GEOPHYSICAL RESEARCH LETTERS, VOL. ???, XXXX, DOI:10.1029/,

Predicting the cloud patterns of the Madden-Julian Oscillation through a low-order nonlinear stochastic model

N. Chen,[1] A. J. Majda,[1] and D. Giannakis[1]

Corresponding author: N. Chen, Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA. (chennan@cims.nyu.edu)

[1]Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, New York, USA.

We assess the limits of predictability of the large scale cloud patterns in the boreal winter Madden-Julian Oscillation (MJO) as measured through outgoing longwave radiation (OLR) alone, a proxy for convective activity. A recent advanced nonlinear time series technique, Nonlinear Laplacian Spectral Analysis, is applied to the OLR data to define two spatial modes with high intermittency associated with the boreal winter MJO. A recent data driven physics constrained low-order stochastic modeling procedure is applied to these time series. The result is a four dimensional nonlinear stochastic model for the two observed OLR variables and two hidden variables involving correlated multiplicative noise defined through energy conserving nonlinear interaction. Systematic calibration and prediction experiments show the skillful prediction by these models for 40, 25 and 18 days in strong, moderate and weak MJO winters, respectively. Furthermore, the ensemble spread is an accurate indicator of forecast uncertainty at long lead times.

1. Introduction

The dominant mode of tropical intraseasonal variability is the Madden-Julian Oscillation (MJO) which is a slow moving planetary scale envelope of convection propagating eastward typically from the Indian Ocean through the Western Pacific. The MJO effects tropical precipitation, the frequency of tropical cyclones, and extratropical weather patterns [*Lau and Waliser*, 2012]. Understanding and predicting the MJO is a central problem in contemporary meteorology with large societal impacts [*Zhang et al.*, 2013]. Predicting the MJO is a major enterprise through either low-order statistical models [*Jiang et al.*, 2008; *Seo et al.*, 2009; *Kang and Kim*, 2010; *Kondrashov et al.*, 2013] or operational dynamical models [*Gottschalck et al.*, 2010; *Vitart and Molteni*, 2010; *Zhang et al.*, 2013]. The popular metric for assessing large scale skill in MJO predictions [*Wheeler and Hendon*, 2004] involves both the winds at the top and bottom of the troposphere and the outgoing longwave radiation (OLR) which is a proxy for convective activity. While the use of this index has stimulated significant improvements and developments in MJO prediction, recent case studies [*Straub*, 2013; *Kiladis et al.*, 2014] have pointed out its limitations in measuring the OLR activity in some MJO events.

Here we assess the limits of predictability of the large scale cloud patterns in the boreal winter MJO [*Kang and Kim*, 2010] as measured through OLR activity alone. This is achieved in two steps. In the first step, a recent advanced nonlinear time series technique, Nonlinear Laplacian Spectral Analysis (NLSA) is applied directly to the OLR data to define two spatial modes associated with the boreal winter MJO. NLSA by design requires no ad hoc detrending or spatial-temporal filtering of the full OLR data set and captures both intermittency and low frequency variability [*Giannakis and Majda*, 2012a, b, 2013; *Giannakis et al.*, 2012]. The resulting time series for the two spatial modes representing the boreal winter MJO are depicted in Figure 1 and are highly intermittent with large variation in amplitude from year to year in the winter season. In the second stage, a recent systematic strategy for data driven physics constrained low-order stochastic modeling of time series [*Majda and Harlim*, 2013; *Harlim et al.*, 2014] is applied to the time series in Figure 1. The result is a four dimensional nonlinear stochastic model for the two variables in Figure 1 and two hidden variables. This low-order model involves correlated multiplicative noise defined through energy conserving nonlinear interactions between the observed and hidden variables as well as additive stochastic noise. The remainder of the paper as well as the auxiliary material demonstrate that this low-order stochastic model has high predictive skill and captures the limits of prediction for the OLR patterns of the boreal winter MJO as depicted in Figure 1.

2. The Boreal Winter MJO through NLSA

We analyze multi-satellite infrared brightness temperature (T_b) data from the Cloud Archive User Service (CLAUS) Version 4.7 (e.g., [*Hodges et al.*, 2000]). Brightness temperature is a measure of the earth's infrared emission in terms of the temperature of a hypothesized blackbody emitting the same amount of radiation at the same wavelength ($\sim$ 10-11 μm in CLAUS). It is a highly correlated variable with the total terrestrial longwave emission. In the tropics, positive (negative) T_b anomalies are associated with reduced (increased) cloudiness, hence suppressed (enhanced) deep convection. The global CLAUS T_b data are on a 0.5° longitude by 0.5° latitude fixed grid, with three-hour time resolution from 00 UTC to 21 UTC, spanning July 1, 1983 to June 30, 2006. The values of T_b range from 170 K to 340 K at approximately 0.67 K resolution.

