection between the MJO and tropical waves.

It has been pointed out in many previous studies that convective parameterization in an atmospheric GCM is one of the most important factors determining whether a reasonable MJO variability can be simulated (e.g., Inness and Gregory 1997; Wang and Schlesinger 1999; Liu et al. 2005). It is possible that the different behavior in simulating the tropical waves and the MJO between GEM and GCM3 is related to different convective parameterization schemes utilized in these two models. Using the NCAR Community Climate Model version 3.6 (CCM3), which employs the deep convective parameterization of Zhang and McFarlane (1995), Maloney and Hartmann (2001) also found that the simulated MJO is much weaker than in the observations. Using the revised Zhang-McFarlane convection parameterization scheme which implements a new closure assumption and convection trigger, Zhang and Mu

(2005) were able to obtain an improved simulation of the MJO in CCM3. It is likely that the Kain-Fritsch (KF) convective parameterization implemented in GEM contributes to the reasonable simulation of the tropical waves and MJO variability. However, among many studies, an agreed result on how the MJO is sensitive to the convective parameterization scheme has not been achieved. For example, Slingo et al. (1996) suggested that convective schemes closed on buoyancy tend to produce better MJO signals than those closed on moisture convergence, whereas some other studies (e.g., Chao and Deng 1998; Lee et al. 2003) demonstrated that the moist convective adjustment (Manabe et al. 1965) scheme produces a stronger MJO variability. Wang and Schlesinger (1999) found that the simulated MJO becomes stronger when the relative humidity criterion for convection increases, but Maloney and Hartmann (2001) pointed out that this relationship does not hold in CCM3. Obviously, further studies are needed to clarify the sensitivity of the MJO to convective parameterization schemes.

In the HFP2, which was designed for seasonal forecast studies, the atmospheric initial conditions were taken from the NCEP/NCAR reanalysis. It can be expected that an increase of MJO forecast skill in GEM and GCM3 could benefit from an improved initialization. The skill could be further improved if a carefully designed ensemble approach were applied that takes into account the uncertainty in the initial conditions and model errors.

The lack of air-sea coupling in the HFP2 experiment may also affect the forecast skill of the MJO. Several observational studies have reported that the MJO oscillation is accompanied by significant changes in surface heat fluxes and SST (e.g., Zhang 1996; Woolnough et al. 2000). However, as demonstrated in most of these studies, the MJO circulation is likely a driving force for the intraseasonal variability in the SST. How the SST feeds back to the MJO and influences its forecast skill is less clear.

REFERENCES

Boer, G. J., N. A. McFarlane, R. Laprise, J. D. Henderson and J.-P. Blanchet, 1984: The Canadian Climate Centre spectral atmospheric general circulation model. *Atmosphere-Ocean*, **22**, 397–429.

Brunet, G., P. H. Haynes, 1996: Low-latitude reflection of Rossby wave trains. *J. Atmos. Sci.*, **53**, 482–496.

Chao, W. C., and L. Deng, 1998: Tropical intraseasonal oscillation, super cloud clusters, and cumulus convection scheme. Part II: 3D aquaplane simulations. *J. Atmos. Sci.*, **55**, 690–709.

Chen, T.-C., and J. C. Alpert, 1990: Systematic errors in the annual and intraseasonal variations of the planetary-scale divergent circulation in NMC medium-range forecasts. *Mon. Wea. Rev.*, **118**, 2607–2623.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998a: The operational CMC-MRB Global Environmental Multiscale (GEM) model: Part I - Design considerations and formulation. *Mon. Wea. Rev.*, **126**, 1373–1395.

——, J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998b: The operational CMC-MRB Global Environmental Multiscale (GEM) model: Part II - Results. *Mon. Wea. Rev.*, **126**, 1397–1418.

Derome, J., and Coauthors, 2001: Seasonal predictions based on two dynamical models. *Atmosphere-Ocean*, **39**, 485–501.

Flatau, M., P. J. Flatau, P. Phoebus, and P. P. Niiler, 1997: The feedback between equatorial convection and local radiative and evaporative process: The implications for

intraseasonal oscillations. *J. Atmos. Sci.*, **54**, 2373–2386.

Hendon, H. H., B. Liebmann, M. Newman, and J. Glick, 2000: Medium-range forecast errors associated with active episodes of the Madden-Julian oscillation. *Mon. Wea. Rev.*, **128**, 69–86.

Hoskins, B. J., and G.-Y. Yang, 2000: The equatorial response to higher-latitude forcing. *J. Atmos. Sci.*, **57**, 1197–1213.

Huffman, G. J., R. F. Adler, M. M. Morrissey, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at 1° daily resolution from multisatellite observations. *J. Hydrometeor.*, **2**, 36–50.

Inness, P. M., and D. Gregory, 1997: Aspects of the intraseasonal oscillation simulated by the Hadley Centre Atmosphere Model. *Climate Dyn.*, **13**, 441–458.

Jones, C., D. E. Waliser, J. K. Schemm, and K. M. Lau, 2000: Prediction skill of the Madden and Julian oscillation in dynamical extended range forecasts. *Climate Dyn.*, **16**, 273–289.

——, L. M. V. Carvalho, R. W. Higgins, D. E. Waliser, and J. K. Schemm, 2004: A statistical forecast model of tropical intraseasonal convective anomalies. *J. Climate*, **17**, 2078–2095.

Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining / detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteo. Soc.*, *77*, 437–471.

Kharin, V. V., and F. W. Zwiers, 2004: Skill as a function of time scale in ensembles of seasonal hindcasts. *Climate Dyn.*, **17**, 127–141.

Lau, K. M., and P. H. Chan, 1986: Aspects of the 40-50-day oscillation during the northern summer as inferred from outgoing longwave radiation. *Mon. Wea. Rev.*, **114**, 1354–1367.

——, and T. J. Phillips, 1986: Coherent fluctuations of extratropical geopotential height and tropical convection in intraseasonal time scales. *J. Atmos. Sci.*, **43**, 1164–1181.

Lee, M.-I., I.-S. Kang, and B. E. Mapes, 2003: Impacts of cumulus convection parameterization on aqua-planet AGCM simulations of tropical intraseasonal variability. *J. Meteor. Soc. Japan*, **81**, 963–992.

Lin, H., J. Derome and G. Brunet, 2005: Tropical Pacific link to the two dominant patterns of atmospheric variability. *Geophys. Res. Lett.*, **32**, L03801, doi:10.1029/2004GL021495.

——, G. Brunet, and J. Derome, 2007: Seasonal forecasts of Canadian winter precipitation by post-processing GCM integrations. *Mon. Wea. Rev.*, in press.

