

# Real-time localised forecasting of the Madden–Julian Oscillation using neural network models

Barnaby S. Love<sup>a,\*</sup> and Adrian J. Matthews<sup>a,b</sup>

<sup>a</sup>*School of Environmental Sciences, University of East Anglia, UK*

<sup>b</sup>*School of Mathematics, University of East Anglia, UK*

**ABSTRACT:** Existing statistical forecast models of the Madden–Julian Oscillation (MJO) are generally of very low order and predict the evolution of a small number (typically two) of principal components (PCs). While such models are skilful up to 25 days lead time, by design they only predict the very largest-scale features of the MJO. Here we present a higher-order MJO statistical forecast model that is able to predict MJO variability on smaller, more localised scales, that will be of more direct benefit to national weather agencies and regional government planning. The model is based on daily outgoing long-wave radiation (OLR) data that are intraseasonally filtered using a recently developed technique of empirical mode decomposition that can be used in real time. A standard truncated PC analysis is then used to isolate the maximum amount of variance in a finite number of modes. The evolution of these modes is then forecast using a neural network model, which does not suffer from the parametrisation problems of other statistical forecast techniques when applied to a higher number of modes. Compared to a standard 2-PC model, the higher-order PC model showed improved skill over the whole MJO domain, with substantial improvements over the western Pacific, Arabian Sea, Bay of Bengal, South China Sea and Phillipine Sea. Copyright © 2009 Royal Meteorological Society

**KEY WORDS** MJO; OLR; principal components; statistical forecasting

*Received 18 December 2008; Revised 21 May 2009; Accepted 22 May 2009*

## 1. Introduction

The Madden–Julian Oscillation (MJO) is the dominant mode of intraseasonal tropical convective variability (Madden and Julian, 1994; Zhang, 2005). It has a significant influence on precipitation patterns over the tropical Indian ocean, the Maritime Continent and the Western Pacific warm pool region, in addition to influencing a number of other atmospheric phenomena such as the Asian monsoon (Goswami, 2005) and El Niño (McPhaden, 1999). The MJO can essentially be characterized as an eastward propagation of tropical deep convective precipitation anomalies over the warm pool from the equatorial Indian Ocean over the Maritime Continent into the western Pacific region with one complete cycle of the MJO lasting between 30 and 60 days. The ability to accurately forecast such a significant tropical mode as the MJO is not only important in its own right but will be crucial to the success of medium- to extended-range numerical weather forecasts (Hendon *et al.*, 2000).

Numerical weather prediction (NWP) models are not attaining the useful skill we might expect in predicting the MJO (Jones *et al.*, 2000; Waliser, 2005), though recently there have been some improvements (Woolnough *et al.*, 2007). These errors in NWP models, associated with the poor representation of the MJO, are a significant limiting

factor in the forecast quality and range of synoptic forecasts in the Tropics. However, the current limitations are a symptom of the numerical model's inability to correctly represent deep tropical convection, rather than the reaching of some intrinsic limit of predictability of the MJO. Ferranti *et al.* (1990) and Hendon *et al.* (2000) have demonstrated that the medium- to extended-range forecast skill of numerical models in the Tropics can be significantly improved if errors associated with the representation of the MJO are eliminated.

With the problems encountered by NWP models, alternative approaches to representing and forecasting the MJO have been sought; one such approach has been the use of statistical models. Much of the work using statistical models has focused on forecasting principal component (PC) time series generated from the leading empirical orthogonal functions (EOFs) of intraseasonally filtered outgoing long-wave radiation (OLR) and/or streamfunction data. Several statistical models have been developed that have demonstrated useful forecast skill of these leading EOF patterns out to lead times of up to 25 days (Waliser *et al.*, 1999; Jones *et al.*, 2004; Love *et al.*, 2008). Furthermore, skilful forecasts of tropical cyclone activity can also be made, based on statistical forecasts of the MJO (Leroy and Wheeler, 2008). Waliser *et al.* (1999) and Jones *et al.* (2004) used variations of multiple linear regression, while Love *et al.* (2008) used an auto-regressive, moving-average approach to forecast the PC time series. This leading EOF methodology is

\*Correspondence to: Dr Barnaby Love, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK.  
E-mail: b.love@uea.ac.uk

based on the widely accepted assumption that the MJO is adequately represented by the first two EOFs of OLR data (Salby and Hendon, 1994; Wheeler and Hendon, 2004).

(The method of Wheeler and Hendon (2004) differs from the method presented here. Given that meridional variability dominates the seasonal cycle of the MJO, Wheeler and Hendon (2004) use meridional averaging to remove this seasonality and represent the MJO with just two indices. They include seasonal variability at the presentation stage by projecting the indices onto seasonally varying regression maps. Our method has no meridional averaging, thus the seasonal meridional variation of the MJO is retained through the inclusion of the higher-order EOFs. We also project onto seasonally varying regression maps at the presentation stage.)

The aim of this paper is not to necessarily contradict this assumption, but to create models that have more degrees of freedom than the basic leading 2-PC model. To this end, this study will produce forecast models containing multiple-PC time series with the aim of encapsulating the large-scale organized behaviour of the MJO as the leading 2-PC model does, but with a much greater spatial resolution that will allow us to produce more detailed forecasts. In this regard, the models have some similarity to linear inverse model techniques (e.g. Alexander *et al.*, 2008). Such models could be used as a pure MJO forecast or as an input or boundary condition in another model and would be of greater practical use to national weather agencies and national and regional governments.

The statistical models we will use are neural network models; these models have been used extensively in a number of fields to learn complex nonlinear patterns in large sets of data. They are strongly nonlinear, and are potentially a significant advancement on the linear or partially nonlinear, multiple linear regression and autoregressive moving average (ARMA) models (Box and Jenkins, 1970) that have been employed in previous studies. They are also able to handle large numbers of inputs and outputs without suffering significant degradation in model accuracy, making them ideally suited to this kind of experiment. In contrast, ARMA-type models become more difficult to consistently parametrise as the number of inputs and outputs increases.

## 2. Data processing

### 2.1. Real-time intraseasonal data filtering

Producing real-time MJO forecasts with statistical models requires the use of either a projection onto indices or a real-time intraseasonal filtering methodology. In this study we use the empirical mode decomposition (EMD) methodology (Huang *et al.*, 1998) developed in Love *et al.* (2008). This method is able to efficiently extract the MJO signal while preserving nonlinearities in the input data and without introducing large unwanted end effects. The methodology is discussed comprehensively in Love *et al.* (2008), so we include only a brief review here.

EMD works on the assumption that the raw data consist of a number of simple, intrinsic modes of oscillation. These intrinsic mode functions (IMFs) have a simple empirical definition. If  $x(t)$  is a time series of raw data at time  $t$ , then two cubic spline-fitted functions can be found that pass through all the local maxima and local minima in this time series, respectively. The mean function  $m_1(t)$  of these two spline-fitted functions can then be calculated. The first IMF  $e_1(t)$  is then equal to the difference between the raw time series and the mean function  $m_1(t)$  of the two cubic spline-fitted functions.

$$e_1(t) = x(t) - m_1(t). \quad (1)$$

The mean function is then recycled and becomes the raw data for the calculation of the second IMF. Maximum and minimum spline-fitted functions are again fitted and the mean function  $m_2(t)$  calculated. This mean function is then subtracted from the input time series to give the second IMF,

$$e_2(t) = m_1(t) - m_2(t). \quad (2)$$

The remainder  $m_2(t)$  then once more becomes the input data for the next cycle, and so on, until the remaining data are either a constant or a simple monotonic function with some additional insignificant noise.

An IMF is a simple oscillatory mode that consists of the locally highest frequencies of the time series input to each decomposition, but its form is more general than a normal oscillatory mode and can have an amplitude and frequency that vary continuously in time. A consequence of the EMD method selecting the locally highest-frequency components from a time series to create each IMF is that the broadband MJO signal tends to be split between the first three IMFs. In addition, the basic method produces IMFs which have non-negligible end effects, so some empirical adaptations to the EMD process are required.

These empirical adaptations aim to isolate the MJO signal in a single IMF, typically using temporal running means and changes in the data time interval (i.e. daily to pentad means) of the input data, and by adjusting the criteria for identifying the local maxima and minima (i.e. over nearest neighbours or over next-nearest neighbours, etc.). The end effects are reduced by increasing the length of the record of extrema over which the cubic spline-fitted functions are fitted, therefore mitigating against the sensitivity at the end point of a cubic spline.