We apply the NLSA algorithm to the full CLAUS data set restricted to the tropical belt 15°N–15°S, with a lagged embedding window of 60 days. A variety of extended spatial cloud patterns emerge from the analysis but the focus here is on the two spatial cloud patterns with time series depicted in Figure 1. It is evident from Figure 1 that these patterns are active from December through April of each year corresponding to boreal winter. Animation 1 in the auxiliary material presents the evolution of the cloud patterns from NLSA associated with these two time series from November 1992 through March 1993. The video shows two large scale MJO-like cloud patterns coinciding in time with the two boreal winter MJO's observed during the TOGA-COARE field experiment of 1992-1993 [*Webster and Lukas*, 1992; *Yanai et al.*, 2000]. This indicates that the time series depicted in Figure 1 give a reasonable representation of the movement of global cloud patterns associated with the boreal winter MJO. The details of the NLSA algorithm are not provided here since they are readily available [*Giannakis and Majda*, 2012a, b, 2013] and there is even a similar application of NLSA to study the MJO which utilizes the same lagged embedding window and compressed symmetric meridional averages of the CLAUS data [*Giannakis et al.*, 2012]. We use the terminology, MJO indices, for the two time series in Figure 1.

3. The Low-Order Nonlinear Stochastic Model

Denote by u_1 and u_2 the two components, MJO 1 and MJO 2, depicted in Figure 1. The probability distribution functions (PDFs) for u_1 and u_2 are highly non-Gaussian with fat tails indicative of the temporal intermittency in the large scale cloud patterns associated with boreal winter MJO. We propose the following family of low-order stochastic models to describe the intermittent variability of the time series u_1 and u_2:

$$\frac{du_1}{dt} = (-d_u\, u_1 + \gamma\,(v + v_f(t))\, u_1 - (a + \omega_u)\, u_2) + \sigma_u\, \dot{W}_{u_1}, \tag{1}$$
$$\frac{du_2}{dt} = (-d_u\, u_2 + \gamma\,(v + v_f(t))\, u_2 + (a + \omega_u)\, u_1) + \sigma_u\, \dot{W}_{u_2}, \tag{2}$$
$$\frac{dv}{dt} = (-d_v\, v - \gamma\,(u_1^2 + u_2^2)) + \sigma_v\, \dot{W}_v, \tag{3}$$
$$\frac{d\omega_u}{dt} = (-d_\omega \omega_u + \hat{\omega}_u) + \sigma_\omega\, \dot{W}_\omega, \tag{4}$$

with

$$v_f(t) = f_0 + f_t \sin(\omega_f t + \phi). \tag{5}$$

Besides the two observed MJO variables, u_1, u_2, the other two variables v and ω_u are hidden unobserved variables which represent the stochastic damping and stochastic phase, respectively. In (1)–(4), $\dot{W}_{u_1}, \dot{W}_{u_2}, \dot{W}_v$ and $\dot{W}_\omega$ are independent white noise. The time periodic damping in the equations in (1) and (2) is utilized to crudely model the active phase of the boreal winter MJO and the quiescent summer season in the seasonal cycle. The hidden variables v, ω_u interact with the observed MJO variables u_1, u_2 through energy conserving nonlinear interactions following the systematic physics constrained nonlinear regression strategies for time series developed recently [*Majda and Harlim*, 2013; *Harlim et al.*, 2014]. The low-order stochastic nonlinear models in (1)–(4) are fundamentally different from those utilized earlier [*Kondrashov et al.*, 2013; *Kravtsov et al.*, 2005] which allow for nonlinear interactions only between the observed variables u_1, u_2 and only special linear interactions with layers of hidden variables. Further motivation for the models in

(1)–(4) is provided by the stochastic skeleton model which predicts key features of the MJO [*Majda and Stechmann*, 2009, 2011; *Thual et al.*, 2013]; these are coupled nonlinear oscillator models of the MJO where if we identify the OLR variables with the envelope of synoptic scale convective activity, the hidden variables v, ω_u and their dynamics become phenomenological surrogates for the energy conserving interactions in the skeleton model involving the synoptic scale activity and the equatorial convective dynamic equations for temperature, velocity, and moisture.