Lin, J.-L., and Coauthors, 2006: Tropical intraseasonal variability in 14 IPCC AR4 climate models: Part I: Convective signals. *J. Climate*, **19**, 2665–2690.

Lindzen. R. S., 1967: Planetary waves on beta planes. *Mon. Wea. Rev.*, **95**, 441–451.

Liu, P., B. Wang, K. R. Sperber, T. Li, and G. A. Meehl, 2005: MJO in the NCAR CAM2 with the Tiedtke convective scheme. *J. Climate*, **18**, 3007–3020.

Lo, F., and H. H. Hendon, 2000: Empirical extended-range prediction of the Madden-Julian oscillation. *Mon. Wea. Rev.*, **128**, 2528–2543.

Manabe, S., J. Smagorinsky, and R. F. Strickler, 1965: Simulated climatology of a general circulation model with a hydrological cycle. *Mon. Wea. Rev.*, **93**, 769–798.

Madden, R. A., 1986: Seasonal variations of the 40-50 day oscillation in the tropics, *J.*

Atmos. Sci., **43**, 3138–3158.

——, and P. R. Julian, 1971: Description of a 40-50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708.

Majda, A. J., and J. A. Biello, 2005: A new multiscale model for the Madden-Julian oscillation. *J. Atmos. Sci.*, **62**, 1694–1721.

Maloney, E. D., and D. L. Hartmann, 2001: The sensitivity of intraseasonal variability in the NCAR CCM3 to changes in convective parameterization. *J. Climate*, **14**, 2015–2034.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Meteor. Soc. Japan*, **44**, 25–43.

Matthews, A. J., B. J. Hoskins, and M. Masutani, 2004: The global response to tropical heating in the Madden-Julian Oscillation during Northern winter. *Quart. J. Roy. Meteor. Soc.*, **130**, 1991-2011.

McFarlane, N.A., G. J. Boer, J.-P. Blanchet and M. Lazare, 1992: The Canadian Climate Centre second-generation general circulation model and its equilibrium climate. *J. Climate*, **5**, 1013–1044.

Mo, K. C., 2001: Adaptive filtering and prediction of intraseasonal oscillations. *Mon. Wea. Rev.*, **129**, 802–817.

Moncrieff, M. W., 2004: Analytic representation of the large-scale organization of tropical convection. *J. Atmos. Sci.*, **61**, 1521–1538.

Murphy, A. H., 1998: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.*, **116**, 2417–2424.

North, G. R., T. L. Bell, R.F., Calahan, and F. J. Moeng, 1982: Sampling errors in the

estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699–706.

Reichler, T., and J. Roads, 2005: Long-range predictability in the tropics. Part II: 30-60-day variability. *J. Climate*, **18**, 634–649.

Sheng, J., 1995: The Madden-Julian oscillation in the Canadian Climate Centre general circulation model. *Clim. Dyn.*, **12**, 125–140.

Slingo, J. M., and Coauthors, 1996: Intraseasonal oscillations in 15 atmospheric general circulation models: Results from an AMIP diagnostic subproject. *Clim. Dyn.*, **12**, 325–357.

Shukla J. S., and Coauthors, 2000: Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.*, **81**, 2593–2606.

van Den Dool, H. M., 1994: Long-range weather forecasts through numerical and empirical methods. *Dyn. Atmos. Oceans.*, **20**, 247–270.

von Storch, H., and J. Xu, 1990: Principal Oscillation Pattern analysis of the tropical 30-60 day oscillation. Part I: Definition of an index and its prediction. *Clim. Dyn.*, **4**, 179–190.

Waliser, D. E., C. Jones, J. K. Schemm, and N. E. Graham, 1999: A statistical extended-range tropical forecast model based on the slow evolution of the Madden-Julian oscillation. *J. Climate*, **12**, 1918–1939.

——, K. M. Lau, W. Stern, and C. Jones, 2003: Potential predictability of the Madden-Julian oscillation. *Bull. Amer. Meteor. Soc.*, **84**, 33–50.

——, and Coauthors, 2006: The experimental MJO prediction project. *Bull. Amer. Meteor. Soc.*, **87**, 425–431.

Wang, W., and M. E. Schlesinger, 1999: The dependence on convection parameterization of the tropical intraseasonal oscillation simulated by the UIUC 11-layer atmospheric GCM. *J. Climate*, **12**, 1423–1457.

Webster, P. J., and J. R. Holton, 1982: Cross-equatorial response to middle-latitude forcing in a zonally varying basic state. *J. Atmos. Sci.*, **39**, 722–733.

Wheeler, M., and G. N. Kiladis, 1999: Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain. *J. Atmos. Sci.*, **56**, 374–399.

——, and K. Weickmann, 2001: Real-time monitoring and prediction of modes of coherent synoptic to intraseasonal tropical variability. *Mon. Wea. Rev.*, **129**, 2677–2694.

——, and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932.

Woolnough, S. J., J. M. Slingo, and B. J. Hoskins, 2000: The relationship between convection and sea-surface temperature on intraseasonal timescales. *J. Climate*, **13**, 2086–2104.

Zhang, C., 1996: Atmospheric intraseasonal variability at the surface in the tropical western Pacific Ocean. *J. Atmos. Sci.*, **53**, 739–758.

——, 2005: Madden-Julian Oscillation. *Rev. Geophys.*, **43**, RG2003, doi:10.1029/2004RG000158.

Zhang, G. J., and N. A. McFarlane, 1995: Sensitivity of climate simulation to the parameterization of cumulus convection in the Canadian Climate Centre General Circulation Model. *Atmos.-Ocean*, **33**, 407–446.