### 2.2. Data

This real-time intraseasonal filtering methodology is applied to the interpolated  $2.5^\circ$  longitude by  $2.5^\circ$  latitude gridded set of daily means of OLR (Liebmann and Smith, 1996), in a region from  $40^\circ\text{E}$  to  $240^\circ\text{E}$  and from  $25^\circ\text{N}$  to  $25^\circ\text{S}$ , representing the tropical regions of major MJO activity. The annual cycle at each grid point (calculated from the mean and the leading three annual harmonics) was subtracted from the raw data to produce daily

anomaly maps. This satellite-measured dataset is a good proxy for precipitation in the Tropics, and has been used many times as a measure of the state of the MJO (e.g. Waliser *et al.*, 1999; Lo and Hendon, 2000; Matthews, 2008).

The dataset runs for 27 years, from 1 January 1979 to 31 December 2005. The data were split into an 18.5-year dependent dataset from 1 January 1979 to 30 June 1997, and an 8.5-year independent-validation dataset from 1 July 1997 to 31 December 2005. This adapted-EMD methodology was then used at each grid point independently, in each of the datasets. This is a different approach from the single methodology applied at all points used in Love *et al.* (2008) and is taken because the focus of this study is to forecast the MJO with a high degree of spatial resolution. However the principles developed in Love *et al.* (2008) that are used to calculate the most appropriate empirical adaptations remain the same.

From the empirical mode decompositions, the IMFs containing the MJO signal are identified and averaged into pentad means. All the grid point data were then combined to make 18.5 years of dependent pentad maps of MJO data and 8.5 years of independent-validation pentad maps of MJO data. At this point, the first 12 and last 6 months of the dependent dataset and the first 6 and last 12 months of the independent-validation dataset were discarded because of the end effects from the EMD analysis. This was so they did not influence the model parametrisation in the case of the dependent dataset, and so that they did not affect the verification of forecasts in the case of the independent-validation dataset. For the actual forecasts, only data up to the respective forecast date were used in the filtering process so the forecasts are considered to be true forecasts. Finally, the dependent dataset was divided into a 12-year dependent-training dataset (1 January 1980 to 31 December 1991) and a 5-year dependent-testing dataset (1 January 1992 to 31 December 1996).

The effectiveness of the adapted-EMD method as an intraseasonal filter was illustrated in Figure 5 of Love *et al.* (2008), by comparing two-dimensional space–time power spectra of 30–70-day conventional bandpass-filtered OLR data and of the adapted-EMD filtered OLR data. They were very similar, indicating the effectiveness of the EMD method as an intraseasonal filter.

### 2.3. Principal component time series

The statistical models used PC time series that were calculated from an EOF analysis performed on the adapted-EMD filtered dependent-training dataset over the domain and periods used in filtering and stated in the previous section. The PC time series used in each of the models, along with the proportion of the total intraseasonal variance they account for, is shown in Figure 1. Although previous MJO statistical forecast models have tended to use the leading two PCs, this only accounts for 16.5% of the variance. This rises rapidly to 37.9% when the leading ten PCs are included, and then

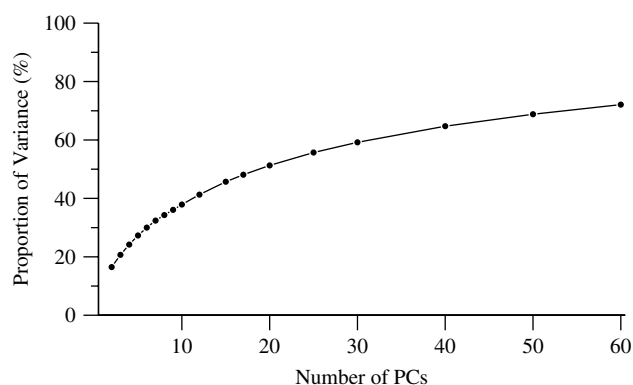


Figure 1. Proportion of variance (%) accounted for, as a function of the number of PCs used in the neural network models. The first point is the 2-PC case.

more slowly to 72.1% by 60 PCs. There are 1701 grid points in the domain, so eventually 100% of the variance is accounted for by all 1701 PCs.

## 3. Neural network modelling

### 3.1. Methodology

Real-time forecasts (hindcasts) of each set of time series can be made using neural network statistical models. Modelling using neural networks is a large and diverse field, with a wide variety of different types of neural network and different developmental methodologies. The networks used here are layered feed forward networks (Rosenblatt, 1962). These are chosen because the process of learning patterns in the input data is relatively simple, making them ideal for model development.

The network consists of an input layer where input data patterns are presented to the network, a single ‘hidden’ layer (so called because it provides a hidden connection between the inputs and outputs) and an output layer where the final network outputs are calculated. Theoretically the network can have any number of these hidden layers but only one is necessary as any continuous valued function can be represented by a three-layer network with a single hidden layer of nonlinear neurons, given enough neurons in the hidden layer (Cybenko, 1989). While this may require an infinite number of neurons in practice, limiting the neural network to a single layer on the basis of this theory is a concise way of reducing the vast number of potential options available when deploying neural network models.

The architecture of the networks thus takes the general form shown in Figure 2. The model uses data from the last 5 pentads; this time lag was identified as optimum for including all information pertinent to the forecasts in Love *et al.* (2008), and was also used by Jones *et al.* (2004). Therefore for the  $N = 5$  time series model, there would be  $I = 25$  inputs in the input layer (values from the five time series, at 1–5 pentad lags), a hidden layer (number of units ( $J$ ) identified in section 3.3) and  $K = 25$  outputs in the output layer (values for the five time series, for 1–5 forecast pentads).

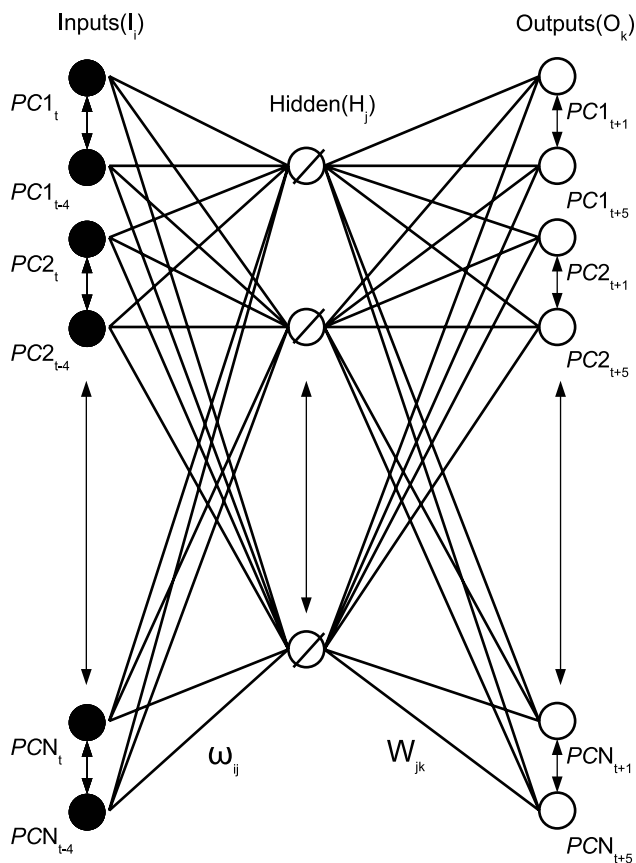


Figure 2. Network architecture of the multiple-PC neural network models.  $N$  is the number of selected time series, the subscript  $t$  indicates the current time,  $t - \alpha$  is  $\alpha$  pentads previously and  $t + \alpha$  is the  $\alpha$ -pentad forecast.

The values of the network parameters  $\omega_{ij}$  and  $W_{jk}$ , shown in Figure 2, are calculated using a supervised learning process in conjunction with a back-propagation method (Rumelhart *et al.*, 1986a,b) which is described here with reference to Figure 2. In a supervised learning process, a portion of the dependent-training dataset is divided into input and correct output data pairs, which are presented iteratively to the network. Initially a set of inputs  $I$  are fed forward through the network by multiplication with the input to hidden layer parameters  $\omega_{ij}$ , which are initially assigned small random values. Thus, the hidden units receive the inputs,

$$H_j^{\text{Input}} = \sum_{i=1}^I \omega_{ij} I_i. \quad (3)$$

At the hidden layer, a nonlinear activation function is applied. The activation function used here is a sigmoid function (Figure 3). Therefore, the hidden units produce the outputs,

$$H_j^{\text{Output}} = g(H_j^{\text{Input}}) = \frac{1}{1 + \exp(-H_j^{\text{Input}})}. \quad (4)$$

The values of these hidden layer neurons are then fed forward by multiplication by the hidden-to-output layer