3.1. Calibration of the Nonlinear Stochastic Models

The parameters of the stochastic model in (1)–(5) are calibrated by fitting the highly non-Gaussian PDFs and autocorrelations of the two MJO variables u_1, u_2. Table 1 records the optimal parameter values while Figure 2 displays the skill of the stochastic model with these parameters in recovering the statistics of the two MJO indices. Panels (a) and (b) show that the stochastic model from (1)–(4) succeeds in capturing the autocorrelations almost perfectly for a three month duration and even the wiggles that appears with lags around one year. Panel (c) shows that the stochastic model captures the fat tailed highly non-Gaussian PDF's of the two MJO indices due to intermittency. Panel (d) shows that the power spectrum of the two MJO indices from the data and those from the stochastic model match very well. The optimal parameters in the stochastic model from Table 1 have been determined by systematically minimizing the information distance of the equilibrium PDF of the stochastic model compared with that of the actual data [*Majda and Gershgorin*, 2010, 2011]. Details are presented in the auxiliary material which also demonstrates the robustness of these optimal parameters to their variation.

3.2. Prediction algorithm and Data Assimilation of the Hidden Variables

As shown in Figure 1, the stochastic model in (1)–(4) is trained on the first seventeen years of data from 1983 through 1999 and forecasts are made for the last six years of data from January 1, 2000 until the end of 2005. The estimates of the hidden parameters v, ω_u during the training period and initialization of these parameters during the prediction phase exploit the special structure of the low-order nonlinear stochastic model; the equations in (1)–(4) are a conditional Gaussian system with respect to the observation of u_1 and u_2, meaning that once u_1 and u_2 are given, there are closed analytic equations for the conditional Gaussian distributions of the hidden parameters v, ω_u [*Liptser and Shiryaev*, 2001]. Thus, we have conditional Gaussian distributions for the hidden variables, v, ω_u, during the training phase. The auxiliary material contains the details and explicit equations. We utilize this fact to construct an initial ensemble for forecasting in the prediction phase. Take the initial data $\vec{U}_0 = (u_1, u_2)$ which is given at time t from the observed data; consider all data in the training period $\vec{U}_\epsilon$ so that $|\vec{U}_0 - \vec{U}_\epsilon| < \epsilon$ with their corresponding conditional Gaussian distribution for the hidden variables $p_0(\Gamma|\vec{U}_\epsilon)$ where $\Gamma = (v, \omega_u)$. Construct an ensemble PDF $p_0(\Gamma|\vec{U}_0)$ by collecting all the PDF's for $p_0(\Gamma|\vec{U}_\epsilon)$ with equal weights; make an initial ensemble for prediction for (1)–(4) by using $\vec{U}_0$ from the observations and drawing N-samples for Γ_0 from the distribution $p_0(\Gamma|U_0)$. In practice, we start with $\epsilon = .01$; if there is no $\vec{U}_\epsilon$ in the historic training period with $|\vec{U}_0 - \vec{U}_\epsilon| < \epsilon$, increase ϵ to $\epsilon = .02$, etc. Typically $\epsilon = .01$ or $.02$ is large enough to generate a few $\vec{U}_\epsilon$ in the historic training period while occasionally $\epsilon = .05$ is needed for the active MJO phase. In the predictions below with (1)–(4) we use N-ensemble members with $N = 50$.