——, and M. Mu, 2005: Simulation of the Madden-Julian Oscillation in the NCAR CCM3 using a revised Zhang-McFarlane convection parameterization scheme. *J. Climate*, **18**, 4046–4064.
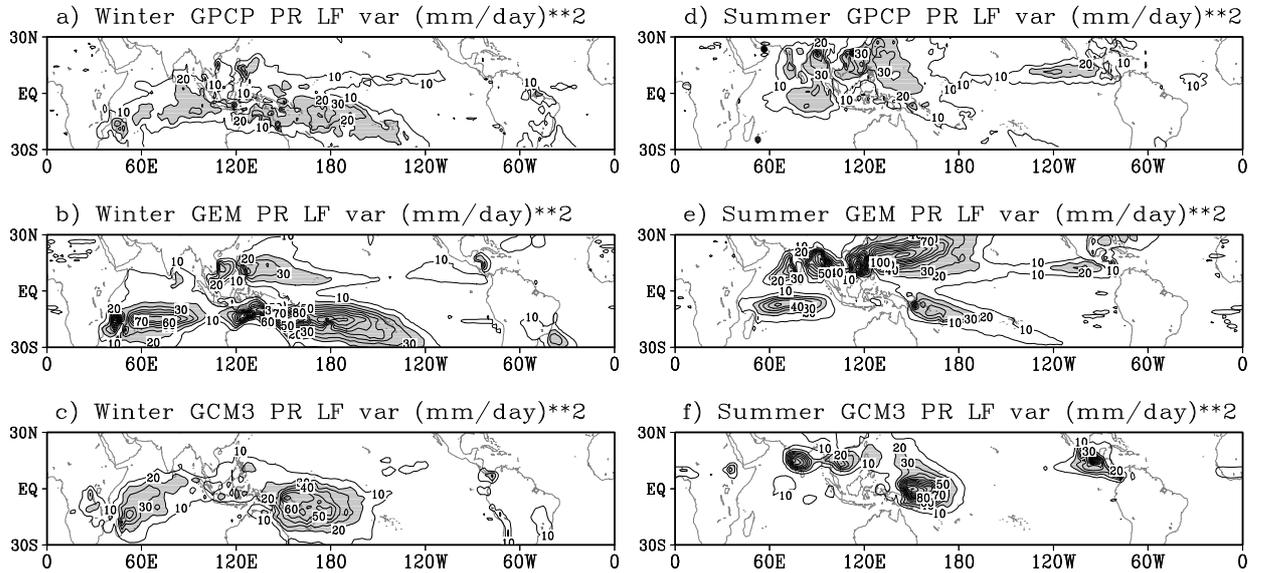
List of Figures

Figure 1: Variance of low-frequency (20-90 days) variability of u200 for NCEP reanalysis, GEM and GCM3. Left panels are for boreal winter, and the right panels for boreal summer. The contour interval is 8 m^2/s^2. Areas with variance greater that 24 m^2/s^2 are shaded.

Figure 2: Variance of low-frequency (20-90 days) variability of u850 for NCEP reanalysis, GEM and GCM3. Left panels are for boreal winter, and the right panels for boreal summer. The contour interval is 3 m^2/s^2. Areas with variance greater that 6 m^2/s^2 are shaded.