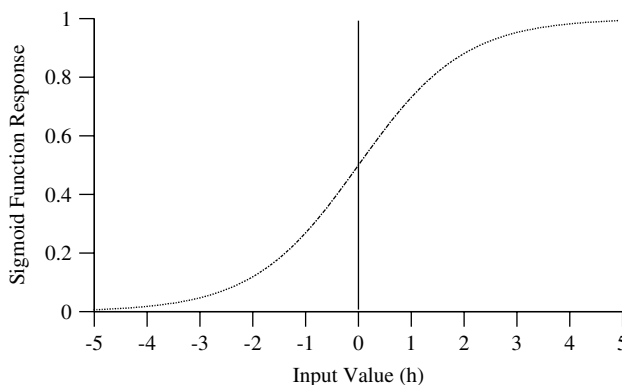


Figure 3. The sigmoid activation function.

parameters  $W_{jk}$  to create the values of the output layer units

$$O_k = \sum_{j=1}^J W_{jk} H_j^{\text{Output}}. \quad (5)$$

This is the feed forward process; the next stage is the back propagation part of the learning method. Back propagation is essentially a form of gradient descent modified to work with multi-layer nonlinear networks. It gets its name from the manner in which errors between forecast output and observed output are propagated back through the network and are subsequently used to make adjustments to the network parameters. To calculate these errors we use the error measure

$$E = \frac{1}{2} \sum_{p=1}^P \sum_{k=1}^K (\zeta_k^p - O_k^p)^2, \quad (6)$$

where  $\zeta_k^p$  is the correct or observed output  $k$  for pattern  $p$ , and the squared errors are summed over  $P$  patterns (days on which the forecast is initiated). Substituting the network output into the error measure equation yields

$$E = \frac{1}{2} \sum_{p=1}^P \sum_{k=1}^K \left[ \zeta_k^p - \left\{ \sum_{j=1}^J W_{jk} g \left( \sum_{i=1}^I \omega_{ij} I_i^p \right) \right\} \right]^2. \quad (7)$$

As this is a continuously differentiable function of all its parameters, we are able to use the gradient descent algorithm to propagate the errors back through the network and consequently calculate error-related modifications to the network parameters  $\omega_{ij}$  and  $W_{jk}$ . Through the continual presentation of these dependent-training dataset pairs, the parameters are forced to incrementally converge to a solution. During this process, the decrease in network error can be monitored for progress and signs of overfitting as the parameter updates are allowed to adapt to the local gradient of the network error. Before we consider the results of the neural network statistical models, a brief description of the practical process of parametrising them is required. All of the values of the variables considered were established in a pragmatic manner from practical experimentation with the models.

### 3.2. Practical application

The network error is calculated after the presentation of every 73 input/correct-output data pairs (one year's worth of data). If the network error is less than all of those previously calculated, then the current set of parameters is stored as the optimal set of model parameters.

Due to the high number of degrees of freedom in a neural network, noise is added to the dependent-training data by selecting each input/correct-output pair from the dependent-training dataset randomly with replacement. This ensures that the solution found is a global rather than a local minimum. To ensure the generalisation of the forecast model, the network is tested for overfitting by calculating the network errors of the dependent-training and dependent-testing datasets after a minimum of 10 epochs of data have been presented to the network. An epoch is defined as the number of presentations of the input/correct-output patterns that equals the number of pairs in the dependent-training dataset, in this case 876. The two network errors are then normalised by dividing them by the number of patterns in each set and then compared as a ratio (dependent-testing divided by dependent-training). If this figure exceeds 1.05 twice during the presentation of the data, then the network is considered to be overfitting, the pattern presentation process is stopped, and the final set of saved parameters are selected for the finished model.

Overfitting may not necessarily occur, so an upper limit on the capacity of the model to run is set at 1000 epochs. In addition, if the model has run for a number of epochs equal to or greater than the number taken to reach its last network error minimum with no further improvement, then the parametrisation process is terminated to prevent unnecessary computation.

### 3.3. Number of units in the hidden layer

While the number of units in the input and output layers are fixed variables, the number of units in the hidden layers of each network is a variable that can be adjusted so as to most accurately represent the function being modelled whilst retaining model parsimony. The most appropriate number for each network was identified by parametrisation each network with a range of numbers of units in the hidden layer. For each of the multiple-PC models (Figure 1), 20 networks were created, with 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 17, 20, 22, 25, 27, 30, 32, 35 and 40 units in the hidden layer, respectively. For each of these networks, the network error of the final parametrised model was plotted against the number of units in the hidden layer. An example of such a graph for the PC 1–7 time series model is shown in Figure 4. The network error can be seen to decrease as the number of units in the hidden layer increases, it then begins to asymptote before rising again as the model begins to overfit the data. The optimum number of units in the hidden layer can then be identified at the point the graph appears to asymptote, judged to be at approximately 8 units.

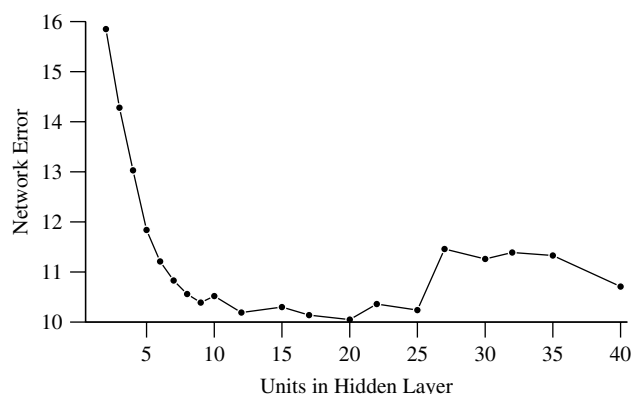


Figure 4. Network error as a function of the number of units in the hidden layer for the 7-PC model.

This is quite a conservative choice; we could easily choose to have more units in the hidden layer. However, to ensure model parsimony and prevent overfitting, the number of units in the hidden layer and consequently the complexity of the model is kept to a minimum. This choice is related to the decision not to remove irrelevant network connections after parametrisation through the use of 'pruning' techniques (Karnin, 1990). Pruning is a common practice in neural network modelling, but experimentation showed no significant advantages in its use with these relatively simple models if the choice of the number of units in the hidden layer is made conservatively.

## 4. Model forecast results

### 4.1. Individual model results

The parametrised networks were then used to produce real-time forecasts (hindcasts) of the MJO from the independent-validation dataset. Forecasts were made of the sets of PCs, which were then projected onto sets of regression maps. These regression maps were previously calculated from the dependent-training dataset, between maps of the total intraseasonally filtered anomaly field and the individual PC time series making up the model. Because of the strong seasonal cycle of the MJO, separate regression maps were calculated for each calendar month (i.e. 12 in total for each PC). The relevant regression maps from the month in which the forecast was valid were then selected, multiplied by the amplitude of their respective forecast PCs and summed to make the forecast regression map. A similar verification regression map was created by regressing the observed values of the PCs onto the regression maps and summing them.

To ensure that any forecast skill statistics calculated are representative of intraseasonal behaviour, a mask was applied to the maps prior to their calculation. This mask was created by calculating the variance of the adapted-EMD filtered dependent-training dataset over the whole domain (Figure 5). Any grid points that had an intraseasonal variance less than  $100 \text{ W m}^{-2}$  were masked out, as the model was judged to be forecasting noise there.

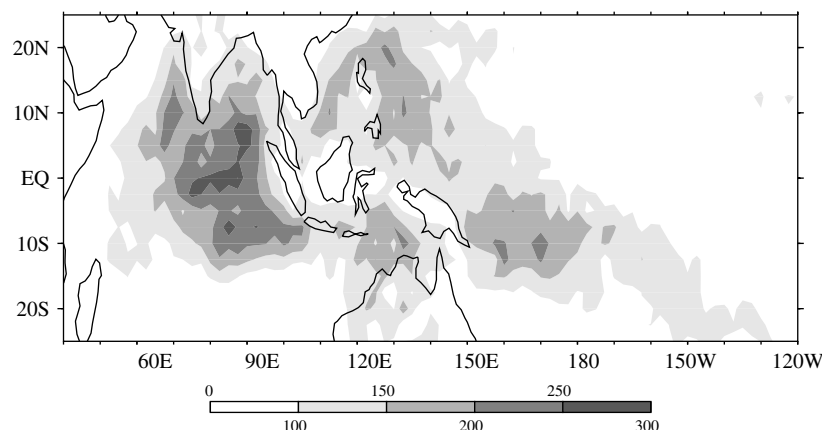


Figure 5. Variance of  $2.5^\circ$  gridded adapted-EMD filtered OLR data over the MJO's principal region of influence. The contour interval is  $50 \text{ W}^2\text{m}^{-4}$ . See legend for shading. Grid points where the variance is under  $100 \text{ W}^2\text{m}^{-4}$  (i.e. no shading) will be masked.