4. Results and Discussion

We report the prediction skill of the stochastic model in (1)–(4) with the optimal parameters from Table 1 and the ensemble initialization scheme described above for the six year prediction period from January 1, 2000 to the end of December, 2005. The comparison of the ensemble mean prediction and the truth at lead times of 15 and 25 days for MJO index 1 for all six years are shown in Figure 3. The 15 day predictions are very skillful and even the 25 day predictions have highly significant skill. It is evident from Figure 3 that the years 2001, 2002, 2004 have strong boreal winter MJO's while the years 2003 and 2005 have moderate MJO's and the year 2000 has weak boreal winter MJO's. Figure 4 presents the RMS errors in prediction of the two MJO indices as a function of lead time in the six years as well as the bivariate correlation patterns. In the strong MJO years, 2001, 2002, 2004, there is significant prediction skill out to roughly forty days; for the moderate MJO years, 2003, 2005, there is skillful prediction until 25 days while for the weak MJO years, 2000, there is skillful prediction out to 18 or 19 days. Figure 5 shows the ensemble predictions including the ensemble spread for the six years, beginning at the three dates, November 1, January 10 and March 1. November 1 is a time at the transition between the quiescent phase and the active phase of the boreal winter MJO indices; January 10 is a starting date in the active mature phase while March 1 is a starting date in the decaying phase of MJO activity. As shown in Figure 5, the ensemble mean predictions for the November 1 starting date do not have any long range skill but the ensemble spread automatically predicts this lack of skill and the envelope of the ensemble predictions contains the true signal for all years and forecast times including the return to skill in the summer quiescent phase. The forecasts from January 10 obviously have skill from both the mean and ensemble spread for all years for long lead times. The forecasts starting from March 1 have both an accurate mean and small ensemble spread for all six years and for very long times. The auxiliary material shows that the prediction skill of the nonlinear stochastic model is robust to suboptimal parameters.

The auxiliary material also contains the results of twin prediction experiments with the perfect nonlinear stochastic model in (1)–(4) where 17 year training segments of the data generated from the model are utilized to make 6 year forecasts. It is significant that this internal prediction skill of the stochastic model is comparable to its skill in predicting the two boreal winter MJO indices from observations. This lends support to the fact that the nonlinear stochastic model in (1)–(4) can accurately determine the predictability limits of the two OLR MJO indices for boreal winter developed here.

5. Conclusions

A recently developed technique for nonlinear time series analysis NLSA [*Giannakis and Majda*, 2012a, b, 2013] has been utilized to define two MJO indices of the boreal winter MJO for the large scale cloud patterns based only on OLR from the CLAUS data set without detrending or spatial-temporal filtering. The observed time series have non-Gaussian fat-tailed PDF's as a consequence of intermittency. Both systematic strategies for physics constrained regression models [*Majda and Harlim*, 2013; *Harlim et al.*, 2014] and the dynamic stochastic skeleton model for the MJO [*Majda and Stechmann*, 2009, 2011; *Thual et al.*, 2013] suggest a four dimensional stochastic model with two hidden variables representing stochastic damping and random phasing with energy conserving nonlinear feedback interaction. In a calibration phase, these models can successfully capture the observed non-Gaussian PDFs and autocorrelations (Figure 2). The models have a special structure which leads to efficient data assimilation and ensemble initialization algorithms for the hidden variables. The low-order nonlinear stochastic model has been applied to prediction of the OLR-based indices for boreal winter MJO's with forecasting skill up to 40 days in strong MJO years, 25 days in moderate MJO years and roughly 18 or 19 days in weak MJO years (Figure 3, 4 and 5); furthermore, the ensemble spread in the stochastic model has been shown to be an accurate predictive indicator of forecast uncertainty at long range (Figure 5). It is shown in the auxiliary material that perfect twin experiments with the stochastic model have the comparable skill as with the observed data suggesting that the low-order nonlinear stochastic model has significant skill for determining the predicability limits of the large scale cloud patterns of the boreal winter MJO.

Acknowledgments. The authors thank Wen-Wen Tung and Eniko Szekely for discussion on NLSA applied to CLAUS data. This research of A.J.M and D.G is partially supported by the Office of Naval Research grant ONR MURI N00014-12-1-0912. N.C. is supported as a graduate research assistant on this grant.

References

Giannakis, D., and A. J. Majda (2012a), Comparing low-frequency and intermittent variability in comprehensive climate models through Nonlinear Laplacian Spectral Analysis, *Geophysical Research Letters*, *39*(10), doi: 10.1029/2012GL051575.

Giannakis, D., and A. J. Majda (2012b), Nonlinear Laplacian Spectral Analysis for time series with intermittency and low-frequency variability, *Proceedings of the National Academy of Sciences*, *109*(7), 2222–2227.

Giannakis, D., and A. J. Majda (2013), Nonlinear laplacian spectral analysis: capturing intermittent and low-frequency spatiotemporal patterns in high-dimensional data, *Statistical Analysis and Data Mining*, *6*(3), 180–194.

Giannakis, D., W.-w. Tung, and A. J. Majda (2012), Hierarchical structure of the Madden-Julian Oscillation in infrared brightness temperature revealed through Nonlinear Laplacian Spectral Analysis, in *Intelligent Data Understanding (CIDU), 2012 Conference on*, pp. 55–62, IEEE.