Figure 3: Variance of low-frequency (20-90 days) variability of precipitation rate for GPCP data, GEM and GCM3. Left panels are for boreal winter, and the right panels for boreal summer. The contour interval is 10 mm^2/day^2. Areas with variance greater that 20 mm^2/day^2 are shaded.
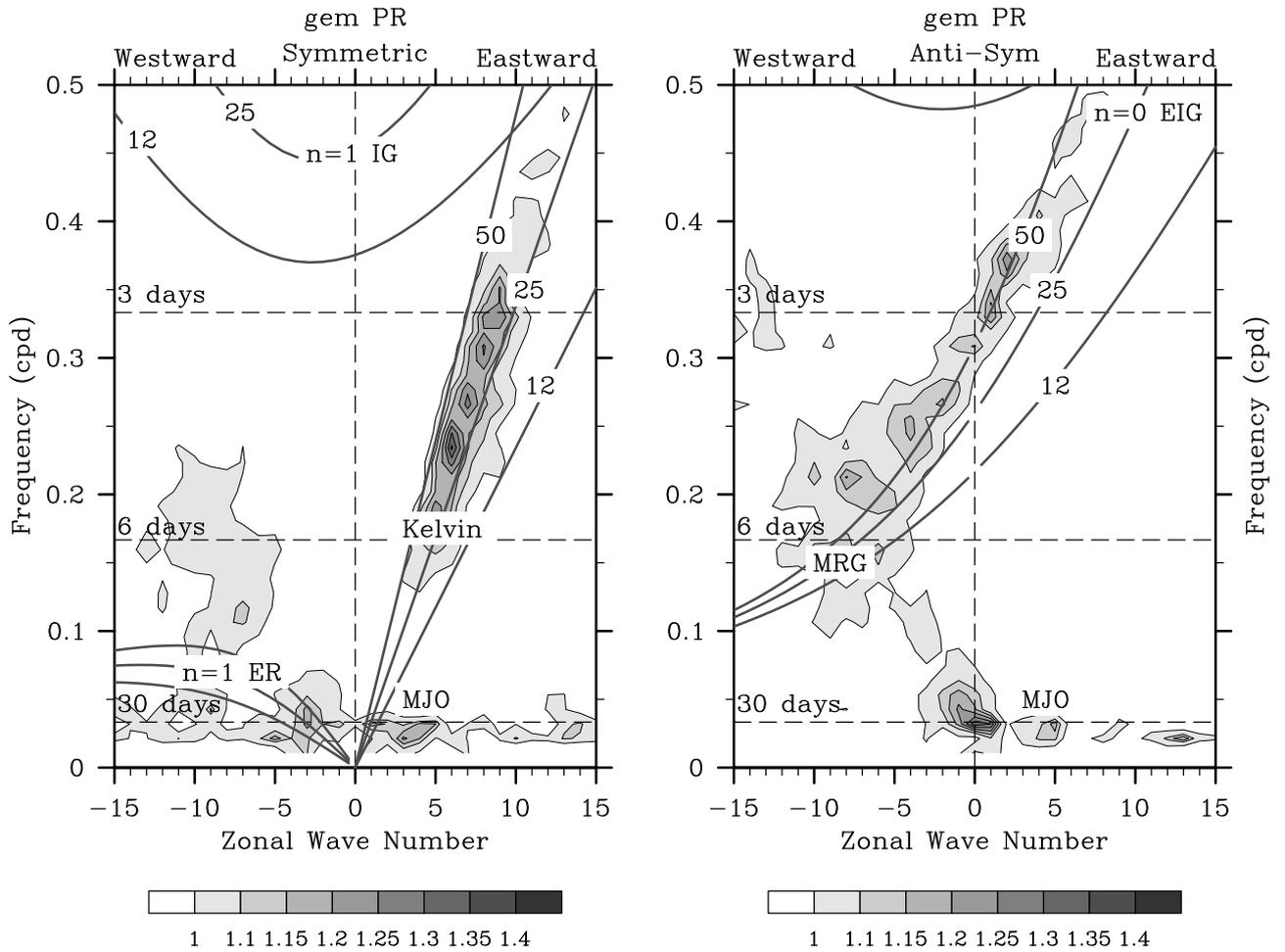
Figure 4: Left (right) panel is wavenumber-frequency spectrum of 15°S-15°N symmetric (antisymmetric) component of precipitation rate divided by the background spectrum for GEM. Superimposed are the dispersion curves of the odd (even) meridional mode numbered equatorial waves for the equivalent depths of $h = 12$, 25 and 50 m.
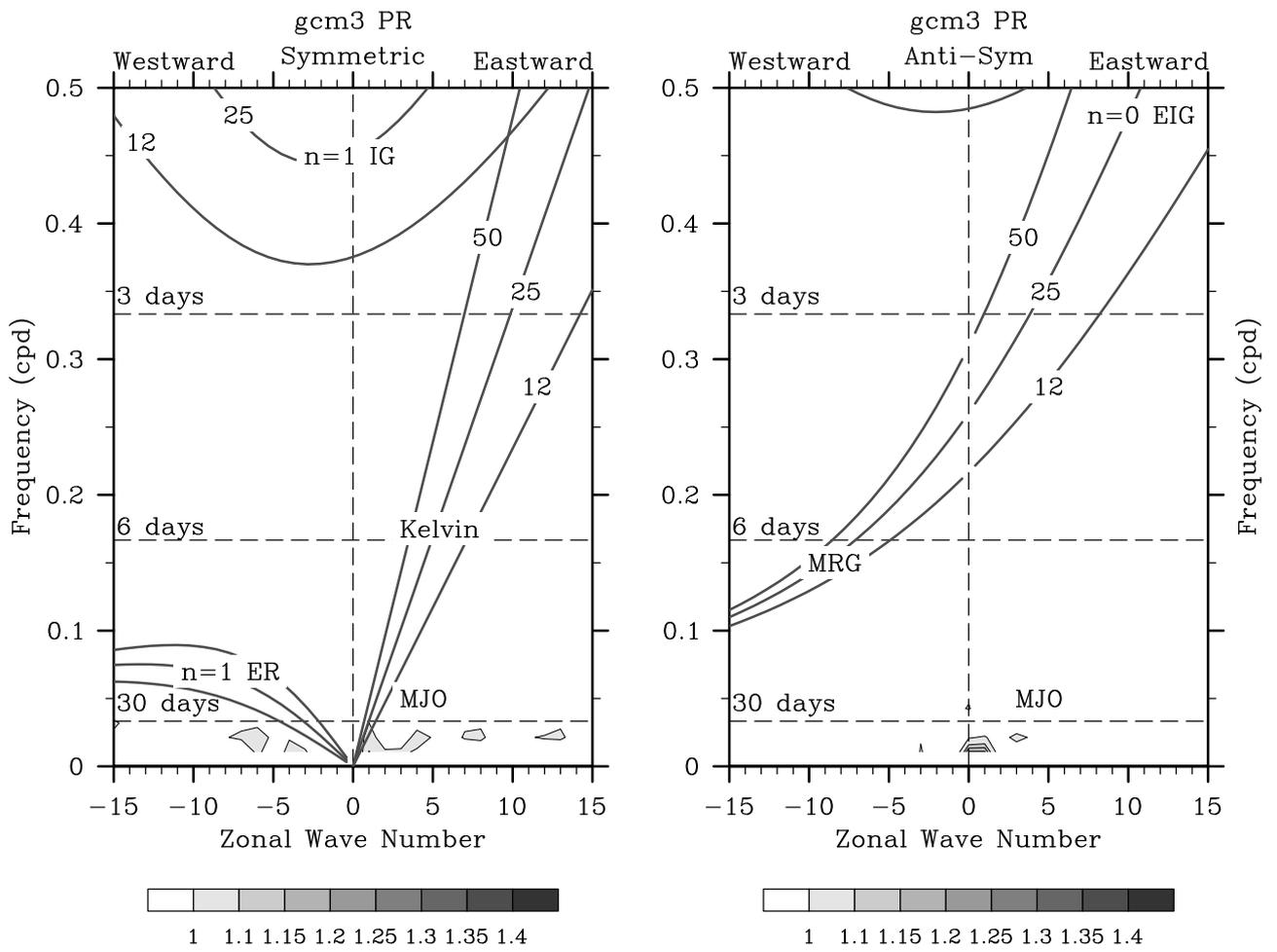
Figure 5: Same as Fig. 4, except for GCM3.