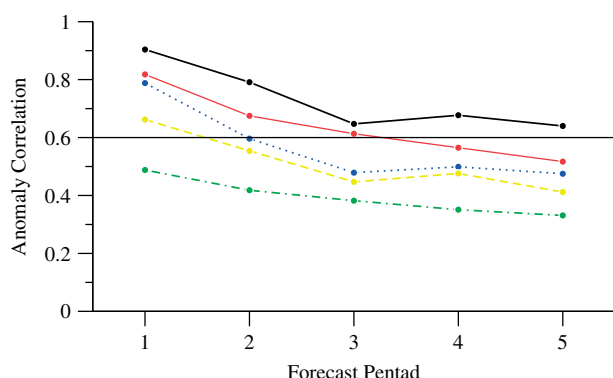


Figure 6. The median of the distribution of anomaly correlations between forecast regression maps and verification regression maps of real-time data as a function of forecast lead time, for the multiple-PC models containing 2 (bold solid), 5 (thin solid), 8 (dotted), 15 (dashed) and 25 (dot-dashed) PCs. This figure is available in colour online at [www.interscience.wiley.com/journal/qj](http://www.interscience.wiley.com/journal/qj)

The pattern correlation was then calculated between each pair of forecast regression maps and verification regression maps. These pattern correlations were then ordered, and the median and upper and lower quartiles calculated. The pattern correlation distributions from the multiple-PC models were then compared to that of the leading 2-PC model (Figure 6). The leading 2-PC neural network model has almost identical skill to the leading 2-PC VARMA model in Love *et al.* (2008). Following Wilks (2005), useful skill is defined here as a median pattern correlation above 0.6. Using this global measure of skill, the leading 2-PC model appears to be the best model, with useful forecast skill out to a lead time of 5 pentads. The skill of the multiple-PC models gradually decreases as the number of PCs included increases, resulting in a shortening of the forecast lead time over which useful skill is displayed (down to 3 pentads for the PCs 1–5 model). The large spread in model performance should also be noted; the upper quartile forecasts for all the models give excellent forecasts (pattern correlations above 0.85) out to a lead time of 5 pentads, while the lower quartile forecasts do not have useful forecast skill at even one pentad lead time.

However, the important question that arises from these results is whether the reductions in forecast skill, when the multiple-PC models are compared to the leading 2-PC model, is due to a poorer representation of the MJO, or simply because they contain a greater number of degrees of freedom in both the forecast regression maps and verification regression maps. Therefore to make a true comparison of the models, they all need to be compared to the same, standard reference dataset.

#### 4.2. Comparison to total anomaly field

The models can be compared by using the total anomaly field obtained from the adapted-EMD filtered OLR data as a standard, reference dataset. While this dataset does contain much information that is not traditionally defined as being part of the MJO (the leading 2-PC definition accounts for only 16% of the variance), it allows for a fair verification process.

Pattern correlations were calculated between forecast regression maps and the total anomaly field maps for each model and are shown in Figure 7. The values of the pattern correlations are lower than what would normally be considered to be a skilful forecast, but are adequate for the purpose of comparing the forecasts.

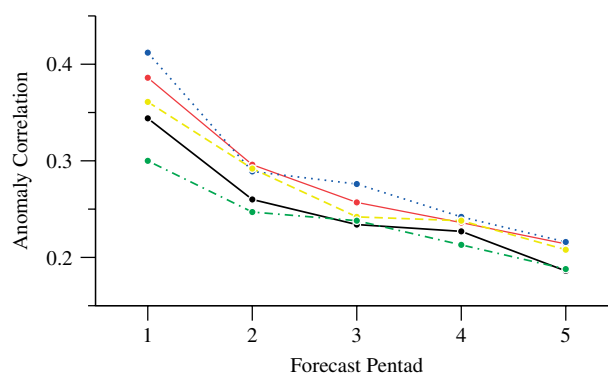


Figure 7. As Figure 6, but for the anomaly correlations between the forecast regression maps and the total anomaly field verification maps. This figure is available in colour online at [www.interscience.wiley.com/journal/qj](http://www.interscience.wiley.com/journal/qj)

The multiple-PC models with 5, 8 and 15 PCs are superior to the leading 2-PC model. However, the relationship between the skill of the forecasts and the number of PCs they contain is not a simple monotonic one, as the 25-PC model has lower forecast skill. There are two reasons for this relationship. Firstly, while adding some additional PCs to the basic leading 2-PC model produces an increase in forecast skill, there is a point at which adding extra PCs to the model provides no new information on organised intraseasonal behaviour; including them simply adds noise to the model. Secondly, including additional PCs increases the complexity of the network, which means it becomes progressively harder to accurately parametrise.

However, the nature of the relationship between the number of PCs included in the model and the forecast skill means that it may be possible to identify the optimum number of PCs required to forecast the MJO with a neural network model. To calculate this optimum number of PCs, the median pattern correlations between forecast regression maps and the total anomaly field verification maps were averaged over the five forecast pentads and over the 20 networks with different numbers of units in the hidden layer (section 3.3), for each of the multiple-PC models (Figure 1). Ideally, averaging over a number of networks would not be required and we would use a single network with the optimum number of hidden units for each model. However, in practice this averaging was required to remove sample noise from the data. The results are then plotted against the number of PC time series used in the model (Figure 8, solid line).

This figure shows that there is a clear advantage to be gained from using more than the leading two PCs when forecasting the MJO. The averaged pattern correlations increase with the number of PCs in the forecast model, before levelling off between 7 and 9 PCs. They then decrease rapidly as the number of PCs is increased further. Using the principle of parsimony, and after examining skill scores from models with up to 25 PCs, a model containing 7 PCs was decided to be the optimum choice.

The dotted line in Figure 8 shows the averaged pattern correlations based on forecast regression maps from just

the leading two PCs in each of the models. For the 2-PC model, this is identical to the solid line from the full model. For multiple-PC models up to 40 PCs, the anomaly correlation from the full model forecast (Figure 8, solid line) is higher than the anomaly correlation from just the leading two PCs of the model forecast (Figure 8, dotted line). This indicates that the additional PCs increase our ability to forecast the leading two PCs. However, above 40 PCs the increased complexity of the model reduces our ability to accurately parametrise the network, and the skill of the full model is lower than that based on just its leading two PCs.

#### 4.3. Model comparisons

With the optimum network identified (7 PCs, 8 hidden layer units), we can compare the ability of this model to the conventional leading 2-PC model over the MJO domain, during both winter and summer seasons. This is done firstly by comparing maps of temporal correlations at each forecast pentad for northern winter (October–March) and northern summer (April–September). It should be noted that although the forecasts are separated seasonally for analysis, they are generated from a year-round model. These are created by calculating temporal correlations at each grid point between forecast and total anomaly field verification maps for each season using the independent-validation dataset (Figure 9). For each season, the 7-PC model correlation maps are similar to the 2-PC model correlation maps but the 7-PC model has slightly greater correlations over the majority of the domain, with the starkest difference between the two models found in northern summer. Two regions are of particular note in the northern summer correlation maps – the northern parts of the domain (Arabian Sea, Bay of Bengal, South China Sea and Phillipine Sea), and the equatorial western Pacific. In these regions the optimum 7-PC model has relatively higher correlations than the conventional leading 2-PC model at both forecast lead times illustrated.

At this juncture it is important to verify that the fully nonlinear network model provides superior performance to the linear multiple regression models that have been used to forecast the MJO in previous studies (Jones *et al.*, 2004). This was achieved by creating multiple regression models using the same inputs, and forecasting the same outputs as the 2-PC and optimum 7-PC neural network models. Maps of temporal correlations between forecast and total anomaly field verification maps for winter and summer seasons, for each model are compared in Figure 10. The maps show that, as with the neural network models, there is an increase in the correlations with the addition of extra PCs and that this increase is most pronounced during northern summer. Comparing Figure 10 to Figure 9 illustrates that the neural network models perform better than the linear multiple regression models in general and that this improvement is most pronounced for the 7-PC model and also during northern summer.

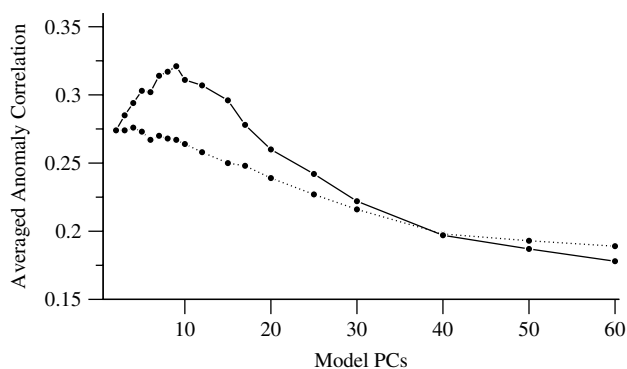


Figure 8. Median anomaly correlations between forecast regression maps and total anomaly field verification maps averaged over 5 pentad forecasts from an ensemble of 20 models with a range of units in the hidden layer, plotted against the number of PCs used in the model (solid line). The dotted line shows the median anomaly correlation when the forecast regression map is constructed using just the leading two PCs.