Gottschalck, J., M. Wheeler, K. Weickmann, F. Vitart, N. Savage, H. Lin, H. Hendon, D. Waliser, K. Sperber, M. Nakagawa, et al. (2010), A framework for assessing operational Madden-Julian Oscillation forecasts: A clivar MJO working group project, *Bulletin of the American Meteorological Society*, *91*(9), 1247–1258.

Harlim, J., A. Mahdi, and A. J. Majda (2014), An ensemble kalman filter for statistical estimation of physics constrained nonlinear regression models, *Journal of Computational Physics*, *257*, 782–812.

Hodges, K., D. Chappell, G. Robinson, and G. Yang (2000), An improved algorithm for generating global window brightness temperatures from multiple satellite infra-red imagery. *Journal of Atmospheric and Oceanic Technology*, *17*, 1296–1312.

Jiang, X., D. E. Waliser, M. C. Wheeler, C. Jones, M.-I. Lee, and S. D. Schubert (2008), Assessing the skill of an all-season statistical forecast model for the Madden-Julian oscillation, *Monthly Weather Review*, *136*(6), 1940–1956.

Kang, I.-S., and H.-M. Kim (2010), Assessment of MJO predictability for boreal winter with various statistical and dynamical models, *Journal of Climate*, *23*(9), 2368–2378.

Kiladis, G. N., J. Dias, K. H. Straub, M. C. Wheeler, S. N. Tulich, K. Kikuchi, K. M. Weickmann, and M. J. Ventrice (2014), A comparison of OLR and circulation based indices for tracking the MJO, *Monthly Weather Review*, *142*(5), 1697–1715.

Kondrashov, D., M. Chekroun, A. Robertson, and M. Ghil (2013), Low-order stochastic model and past-noise forecasting of the Madden-Julian Oscillation, *Geophysical Research Letters*, *40*(19), 5305–5310, doi: 10.1002/grl.50991.

Kravtsov, S., D. Kondrashov, and M. Ghil (2005), Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability, *Journal of Climate*, *18*(21), 4404–4424.

Lau, W. K., and D. E. Waliser (2012), *Intraseasonal variability in the atmosphere-ocean climate system*, Springer.

Liptser, R. S., and A. N. Shiryaev (2001), *Statistics of Random Processes II: II. Applications*, vol. 2, Springer.

Majda, A. J., and B. Gershgorin (2010), Quantifying uncertainty in climate change science through empirical information theory, *Proceedings of the National Academy of Sciences*, *107*(34), 14,958–14,963.

Majda, A. J., and B. Gershgorin (2011), Improving model fidelity and sensitivity for complex systems through empirical information theory, *Proceedings of the National Academy of Sciences*, *108*(25), 10,044–10,049.

Majda, A. J., and J. Harlim (2013), Physics constrained nonlinear regression models for time series, *Nonlinearity*, *26*(1), 201–217.

Majda, A. J., and S. N. Stechmann (2009), The skeleton of tropical intraseasonal oscillations, *Proceedings of the National Academy of Sciences*, *106*(21), 8417–8422.

Majda, A. J., and S. N. Stechmann (2011), Nonlinear dynamics and regional variations in the MJO skeleton, *Journal of the Atmospheric Sciences*, *68*(12), 3053–3071.

Seo, K.-H., W. Wang, J. Gottschalck, Q. Zhang, J.-K. E. Schemm, W. R. Higgins, and A. Kumar (2009), Evaluation of MJO forecast skill from several statistical and dynamical forecast models, *Journal of Climate*, *22*(9), 2372–2388.

Straub, K. H. (2013), MJO initiation in the real-time multivariate MJO index, *Journal of Climate*, *26*(4), 1130–1151.

Thual, S., A. J. Majda, and S. N. Stechmann (2013), A stochastic skeleton model for the MJO, *Journal of the Atmospheric Sciences*, *71*, 697–715.

Vitart, F., and F. Molteni (2010), Simulation of the Madden-Julian Oscillation and its teleconnections in the ecmwf forecast system, *Quarterly Journal of the Royal Meteorological Society*, *136*(649), 842–855.

Webster, P. J., and R. Lukas (1992), TOGA COARE: The coupled ocean-atmosphere response experiment, *Bulletin of the American Meteorological Society*, *73*(9), 1377–1416.

Wheeler, M. C., and H. H. Hendon (2004), An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction, *Monthly Weather Review*, *132*(8), 1917–1932.