41
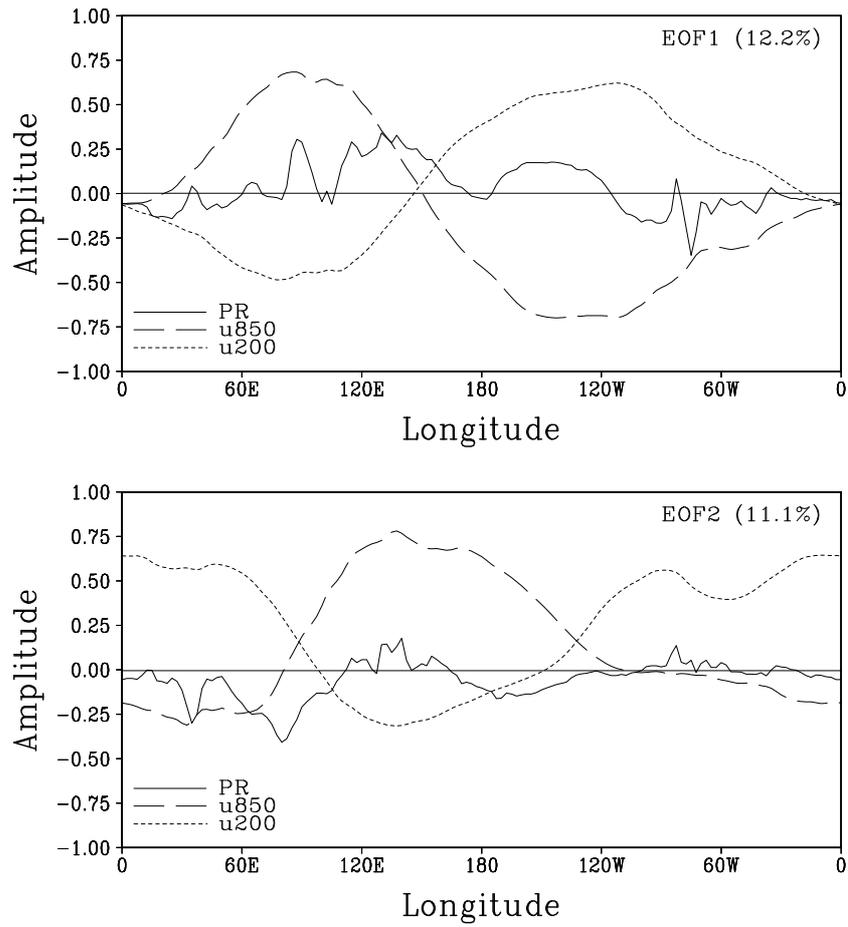
Figure 6: Longitudinal distributions of EOF1 and EOF2 of the combined analysis of PR, u850 and u200.
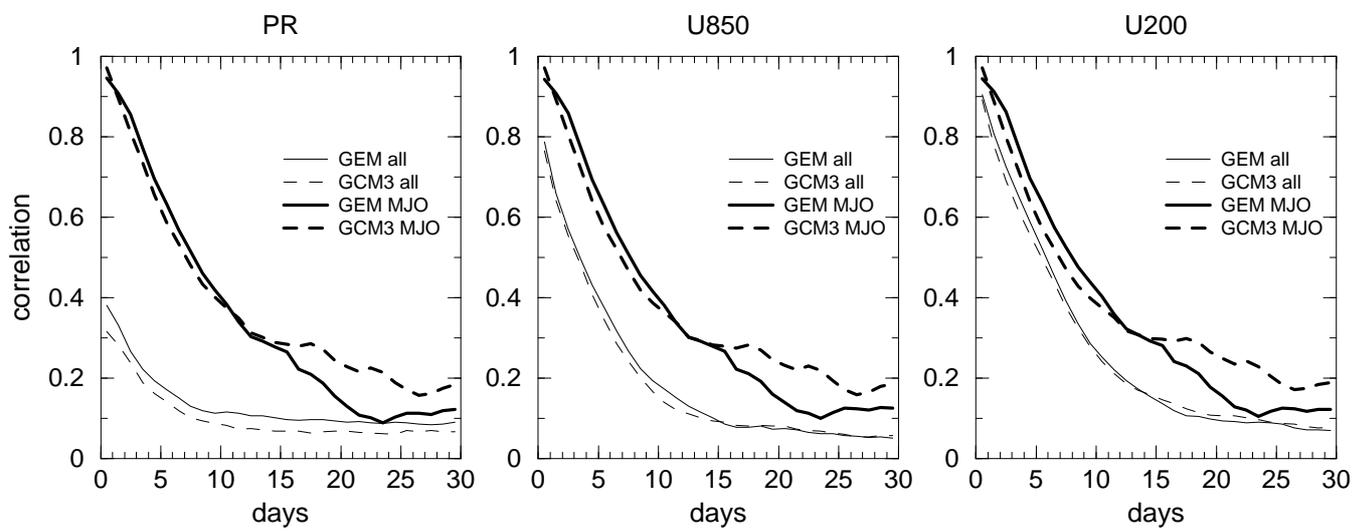
Figure 7: Area averaged anomaly correlation coefficient between the forecasts and the observations for PR, u850 and u200. Solid curves are for GEM, and dashed curved for GCM3. Thick curves are for MJO forecasts, whereas thin curves for forecasts of total flow.
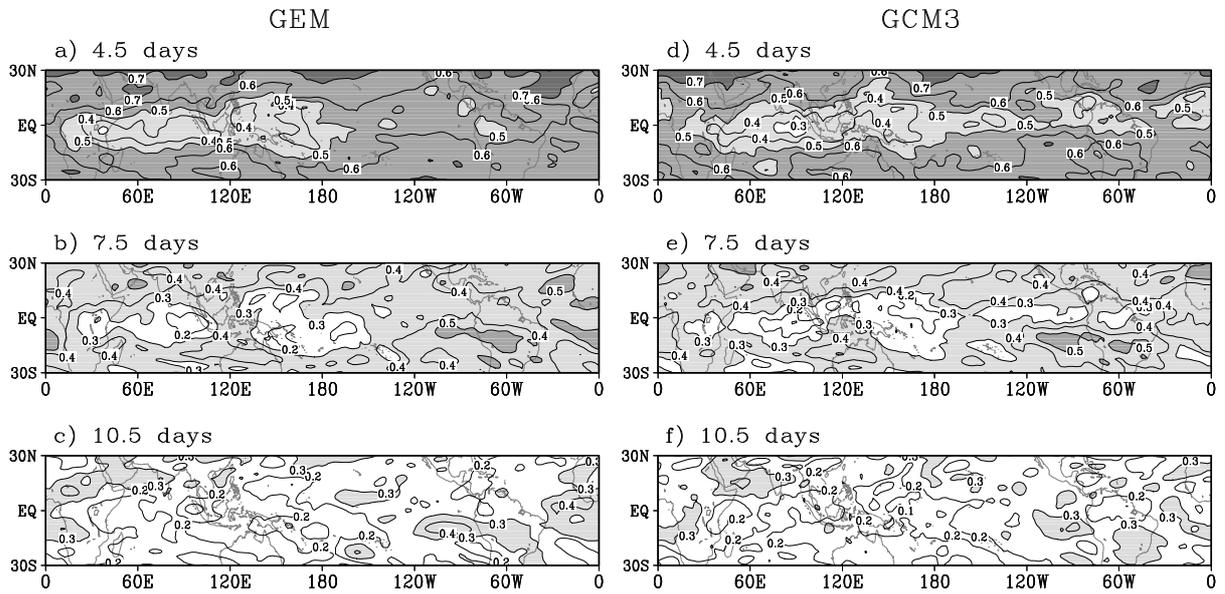
Figure 8: Temporal correlation between the total field of forecast u200 and that of the observations at lead times of 4.5, 7.5 and 10.5 days for GEM (left panels) and GCM3 (right panels). Light shading areas are for correlations greater than 0.3, medium shading areas for those greater than 0.5, and heavy shading areas for correlations greater than 0.7.