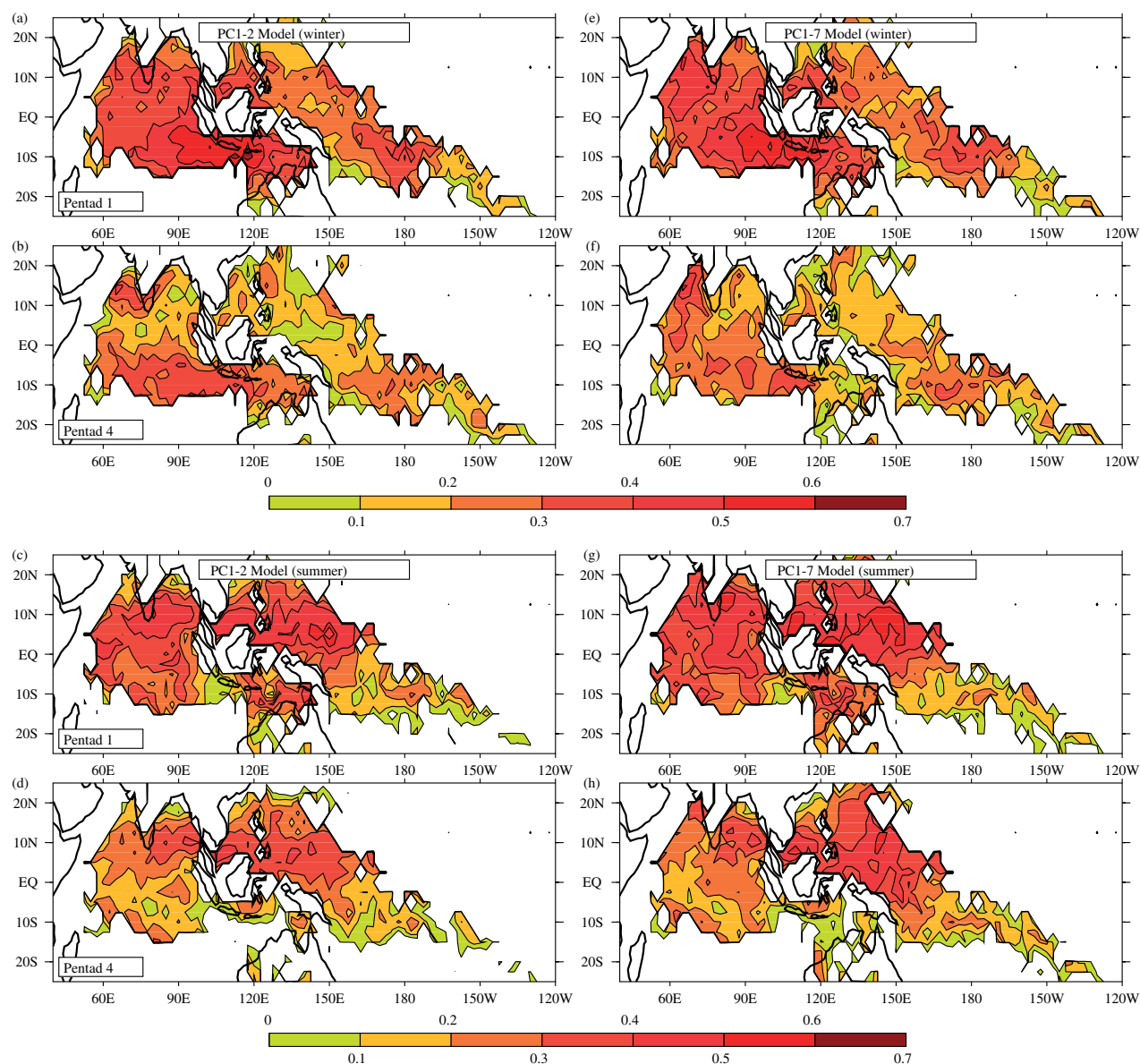


Figure 9. Masked maps of the correlation coefficient between forecasts and the total intraseasonal anomaly field from the 2-PC model in the northern winter at lead times of (a) 1 and (b) 4 pentads. (c, d) are as (a, b), but for the northern summer. (e–h) are as (a)–(d), but from the 7-PC model. The contour interval is 0.1. This figure is available in colour online at [www.interscience.wiley.com/journal/qj](http://www.interscience.wiley.com/journal/qj)

The behaviour of the optimum 7-PC neural network model can be analysed in greater detail by calculating skill scores (Wilks, 2005) for the optimum model containing 3, 4, 5, 6 and 7 PCs, with the conventional leading 2-PC model used as a reference, comparison forecast for each season. The skill scores  $S$  are calculated using the equation.

$$S = \left( \frac{A_{\text{for}} - A_{\text{ref}}}{A_{\text{per}} - A_{\text{ref}}} \right) \times 100, \quad (8)$$

where  $A_{\text{for}}$  is the measure of the multiple-PC model's forecast ability (e.g. temporal correlation),  $A_{\text{per}}$  is the value of the measure for a perfect forecast (i.e. 1 for a correlation), and  $A_{\text{ref}}$  is the value of the measure for the reference 2-PC model forecast.

In northern winter the reference 2-PC model will have skill in the regions where its constituent EOF1

(Figure 11(a)) and EOF2 (Figure 11(b)) patterns have large amplitude. This is the core MJO region, from the equatorial Indian Ocean eastwards to the western Pacific. With the addition of one extra PC (PC3), the model skill shows an improvement over the western equatorial Pacific (Figure 11(c)). This is a region where the corresponding EOF3 pattern has significant amplitude (Figure 11(d)). Hence part of the extra variance accounted for by including PC3 is coherently related to the MJO and improves the skill of the forecast. This may be related to the movement of the MJO convective envelope eastward over the Pacific during El Niño events (Kessler, 2001). The addition of PC4 to the model leads to a further increase in skill, over the Phillipine Sea (Figure 11(e)). This corresponds to a region of large amplitude in the corresponding EOF4 (Figure 11(f)). PC5 sees an increase in skill over the Bay of Bengal and the eastern