Yanai, M., B. Chen, and W.-w. Tung (2000), The Madden-Julian Oscillation observed during the TOGA COARE IOP: Global view, *Journal of the atmospheric sciences*, *57*(15), 2374–2396.

Zhang, C., J. Gottschalck, E. D. Maloney, M. W. Moncrieff, F. Vitart, D. E. Waliser, B. Wang, and M. C. Wheeler (2013), Cracking the MJO nut, *Geophysical Research Letters*, *40*(6), 1223–1230, doi: 10.1002/grl.50244.

Table 1. Parameters for low-order stochastic model (1)–(5).

d_u	a	f_0	f_t	ϕ	ω_f	σ_u	d_v	σ_v	γ	d_ω	$\hat{\omega}_u$	σ_ω
0.9	3	1	4.9	-1	$2\pi/12$	0.3	0.9	1	0.3	0.5	0	1.1

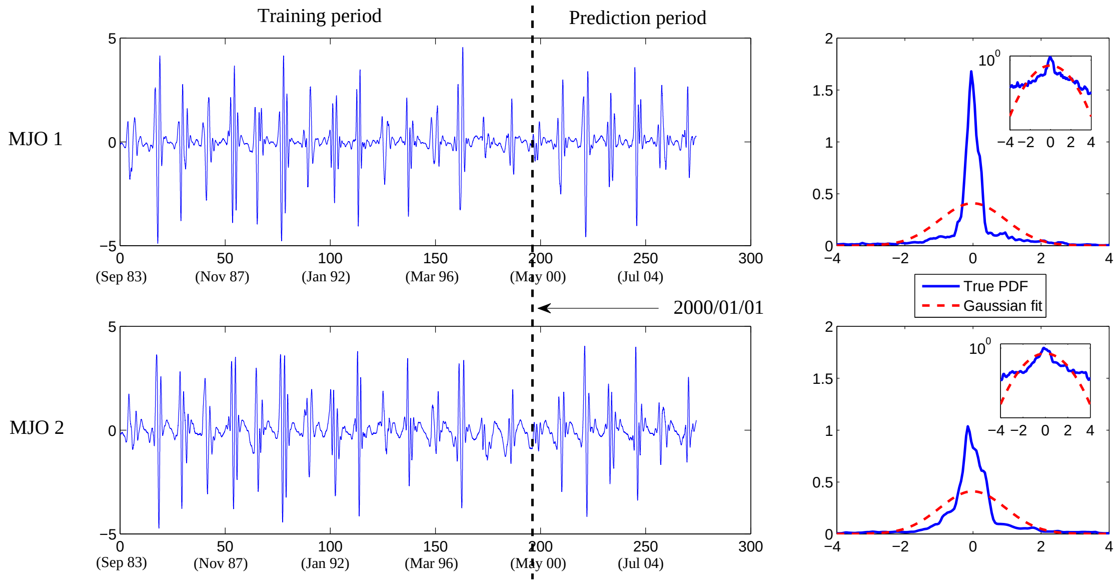

Figure 1. Left: MJO indices from NLSA (modes 8 and 9) ranging from 1983/09/03 to 2006/06/30. The time-series before 2000/01/01 is utilized as training period to get the statistics and that after 2000/01/01 represents the prediction period using the low-order stochastic model. Right: The associated PDF of each index and the Gaussian fit. The small panel inside each subplot shows the PDF in the logarithm scale.