44
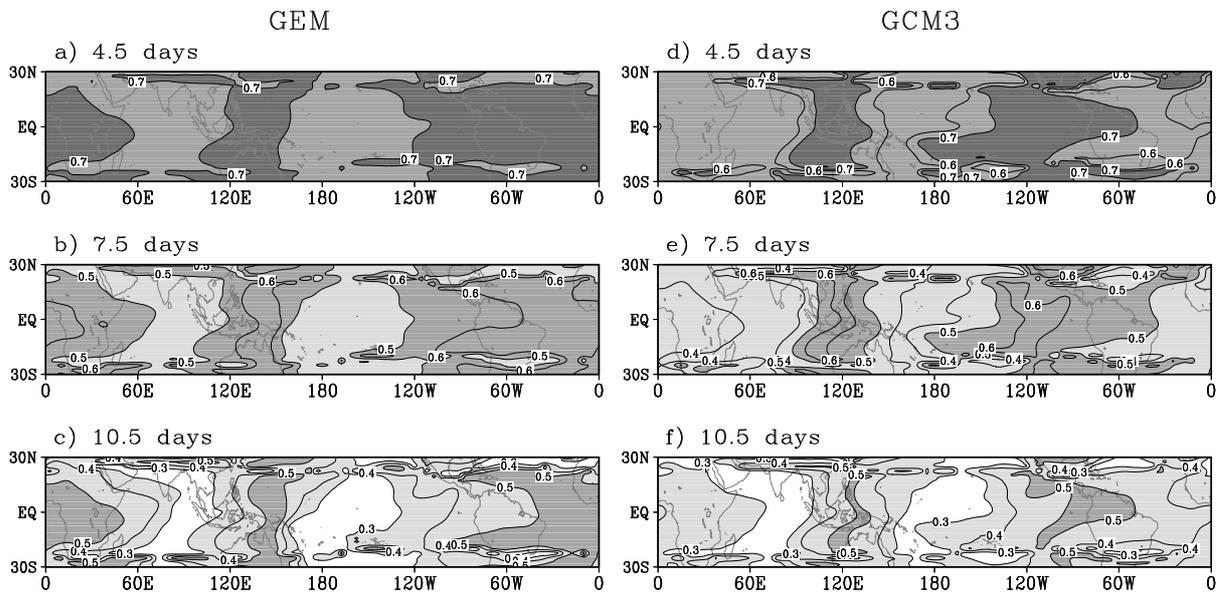
Figure 9: Same as Fig. 8, except for forecasts of the MJO u200.
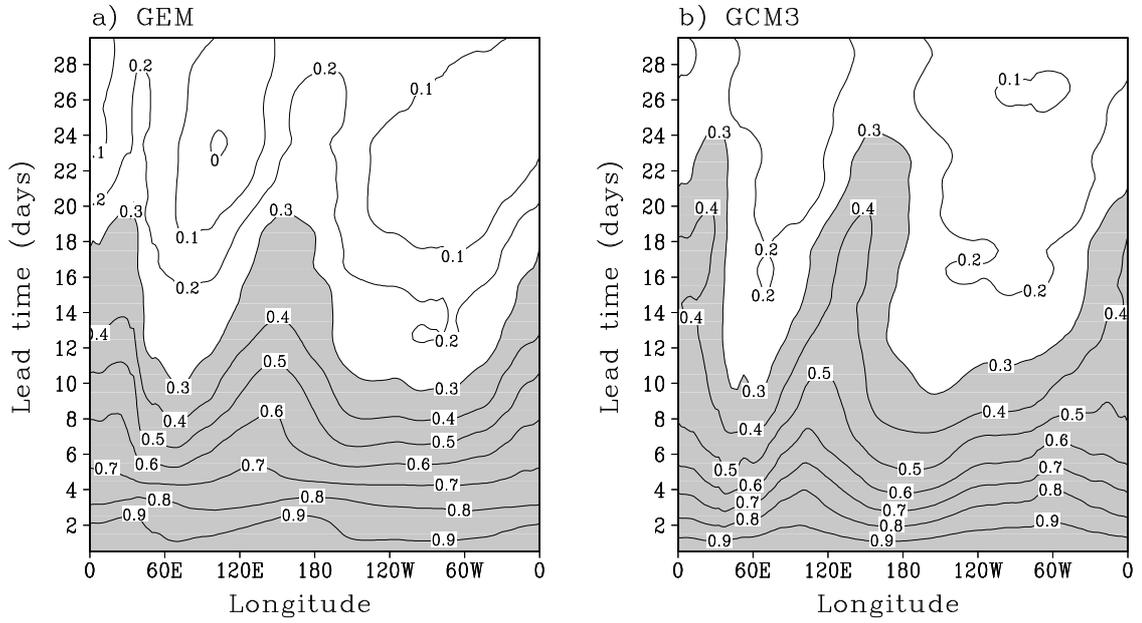
Figure 10: Lead time-longitude distribution of 15°S-15°N averaged correlation between the MJO reconstructed forecast u850 and that of the observations for a) GEM; b) GCM3. The