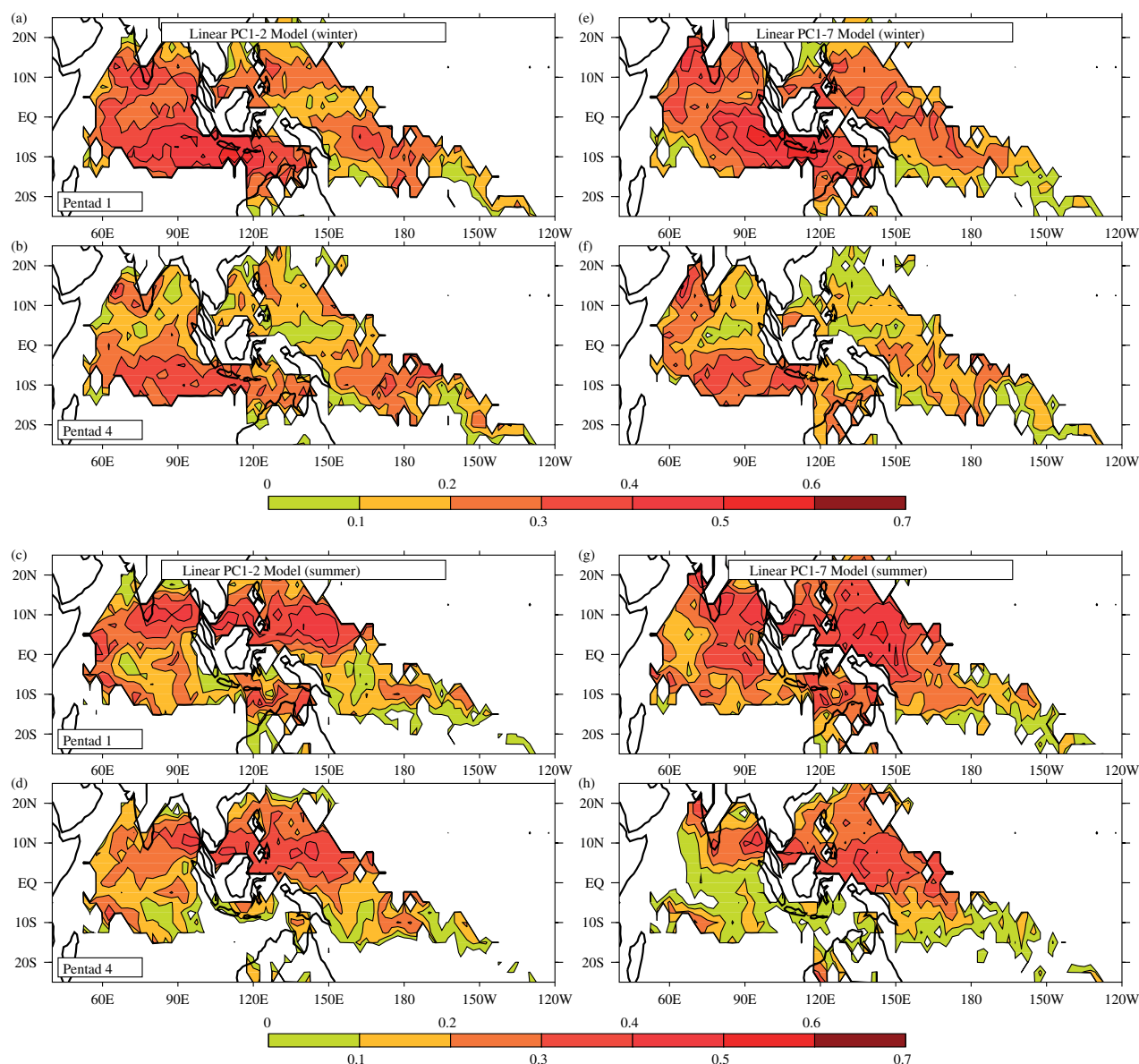


Figure 10. Masked maps of the correlation coefficient between forecasts and the total intraseasonal anomaly field from the 2-PC model in northern winter at lead times of (a) 1 and (b) 4 pentads. (c, d) are as (a, b) but for the northern summer. (e–h) are as (a)–(d), but from the 7-PC multiple linear regression model. The contour interval is 0.1. This figure is available in colour online at [www.interscience.wiley.com/journal/qj](http://www.interscience.wiley.com/journal/qj)

Arabian Sea (Figure 11(g)), corresponding to an anti-node in EOF5 (Figure 11(h)). With the addition of PC6 (Figure 11(i)), there are marginal increases in skill over these regions, consistent with regions of large amplitude in EOF6 (Figure 11(j)). With the addition of EOF7 (Figure 11(l)) there is little change in the skill (Figure 11(k)).

However, it should be noted that the increase in skill is not monotonic with the addition of each further PC. Not all of the variance associated with each EOF is coherently and predictably related to the MJO. The addition of this noise eventually degrades the forecast, such that no further advantage is gained by including more than 7 PCs. Moreover, of particular note is the region of negative skill in the 7-PC model over the eastern Indian Ocean, and the Banda Sea, Timor Sea and Arafura Sea just to the north of Australia. This feature becomes coherent when

PC4 is added (Figure 11(e)) and is presumably related to the part of the EOF4 pattern (Figure 11(f)) in this region that is not coherently related to the MJO.

In northern summer, the superior performance of the 7-PC model compared to the 2-PC model is much more pronounced. The addition of one extra PC (PC3), results in an improvement in model skill over the western equatorial Pacific (Figure 12(a)). This improvement is over a greater area than that of the northern winter forecasts but is once again a region where the corresponding EOF3 pattern has significant amplitude (Figure 11(d)). The addition of PC4 to the model leads to a further increase in skill, over the Philippine Sea and an extension and reinforcing of the improved area over the western equatorial Pacific (Figure 12(b)). This again corresponds to a region of large amplitude in the corresponding EOF4 (Figure 11(f)), while the addition of PC5 sees an

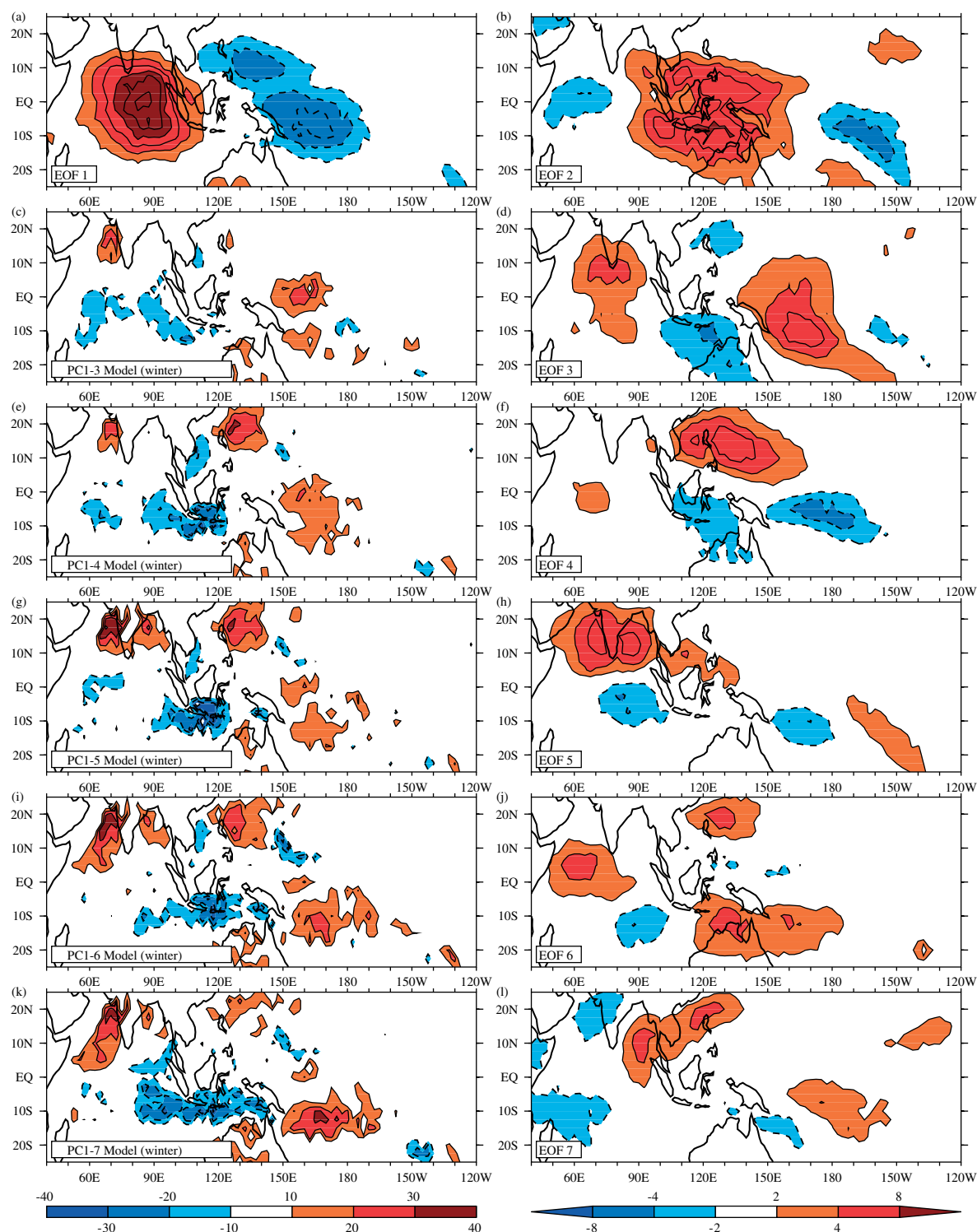


Figure 11. The spatial patterns of the leading EOFs of OLR: (a) EOF1, (b) EOF2, (d) EOF3, (f) EOF4, (h) EOF5, (j) EOF6, and (l) EOF7. The contour interval is  $2 \text{ W m}^{-2}$ , the zero contour is omitted and negative contours are dotted. The right-hand legend gives the shading. Maps of the skill score of the multiple-PC models compared to the 2-PC model for the northern winter with correlation as the measure of forecast ability, for one pentad lead time forecasts: (c) 3 PCs, (e) 4 PCs, (g) 5 PCs, (i) 6 PCs, and (k) 7 PCs. The contour interval is 10%, the zero contour is omitted and negative contours are dashed, and the left-hand legend gives the shading. This figure is available in colour online at [www.interscience.wiley.com/journal/qj](http://www.interscience.wiley.com/journal/qj)

increase in skill over the Bay of Bengal and the eastern Arabian Sea (Figure 12(c)), which, as for the northern winter forecasts, corresponds to an anti-node in EOF5 (Figure 11(h)). The addition of PC6 (Figure 12(d)) does not result in any significant changes in skill, but the

addition of EOF7 (Figure 11(l)) results in an increase in skill over the Celebes Sea and a significant reinforcing of the existing improved areas skill (Figure 12(e)).