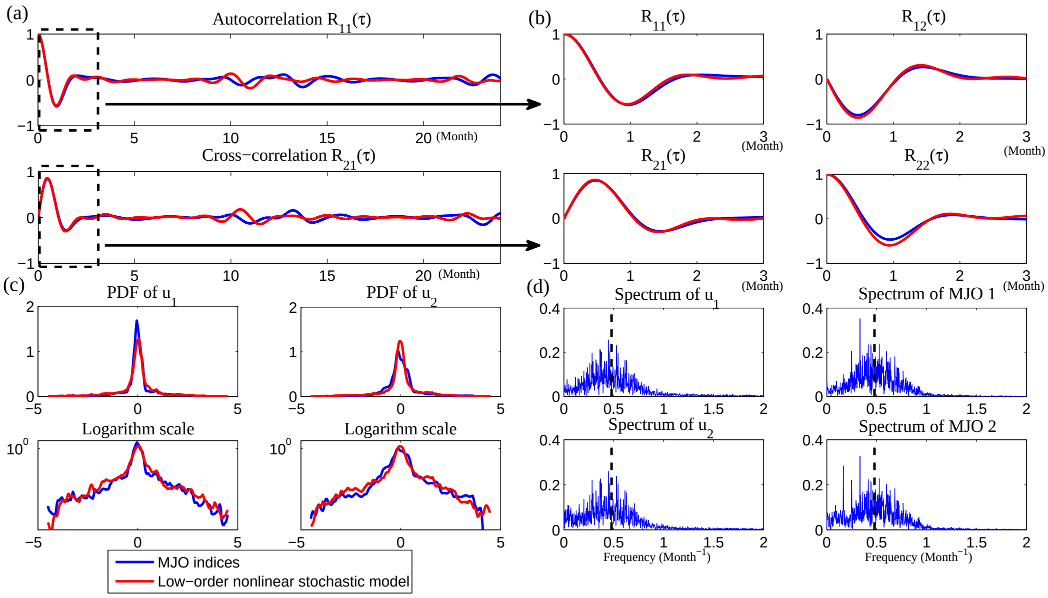

Figure 2. Statistics of the stochastic model with the optimal parameters in Table 1. (a) Long-term autocorrelation function $R_{11}(\tau)$ and cross-correlation function $R_{21}(\tau)$ from 0 to 24 months. (b) Short-term autocorrelation functions $R_{11}(\tau)$, $R_{22}(\tau)$ and cross-correlation functions $R_{21}(\tau)$, $R_{12}(\tau)$ from 0 to 3 months. (c) Equilibrium PDFs of the signal u_1, u_2 from stochastic model compared with that of the MJO indices. (d) Spectrum of u_1, u_2 compared with that of MJO indices. Here, the black dashed line indicates the frequency $a/(2\pi)$.

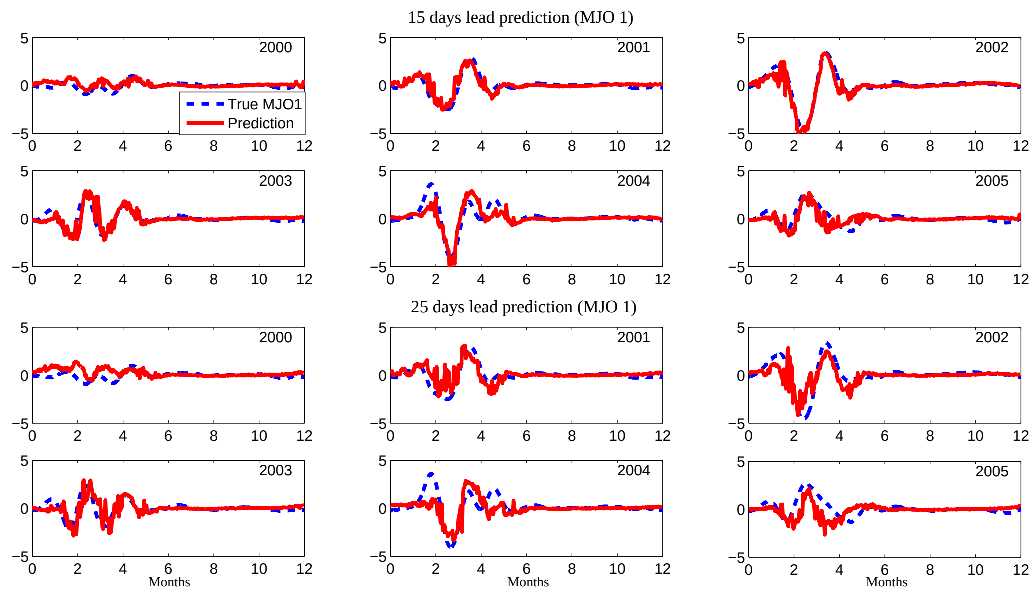

Figure 3. Prediction of MJO 1 at a 15 (top) and 25 (bottom) days lead. The blue line shows the true signal and the red line shows the ensemble average of the predicted signal with 50 ensemble members.