contour interval is 0.1. Areas with correlation greater than 0.3 are shaded.
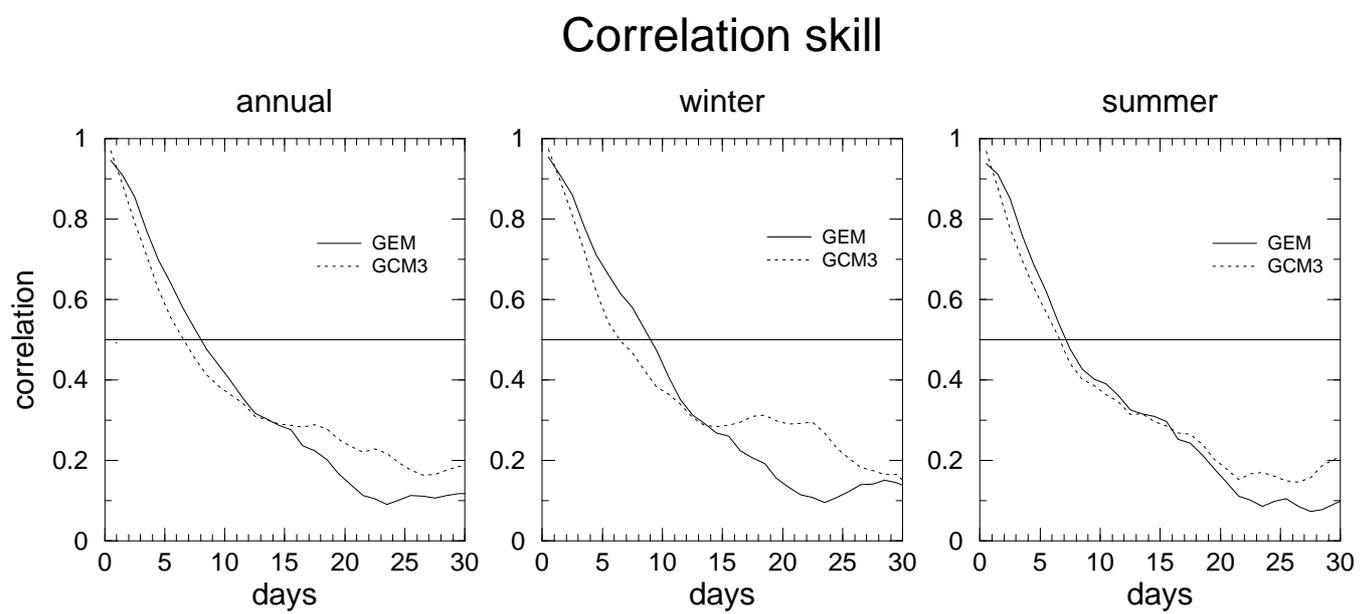
Figure 11: Correlation skill for the bivariate MJO index. Solid curves are for GEM, and dashed curves for GCM3. The horizontal line represents correlation skill of 0.5.
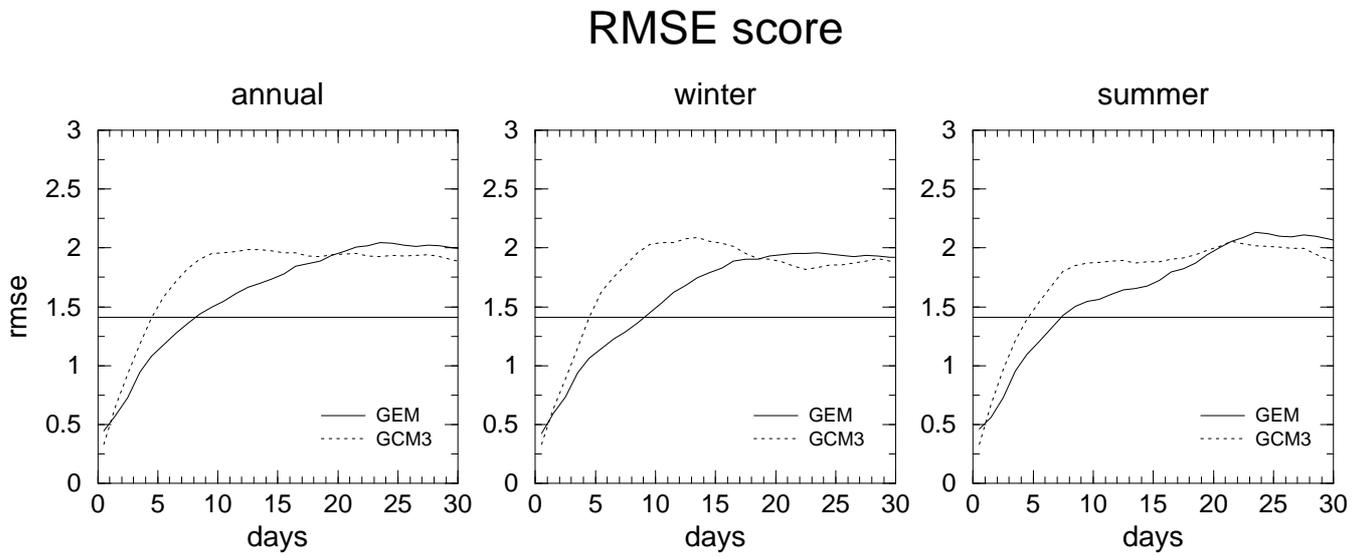
47