The effect of this regional increase in skill, particularly in northern summer, from the addition of EOFs 3–7

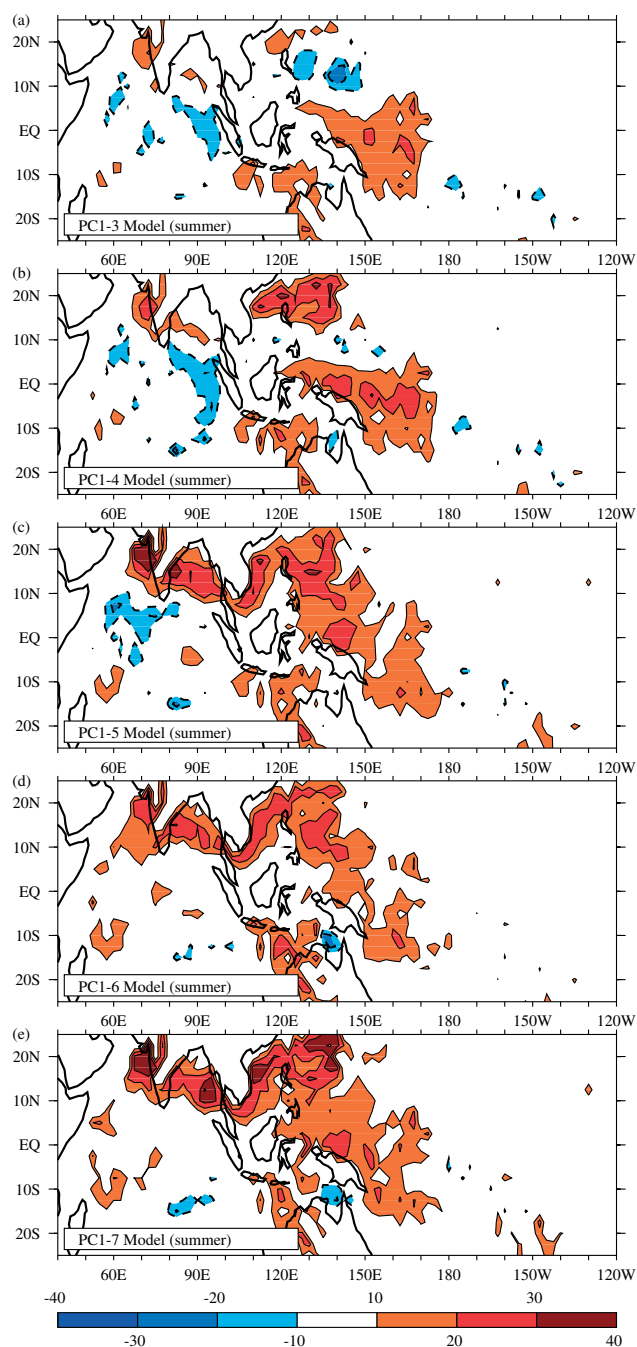


Figure 12. Maps of the skill score of the multiple-PC models compared to the 2-PC model for northern summer with correlation as the measure of forecast ability, for one pentad lead time forecasts: (a) 3 PCs, (b) 4 PCs, (c) 5 PCs, (d) 6 PCs, and (e) 7 PCs. The contour interval is 10%, the zero contour is omitted and negative contours are dashed. This figure is available in colour online at [www.interscience.wiley.com/journal/qj](http://www.interscience.wiley.com/journal/qj)

is further demonstrated in a sample forecast initiated on the 5 August 2001 (Figure 13). The 7-PC model performs better than the 2-PC model with higher anomaly correlations at both forecast lead times. The values of the anomaly correlations should be considered in the context of the amount of variance accounted for by the EOFs used in each model: 16% for the 2-PC model, and 32% for the 7-PC model (Figure 1). Comparing this sample forecast to the map of the skill of the 7-PC model compared

to the 2-PC model in Figure 12(e), it can be seen that the areas where the 7-PC model performs well, and the 2-PC model does not, correspond approximately to the areas of positive skill. It should also be noted that the 7-PC model forecast not only forecasts the phase of the anomalies well, but forecasts their amplitude accurately as well, while the forecast from the 2-PC model is of low amplitude when compared to the observations.

## 5. Conclusions and discussion

This study has assessed the ability of multiple-PC neural network models in forecasting the MJO and has compared their performance to conventional MJO forecasts made up of the leading two PCs. By creating models containing increasing numbers of PCs we were able to identify the optimum number of 7 PCs to include in the neural network model. This 7-PC model produced better forecasts of the total intraseasonal anomaly field than the conventional leading 2-PC model over the majority of the MJO's principle region of influence and at all forecast lead times. These improvements were most significant during northern summer particularly over the northernmost extent of the domain and over the equatorial Western Pacific (Figure 12). The improvements were also present in the northern winter forecasts, although there was also a reduction in skill over the eastern Indian Ocean. The scale of the improvements in skill should not be underestimated, as they must be considered in the context of the chosen 7-PC model accounting for approximately twice the amount of variance in the data (32% of the total intraseasonal anomaly field; Figure 1) that its forecasts are compared to, than the 2-PC model (16%).

There are probably a number of reasons for the improvements in forecasting ability. The reason that the multiple-PC models have superior forecast skill in the equatorial Western Pacific is likely due to the weaker variance and less coherent spatial structure of the MJO in this region, that is poorly represented by the leading two PCs alone. In the northernmost extent of the domain, it is probably a case of the additional PCs providing information that is not present in the leading two PCs. The analysis of this additional forecast skill also considered which PCs contributed additional skill in particular areas. By calculating the skill of the multiple-PC models, compared to the leading 2-PC model, and then comparing the results to the spatial patterns of each PC, we found that PC3 provides the increase in skill over the equatorial western Pacific while PCs 4–7 provide the increases in skill over the northern extent of the domain.

There is an area of negative skill over the eastern Indian Ocean where the leading 2-PC model performs better than the multiple-PC models in northern winter. This result is likely due to the inclusion of variance in EOF4 that is uncorrelated with the MJO, and therefore degrades the forecast. However, despite the negative skill here, the forecast correlations in this region (Figure 9) are still much higher than those over the western Pacific and the