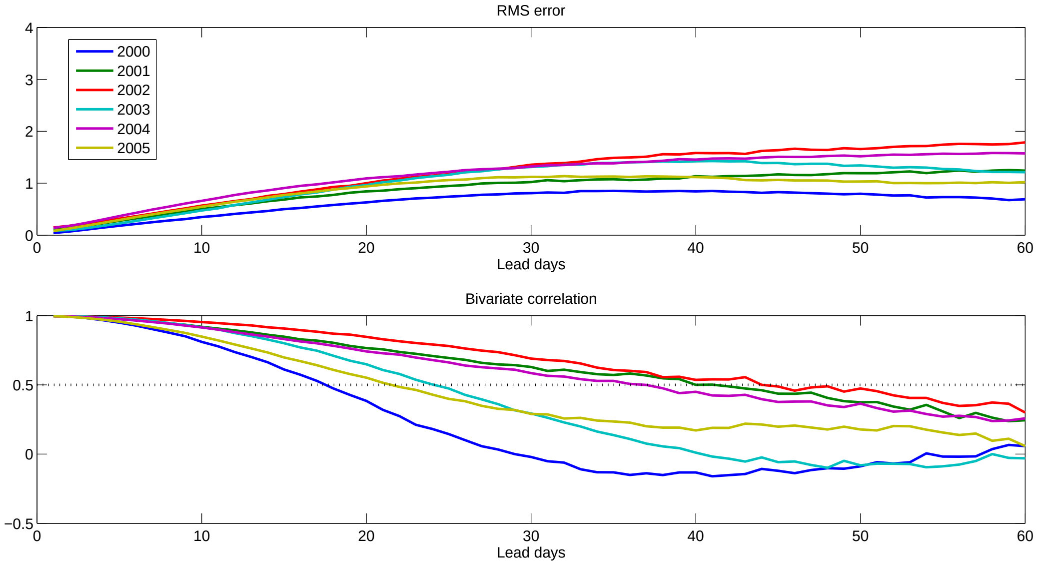

Figure 4. Skill scores with RMS error (top) and bivariate correlation (bottom) for prediction in different years.

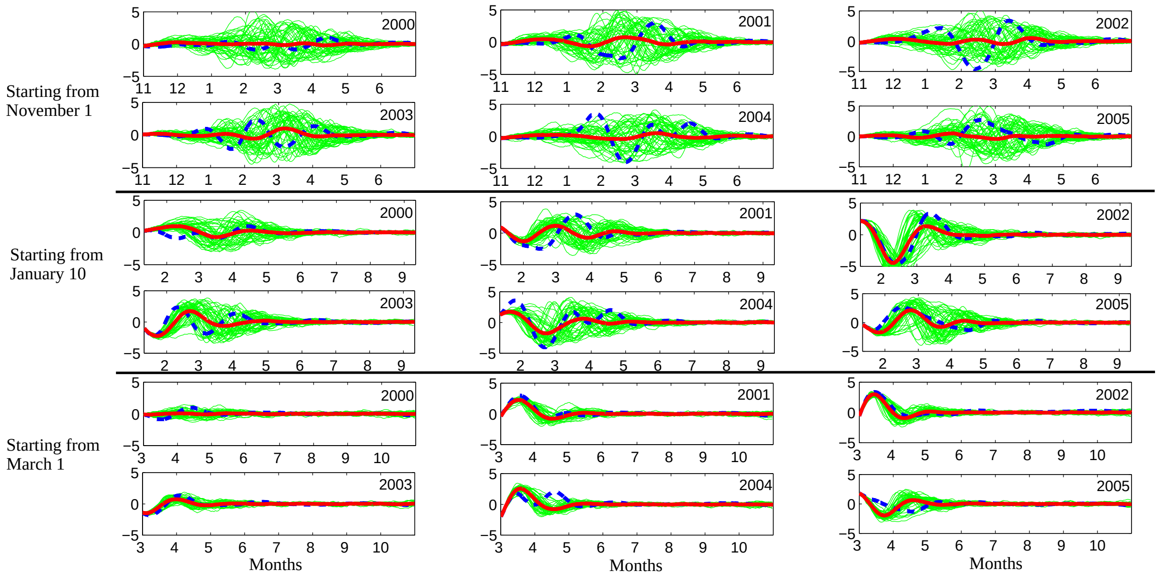

Figure 5. First and second rows: Prediction of MJO 1 starting from November 1 for different years. Each panel show the prediction skill of 8 months with the label in x-axis indicating the month. Third and forth rows: Same but starting from January 10. Fifth and Sixth rows: Same but starting from March 1. The thick blue dashed line is the MJO 1 index. The thick red solid line is the ensemble mean with 50 members, which are shown by the thin solid lines.