RMSE score

Figure 12: RMSE skill for the bivariate MJO index. Solid curves are for GEM, and dashed curves for GCM3. The horizontal line represents RMSE skill of $\sqrt{2}$.
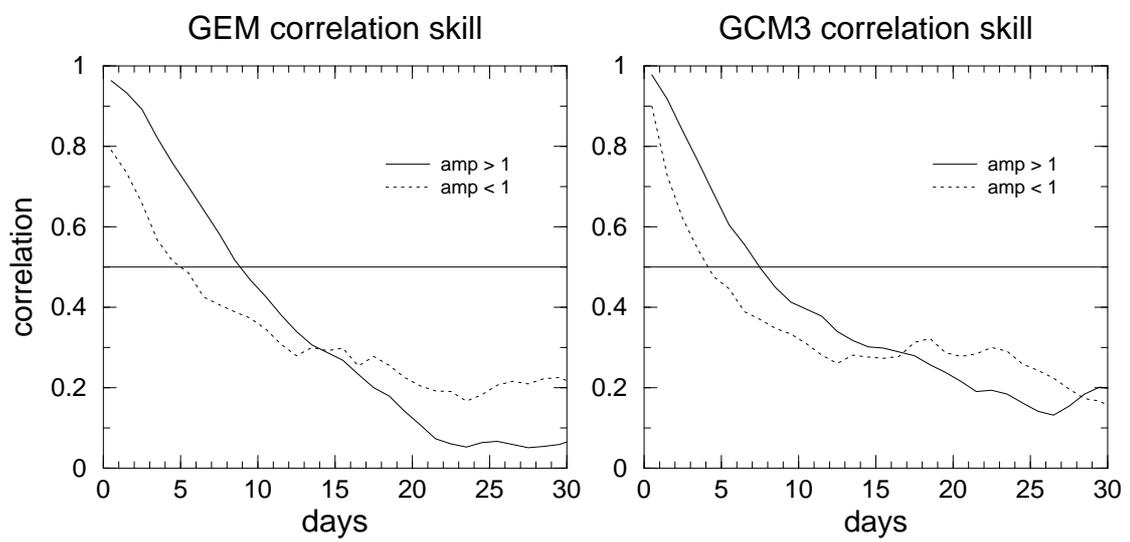
Figure 13: Correlation skill of the bivariate MJO index for strong MJOs (solid curves) and weak MJOs (dashed curves).
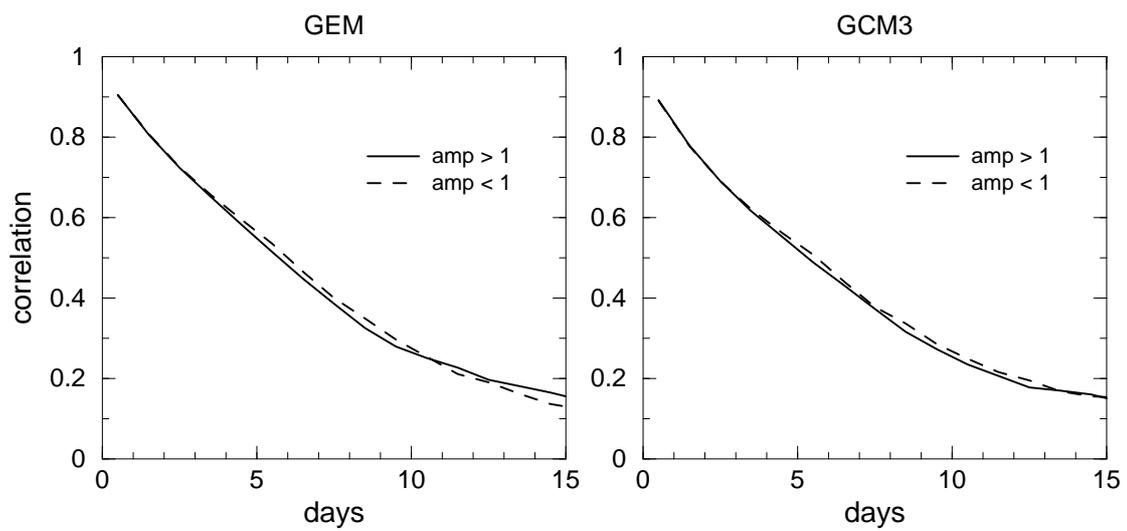
Figure 14: Area averaged anomaly correlation coefficient between the total field of forecast u200 and that of the observations for strong MJOs (solid curves) and weak MJOs (dashed curves).
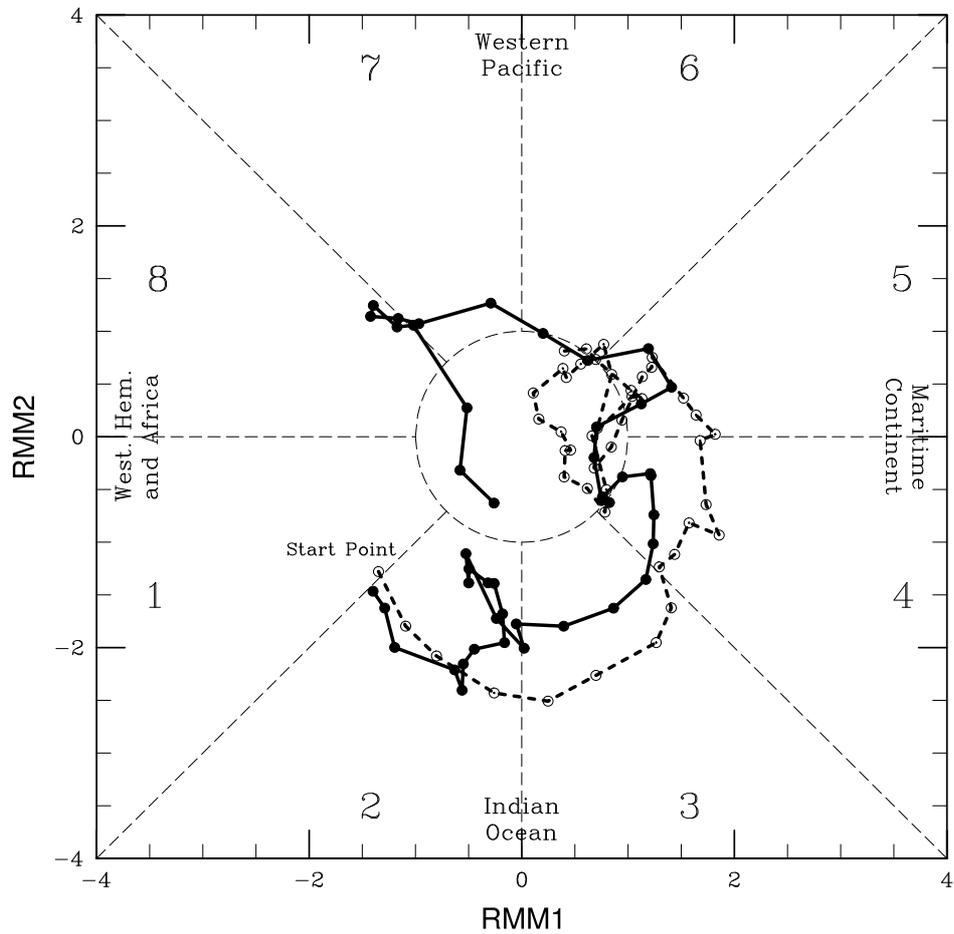
Figure 15: Trajectories of the observed (solid curve with closed circles) and forecast (dashed curve with open circles) MJO in the RMM1-RMM2 phase space at 00Z daily during 45 days starting from 1 December 1997. The forecast is initialized at 12Z 30 November, 1997. Eight defined regions of the phase space are marked, along with the region of weak MJO.
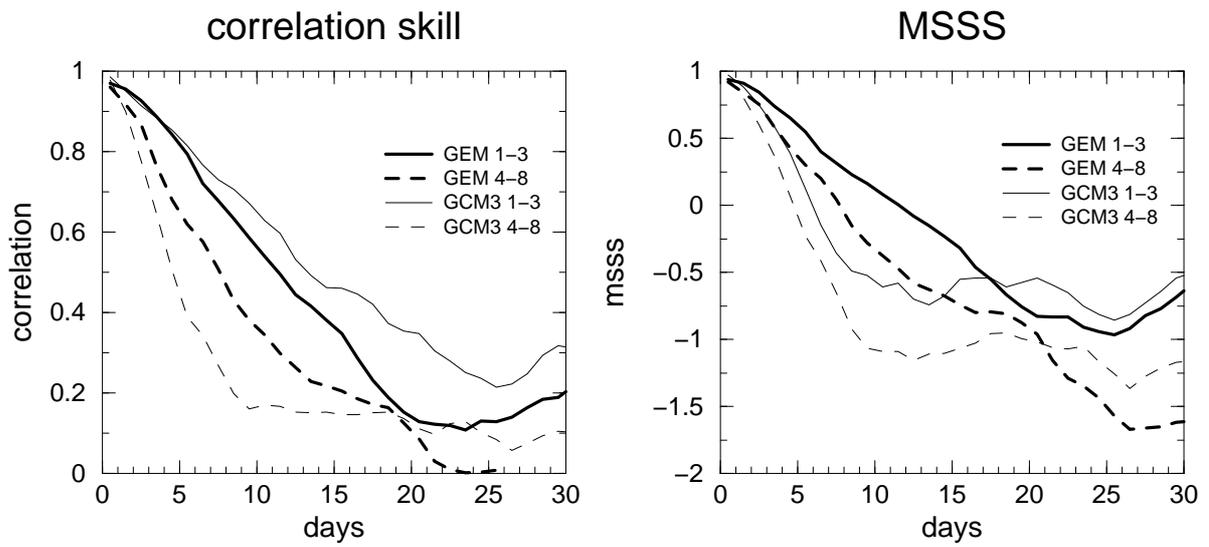
Figure 16: Left panel: correlation skill of the bivariate MJO index for MJOs in phase space regions 1, 2 and 3 (solid curves) and in phase space regions 4-8 (dashed curves). Thick curves are for GEM and thin curves for GCM3. Right panel: Same as in the left panel, except for MSSS skill.