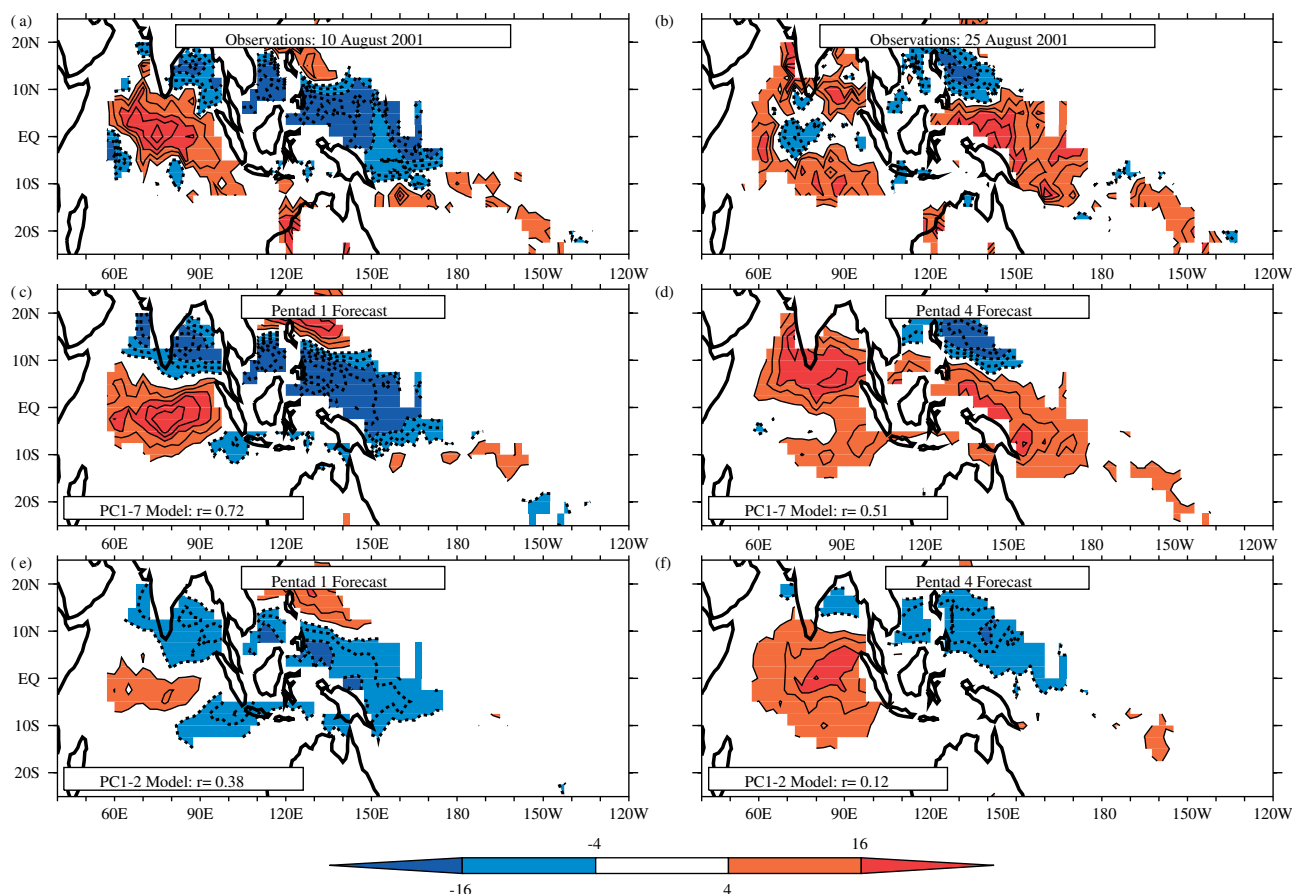


Figure 13. Observed masked OLR anomaly field on (a) 10 August 2001 and (b) 25 August 2001. Corresponding forecast regression maps from forecasts initiated on 5 August 2001 for the 7-PC model at lead times of (c) 1 pentad and (d) 4 pentads, and for the 2-PC model at lead times of (e) 1 pentad and (f) 4 pentads. The pattern correlation  $r$  between forecast and observation maps is shown at the bottom left of (c)–(f). The contour interval is  $4 \text{ W m}^{-2}$ , the zero contour is omitted and negative contours are dotted. This figure is available in colour online at [www.interscience.wiley.com/journal/qj](http://www.interscience.wiley.com/journal/qj)

northernmost extent of the domain, so the optimum 7-PC forecast model is still valid.

The decrease in skill over the eastern Indian Ocean is best explained by consideration of the seasonal cycle in the MJO. As the MJO is strongest in northern winter and spring, the leading two PCs are biased towards these seasons, when the MJO convective envelope is at its furthest south. Hence the 2-PC model performs well in the eastern Indian Ocean (high correlations in Figure 9(a–e)), and may be the optimum model here. Addition of further PCs then only adds noise here, degrading the forecast in this region.

Conversely, the leading two PCs do not account for much of the variance over the northernmost region, where the MJO is active in northern summer: the Arabian Sea, Bay of Bengal, South China Sea and Phillipine Sea. The forecast in these areas is improved when additional PCs are included.

The example forecast highlights the situations where the skill of the 7-PC model exceeds that of the 2-PC model. The observations in Figure 13(a,b) show that the total intraseasonal anomaly field is dominated by the quadrupole normally associated with the behaviour of the MJO and the Asian monsoon in northern summer (Annamalai and Slingo, 2001). The 7-PC model forecast

also shows that this model can capture the northward deviation of the MJO's propagation in northern summer, whereas the ability of the 2-PC model is constrained by the fact that the variance in the total intraseasonal anomaly field accounted for by the first two EOFs is dominated by the strong variability of the MJO during northern winter. Therefore whilst the inclusion of higher-order EOFs will have less impact on forecasting the MJO during northern winter, it can substantially improve forecasts (compared to the conventional 2-PC model) when the MJO is in its weaker northern summer phase. This is important since it could potentially lead on to improvements in forecasting the Asian summer monsoon. However, while the 7-PC neural network model gives improvements in forecast skill, these improvements come at the expense of not knowing physically where these improvements come from, due to the 'black box' nature of neural networks. This point needs to be considered in any applications of this work or similar studies.

In summary, the 7-PC neural network model produces more skilful forecasts of the total intraseasonal anomaly field than a conventional leading 2-PC model. The improved forecasting ability of this multiple-PC model has important ramifications for the wider field of MJO forecasting and indicates that forecasts of the MJO need



not be limited to the leading two PCs. Being able to produce forecasts that have a greater degree of accuracy over certain regions could be extremely useful to national weather services as a variable in their forecast models or as an indicator of mean rainfall over the next few weeks, which would be of considerable use in agricultural planning and for the tourism industries.

### Acknowledgements

BSL was supported by a studentship from the UK Natural Environment Research Council (EMS/51/2004/17). The OLR data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA from their website at <http://www.cdc.noaa.gov/>.

### References

- Alexander MA, Matrosova L, Penland C, Scott JD, Chang P. 2008. Forecasting Pacific SSTs: Linear inverse model predictions of the PDO. *J. Climate* **21**: 385–402.
- Annamalai H, Slingo JM. 2001. Active/break cycles: Diagnosis of the intraseasonal variability of the Asian summer monsoon. *Climate Dyn.* **18**: 85–102.
- Box GEP, Jenkins GM. 1970. *Time series analysis*. Holden-Day: San Francisco.
- Cybenko G. 1989. Approximation by superpositions of a sigmoidal function. *Math. Control Signal Syst.* **2**: 303–314.
- 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. Pp 19–62 in *Intraseasonal variability in the atmosphere–ocean climate system*. Lau WK-M, Waliser DN. (eds.) Springer-Praxis: Berlin.
- Hendon HH, Liebmann B, Newman M, Glick JD, Schemm J. 2000. Medium range forecast errors associated with active episodes of the Madden–Julian oscillation. *Mon. Weather Rev.* **128**: 69–86.
- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London, Series A* **454**: 903–995.
- Jones C, Waliser DE, Schemm JK, Lau WK. 2000. Prediction skill of the Madden–Julian oscillation in dynamical extended range forecasts. *Climate Dyn.* **16**: 273–289.
- Jones C, Carvalho LMV, Higgins RW, Waliser DE, Schemm JK. 2004. A statistical forecast model of tropical intraseasonal convective anomalies. *J. Climate* **17**: 2078–2095.
- Karnin ED. 1990. A simple procedure for pruning back-propagation trained neural networks. *IEEE Trans. Neural Networks* **1**(2): 239–242.
- Kessler WS. 2001. EOF representations of the Madden–Julian Oscillation and its connection with ENSO. *J. Climate* **14**: 3055–3061.
- 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. Amer. Meteorol. Soc.* **77**: 1275–1277.
- Lo F, Hendon HH. 2000. Empirical extended-range prediction of the Madden–Julian oscillation. *Mon. Weather Rev.* **128**: 2528–2543.
- Love BS, Matthews AJ, Janacek GJ. 2008. Real-time extraction of the Madden–Julian Oscillation using empirical mode decomposition and statistical forecasting with a VARMA model. *J. Climate* **21**: 5318–5335.
- Madden RA, Julian PR. 1994. Observations of the 40–50-day tropical oscillation – A review. *Mon. Weather Rev.* **122**: 814–837.
- Matthews AJ. 2008. Primary and successive events in the Madden–Julian oscillation. *Q. J. R. Meteorol. Soc.* **134**: 439–453.
- McPhaden MJ. 1999. Genesis and evolution of the 1997–98 El Niño. *Science* **283**: 950–954.
- Rosenblatt F. 1962. *Principles of Neurodynamics*. Spartan: New York.
- Rumelhart DE, Hinton GE, Williams RJ. 1986a. Learning representations by back-propagating errors. *Nature* **323**: 533–536.
- Rumelhart DE, Hinton GE, Williams RJ. 1986b. Learning internal representations by error propagation. Chapter 8 in *Parallel Distributed Processing 1*. Rumelhart DE, McClelland JL. (eds.) MIT Press: Cambridge, Mass.
- Salby ML, Hendon HH. 1994. Intraseasonal behaviour of clouds, temperature, and motion in the tropics. *J. Atmos. Sci.* **51**: 2207–2224.
- Waliser DE. 2005. Predictability and forecasting. In *Intraseasonal variability in the atmosphere–ocean climate system*. Lau WK-M, Waliser DE. (eds.) Springer: Berlin.
- Waliser DE, Jones C, Schemm JK, Graham NE. 1999. A statistical extended-range tropical forecast model based on the slow evolution of the Madden–Julian oscillation. *J. Climate* **12**: 1918–1939.
- Wheeler M, 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.
- Wilks DS. 2005. *Statistical methods in the atmospheric sciences*. Academic Press.
- Woolnough SJ, Vitart F, Balmaseda MA. 2007. The role of the ocean in the Madden–Julian Oscillation: Implications for MJO prediction. *Q. J. R. Meteorol. Soc.* **133**: 117–128.
- Zhang C. 2005. Madden–Julian Oscillation. *Rev. Geophys.* **43**: RG2003, DOI: 10.1029/2004RG000158.